

Universidad Autónoma Metropolitana-Iztapalapa

Detección de Errores en la Medición de Espectroscopía de Impedancia Tisular Gástrica

Tesis que presenta para obtener el grado de:

Maestría en Ciencias (Ingeniería Biomédica)

I.B. María de Montserrat Godínez García

Mayo 2010

Asesores:

Dr. Emilio Sacristán Rock
Universidad Autónoma Metropolitana

M.I. Óscar Yañez Suárez
Universidad Autónoma Metropolitana

Sinodales:

Dr. César Antonio González Díaz
Escuela Militar de Graduados de Sanidad

Dr. John Charles Goddard Close
Universidad Autónoma Metropolitana

M.I. Óscar Yañez Suárez
Universidad Autónoma Metropolitana

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo financiero brindado para la realización de este proyecto.

A la empresa Innovamédica S.A.P.I de C.V. por la flexibilidad de horario y acceso a toda la información necesaria para la realización de este proyecto.

A mis asesores Dr. Emilio Sacristán Rock y M. en I. Óscar Yañez Suárez por su asesoría, aprendizaje continuo, apoyo y reflexiones que enriquecen mi formación profesional.

A todo el equipo de trabajo de Innovamédica, por escuchar y aportar continuamente sus puntos de vista.

A todos mi amigos y familiares por comprender y apoyarme incondicionalmente todo el tiempo necesario.

Gracias especiales a Luis Felipe y Leonardo por aguantar todos los desvelos, momentos de estrés, desesperación y frustración.

Gracias especiales a mis padres que hicieron de mí una persona con la capacidad de discernir el mundo.

Resumen

En el Centro de Investigación en Imagenología e Instrumentación Médica de la UAM-I se diseñó un nuevo sistema mínimamente invasivo para el monitoreo clínico de pacientes en estado crítico, que mide las características bioeléctricas (amplitud, fase, resistencia y reactancia) de la mucosa gástrica mediante la técnica de espectrometría de impedancia, para determinar el daño causado por isquemia. Durante la validación del sistema se detectó que un 60 % de las mediciones tuvieron errores por factores inherentes al ámbito clínico. La utilidad de la medición de la impedancia gástrica está en observar el comportamiento continuo de la variable, por lo que es importante evitar la pérdida de mediciones. A través de la identificación de los errores en tiempo real, el monitor podría enviar mensajes de aviso al personal de la terapia intensiva para tomar acciones inmediatas y recuperar la medición.

La fase experimental de este trabajo es parte de la etapa de validación del desarrollo del Espectrómetro de Impedancia Tisular Gástrico. La contribución de esta tesis fue desarrollar un algoritmo que detecta errores por: mala conexión, mala posición o mal contacto de la sonda, en la medición de la impedancia gástrica, para su implementación en la clínica. Se parte de una base de espectros de un estudio multicéntrico, la cual fue pre-clasificada por personal experto con base a evidencia clínica de 20908 registros x 92 mediciones espectrales y priorizada por tipo de error. La base de datos fue redimensionada dividiendo el rango espectral en cuatro bandas obteniendo de cada una su media y desviación estándar en amplitud, fase, resistencia y reactancia. Para la implementación del algoritmo de detección de errores se realizó un análisis de conglomerados por el método de K-medias para observar el comportamiento de la distribución natural de los datos y para determinar el número de clases se utilizó el valor promedio de la gráfica de silueta. Como resultado de lo anterior se obtuvo una nueva base de datos redimensionada y reclasificada, a partir de la cual se realizó la selección de características a través del algoritmo SFS (*Sequential Forward Selection*). Con las mejores características se diseñaron los clasificadores ocupando MSV (*Máquinas de Soporte Vectorial*) y éstos fueron evaluados con medidas de desempeño. En cada clasificador se obtiene una tasa de aciertos mayor al 95 % y una área bajo la curva ROC (A_z) de 0.99. Se realizó una prueba retrospectiva con dos bases de datos de espectros buenos clasificados por personal experto, no usados para el entrenamiento de los clasificadores, obteniendo un tasa de error de 2.3 % y 8.4 % respectivamente.

Índice general

Agradecimientos	I
Resumen	III
1 Introducción	9
2 Antecedentes	13
2.1 Estado de Choque y Mucosa Gástrica	13
2.2 Propiedades Eléctricas del tejido	15
2.2.1 Propiedades Eléctricas del tejido sano	16
2.3 Espectroscopía de Impedancia	18
2.3.1 Sistema de Espectroscopía de Impedancia	19
2.4 Reconocimiento de patrones	23
3 Identificación del Problema	27
3.1 Planteamiento del problema	27
3.2 Hipótesis	28
3.3 Objetivos	28
3.3.1 Objetivo General	28
3.3.2 Objetivos específicos	28
4 Metodología	29
4.1 Adquisición de datos	29
4.2 Pre-clasificación de datos	30
4.3 Pre-procesamiento de datos	30
4.4 Selección de Clases	33
4.4.1 Análisis de Conglomerados	33
4.4.2 Gráfica de Siluetas	34
4.5 Selección de características	36
4.6 Clasificación	37

4.7	Evaluación del desempeño	41
4.8	Resumen	42
5	Resultados	45
5.1	Selección de clases	46
5.2	Selección de características	51
5.3	Clasificación	51
5.3.1	Errores en las conexiones	55
5.3.2	Errores en la Posición	59
5.3.3	Errores en el Contacto	60
5.3.4	Espectros Buenos	63
5.4	Validación con otras bases de datos	64
5.5	Resultados Globales	67
6	Discusión	71
6.1	Selección de Clases	71
6.2	Selección de características	73
6.3	Clasificación	74
7	Conclusiones	77
7.1	Trabajo futuro	78
	Referencias Bibliográficas	84
	A Tipos de Errores	85
	B Conformación de M_3 y m_3	101
	C Características de Espectros por Banda	105

D Códigos

119

Índice de Tablas

3.1	Incidencia de errores por hospital.	27
4.1	Base de Datos pre-clasificada (M_1)	31
4.2	Base de datos agrupada por clase (m_1).	31
4.3	División de bandas y sus rangos.	32
4.4	Base de Datos reorganizada (M_2)	32
4.5	Base de datos agrupada por clase (m_2).	32
4.6	Ensayos propuestos para el análisis de conglomerados.	34
5.1	Ensayo 1 con M_2 con las variables: A, F, R, X (Media \pm DS)	47
5.2	Ensayo 2 con M_2 con las variables: A y F (Media \pm DS)	48
5.3	Ensayo 3 con M_2 con las variables: R y X (Media \pm DS)	48
5.4	Ensayo 3 con m_2 con las variables: R y X (Media \pm DS)	48
5.5	Ensayo 4 con m_2 con las variables: R y X (Media y DS)en las bandas 1 y 2	49
5.6	Ensayo 5 con m_2 con las variables: R y X (Media \pm DS)en las bandas 2 y 3	49
5.7	Ensayo 6 con m_2 con las variables: R y X (Media \pm DS)en las bandas 3 y 4	50
5.8	Características relevantes para cada clase de m_3	51
5.9	Parámetros y desempeño para Circuito Abierto	55
5.10	Parámetros y desempeño para Corto Circuito	57
5.11	Parámetros y desempeño para Sonda Descompuesta	58
5.12	Parámetros y desempeño para Mala Posición	59
5.13	Parámetros y desempeño para Mal Contacto	60
5.14	Parámetros y desempeño para Contenido Gástrico	61
5.15	Parámetros y desempeño para Contacto Pobre	62
B.1	Base de Datos reclasificada y redimensionada (M_3)	102
B.2	Base de Datos reclasificada y redimensionada (m_3)	103
B.3	Características de Base de datos (M_3 y m_3)	103

C.1	Espectros sin error $N = 2500$.	109
C.2	Cable Desconectado $N = 3443$.	110
C.3	Cable Descompuesto 1 $N = 958$.	110
C.4	Cable Descompuesto 2 $N = 2815$.	111
C.5	Sonda Descompuesta $N = 24$.	111
C.6	Cable Descompuesto 3 $N = 386$.	112
C.7	Cable Descompuesto 4 $N = 21$.	112
C.8	Cable Descompuesto 5 $N = 918$.	113
C.9	Boca-Esófago 5 $N = 3232$.	113
C.10	Contenido Gástrico 1 $N = 1897$.	114
C.11	Contenido Gástrico 2 $N = 698$.	114
C.12	Plano $N = 2500$.	115
C.13	Alto $N = 254$.	115
C.14	Aire $N = 7$.	116
C.15	Offset $N = 344$.	116
C.16	Altas Variaciones $N = 911$.	117

Índice de figuras

2.1	Patogénesis de falla orgánica múltiple	14
2.2	Progresión de daño isquémico en la pared intestinal.	16
2.3	Regiones de dispersión de los tejidos biológicos.	17
2.4	Membrana celular.	18
2.5	Membrana celular y su equivalente eléctrico.	18
2.6	Esquema paso de corriente en tejido.	20
2.7	Sistema de cuatro electrodos.	21
2.8	Espectrómetro de impedancia tisular gástrico (EITG).	21
2.9	Configuración de EITG para la realización de los estudios de validación clínica.	22
2.10	Etapas de un sistema de reconocimiento de patrones.	23
2.11	Esquema general de análisis de reconocimiento de patrones.	24
2.12	Técnicas de análisis de reconocimiento de patrones.	25
2.13	Esquema general de métodos de selección de características.	26
4.1	Metodología propuesta	29
4.2	Elementos para obtener la silueta, donde el dato i pertenece al conglomerado A.	35
4.3	Conjunto de datos Iris ejemplificando el uso de la gráfica de silueta.	36
4.4	Fronteras de decisión y vectores de soporte.	38
4.5	Parámetro ξ_i (tolerancia) en el error de clasificación.	39
4.6	Ejemplo de transformación de dominio con la función $\phi()$	40
4.7	Matriz de confusión de un clasificador binario.	41
4.8	Tipo de Espectros a identificar.	43
4.9	Metodología para la selección de mejores parámetros y mejor clasificador.	43
4.10	Diagrama general de la metodología propuesta	44
5.1	Diagrama de resultados.	45
5.2	Valor promedio de silueta en un barrido de k de 10 a 20 conglomerados	46

5.3	Gráfica de Silueta para $k = 14$ con la base de datos M_2	47
5.4	S_{RDS} vs. S_{XM} mostrando todas las clases a clasificar.	52
5.5	S_{RDS} vs. S_{XM} mostrando todas las clases con zoom.	53
5.6	S_{RDS} vs. S_{XM} de las clases CA, B, CC, MP y MC.	55
5.7	Clasificador lineal para CA	56
5.8	Clasificador lineal con un poco de zoom para CA	56
5.9	S_{RDS} vs. S_{XM} de las clases CC, B, MP y MC.	57
5.10	S_{RDS} vs. S_{XM} de las clases SoD y CD.	58
5.11	S_{RDS} vs. S_{XM} de las clases MP, B, y MC.	59
5.12	S_{RDS} vs. S_{XM} de las clases MC y B.	60
5.13	S_{RDS} vs. S_{XM} de las clases CG y Todos.	61
5.14	S_{RDS} vs. S_{XM} de las clases CP y AF.	62
5.15	S_{RDS} vs. S_{XM} de las clases Buenos y MC	63
5.16	Comparación de espectros buenos de dos diferentes bases de datos.	65
5.17	Comparación con espectros de INC en S_{RDS} vs. S_{XM}	65
5.18	Comparación con espectros de Multicéntrico en S_{RDS} vs. S_{XM}	66
5.19	Desempeños de nivel uno.	68
5.20	Desempeños de nivel dos.	69
A.1	Espectros sin error	85
A.2	Cable Desconectado	86
A.3	Cable Descompuesto 1	87
A.4	Sonda Descompuesta	88
A.5	Cable Descompuesto 2	89
A.6	Cable Descompuesto 3	90
A.7	Cable Descompuesto 4	91
A.8	Cable Descompuesto 5	92
A.9	Boca-Esófago	93
A.10	Contenido Gástrico 1	94
A.11	Contenido Gástrico 2	95
A.12	Plano	96

A.13 Alto	97
A.14 Aire	98
A.15 Offset	99
A.16 Altas Variaciones	100
C.1 R_M vs. muestras por banda	105
C.2 R_{DS} vs. muestras por banda	106
C.3 X_M vs. muestras por banda	106
C.4 X_{DS} vs. muestras por banda	107
C.5 $R_M vs X_M$	107
C.6 $R_M vs R_{DS}$	108
C.7 $R_{DS} vs X_{DS}$	108
C.8 $X_M vs X_{DS}$	109

Capítulo 1

Introducción

El estado de choque se define como aquella condición en la cual el suministro de oxígeno cae por debajo de la demanda energética de los tejidos del cuerpo. Se caracteriza por falta de oxígeno, consumo de oxígeno dependiente del suministro y metabolismo anaeróbico con sobreproducción de ácido láctico y bióxido de carbono. Se ha confirmado que el daño isquémico en la mucosa gastrointestinal es de relevancia en pacientes en estado de choque, ya que es el primer tejido afectado en la hipoxia sistémica por los mecanismos compensatorios del cuerpo [1], [2]. Se ha reportado que la isquemia en la mucosa gastrointestinal ocurre en más del 56 % de los pacientes en terapia intensiva y se ha relacionado con el 80 % de las muertes [3]. En el cuidado de los pacientes en estado crítico asegurar la suficiente oxigenación tisular es vital, ya que la hipoxia lleva al metabolismo anaeróbico, la acidosis celular y la muerte [4].

En otros grupos de investigación se han desarrollado y evaluado terapias enfocadas a evitar o eliminar los efectos negativos del daño isquémico en la mucosa, logrando reducir la mortalidad y morbilidad en este tipo de pacientes [5]; pero su efectividad depende del estado de la mucosa en el momento de la intervención [6], [7] y pueden llegar a tener un efecto negativo si se aplican fuera de tiempo. Además, ninguna de estas terapias permite hacer un diagnóstico del daño tisular, ni permite guiar efectivamente una terapia de resucitación [8]. Existen varios métodos de monitoreo para detectar hipoperfusión, hipoxia e isquemia local de la mucosa, como la tonometría gástrica [9], [3], [10], [11], [12] y la capnometría sublingual [13], [14], [15]. Ambas técnicas miden CO_2 como marcador metabólico; pero ninguna es sensible al daño de la mucosa, ni pueden estimar el nivel de daño causado por una isquemia prolongada.

La espectroscopia de impedancia puede obtener los componentes resistivos y reactivos del tejido a diferentes frecuencias, generando mediciones, que pueden ser aplicadas en la investigación médica como la información de la estructura eléctrica del tejido [16] y se ha reportado su uso en la medición de cambios en isquemia o perfusión de diferentes órganos [17], [18], [19], [20], [21], [22]; así como que es una técnica capaz de caracterizar tejidos vivos, mediante mediciones estáticas y la evolución de isquemia, mediante mediciones

dinámicas [20]. Este método ha sido utilizado para determinar niveles de daño tisular en diferentes órganos, en una aplicación intra-operatoria, continua y sin daño al tejido [17]. La importancia de monitorizar el estado de la mucosa gástrica cuando ha sufrido un daño irreversible debida a que el intestino pierde su función de barrera produciendo una traslocación de endotoxinas y bacterias a los diferentes órganos. Sin embargo, si el daño por isquemia en la mucosa gástrica es leve, se puede revertir restaurando el flujo sanguíneo y el suministro de oxígeno a los tejidos; pero si el daño por isquemia ha sido en un tiempo prolongado, el tejido puede estar muy dañado y en ese caso la terapia se debe enfocar a limitar el flujo sanguíneo para evitar el daño por reperfusión y la traslocación de endotoxinas y bacterias.

En el Centro de Investigación en Imagenología e Instrumentación Médica (CI^3M) de la UAM-Iztapalapa se desarrolló un método innovador mínimamente invasivo para determinar la condición de la mucosa gástrica a través de la medición de impedancia tisular gástrica en un rango de frecuencias de 215 Hz a 1 MHz [23]. Para realizar esta medición se modifica una sonda nasogástrica colocando cuatro electrodos en su parte distal, por los electrodos de los extremos se inyecta corriente y por los electrodos internos se adquiere el voltaje del tejido; lo cual permite la adquisición del espectro de impedancia de la mucosa gástrica. El equipo puede utilizarse para identificar y monitorizar de manera continua el nivel de daño tisular causado por isquemia [24], [25].

Esta investigación tiene raíces en la realización de un estudio en un modelo de choque hemorrágico en conejos, en donde se encontraron cambios en los espectros de impedancia asociados a tiempos de isquemia prolongados [26], continuando con el desarrollo de un modelo matemático de las características eléctricas de la mucosa gástrica sometida a espectroscopia de impedancia, para evaluar el efecto de la estimulación y el diseño de la sonda, en donde se demostró que la estimulación utilizada es segura y no causa daño al tejido [27]. En las pruebas pre-clínicas, se obtuvo una base de datos de espectros de impedancia en un modelo de isquemia intestinal ocluyendo la arteria mesentérica superior en cerdos, a partir de la cual se generó una clasificación de daño isquémico [28]. Adicionalmente, se realizó una validación fisiológica de las mediciones espectrales mediante imágenes de resonancia magnética en conejos, en donde se observó que el aumento de impedancia refleja disoxia tisular y no miden perfusión directamente [29]. Se realizó un estudio comparativo de una terapia guiada convencionalmente y una terapia guiada a partir de los datos suministrados por el espectrómetro de impedancia en cerdos [30]. En este estudio se obtuvieron mejores resultados de sobrevivencia en el grupo guiado mediante espectroscopia de impedancia, pero los datos no fueron estadísticamente significativos al compararlos con los resultados obtenidos mediante un tratamiento convencional. Con las pruebas clínicas se validó el uso del espectrómetro de impedancia tisular gástrico (EITG) en humanos, mediante la realización de tres estudios consecutivos. El primero, consistió en obtener espectros de la mucosa gástrica de voluntarios sanos para caracterizar los espectros en humanos, determinar si hay reproducibilidad de las mediciones bajo las mismas condiciones de registro

y evaluar el uso de la succión activa y de la alimentación enteral en la adquisición de espectros de impedancia. El segundo, consistió en la obtención de espectros de la mucosa gástrica en pacientes sometidos a cirugía cardiovascular electiva del Instituto Nacional de Cardiología Ignacio Chávez (INCICh) para caracterizar los cambios espectrales en pacientes con hipoperfusión gástrica evaluando el valor clínico de la información obtenida con el EITG; comparando la información espectral obtenida con variables hemodinámicas y de perfusión regional y evaluando su valor pronóstico. El tercero, consistió en la realización de un estudio en donde se evaluó la diferencia de impedancia de la mucosa gástrica entre pacientes críticos que sobreviven y fallecen, en diferentes terapias intensivas de la Ciudad de México. Se monitorizaron 172 pacientes de 7 Unidades de Terapia Intensiva donde se obtuvieron mediciones por un máximo de 3 días en donde se calculó la diferencia entre la medición de impedancia del paciente y los valores normales durante las primeras 12 h de monitoreo [31], sin embargo sólo se pudieron reportar 100 pacientes dada la incidencia de errores por factores inherentes al ámbito clínico como errores de conexión, posición y contacto. La medición de impedancia gástrica realmente útil es la de su tendencia en el tiempo ya que no es una variable que cambie rápidamente, por lo que evitar la pérdida de mediciones es crucial para el monitoreo del paciente en estado crítico.

El propósito de esta tesis fue diseñar un algoritmo que permita identificar los errores en tiempo real pudiendo evitar la pérdida de mediciones mandando mensajes al personal de la terapia intensiva para que tome acciones y se restablezcan las mediciones sin tener que llamar al soporte técnico para resolver el problema. Para la implementación del algoritmo de detección de errores se realizó un análisis de conglomerados por el método de K-medias para observar el comportamiento de la distribución natural de los datos y para determinar el número de clases se utilizó el valor promedio de la gráfica de silueta. Como resultado de lo anterior se obtuvo una nueva base de datos redimensionada y reclasificada, a partir de la cual se realizó la selección de características a través del algoritmo SFS (*Sequential Forward Selection*). Con las mejores características se diseñaron los clasificadores ocupando MSV (*Máquinas de Soporte Vectorial*) y éstos fueron evaluados con medidas de desempeño. En cada clasificador se obtiene una tasa de aciertos mayor al 95 % y una área bajo la curva ROC (A_z) de 0.99. Se realizó una prueba retrospectiva con dos bases de datos de espectros buenos, clasificados por personal experto, no usados para el entrenamiento de los clasificadores, obteniendo un tasa de error de 2.3 % y 8.4 % respectivamente. Actualmente esta investigación se encuentra en la fase de implementación del monitor clínico, el que todavía está por ser validado clínicamente para dar inicio a la etapa de transferencia para su comercialización; al implementar el algoritmo propuesto en esta tesis en el monitor clínico aumentará el valor agregado ya que quedará un sistema robusto para la detección de errores permitiendo su resolución inmediata.

Capítulo 2

Antecedentes

Para comprender de una mejor manera el problema a resolver, se presentan a continuación los conceptos fundamentales para esta investigación: entorno clínico de aplicación de la medición de espectroscopía de impedancia tisular gástrica, pruebas clínicas realizadas y herramientas de clasificación.

2.1. Estado de Choque y Mucosa Gástrica

El estado de choque se define como aquella condición en la cual el suministro de oxígeno cae por debajo de la demanda energética de los tejidos del cuerpo. Se caracteriza por falta de oxígeno, consumo de oxígeno dependiente del suministro y metabolismo anaeróbico con sobreproducción de ácido láctico y bióxido de carbono. El choque se divide en fase compensatoria, cuando la respuesta autonómica del individuo es capaz de sostener variables hemodinámicas como presión arterial media y gasto cardíaco a pesar de una situación patológica sostenida y fase no compensada, cuando la situación patológica ha sobrepasado la capacidad autonómica del individuo, e incluso puede superar los actos de reanimación del médico tratante, por lo cual conduce a una sub-fase final, llamada irreversible, cuando el paciente ha entrado en falla orgánica múltiple. La compensación (resucitación) solo se puede completar cuando toda evidencia de la mala oxigenación del tejido degradado se haya eliminado. En el paciente que presenta un estado de choque compensado, la presión sanguínea, producción de orina, perfusión cutánea y gases sanguíneos suelen ser normales. La insuficiente oxigenación del tejido durante el choque compensado se hace difícilmente evidente mediante un examen clínico, ya que se concentra en los órganos espláncnicos y especialmente en la mucosa intestinal. La importancia de detectar tempranamente el choque radica en la reducción de la mortalidad y/o costos de los hospitales [32]. La falla orgánica múltiple (ver fig. 2.1) es una causa importante de mortalidad en pacientes en estado crítico, su origen es multifactorial, pero la hipoperfusión esplácnica es de importancia en la patogénesis de esta condición [9]; también es consecuencia de la hipoxia tisular generada por un suministro inadecuado de oxígeno, aumentado por un daño microcirculatorio y por

un incremento en las demandas metabólicas tisulares [33]. Se ha confirmado que el daño isquémico en la mucosa gastrointestinal es de relevancia en pacientes en estado de choque, ya que es el primer tejido afectado en la hipoxia sistémica por los mecanismos compensatorios del cuerpo [1, 2]. Se ha reportado que la isquemia en la mucosa gastrointestinal ocurre en más del 56 % de los pacientes en terapia intensiva y se ha relacionado con el 80 % de las muertes [3]. En el cuidado de los pacientes en estado crítico asegurar la suficiente oxigenación tisular es vital, ya que la hipoxia lleva al metabolismo anaeróbico, la acidosis celular y la muerte [4].

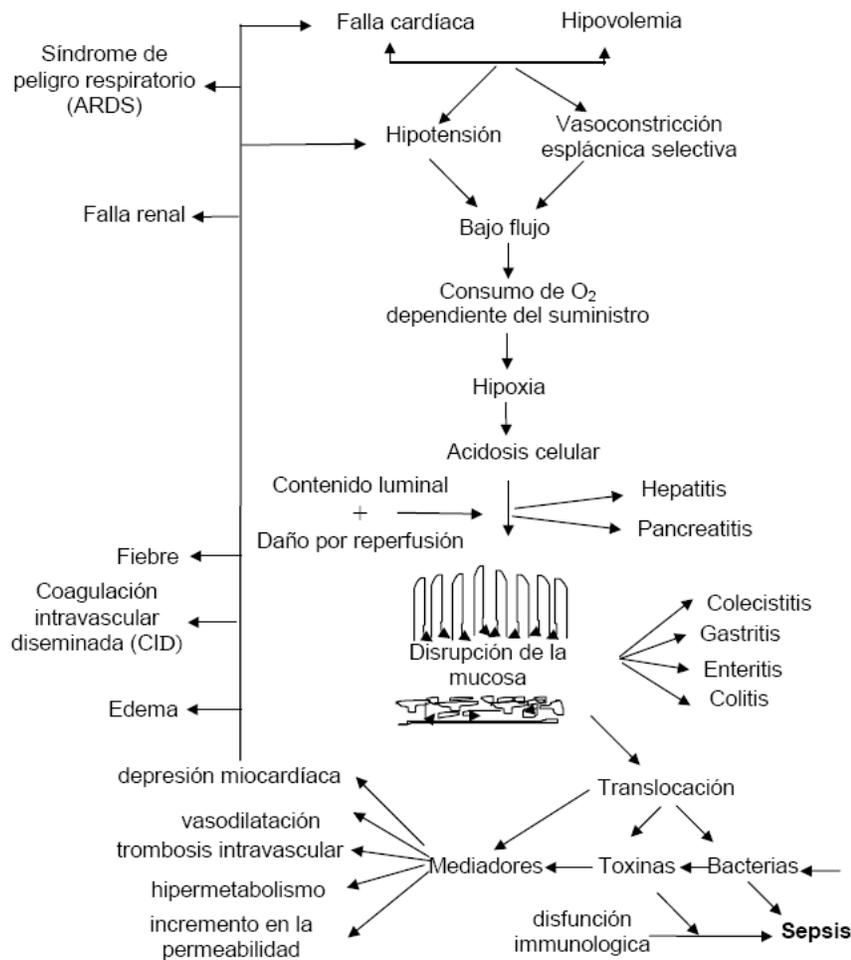


Figura 2.1: Patogénesis de falla orgánica múltiple y el papel motor de isquemia en la mucosa intestinal. Reproducido y traducido de [34].

En un tejido isquémico, hay una disminución de la disponibilidad de oxígeno y glucosa, así como una disminución del proceso de eliminación del CO_2 debido a un inadecuado

flujo sanguíneo. El daño tisular isquémico se caracteriza por un cambio a metabolismo anaeróbico, un consumo de O_2 limitado al suministro, evacuación celular insuficiente y reducción del volumen extracelular. La acumulación de ácido láctico y ácido carbónico (por exceso de CO_2) produce una caída en el pH. La isquemia produce una acumulación de iones y un aumento en la osmolaridad intracelular. Este incremento osmótico induce edema en el tejido. La isquemia también se caracteriza por un incremento en la permeabilidad de las membranas y reducción de los procesos activos de transporte. Después de un cierto periodo de isquemia en el cual se acumula la hipoxantina, el tejido se vuelve vulnerable al daño por reperfusión en el cual si se restaura el suministro de oxígeno se crean altas cantidades de radicales libres derivados del O_2 , agudizando aún más el deterioro tisular [19].

Cuando el daño celular ya es profundo, la mucosa permite primero el paso de endotoxinas y en particular lipopolisacáridos (*LPS*) y más tarde permite la translocación de bacterias enteras a la circulación linfática y portal, empeorando significativamente la condición del paciente. El daño isquémico afecta primero a las puntas de las vellosidades de la mucosa, y progresa gradualmente a capas cada vez más profundas de la pared intestinal (ver fig. 2.2). Cuando la base de la mucosa muere, el daño a la mucosa es irreversible y el paciente muere. El cuidado del paciente crítico en peligro de choque puede incluir una gran variedad de medidas preventivas y correctivas enfocadas a limitar o revertir el daño isquémico de la mucosa. Cuando el daño isquémico es aún leve, es importante limitar la isquemia en la mucosa con medidas de resucitación hemodinámica agresivas, vasodilatadores para aumentar la perfusión local, oxigenación intraluminal y reducción de las necesidades de O_2 eliminando focos de infección y enfriando el tejido. La autodigestión de la pared intestinal puede ser prevenida con el aumento del pH luminal y removiendo o inhibiendo proteasas pancreáticas. Una vez que el daño isquémico ha llegado a cierto punto es importante prevenir el daño adicional por reperfusión y radicales libres con eliminadores farmacológicos de radicales y limitar la perfusión (por ejemplo con inhibidores de *NO*). El peligro debido a endotoxinas y bacterias intestinales puede ser reducido por tratamiento de absorbentes de endotoxinas (Polimicina B), antibióticos, lavado intestinal y limpieza de focos de infección. También es importante mantener la capacidad de reparación de tejidos dañados con nutrición enteral adecuada. Finalmente, hay que reconocer el potencial letal de la mucosa dañada y muerta; y evitar los intentos de salvar tejidos de viabilidad dudosa, prefiriendo aislar y extirpar cuando sea posible [34].

2.2. Propiedades Eléctricas del tejido

La impedancia eléctrica en los tejidos cambia en función de la frecuencia a la cual se está estimulando. Al hacer un barrido en frecuencia se puede caracterizar la impedancia de los tejidos [36]. Cuando el valor de impedancia cambia rápidamente en un pequeño rango de frecuencias, se dice que existe una región de dispersión. En los tejidos biológicos

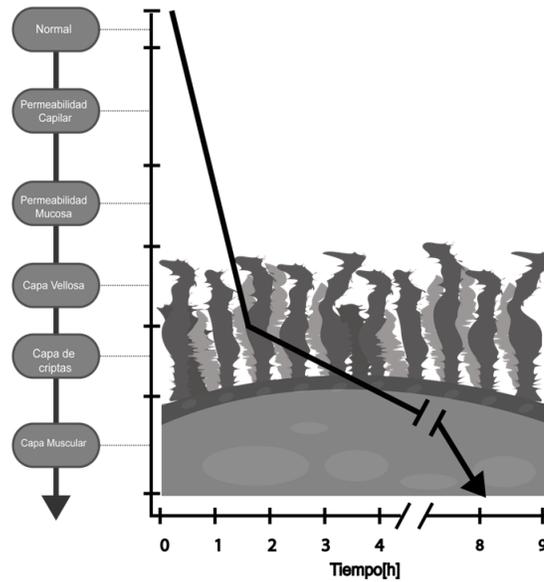


Figura 2.2: Progresión de daño isquémico en la pared intestinal después de una oclusión vascular total. Reproducido y traducido de [35].

debido a las propiedades eléctricas, existen 3 regiones de dispersión llamadas α , β y γ (ver figura 2.3).

- α – La primera región de dispersión se presenta a bajas frecuencias, generalmente entre 10 Hz y 1KHz. En esta región los cambios de impedancia son pequeños y se deben principalmente a procesos de difusión iónica a un lado de la membrana celular.
- β – La segunda región de dispersión se presenta entre 10KHz y 10 MHz. En esta región los cambios se relacionan principalmente con la polarización de la membrana celular, la cual actúa como barrera al flujo de iones entre los medio intra y extra celulares. Además, se produce una polarización de proteínas y macromoléculas orgánicas.
- γ – La tercera región de dispersión se presenta a altas frecuencias (mayores de 100 MHz). En esta región los cambios se asocian al contenido de agua, electrolitos y proteínas de los tejidos .

2.2.1. Propiedades Eléctricas del tejido sano

Existen algunas diferencias entre las propiedades eléctricas de los tejidos biológicos y las propiedades eléctricas de los circuitos eléctricos. La principal diferencia es que en los tejidos biológicos los iones son los principales portadores de carga mientras que en los circuitos eléctricos la corriente se genera por el movimiento de electrones. Por lo anterior,

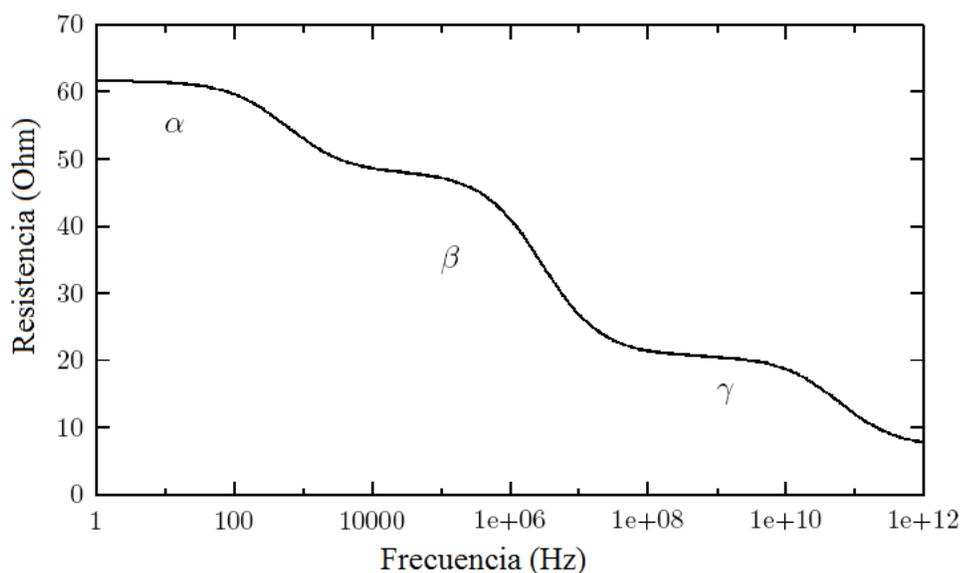


Figura 2.3: *Regiones de dispersión de los tejidos biológicos. Región α está por debajo de 1 KHz, región β entre 10 KHz y 10 MHz y la región γ por encima de 100MHz*

la conductividad de los tejidos biológicos es altamente dependiente de factores como la concentración, carga efectiva, coeficientes de difusión y del tipo de iones [37]. La linealidad de las propiedades de los tejidos se mantiene si la densidad de corriente inyectada es menor a $1\text{mA}/\text{cm}^2$ [38]. A pesar de las diferencias entre las propiedades eléctricas de los tejidos biológicos y las de los circuitos eléctricos, los modelos de circuitos pueden usarse para describir las propiedades eléctricas de los tejidos biológicos. Una de las formas de describir las propiedades eléctricas de un tejido biológico es a través de la impedancia compleja Z :

$$\begin{aligned}
 Z &= R + jX \\
 R &= \text{Re}\{Z\} = \text{resistencia} \\
 X &= \text{Im}\{Z\} = \text{reactancia} \\
 j &= \sqrt{-1}
 \end{aligned}
 \tag{2.1}$$

La impedancia de los tejidos biológicos es el resultado de la interacción de la corriente eléctrica con el tejido a nivel celular o molecular. El comportamiento eléctrico de un tejido puede modelarse mediante redes de circuitos RC.

El circuito equivalente se muestra en la figura 2.5, en donde la resistencia eléctrica del espacio extracelular se modela como una resistencia en paralelo con la resistencia y la capacitancia que modelan el espacio intracelular. Si se inyecta una corriente de excitación de alta frecuencia se genera un efecto de cortocircuito en la membrana celular, y la corriente

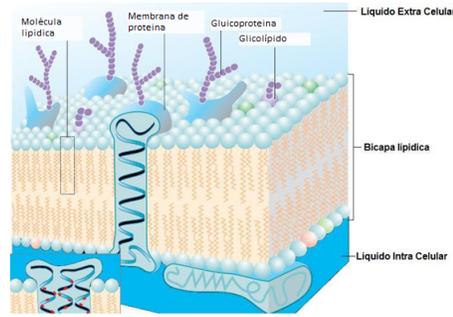


Figura 2.4: Membrana celular (traducido de [39]).

pasa por los medios extra e intracelulares. Si se inyecta una corriente de baja frecuencia, sólo atraviesa el medio extracelular y crece el valor de la impedancia a bajas frecuencias [20]. Teóricamente, la reactancia mide el volumen de la capacitancia de la membrana celular y es una medida indirecta del volumen intracelular.

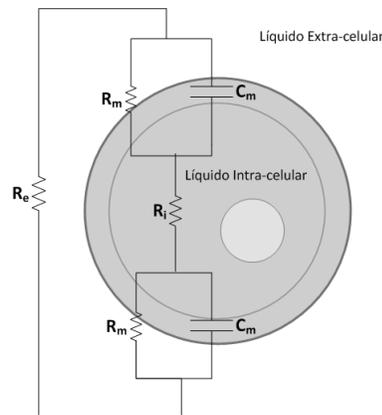


Figura 2.5: Membrana celular y su equivalente eléctrico.

2.3. Espectroscopía de Impedancia

La impedancia eléctrica está compuesta por dos componentes ortogonales: la resistencia (R), que es la oposición al paso del flujo de electrones; y la reactancia (X) que es la capacidad de almacenar energía. Su determinación en tejidos biológicos tienen una historia que se remonta a finales de siglo XVII, con los experimentos realizados por Galvani [40]. Las medidas de impedancia eléctrica determinan o infieren información acerca del tejido, siempre y cuando el evento en análisis presente un cambio en dimensión, en sus propiedades dieléctricas o en su conductividad [41]. No fue sino hasta principios del siglo XX cuando se estudió la estructura de los tejidos biológicos basados en sus propiedades

eléctricas pasivas, lo cual demostró que los tejidos biológicos son conductores y su resistencia variaba con la frecuencia [42].

La caracterización de tejidos sanos y patológicos basados en sus características eléctricas en un espectro de frecuencia determinado se denomina Espectroscopia de Impedancia Eléctrica (EIE), algunas aplicaciones se pueden encontrar en [43] [42] [18]. Las propiedades dieléctricas de las membranas celulares y sus interacciones con los electrolitos extra e intracelulares pueden medirse en el rango de 1kHz a 10 MHz [44]. Si analizamos los componentes a nivel celular el citoplasma celular y el espacio extracelular actúan como un medio conductivo aislado uno del otro por la membrana celular. La conductividad de los espacios extra e intracelulares contribuye a la resistencia promedio del tejido mientras la membrana celular contribuye al efecto capacitivo [16, 17]. Se puede hacer la medición de impedancia eléctrica en los tejidos vivos de una manera continua para determinar su evolución fisiopatológica. Algunas patologías como la isquemia, el infarto o la necrosis, implican alteraciones a nivel celular, las cuales son reflejadas en cambios de impedancia [21].

En la figura 2.6 se muestran dos escenarios de tejidos representando el paso de corriente a baja y alta frecuencia. Durante la condición de tejido normóxico, una cantidad significativa de corriente de baja frecuencia es capaz de fluir a través del espacio extracelular; cuando ocurre disoxia, las células no son capaces de generar suficiente energía para alimentar las bombas iónicas y el agua extracelular penetra al interior de la célula. Lo que implica que las células crecen e invaden el espacio extracelular y causan una reducción de la corriente en los fluidos extracelulares, causando un incremento en la impedancia a baja frecuencia [15]. En la célula también existen uniones de hendidura que al cerrarse contribuyen al aumento de la impedancia a esas bajas frecuencias. A altas frecuencias, los cambios de impedancia están influenciados por las impedancias de los fluidos intracelular y extracelular, y la permeabilidad iónica de las membranas celulares [17, 45].

2.3.1. Sistema de Espectroscopía de Impedancia

En el CI3M de la UAM-I en conjunto con la empresa Innovamedica se desarrolla un espectrómetro de impedancia para el monitoreo de la mucosa gástrica, el cual mide la impedancia compleja del tejido en un rango de 215 Hz a 1MHz, a través del método de demodulación paralela y conversión digital descrito por Othman y Sacristán [23]. Como parte del desarrollo, se diseñó una sonda nasogástrica/espectrométrica [24, 25] que permite la adquisición de la impedancia eléctrica del tejido a través de cuatro electrodos en su parte distal (ver fig. 2.7). El equipo cumple con estándares de regulación internacionales como la BS EN 60601-1:1990 y la ANSI/AAMI ES1:1993.

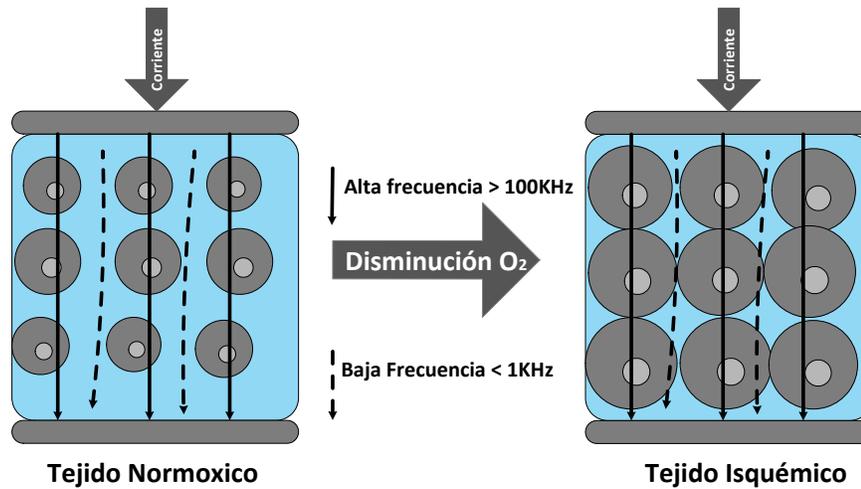


Figura 2.6: Representación esquemática del paso de la corriente a baja y alta frecuencia, y del aumento de impedancia a baja frecuencia debido al edema celular causado por isquemia.

Los electrodos exteriores inyectan una corriente de excitación al tejido en donde se genera un voltaje, el cual es medido por los electrodos interiores. La impedancia de entrada del sistema de medición de voltaje debe ser varios órdenes de magnitud mayor que la impedancia del tejido medido y que aquella de la interfase electrodo-tejido. Por lo tanto la corriente inyectada pasa a través del tejido y una proporción despreciable pasa a través de los electrodos interiores permitiendo mantener la distribución de corriente en el tejido. El voltaje de la interfase electrodo-tejido para los electrodos interiores es despreciable y el voltaje medido es igual al voltaje en el tejido obteniendo la impedancia del tejido únicamente. La sonda es tipo Salem y cuenta con dos lumen para alimentación o succión. El arreglo de electrodos en la parte distal de la sonda tiene un ancho de 1mm y la separación entre ellos también es de 1mm aproximadamente. La densidad de corriente de excitación debe estar en un rango específico menor a $1mA/cm^2$ para mantener la impedancia del tejido en un rango real. La densidad de corriente alta puede inducir potenciales de acción en las células del tejido los cuales pueden causar contracciones e interferir con la medición de la impedancia. Tanto la señal senoidal de corriente de salida como las frecuencias de muestreo son generadas por el reloj maestro del sistema bajo las mismas condiciones permitiendo aplicar una técnica conocida como muestreo sincronizado en el proceso de adquisición de datos. Esta técnica se basa en el principio de sincronizar la frecuencia de muestreo con la frecuencia de la señal de onda senoidal adquirida.

En la figura 2.8 se muestra el espectrómetro utilizado para la realización de tres estudios. El primero, consistió en obtener espectros de la mucosa gástrica de voluntarios sanos para caracterizar los espectros en humanos, determinar si hay reproducibilidad de las mediciones bajo las mismas condiciones de registro, y evaluar el uso de la succión activa y de la alimentación enteral en la adquisición de espectros de impedancia. El segun-

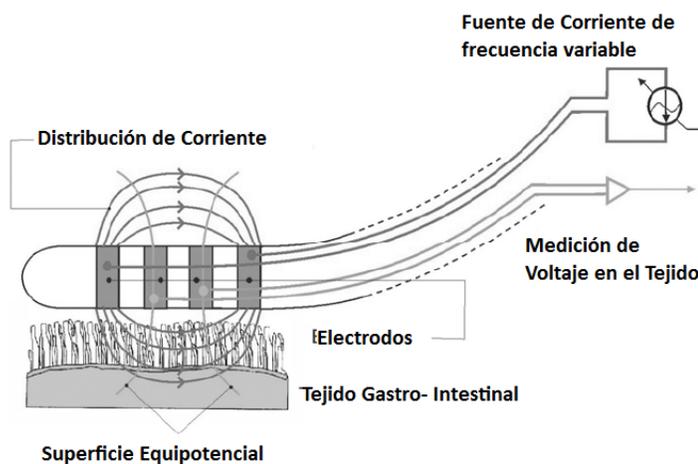


Figura 2.7: Sistema de cuatro electrodos para la medición de impedancia en tejido gástrico. La fuente de corriente produce una excitación senoidal constante en el tejido a través de los electrodos exteriores y la medición de voltaje por los dos interiores.

do, consistió en la obtención de espectros de la mucosa gástrica en pacientes sometidos a cirugía cardiovascular electiva del Instituto Nacional de Cardiología Ignacio Chávez (IN-CICh) para caracterizar los cambios espectrales en pacientes con hipoperfusión gástrica evaluando el valor clínico de la información obtenida con el EITG; comparando la información espectral obtenida con variables hemodinámicas y de perfusión regional y evaluando su valor pronóstico. El tercero, consistió en la realización de un estudio multicéntrico en donde se evalúa la diferencia de impedancia de la mucosa gástrica entre pacientes críticos que sobreviven y fallecen, en diferentes terapias intensivas de la Ciudad de México. En los tres estudios se colocó la sonda nasogástrica/espectrométrica en cada paciente y el sistema se conecto como se ilustra en figura 2.10.



Figura 2.8: Espectrómetro de impedancia tisular gástrico (EITG) y sonda de medición utilizados en la validación clínica.

Entre los tres estudios se monitorizaron 24 voluntarios; 82 pacientes de cirugía cardiovascular electiva de los cuales se excluyeron 5 pacientes por mal posicionamiento de la sonda nasogástrica/espectrométrica, 12 por mal contacto de la sonda con la pared gástrica, 7 por falta de mediciones de tonometría y 3 por mediciones incompletas; 172 pacientes en estado crítico en las 7 Unidades de Terapia Intensiva, donde se obtuvieron mediciones por un máximo de 3 días en donde se calculó la diferencia entre la medición de impedancia del paciente y los valores normales, durante las primeras 12 h de monitoreo [31], sin embargo sólo se pudieron reportar 100 pacientes dado a la incidencia de errores por factores inherentes al ámbito clínico como errores de conexión, posición y contacto. En el apéndice A se muestran espectros típicos de cada tipo de error encontrado por personal experto durante la adquisición del estudio multicéntrico. La base de datos generada por el INCICH se utiliza como base de datos para validar ya que las condiciones en la adquisición fueron controladas y no hubo tanta incidencia en los errores ya que se contaba con personal experto durante todo el experimento.

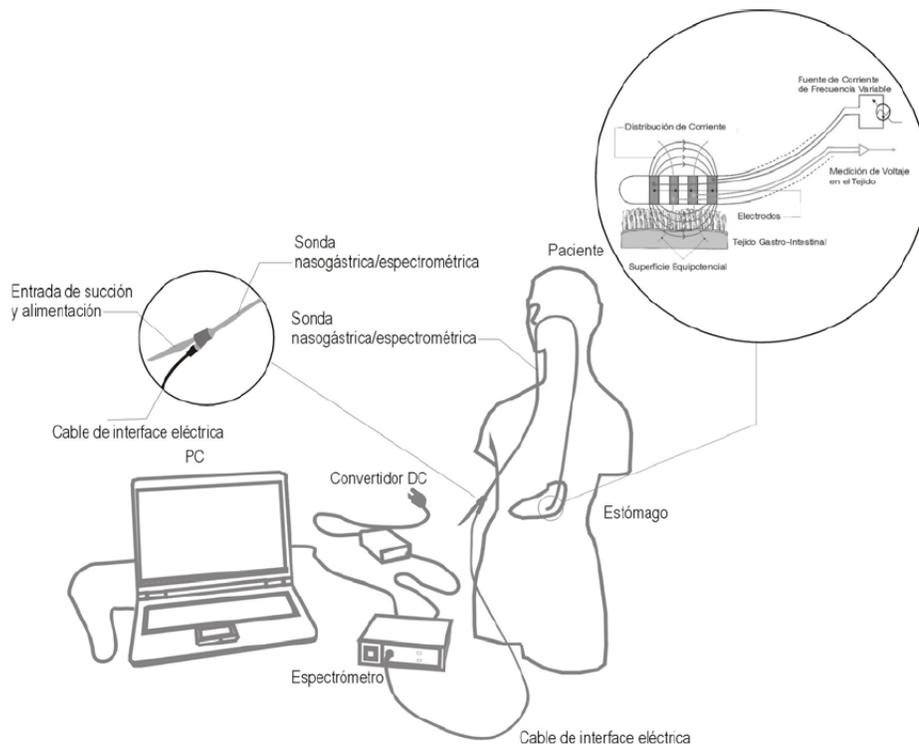


Figura 2.9: Configuración de EITG para la realización de los estudios de validación clínica.

2.4. Reconocimiento de patrones

El objetivo del procesamiento e interpretación de datos sensoriales es la descripción concisa y representativa del universo observado. La información de interés puede incluir nombres, características detalladas, relaciones, modos de comportamiento, etc. que involucran a los elementos del universo (objetos, fenómenos, conceptos). Estos elementos se perciben como patrones y los procesos que llevan a su comprensión son llamados procesos perceptuales. El etiquetado (clasificación, asignación de nombres) de esos elementos es lo que se conoce como reconocimiento de patrones. Por lo tanto, el reconocimiento de patrones es una herramienta esencial para la interpretación automática de datos sensoriales.

Los procesos perceptuales del ser humano pueden ser modelados como un sistema de tres estados:

- Adquisición de datos sensoriales.
- Extracción de características.
- Toma de decisiones.

Por lo tanto es conveniente dividir el problema del reconocimiento automático de una manera similar

- *Sensor*– Su propósito es proporcionar una representación válida de los elementos del universo a ser clasificados. Es un subsistema crucial ya que determina los límites en el rendimiento de todo el sistema. Idealmente se deberán entender completamente las propiedades físicas que distinguen a los elementos en las diferentes clases y usar ese conocimiento para diseñar el sensor, de manera que esas propiedades pudieran ser medidas directamente. En el presente trabajo el sensor es espectrómetro de impedancia gástrica.
- *Extracción de Características*– Esta etapa se encarga, a partir del patrón de representación, de extraer la información discriminadora eliminando la información redundante e irrelevante. Su principal propósito es reducir la dimensionalidad del problema de reconocimiento de patrones.
- *Clasificador*– Es la etapa de toma de decisiones en el sistema. Su rol es asignar los patrones de clase desconocida a la categoría apropiada .

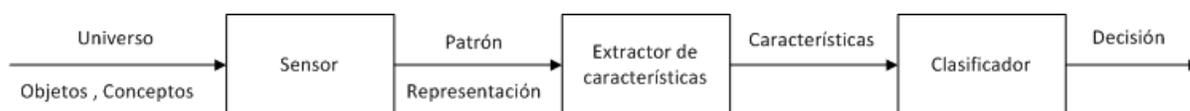


Figura 2.10: *Etapas de un sistema de reconocimiento de patrones.*

Si conociéramos totalmente las características estadísticas del modelo en el proceso de

generación de los patrones, esto es si supiéramos sus distribuciones y su prevalencia; entonces se puede diseñar el sistema de reconocimiento de patrones (SRP) óptimo mediante aplicación directa de la teoría de decisión de Bayes. Sin embargo en la práctica surgen los siguientes problemas: el modelo no se puede conocer totalmente y/o la complejidad del SRP a diseñar está restringida por consideraciones económicas (hardware, tiempo). Por lo general la base de conocimiento disponible para el diseño de un SRP es un conjunto de entrenamiento constituido por observaciones, ya sea etiquetadas o no. En el primer caso asumimos que para cada patrón o vector de observaciones $x_i, i = 1, 2, \dots, N$ en el conjunto de entrenamiento, un experto asigna una etiqueta con la clase correcta t_i . El diseño de un sistema basado en un conjunto de datos clasificados de antemano se conoce como aprendizaje supervisado. Si no se dispone de conocimiento experto sobre el conjunto de datos, o si el etiquetado de los patrones de entrenamiento es impracticable por razones prácticas; entonces el problema de diseño implica la necesidad de una primera etapa de análisis de los datos. Este proceso primario de análisis se conoce como una etapa de aprendizaje no supervisado [46, 47, 48]. Por lo tanto, en el caso general, dado un conjunto de entrenamientos el diseño del SRP implica:

1. Inferencia del modelo a partir de los datos (aprendizaje).
2. Desarrollo de reglas de decisión prácticas.
3. Simulación y evaluación del rendimiento del sistema.

En las figuras 2.11 y 2.12 se encuentra esquematizados de una manera muy general el análisis y técnicas de reconocimiento de patrones [47].

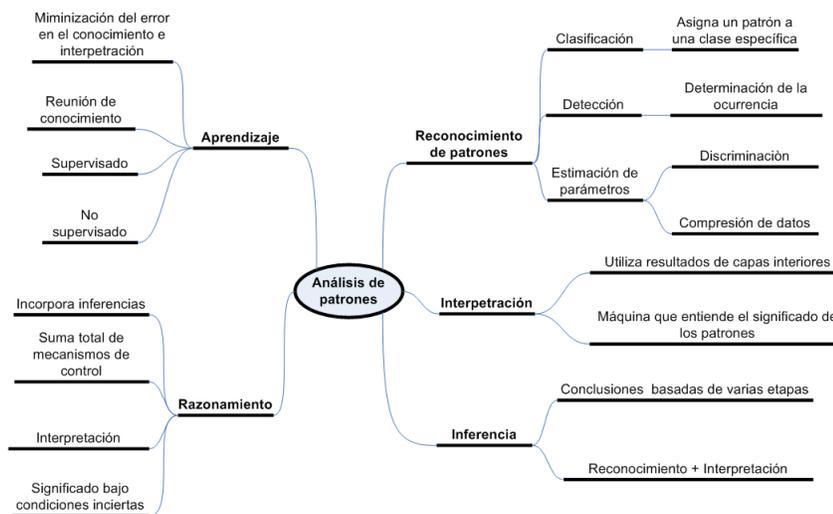


Figura 2.11: Esquema general de análisis de reconocimiento de patrones.

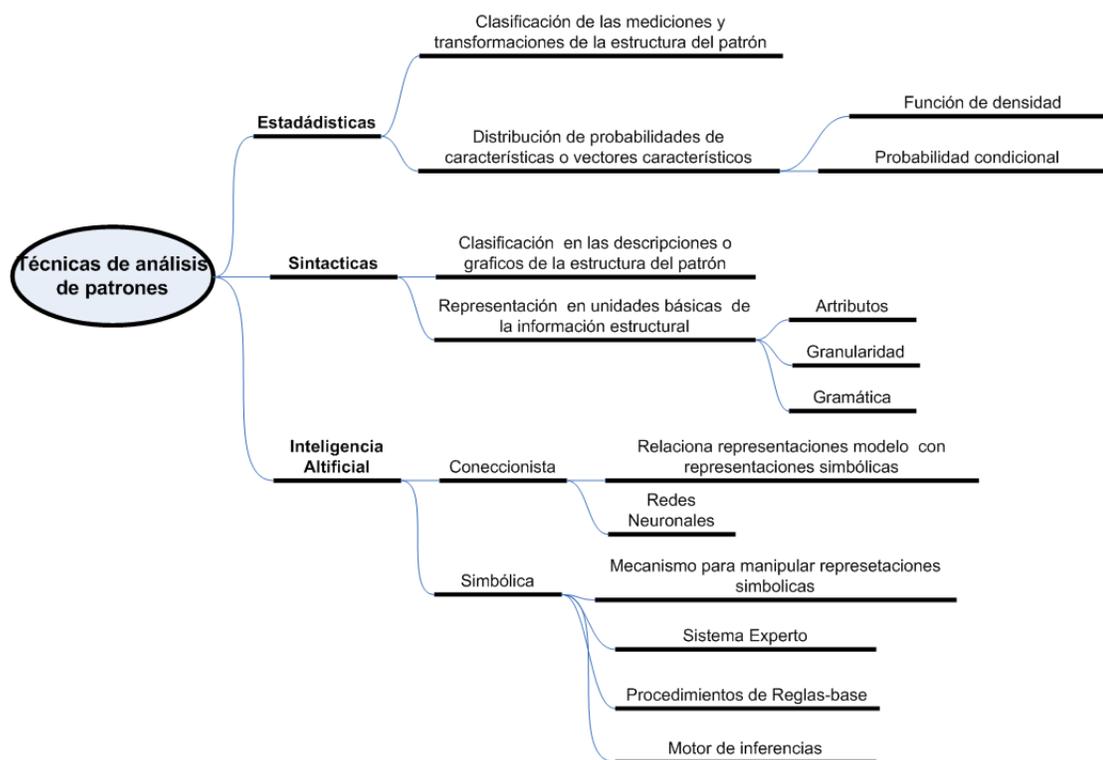


Figura 2.12: *Técnicas de análisis de reconocimiento de patrones.*

Idealmente, antes de diseñar un SRP debemos estudiar el proceso de formación de patrones. En teoría si entendemos las diferencias entre patrones de diferentes clases podremos representar cada una de las clases con un conjunto reducido de medidas que podrán ser usadas para el proceso de clasificación. En general, con muchas clases, este estudio no es factible. A lo sumo podemos arriesgar la hipótesis respecto a cuál es la información más relevante para ser incluida en la representación de los patrones. De esta manera la opción que nos queda es aplicar técnicas de selección de características [46, 49].

Comúnmente la representación de los patrones contiene muchos componentes que reflejan información redundante o irrelevante. En este contexto la selección de característica busca:

- Reducir la dimensionalidad de la representación.
- Minimizar los costos de extracción de las medidas.
- Evaluar el desempeño potencial del SRP.
- Mejorar el desempeño del SRP.

En la figura 2.13 se muestra de manera general un esquema de los métodos de selección de características.

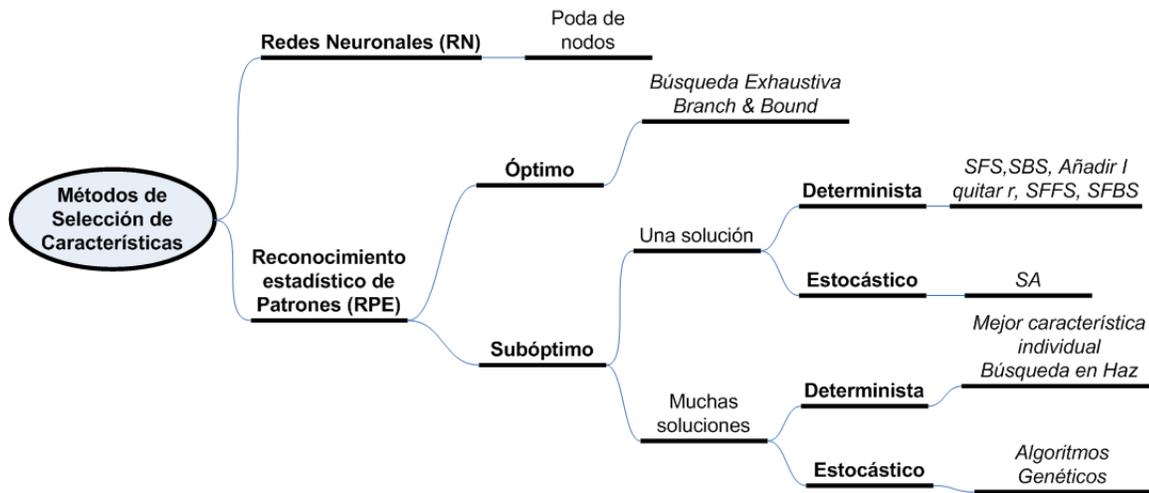


Figura 2.13: Esquema general de métodos de selección de características.

El EITG es un equipo de monitoreo para pacientes en estado crítico siendo su objetivo la obtención de espectros a través de obtener la respuesta en frecuencia ante un estímulo de 1 mA del tejido gástrico. Los espectros obtenidos reflejan el estado de la mucosa gástrica de los pacientes en estado crítico, es decir si el paciente presenta o no isquemia durante su estancia en las unidades de cuidados intensivos, por lo cual es importante tener una tendencia de los espectros obtenida por el sistema. En el estudio multicéntrico se evidenció que hay factores inherentes al ámbito clínico como errores de conexión, posición y contacto de la sonda y cables del EITG, los cuales provocan errores en la medición afectando la tendencia en el monitoreo del paciente. Que el EITG cuente con indicadores para poner en alerta al personal médico al cuidado del paciente permitirá corregir el error para no afectar la tendencia en el monitoreo de la mucosa gástrica, por lo que es importante contar con un algoritmo de detección de errores (ADE) de espectros que puedan clasificarse. El ADE propuesto pretende clasificar espectros buenos (medición esperada de la respuesta del tejido gástrico) y espectros con error (errores asociados a factores inherentes al ámbito clínico). Para desarrollar un algoritmo de clasificación se requiere: caracterizar los espectros buenos y a su vez caracterizar los espectros malos según su naturaleza (conexión, posición y contacto), extraer de información discriminadora eliminando la información redundante e irrelevante, diseñar clasificadores con base a las características relevantes y evaluar el desempeño de los clasificadores propuestos.

Capítulo 3

Identificación del Problema

3.1. Planteamiento del problema

Durante la realización del estudio multicéntrico en siete hospitales para la caracterización de espectros de impedancia gástrica en 172 pacientes en estado crítico, se obtuvo una base de datos de 272,692 espectros, sin embargo sólo el 41 % están asociados a la condición del paciente y el 59 % a errores por mala conexión, mala posición o mal contacto de la sonda. La incidencia de errores y número de mediciones que se obtuvieron por hospital se pueden observar en la tabla 3.1; debido a que la incidencia de error en todos los hospitales es mayor al 50 % se vuelve relevante la identificación de problemas asociados con el uso del Espectrómetro de Impedancia Tisular Gástrico (EITG) en condiciones clínicas. ¿Es posible detectar de una forma automática espectros con error? Esta tesis intenta dar respuesta a la pregunta anterior para ayudar a restablecer las mediciones dando aviso al personal de terapia intensiva para que puedan tomar acción de corrección de dicho evento y/o evitar la pérdida de mediciones del paciente en estado crítico.

Hospital	N	Buenos	Errores
H1	56183	43 %	57 %
H2	26055	41 %	59 %
H3	29568	42 %	58 %
H4	66540	31 %	69 %
H5	17385	47 %	53 %
H6	15810	43 %	57 %
H7	61151	48 %	52 %

Tabla 3.1: *Incidencia de errores por hospital.*

3.2. Hipótesis

Los espectros de impedancia tisular gástrica tienen características particulares en frecuencia, resistencia y/o reactancia que permiten distinguir los espectros buenos (respuesta esperada del tejido gástrico) de cada tipo de espectro con error (errores asociados a factores inherentes al ámbito clínico).

3.3. Objetivos

3.3.1. Objetivo General

- Desarrollar un algoritmo que permita la identificación de espectros buenos y espectros con error.

3.3.2. Objetivos específicos

- Caracterizar y clasificar los diversos tipos de errores de la medición de impedancia gástrica en humanos.
- Implementar un algoritmo de clasificación de espectros.
- Validar el clasificador retrospectivamente para medir el desempeño del algoritmo de clasificación.

Capítulo 4

Metodología

La fase experimental de este trabajo es parte de la etapa de validación del desarrollo del EITG, la contribución de esta tesis es el desarrollo de un algoritmo que detecte errores por: mala conexión, mala posición o mal contacto de la sonda, en la medición de la impedancia gástrica. En este capítulo se detallan los procedimientos hechos para la parte experimental. En el diagrama de la Figura 4.1 se muestra de manera general la metodología propuesta que se siguió para el desarrollo del algoritmo de detección de errores (ADE).

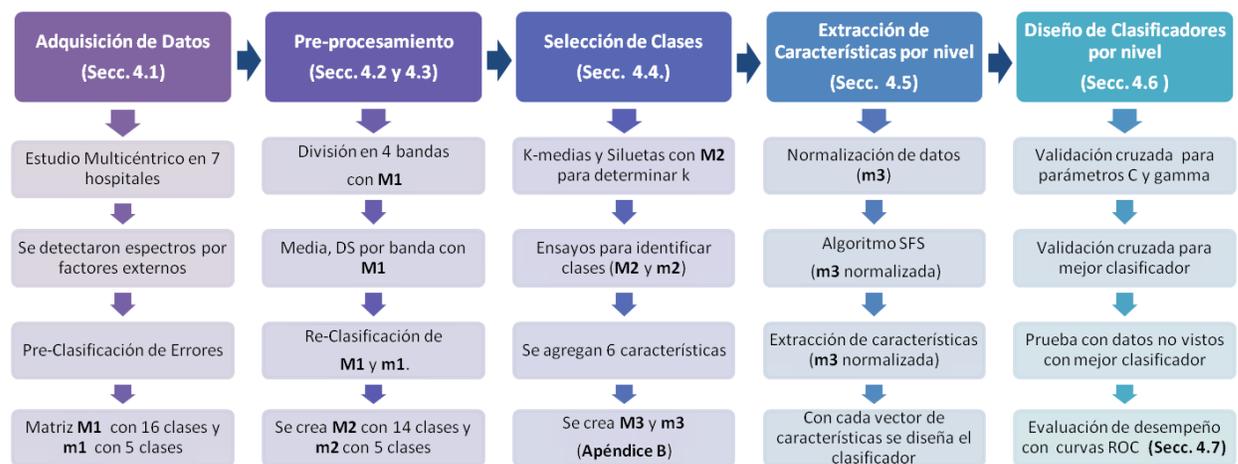


Figura 4.1: Metodología propuesta

4.1. Adquisición de datos

Se parte de una base de datos de espectros en la mucosa gástrica en Amplitud(A), Fase(F), Resistencia(R) y Reactancia(X) obtenida de un estudio multicéntrico que se realizó en siete hospitales en la Ciudad de México, en 172 pacientes en estado crítico. Las

mediciones se hicieron en un barrido de 23 frecuencias en el rango de 215 Hz a 1 MHz cada minuto durante varios días siguiendo el protocolo propuesto para la prueba clínica [31], teniendo un total de 92 mediciones espectrales entre las cuatro variables. A lo largo del estudio se encontró que se presentaban errores de medición y sus fuentes fueron identificadas y anotadas en los registros de cada paciente.

4.2. Pre-clasificación de datos

Personal experto clasificó cada espectro de acuerdo a la evidencia clínica dando origen a una base de datos pre-clasificada de la siguiente manera: los espectros que no tienen error o buenos como error cero, los errores 1-7 se agrupan en la clase conexiones y representan problemas con los cables o con la sonda, los errores 8-10 se agrupan en la clase de mal posicionamiento de la sonda y los errores 11-14 y se agrupan en la clase de mal contacto con la mucosa y como error 15 a la clase de movimiento que producen altas variaciones en los espectros. Cada error tiene asociada una prioridad siendo la más alta uno y la más baja cuatro. Esta prioridad es asignada por el evento que representa, es decir la prioridad cuatro asociada a un artefacto de movimiento es por que supone que: el cable está bien conectado y en buen estado, la sonda está en el estómago y que tiene buen contacto con la mucosa gástrica. Desde esta perspectiva, los datos se agrupan según se resume en la tabla 4.1, representado la base de datos que se identificará en este documento como M_1 (20908 registros x 92 mediciones espectrales); agrupando por clase tenemos la tabla 4.2 en cuyo caso la base de datos será identificada como m_1 (20908 registros x 92 mediciones espectrales).

4.3. Pre-procesamiento de datos

Como primer paso se hace una reducción en la dimensionalidad de los datos, esto se hace dividiendo el espectro en 4 bandas como se muestra en la tabla 4.3, (de los espectros y su morfología y promedio por banda de cada espectro se explica a detalle en el apéndice A). Se reducen las características calculando por banda su media y desviación estándar (DS), obteniendo cuatro promedios con sus desviaciones para cada variable A, F, R y X quedando un total de 32 mediciones. Es necesario mencionar que al observar las características morfológicas del error 6 (N=21) y compararlas con las del error 1 (N=3443) se reconoce que son muy similares, por lo que se fusionan en el error 1 (N=3464). El error 13 con el 9 también comparten características morfológicas por lo que se fusionan en el error 9 (N=1904). Desde este nuevo enfoque los datos se agrupan en 13 tipos de errores según se resume en la tabla 4.4, representando la base de datos que se identificará en este documento como M_2 (20908 registros x 32 mediciones espectrales); agrupando por clase tenemos la

No. Error	Descripción	N	Prioridad	Clase
0	Buenos	2500	-	Bueno
1	Cable desconectado	3443	1	Conexiones
2	Cable descompuesto 1	958	1	Conexiones
3	Cable descompuesto 2	2815	1	Conexiones
4	Sonda descompuesta	24	1	Conexiones
5	Cable descompuesto 3	386	1	Conexiones
6	Cable descompuesto 4	21	1	Conexiones
7	Cable descompuesto 5	918	1	Conexiones
8	Boca-Esófago	3232	2	Posición
9	Contenido Gástrico 1	1897	2	Posición
10	Contenido Gástrico 2	698	2	Posición
11	Plano	2500	3	Contacto
12	Alto	254	3	Contacto
13	Aire	7	3	Contacto
14	Offset	344	3	Contacto
15	Altas Variaciones	911	4	Movimiento

Tabla 4.1: Base de Datos pre-clasificada (M_1)

Clase	Descripción	N
0	Bueno	2500
1	Errores Conexión	8565
2	Mala Posición	5827
3	Mal Contacto	3105
4	Artefacto de Movimiento	911
	Total	20908

Tabla 4.2: Base de datos agrupada por clase (m_1).

tabla 4.5 en cuyo caso la base de datos será identificada como m_2 (20908 registros x 32 mediciones espectrales).

Banda	Rango de Frecuencias
1	215 Hz - 1 kHz
2	1.46 kHz - 10 kHz
3	14.6 kHz - 100 kHz
4	146 kHz - 1 MHz

Tabla 4.3: División de bandas y sus rangos.

No. Error	Descripción	N	Prioridad	Clase
0	Buenos	2500	-	Bueno
1	Cable desconectado	3464	1	Conexiones
2	Cable descompuesto 1	958	1	Conexiones
3	Cable descompuesto 2	2815	1	Conexiones
4	Sonda descompuesta	24	1	Conexiones
5	Cable descompuesto 3	386	1	Conexiones
6	Cable descompuesto 5	918	1	Conexiones
7	Boca Esófago	3232	2	Posición
8	Contenido Gástrico 1	1904	2	Posición
9	Contenido Gástrico 2	698	2	Posición
10	Plano	2500	3	Contacto
11	Alto	254	3	Contacto
12	Offset	344	3	Contacto
13	Altas Variaciones	911	4	Movimiento

Tabla 4.4: Base de Datos reorganizada (M_2)

Clase	Descripción	N
0	Bueno	2500
1	Errores Conexión	8565
2	Mala Posición	5834
3	Mal Contacto	3098
4	Artefacto de Movimiento	911
	Total	20908

Tabla 4.5: Base de datos agrupada por clase (m_2).

4.4. Selección de Clases

En esta sección se detallan los procedimientos que se utilizaron para seleccionar las clases que mejor separan cada tipo de espectro.

4.4.1. Análisis de Conglomerados

El análisis por conglomerados es el proceso de agrupar datos en clases o conglomerados de tal forma que los objetos de un conglomerado tengan una similaridad alta entre ellos y baja (sean muy diferentes) con objetos de otros conglomerados. Algunas de sus características son:

- Escalabilidad: normalmente corren con pocos datos.
- Conglomerados de formas arbitrarias: basados en distancias numéricas tienden a encontrar conglomerados esféricos.
- Capacidad de manejar diferentes tipos de atributos: numéricos, binarios, nominales, ordinales, etc..
- Capacidad de añadir restricciones.
- Manejo del ruido: muchos son sensibles a datos erróneos.
- Puede funcionar eficientemente con alta dimensionalidad.
- Requerimientos mínimos para especificar parámetros, como el número de conglomerados.
- Independientes del orden de los datos.
- Que los conglomerados sean interpretables y utilizables.

En esta tesis se ocupó el algoritmo K-medias [46] ya implementado en el programa MATLAB sobre la base de datos de menor dimensionalidad m_2 y M_2 para realizar el análisis de conglomerados. Este algoritmo sigue un procedimiento simple de clasificación de un conjunto de objetos en un determinado número de K conglomerados, K es determinada a priori. El nombre de K-medias viene porque representa cada uno de los conglomerados por la media (o media ponderada) de sus puntos, es decir, por su centroide. La representación mediante centroides tiene la ventaja de que tiene un significado gráfico y estadístico inmediato. Cada conglomerado por tanto es caracterizado por su centro o centroide que se encuentra en el centro o el medio de los elementos que componen el conglomerado.

El algoritmo K-medias es un algoritmo para poner N datos de un espacio I-dimensional a un espacio de K conglomerados. Cada conglomerado es parametrizado por un vector $m^{(k)}$ llamadas medias. Los datos pueden ser denotados por $x^{(n)}$ donde el superíndice n va

desde 1 hasta N . Cada vector x tiene I componentes. Supongamos que el espacio en donde esta definida x es un espacio real y que tiene definida una distancia entre los puntos, por ejemplo la distancia euclidiana.

Para inicializar este algoritmo, las K medias $m^{(k)}$ son inicializadas de alguna forma por ejemplo aleatoriamente y éste es un algoritmo iterativo de dos pasos: asignación y actualización. En paso de asignación, cada punto n es asignado a la media más cercana y en la actualización cada vez que se reasigna una observación a un conglomerado distinto del que lo contenía se recalculan las medias del conglomerado que pierde el dato y del que lo recibe.

En este trabajo se propone realizar seis ensayos (tabla 4.6) empleando al algoritmo de K-medias en los cuales se hacen variaciones en la clasificación de la base de datos, número de variables y bandas. Esto para poder observar en cada caso la distribución natural de los datos. Los ensayos 1-3 se realizan con M_2 y los ensayos 3-6 con m_2 .

Ensayo	Variables	Bandas	Características	Tipo
1	A,F,R,X	1 a 4	32	Media y DS
2	A,F	1 a 4	16	Media y DS
3	R,X	1 a 4	16	Media y DS
4	R,X	1 y 2	8	Media y DS
5	R,X	2 y 3	8	Media y DS
6	R,X	3 y 4	8	Media y DS

Tabla 4.6: *Ensayos propuestos para el análisis de conglomerados.*

4.4.2. Gráfica de Siluetas

Como se mencionó en la sección anterior para el uso del algoritmo de K-medias es necesario determinar a priori los K conglomerados, pero a veces esto resulta un método de prueba y error, como una herramienta adicional se ocupan las gráficas de silueta [50] para evaluar cuán bien han sido asignados los datos en los conglomerados.

La silueta del dato i se define como:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4.1)$$

Donde

- $a(i) = d(i, A)$ que es el promedio de las distancias del dato i a todos los elementos del conglomerado A . En la figura 4.2 es el promedio de todas las longitudes de líneas dentro de A .

- Sea C cualquier otro conglomerado diferente de A , $C \neq A$.
- $d(i, C)$ es el promedio de las distancias del dato i a todos los elementos del conglomerado C . En la figura 4.2 es el promedio de todas las longitudes de i a C .
- $b(i) = \min_{C \neq A} d(i, C)$.
- El conglomerado B que alcanza el mínimo, es decir, tal que $d(i, B) = b(i)$ se denomina vecindad del dato i . El conglomerado B es el segundo mejor conglomerado para el dato i .
- $b(i) - a(i)$ es la distancia de los promedios. Entre más alejado este el conglomerado B del A el objeto será mejor clasificado.

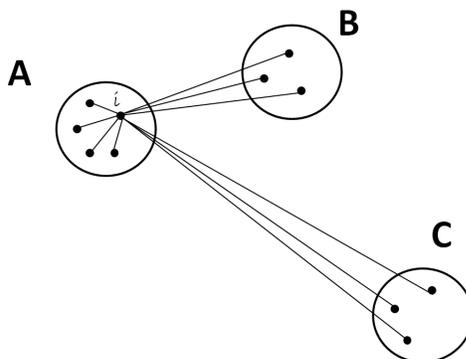


Figura 4.2: Elementos para obtener la silueta, donde el dato i pertenece al conglomerado A .

Hay que observar que $s(i)$ esta definida entre $-1 \leq S(i) \leq 1$. Esto lo podemos interpretar como:

- $s(i) \approx 1$, el dato i está bien clasificado.
- $s(i) \approx 0$, el dato i está entre dos conglomerados.
- $s(i) \approx -1$, el dato i está mal clasificado.

En la figura 4.3 tenemos un ejemplo de una gráfica de silueta con una $k = 3$ con datos del conjunto Iris que es un pequeño grupo de datos clásicos en el área de reconocimiento de patrones. Los datos de este conjunto corresponden a un problema académico tradicional, introducido por *Fisher* en 1936, consistente en la clasificación de flores de la familia iris (liliáceas) con base a las dimensiones de sus pétalos y sépalos. El conjunto de datos consta de 150 ejemplares de tres subclases; Setosa, Versicolor y Virgínica. Este conjunto se utiliza frecuentemente como ejemplo para comparar las prestaciones y posibilidades de distintos métodos.

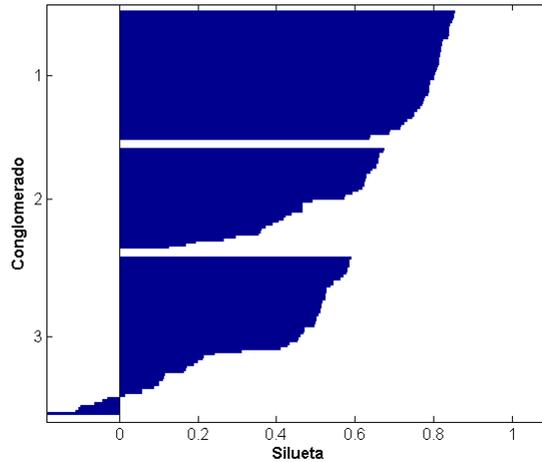


Figura 4.3: *Conjunto de datos Iris ejemplificando el uso de la gráfica de silueta.*

Para establecer la mejor k se propone hacer un barrido de 10 a 20 en M_2 . Se obtendrá la silueta promedio para determinar cuál será el mejor valor de k . Tomando en cuenta que se tiene una base de datos pre-clasificada se realiza el cálculo de las incidencias de cada error por grupo, pudiéndose determinar porcentajes de agrupamiento por el tipo de error. Con esto puede establecerse si la pre-clasificación de los datos tiene alguna relación con la agrupación natural de los datos.

4.5. Selección de características

Para esta etapa es conveniente hacer una normalización de los datos removiendo la media y dividiendo entre la DS cada característica, guardando su valor medio y desviación para la implementación final del algoritmo. Con la base de datos normalizada, el problema de la selección de características se puede definir de la siguiente manera: dado un grupo de características candidatas Y de dimensión n , elegir un subconjunto X de dimensión d , tal que $X \subseteq Y$ que se comporte de forma óptima al ser utilizado con un determinado clasificador. Entre de los diversos tipos de algoritmos de selección, se eligió un método secuencial determinístico de una única solución del cual existen dos variantes: SFS (*Sequential Forward Selection*) y SBS (*Sequential Backward Selection*) [51] [52].

Los métodos SFS parten de la mejor característica individual (aquella que posee el mayor grado de discriminación con base a un criterio establecido) y van añadiendo una característica cada vez. Los métodos SBS parten del conjunto completo de características Y

y van eliminando una característica en cada iteración. Tanto uno como otro no garantizan el subconjunto óptimo ya que no evalúan todos los subconjuntos. Los resultados que se pueden obtener son similares, aunque el SFS es más rápido que el SBS al trabajar con conjuntos más reducidos de características. Para este trabajo se utilizó la herramienta *Rapid Miner* [53] para obtener las características por el método SFS. Se seleccionó un subconjunto aleatorio de 1000 datos por clase para optimizar el tiempo de procesamiento.

4.6. Clasificación

Para la identificación de cada tipo de error, se propone una solución jerárquica en donde se va considerando la clase de mayor prioridad vs. todas las demás, teniendo por cada nivel un problema binario de clasificación a resolver. En cada nivel se utilizaron Máquinas de Soporte Vectorial (MSV o SVM por su nombre en inglés *Support Vector Machines*).

Las MSV mapean los datos de entrada a un espacio de características de una dimensión mayor (por ejemplo, si los datos de entrada están en \mathbb{R}^2 estos son mapeados por la MSV a \mathbb{R}^3) encontrando un hiperplano que separe dos clases, etiquetadas $y_i \in \{-1, +1\}$, de forma que la distancia entre el hiperplano óptimo y el patrón de entrenamiento más cercano — *margen* — sea máxima, con la intención de maximizar la generalización de la máquina de aprendizaje [54][55]. Maximizar el margen es un problema de programación cuadrática y puede ser resuelto en su forma dual introduciendo multiplicadores de Lagrange. La MSV encuentra el hiperplano óptimo utilizando el producto punto en el espacio transformado, pero evaluándolo mediante funciones en el espacio de características que son llamadas *núcleos*. La solución del hiperplano óptimo puede ser escrita como la combinación lineal de productos interiores con unos pocos puntos de entrada que son llamados *vectores de soporte* como se puede apreciar en la Fig. (4.4).

Considerando que se tiene un conjunto de datos para entrenamiento M , cada punto de entrenamiento $x_i \in \mathbb{R}^N$ pertenece a alguna de dos clases $y_i \in \{-1, 1\}$ para $i = 1, \dots, l$; donde l es el número total de datos para entrenamiento. La búsqueda del hiperplano adecuado en un espacio de entrada es demasiado restrictivo por lo que una solución es mapear el espacio de entrada en un espacio de características de una dimensión mayor y buscar el hiperplano que separe optimamente las clases. Siendo x el vector en el espacio de características y $z = \phi(x)$ un mapeo ϕ de \mathbb{R}^N a un espacio de características Z , deseamos encontrar el hiperplano

$$\mathbf{w}^T z + b = 0 \tag{4.2}$$

Definido por el par (\mathbf{w}, b) , tal que se pueda separar el punto x_i de acuerdo a la función

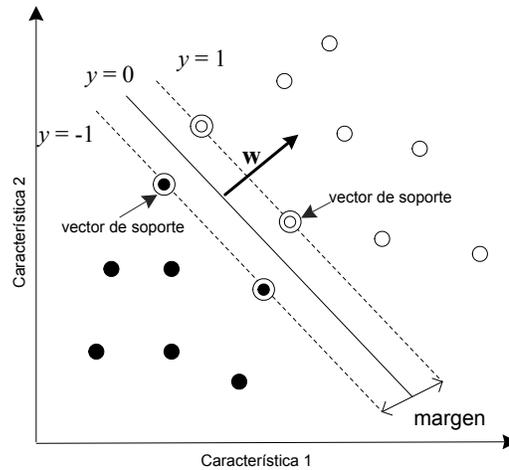


Figura 4.4: Fronteras de decisión y vectores de soporte.

$$f(x_i) = \text{sign}(\mathbf{w}^T z_i + b) = \begin{cases} 1 & y_i = 1 \\ -1 & y_i = -1 \end{cases} \quad (4.3)$$

Para el caso linealmente separable de M , se puede encontrar un único hiperplano óptimo, para el cual, el margen entre las proyecciones de los puntos de entrenamiento de dos diferentes clases es maximizado. Pero si M no es linealmente separable hay que considerar una tolerancia $\xi_i \geq 0$ de tal modo que (4.3) es modificada a

$$y_i (\mathbf{w}^T z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l. \quad (4.4)$$

Los $\xi_i \neq 0$ en (4.4), son aquellos para los cuales el punto x_i no satisface (4.3). Entonces el término $\sum_{i=1}^l \xi_i$ puede ser tomado como algún tipo de medida del error en la clasificación. El problema del hiperplano óptimo se puede escribir como la solución del problema

$$\min \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \right) \quad (4.5)$$

$$y_i (\mathbf{w}^T z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l. \quad \xi_i \geq 0, \quad i = 1, \dots, l. \quad (4.6)$$

Donde C es una constante libre de regularización también llamada *costo*, el ajuste de ésta puede hacer un balance entre la maximización del margen y la violación a la clasificación. Es decir carga el valor de la equivocación hacia un sentido o hacia el otro.

Buscar el hiperplano óptimo en (4.6) es un problema de programación cuadrática que puede ser resuelto construyendo el Lagrangiano y transformándolo en el dual

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j z_i \cdot z_j \quad (4.7)$$

siendo

$$\sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad (4.8)$$

Donde $\alpha = (\alpha_1, \dots, \alpha_l)$ es un vector de multiplicadores de Lagrange positivos asociados con las constantes en (4.4).

De acuerdo al teorema de Kuhn-Tucker [56] la solución de $\bar{\alpha}_i$ del problema (4.7) satisface:

$$\bar{\alpha}_i ((\bar{\mathbf{w}}^T z_i + \bar{b}) - 1 + \bar{\xi}_i) = 0, \quad i = 1, \dots, l \quad (4.9)$$

$$(C - \bar{\alpha}_i) \bar{\xi}_i = 0, \quad i = 1, \dots, l \quad (4.10)$$

De esta igualdad se deduce que los únicos valores $\bar{\alpha}_i \neq 0$ (4.10) son satisfechos con el signo de igualdad. El punto x_i correspondiente con $\bar{\alpha}_i > 0$ es llamado *vector de soporte*. En el caso $0 < \bar{\alpha}_i < C$, el correspondiente vector de soporte x_i satisface a las igualdades $y_i(\bar{\mathbf{w}}^T z_i + \bar{b}) = 1$ y $\bar{\xi}_i = 0$. En el caso $\bar{\alpha}_i = C$, el correspondiente $\bar{\xi}_i$ es diferente de cero y el correspondiente vector de soporte x_i no satisface (4.3). El punto x_i correspondiente con $\bar{\alpha}_i = 0$ es clasificado correctamente y esta claramente alejado del margen de la decisión ver Fig.4.5

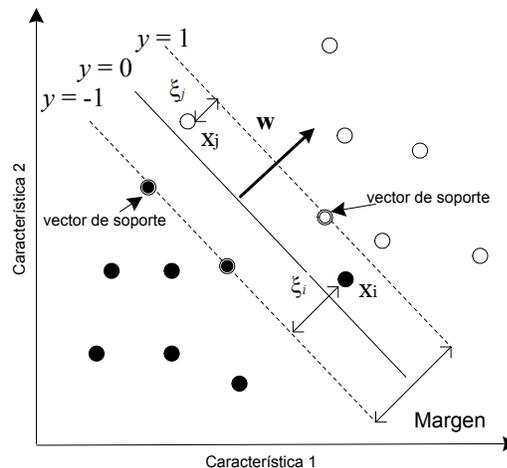


Figura 4.5: Parámetro ξ_i (tolerancia) en el error de clasificación.

Para construir el hiperplano óptimo $\bar{\mathbf{w}}^T z + b$, se utiliza

$$\bar{\mathbf{w}} = \sum_{i=1}^l \bar{\alpha}_i y_i z_i \quad (4.11)$$

El escalar b puede ser determinado de las condiciones de Kuhn-Tucker (4.10). La función de decisión generalizada de (4.3) y (4.11) es tal que

$$f(x) = \text{sign}(\mathbf{w}^T z + b) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i z_i \cdot z + b\right) \quad (4.12)$$

Sin embargo no es posible determinar a priori, una función ϕ que sea óptima por lo que el tratamiento para éstos casos está determinado con ensayos de funciones ϕ conocidas. Por ejemplo:

Lineal:

$$K(x_i, x_j) = x_i \cdot x_j \quad (4.13)$$

Polinomial:

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \quad \gamma > 0, r, d \in \mathbb{R} \quad (4.14)$$

Funciones de Base Radial:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\gamma^2}} \quad (4.15)$$

La figura 4.6 muestra un ejemplo de núcleo polinomial sobre un conjunto de datos dado [46].

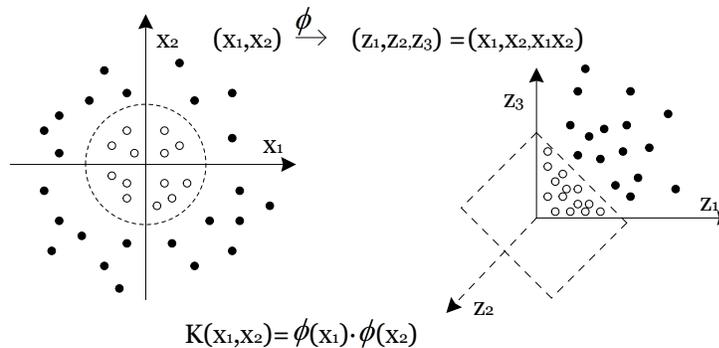


Figura 4.6: Ejemplo de transformación de dominio con la función $\phi()$.

		Clase Verdadera	
		+	-
Clase Predicha	+	Verdadero Positivo <i>clasificado correctamente</i>	Falso Positivo <i>clasificado incorrectamente</i>
	-	Falso Negativo <i>clasificado incorrectamente</i>	Verdadero Negativo <i>clasificado correctamente</i>

Figura 4.7: Matriz de confusión de un clasificador binario.

La herramienta de acceso libre para el diseño de MSV que se utilizó en este trabajo, llamada *LIBSVM* [57], permite modificar los parámetros de costo, diferentes tipos de núcleo, para clasificación y regresión. Se puede encontrar información detallada de su implementación [58], lo mismo que una guía práctica para la clasificación con MSV [59]. Tanto las guías y la herramienta se encuentran disponibles en [57].

4.7. Evaluación del desempeño

Las Curvas Relativas de Operación (CRO o ROC por su nombre en inglés *Receiver Operating Characteristics*), fueron introducidas por primera vez a la comunidad del procesamiento de señales para poder evaluar la capacidad de que un operador humano pudiera distinguir entre señales de radar y ruido. Pero han sido también de gran utilidad en el ámbito médico para la toma de decisiones de pruebas diagnósticas. Consideremos un problema de predicción de dos clases o clasificación binaria, en la que los resultados se etiquetan como: positivos (+) o negativos (-). Si el dato tiene etiqueta positiva y fue clasificado como positivo es contado como *verdadero positivo* (VP) si es clasificado como negativo se cuenta como *falso negativo* (FN). Si el dato original tiene etiqueta negativa y fue clasificado como negativo se cuenta como un *verdadero negativo* (VN), si es clasificado como positivo es contado como *falso positivo* (FP). Estas cuatro medidas se pueden organizar en una matriz de confusión como la que se muestra en la figura 4.7. Los números en la diagonal mayor representan las decisiones hechas correctamente [60].

Esta matriz de confusión pueden calcularse medidas de desempeño como la tasa de Verdaderos Positivos (TVP) o sensibilidad de un clasificador, la tasa de falsos positivos

(TFP) o falsa alarma, 1-especificidad, certeza y precisión.

$$\text{sensibilidad} = TVP = \frac{VP}{VP+FN} \quad (4.16)$$

$$\text{falsa alarma} = TFP = \frac{FP}{VN+FP} \quad (4.17)$$

$$1 - \text{especificidad} = \frac{VN}{FP+VN} = 1 - TFP \quad (4.18)$$

$$\text{certeza} = \frac{VP+VN}{VP+FN+VN+FP} \quad (4.19)$$

$$\text{precisión} = \frac{VP}{VP+FP} \quad (4.20)$$

Para obtener la CRO basta con hacer la gráfica TVP vs. TFP. La sensibilidad mide hasta qué punto un clasificador es capaz de clasificar los casos positivos correctamente y TFP o 1-especificidad nos dice cuántos resultados positivos son incorrectos. La curva CRO revela la capacidad que el clasificador tiene para discriminar las dos clases. El área A_z bajo la curva CRO es una *cifra de mérito* de esta capacidad : $0,5 \leq A_z \leq 1$. Por ejemplo si $A_z = 0,5$ corresponde al caso de una clasificación aleatoria, ya que tiene la misma probabilidad de clasificarlo de una clase u otra. Si $A_z = 1$ entonces la clasificación es perfecta.

4.8. Resumen

Siguiendo con los objetivos de esta tesis para identificar cada tipo de espectro el primer paso es implementar el análisis de conglomerados a la matriz M_2 y m_2 para identificar qué características separan naturalmente a los datos complementando con la gráficas de siluetas que nos ayudarán a elegir la partición más adecuada. Este análisis brinda información relevante acerca de las características que pueden agrupar de una manera natural a cada tipo de espectro pudiendo reducir la dimensionalidad de los datos o encontrar nuevas agrupaciones o fusiones de características.

La base de datos M_2 después de este paso tuvo cambios importantes en cuanto a la clasificación de los espectros y la dimensionalidad de los datos, lo cual se detalla en el apéndice B. Las nuevas bases de datos M_3 y m_3 son de menor dimensionalidad y los tipos de espectros fueron reetiquetados con base a sus características encontradas como se

muestra en la figura 4.8.



Figura 4.8: Tipo de Espectros a identificar.

Una vez teniendo esta nueva matriz se normaliza para poder pasar a la etapa de selección de características, esto se hace de una manera jerárquica tomando por nivel un subconjunto aleatorio de mil datos con todos sus rasgos. El algoritmo SFS indicará qué características son la de mayor relevancia en cada caso.

El siguiente paso es diseñar el clasificador con MSV en cada etapa; los núcleos que se proponen son el lineal y el gaussiano, recordando que se mencionó en la sección 4.6 que hay varios parámetros que gobiernan a las MSV que son el *costo* y γ para optimizarlos se realiza una validación cruzada de cinco vías con datos de entrenamiento [46]. Al hacer esto se elige el *costo* y la γ que tenga la mejor certeza con los datos de entrenamiento. Definidos los mejores parámetros, hay que encontrar el clasificador en donde se obtenga el mejor desempeño por lo que otra vez se recurre a la validación cruzada de cinco vías obteniendo en cada iteración las medidas de desempeño de las CRO así como A_z como se puede observar en las figura 4.9.

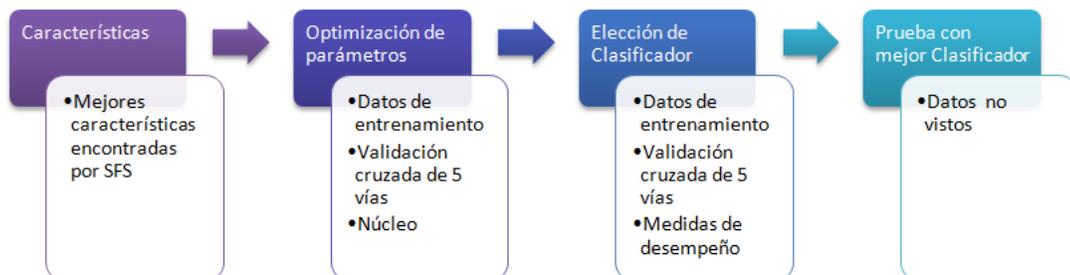


Figura 4.9: Metodología para la selección de mejores parámetros y mejor clasificador.

Finalmente se obtiene el desempeño promedio, lo que nos ayudará a decidir entre los paquetes de características cuál es el mejor. Se selecciona el clasificador que tuvo mejor desempeño en la validación cruzada y con este se hace la prueba con datos no vistos para evaluar el desempeño del clasificador. Como criterio de selección adicional, se considera la obtención del mejor desempeño con el menor número de características.

Una vez definido el clasificador por cada nivel y subnivel es posible implementar el algoritmo para la identificación de errores. En la figura 4.10 se muestra a manera de resumen general la metodología propuesta en esta tesis.



Figura 4.10: Diagrama general de la metodología propuesta

Capítulo 5

Resultados

En esta sección se describen de una forma detallada los resultados de la selección de clases, selección características, diseño de clasificadores y validación con otras bases de datos (ver figura 5.1). En el análisis de selección tuvo como resultado la matriz M3 y m3 (detalle en apéndice B) una vez definidas las mejores características se diseñan los clasificadores y se eligen los que tienen el mejor para desempeño del primer y segundo nivel considerando la prioridad de cada tipo de error (ver figura 4.8). Resuelto el primer nivel (Circuito Abierto-CA, Corto Circuito-CC, Mala Posición-MP, Mal Contacto-MC), se diseñan los clasificadores para el segundo nivel (Sonda Descompuesta-SoD, Contenido Gástrico-CG, Contacto Pobre-CP, Artefacto de Movimiento-AM) quedando completo el algoritmo de detección de errores. Para la validación del algoritmo se realizaron pruebas retrospectivas con datos no vistos para medir el desempeño del algoritmo de clasificación(ADE).

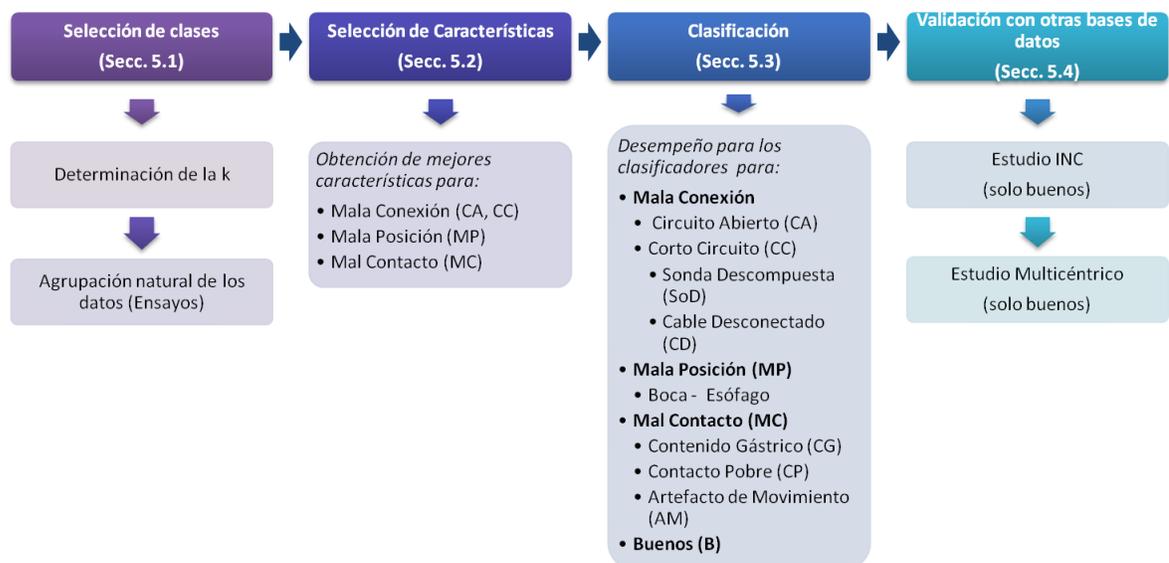


Figura 5.1: Diagrama de resultados.

5.1. Selección de clases

En esta sección se muestran los resultados por el análisis de conglomerados en un conjunto con las gráficas de silueta; con base en los hallazgos encontrados de éste análisis se generó una nueva base de datos que denominaremos en este documento M_3 y m_3 (el detalle de la formación de éstas se encuentra en el apéndice B) es decir, se reclasificó y redimensionó. En la figura 5.2 se muestran los valores promedio de la silueta obtenida haciendo el barrido de k de 10 a 20 conglomerados ocupando M_2 ; podemos observar que el máximo valor promedio de la gráfica se obtiene en $k=14$ que coincide con las clasificación de M_2 . En la gráfica de silueta que se muestra en la figura 5.3 se pueden observar cómo se agrupan los datos en cada conglomerado.

Pero esta gráfica no nos proporciona información de la distribución de los tipos de errores por conglomerado, por lo que se realizan los ensayos con $k = 14$ mostrados en la tabla 4.6. En cada ensayo se podrá observar cómo se están distribuyendo los errores en cada conglomerado con esto se tiene una visión general de la forma natural en la que los errores se agrupan.

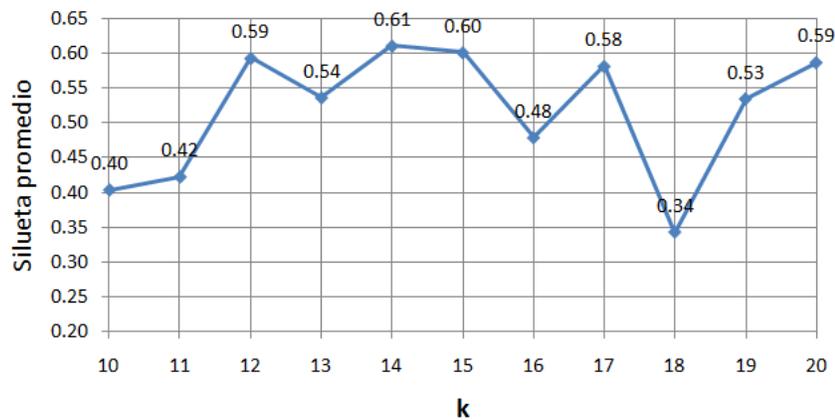


Figura 5.2: Valor promedio de silueta en un barrido de k de 10 a 20 conglomerados

En la tabla 5.1 se muestran los resultados obtenidos para el ensayo 1 con M_2 en donde se prueban las variables: A, F, R, y X con su media \pm DS respectivamente en las cuatro bandas. En la tabla 5.2 se muestran los resultados del ensayo 2 con M_2 para las variables A y F con su media y DS respectivamente en las cuatro bandas. En la tabla 5.3 se muestran los resultados para el ensayo 3 con M_2 para las variables: R y X con su media y DS respectivamente en las cuatro bandas.

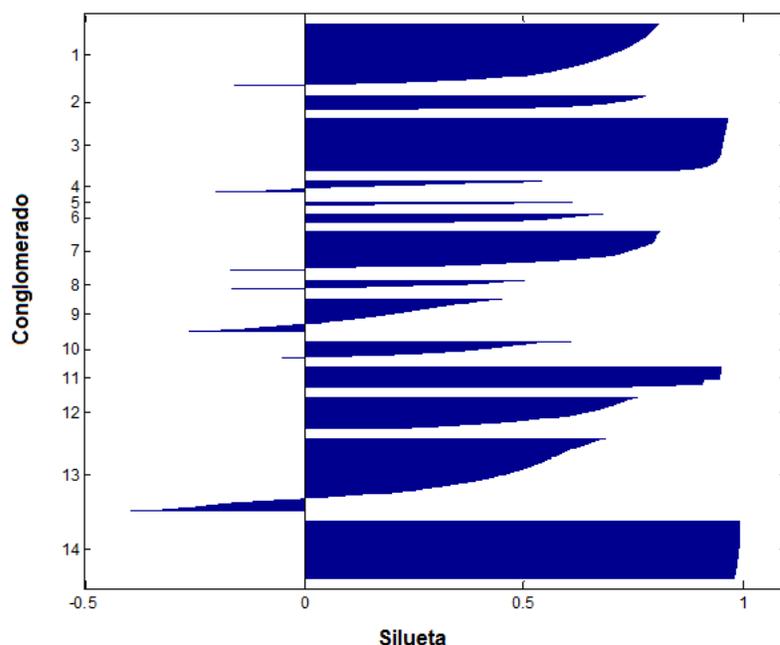


Figura 5.3: Gráfica de Silueta para $k = 14$ con la base de datos M_2

En la tabla 5.4 se muestran los resultados para el ensayo 3 con m_2 para las variables R y X con media y DS respectivamente en las cuatro bandas. Los resultados para el ensayo 4 se muestran en la tabla 5.5 en donde se evalúan variables R y X con su media \pm DS respectivamente ocupando m_2 en las bandas 1 y 2. Los resultados obtenidos para el ensayo 5 se muestran en la tabla 5.6 en donde se evalúan las variables R y X con media \pm DS respectivamente en las bandas 2 y 3 con m_2 . En la tabla 5.7 se muestran los resultados para el ensayo 6 donde se evalúan R y X con media \pm DS respectivamente en las bandas 3 y 4 con m_2 .

Grupo	Error													
	e0	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13
1	99.8 %							1.6%	3.3%	9.9%		7.9%	59.9 %	5.2 %
2		0.1 %						21.9 %						0.2 %
3	0.2 %								0.6 %		100.0 %			0.8 %
4		6.5 %												
5														
6		0.3 %					100.0 %							0.3 %
7														0.2 %
8		2.3 %							91.5 %	17.2 %				
9		46.3 %												
10														90.2 %
11			100.0 %		100.0 %									0.3 %
12		44.6 %						0.3 %						
13														
14				100.0 %					76.5 %	4.6 %	72.9 %		92.1 %	40.1 %
Clase	0	1	1	1	1	1	1	2	2	2	3	3	3	4
N	2500	3464	958	2815	24	386	918	3232	1904	698	2500	254	344	911

Tabla 5.1: Ensayo 1 con M_2 con las variables: A, F, R, X (Media \pm DS)

Grupo \ Error	e0	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13
1		46.8 %					0.4 %							
2							0.2 %		89.1 %	14.1 %				0.2 %
3				56.3 %										
4		0.4 %				100.0 %	1.6 %		0.1 %	0.3 %				0.3 %
5		0.1 %					0.1 %	21.9 %						0.3 %
6	4.6 %								0.5 %		100.0 %			0.8 %
7	95.4 %							0.7 %	2.4 %	5.7 %		7.9 %	59.9 %	88.7 %
8			100.0 %		100.0 %									5.8 %
9								14.8 %	7.9 %	79.6 %		91.7 %	40.1 %	3.0 %
10		8.2 %					78.1 %							0.2 %
11				43.7 %										0.7 %
12								62.6 %		0.3 %		0.4 %		
13		0.1 %					19.2 %							
14		44.5 %					0.3 %							
Clase	0	1	1	1	1	1	1	2	2	2	3	3	3	4
N	2500	3464	958	2815	24	386	918	3232	1904	698	2500	254	344	911

Tabla 5.2: Ensayo 2 con M_2 con las variables: A y F (Media \pm DS)

Grupo \ Error	e0	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13
1		0.1 %						22.0 %						0.2 %
2	99.8 %							1.7 %	3.4 %	9.9 %		7.9 %	59.9 %	5.2 %
3				43.7 %										0.2 %
4														2.4 %
5		6.9 %					40.0 %							
6							16.6 %		83.3 %	6.9 %				0.2 %
7		90.5 %					0.4 %							
8				23.0 %										
9														90.7 %
10				33.4 %										
11		0.3 %				100.0 %	0.3 %							0.3 %
12	0.2 %				100.0 %				0.6 %		100.0 %			0.8 %
13		2.3 %												
14			100.0 %				42.7 %							
Clase	0	1	1	1	1	1	1	2	2	2	3	3	3	4
N	2500	3464	958	2815	24	386	918	3232	1904	698	2500	254	344	911

Tabla 5.3: Ensayo 3 con M_2 con las variables: R y X (Media \pm DS)

Grupo \ Clase	C0	C1	C2	C3	C4
1	100.00 %	0.28 %	1.83 %	87.99 %	4.94 %
2		11.19 %			
3		6.80 %			
4		4.48 %			
5			58.07 %	12.01 %	3.18 %
6		32.87 %			0.33 %
7		13.64 %			
8		2.01 %	27.82 %		0.22 %
9		0.02 %	12.17 %		0.11 %
10		10.50 %			
11			0.10 %		90.78 %
12		5.59 %			0.44 %
13		3.43 %			
14		9.20 %			
N	2500	8565	5827	3105	911

Tabla 5.4: Ensayo 3 con m_2 con las variables: R y X (Media \pm DS). C_0 - Buenos, C_1 - Errores de Conexión, C_2 - Errores de Posición, C_3 - Errores de Contacto, C_4 -Artefactos de Movimiento

Grupo \ Clase	C0	C1	C2	C3	C4					
1	99.96 %	4.68 %	81.78 %	9.07 %	0.33 %					
2		0.04 %			6.04 %					
3		0.68 %			2.20 %					
4		18.52 %			89.24 %					
5						0.11 %				
6							13.49 %			
7								1.65 %		
8		7.01 %								
9					0.04 %	11.47 %			0.36 %	80.70 %
10					4.94 %	12.10 %	0.22 %			
11					1.32 %					
12		5.80 %								
13		14.35 %			0.33 %					
14		16.08 %								
N	2500	8565	5827	3105	911					

Tabla 5.5: *Ensayo 4 con m_2 con las variables: R y X (Media y DS) en las bandas 1 y 2. C0- Buenos, C1- Errores de Conexión, C2- Errores de Posición, C3-Errores de Contacto, C4-Artefactos de Movimiento*

Grupo \ Clase	C0	C1	C2	C3	C4			
1	100.00 %	0.30 %	2.18 %	91.99 %	5.05 %			
2		11.19 %	0.22 %					
3		9.92 %						
4		2.55 %						
5		0.04 %			30.07 %	7.97 %	3.73 %	
6		32.87 %			0.22 %			
7		12.25 %						
8		2.08 %				34.68 %	0.03 %	1.65 %
9		5.03 %				0.44 %		
10		8.90 %						
11		0.01 %			0.02 %		88.69 %	
12		33.06 %						
13		3.25 %						
14		11.63 %						
N	2500	8565		5827	3105	911		

Tabla 5.6: *Ensayo 5 con m_2 con las variables: R y X (Media \pm DS) en las bandas 2 y 3. C0- Buenos, C1- Errores de Conexión, C2- Errores de Posición, C3-Errores de Contacto, C4-Artefactos de Movimiento*

Grupo \ Clase	C0	C1	C2	C3	C4
1		0.63 %	22.90 %		
2		4.40 %			
3		1.16 %			
4		2.28 %	13.08 %	0.03 %	1.98 %
5	100.00 %	0.42 %	62.72 %	99.97 %	8.34 %
6		2.36 %			
7		0.08 %	0.36 %		88.58 %
8		4.54 %			0.11 %
9		14.30 %			0.11 %
10		18.56 %			
11		2.01 %			
12		1.95 %			
13		11.19 %			0.88 %
14		36.12 %	0.94 %		
N	2500	8565	5827	3105	911

Tabla 5.7: *Ensayo 6 con m_2 con las variables: R y X (Media \pm DS) en las bandas 3 y 4. C0- Buenos, C1- Errores de Conexión, C2- Errores de Posición, C3- Errores de Contacto, C4- Artefactos de Movimiento*

5.2. Selección de características

Los resultados que se muestran a continuación son los que se obtuvieron de utilizar SFS con un subconjunto de 1000 muestras aleatorias de m_3 con 22 características (ver tabla B.3) de cada clase manteniendo el modelo binario en cada caso; es decir para el primer nivel se consideró como una clase *Circuito Abierto* (CA) y como la otra clase la fusión de todas las demás. Una vez identificado el clasificador con mejor desempeño se quitan todos los espectros que fueron identificados para la clase CA y se va al siguiente nivel realizando el mismo procedimiento hasta llegar a la última clase.

En la tabla 5.8 se muestran las características que fueron relevantes para cada clase de m_3 . Para las subclases *Contenido Gástrico*(CG), *Contacto Pobre*(CP) y *Artefacto de Movimiento* (AM); se encontraron que las primeras 16 características de la tabla B.3 fueron relevantes en cada caso. En el caso de la clase de *sonda descompuesta* (SoD) no se empleó algoritmo de selección de características ya que es claramente identificable con un umbral por la naturaleza de los datos.

Clase	Características	N
CA	$R4_{DS}, S_{RX_{DS}}$	2
CC	$R4_M, X2_M, X4_M, X2_{DS}, S_{R_M}, S_{RX_M}, S_{RX_{DS}}$	7
MP	$R4_{DS}, X2_M, X4_M, X2_{DS}, S_{X_M}, S_{RX_M}, S_{RX_{DS}}$	7
MC	$R4_M, R4_{DS}, X2_M, X3_{DS}, S_{R_M}, S_{R_{DS}}, S_{X_{DS}}, S_{RX_M}, S_{RX_{DS}}$	9

Tabla 5.8: Características relevantes para cada clase de m_3

Al tener las características relevantes por clase, se realizan ensayos con subconjuntos de características para observar si se puede obtener el mismo desempeño que ocupando todas las características relevantes, en este documento sólo se muestran los resultados obtenidos con el conjunto de características en donde se obtuvo el mejor desempeño.

5.3. Clasificación

En esta sección se muestran los clasificadores que presentaron el mejor desempeño, con sus parámetros de inicialización: número de datos de entrenamiento, tipo de núcleo, características, costo y γ ; así como los resultados del desempeño de cada clasificador ocupando el método de validación cruzada. Recordando algunas consideraciones hechas:

- Se resuelve un problema binario en cada nivel.

- Se quitan los datos de entrenamiento para evaluar sólo con datos no vistos a los clasificadores.
- Se devuelven los datos de entrenamiento de la clase inferior a identificar para evitar la reducción de datos.
- Se quitan los datos identificados como la clase superior.

En la figura 5.4 y 5.5 se muestra un espacio bidimensional en las características S_{RDS} vs. S_{XM} en donde se pueden observar todas las clases a identificar. En las siguientes secciones se muestran los clasificadores para los errores en: conexiones, posición y contacto. Para cada tipo de error se muestran a manera de tabla los parámetros, las características, el tipo de núcleo, la matriz de confusión (ver Fig. 4.7) y las medidas de desempeño del clasificador así como las gráfica S_{RDS} vs. S_{XM} del problema binario de cada nivel.

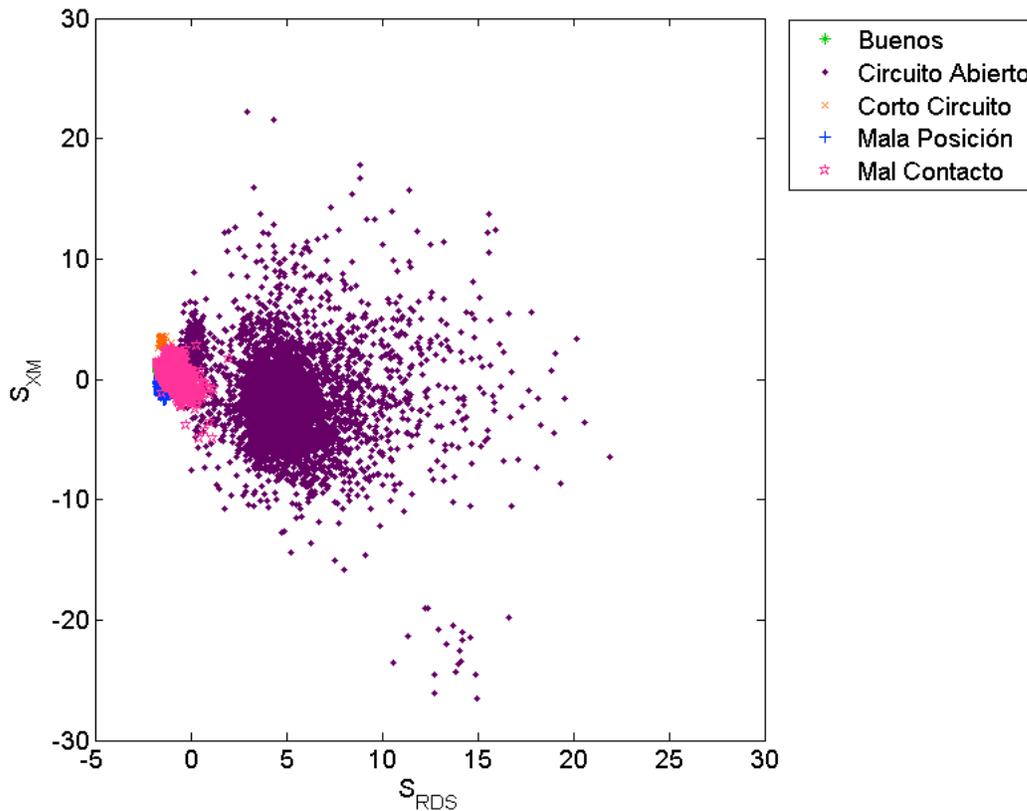


Figura 5.4: S_{RDS} vs. S_{XM} mostrando todas las clases a clasificar.

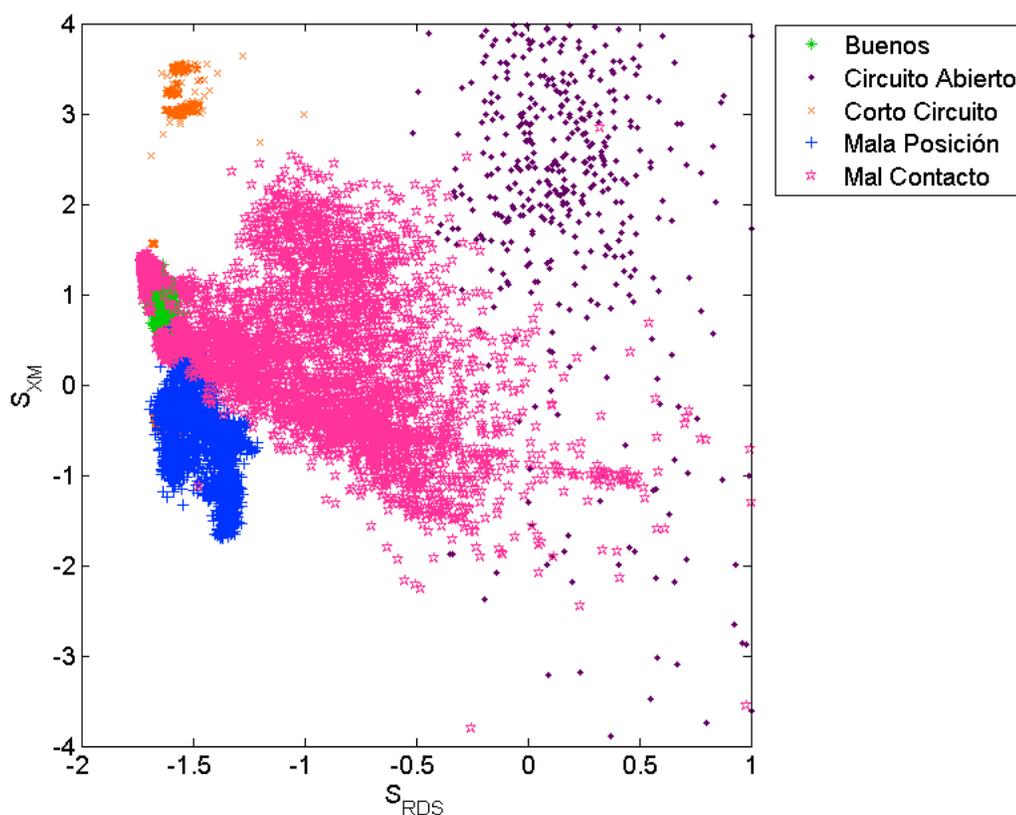


Figura 5.5: S_{RDS} vs. S_{XM} mostrando todas las clases con zoom.

Recordando que los errores asociados a un error en las conexiones con prioridad uno son: circuito abierto (CA) y corto circuito (CC), donde el segundo nivel de CC esta conformado de los errores de sonda descompuesta (SoD) y cable descompuesto (CD). Para entrenar al clasificador de CA se ocupan: 2000 datos aleatorios por clase para el entrenamiento, núcleo lineal y las características: $R4_{DS}$, $S_{RX_{DS}}$. Para medir el desempeño se prueban con datos no vistos 2768 para CA y 14140 para la otra clase, en la tabla 5.9 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.6 se muestra la clase de CA (azul) vs. B, CC, MP y MC (rojo) del problema binario a resolver. En la figura 5.7 se muestra el espacio bidimensional en donde se encuentra la solución con el núcleo lineal con su probabilidad posterior, la figura 5.8 es un acercamiento de la figura 5.7.

Para entrenar al clasificador CC se quitan los datos identificados como CA y se ocupan 2000 datos aleatorios por clase para el entrenamiento, núcleo gaussiano y las características: S_{RM} , S_{RX_M} , $S_{RX_{DS}}$. Para medir el desempeño se prueban con 1797 datos no vistos para CC y 10331 para la otra clase, en la tabla 5.10 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.9 se

muestra la clase de CC (azul) vs. B, MP y MC (rojo) del problema binario a resolver. De subconjunto de datos identificados como CC se resuelve el segundo nivel de SoD y CD por método de umbral por la característica S_{X_M} en la tabla 5.11 se muestra la matriz de confusión para este clasificador. En la figura 5.10 se muestra en SoD(azul)y CD(rojo) en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

Para entrenar al clasificador de MP se quitan los datos identificados como CC y se ocupan 2000 datos aleatorios por clase para el entrenamiento, núcleo gaussiano y las características: R_{4M} , X_{2M} , X_{4M} , X_{2DS} , S_{R_M} , S_{RX_M} , $S_{RX_{DS}}$. Para medir el desempeño se prueban con 1232 datos no vistos para MP y 7099 para la otra clase, en la tabla 5.12 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.11 se muestra la clase MP (azul) vs. B y MC (rojo) del problema binario a resolver en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

Recordando que los errores asociados a MC con prioridad 4 son: contenido gástrico(CG), contacto pobre(CP)y artefacto de movimiento (AM). Primero se identifica el nivel uno MC y posteriormente el segundo nivel: CG, CP, AM. Para entrenar al clasificador de MC se quitan los espectros clasificados como MP y se ocupan 2000 datos aleatorios por clase para el entrenamiento, núcleo gaussiano, características: S_{R_M} , $S_{R_{DS}}$, $S_{X_{DS}}$, S_{RX_M} , $S_{RX_{DS}}$. Para medir el desempeño se prueban con 4598 datos no vistos para MP y 500 para B, en la tabla 5.13 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.12 se muestra la clase de MC (azul) vs. B(rojo) del problema binario a resolver en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

Para resolver el segundo nivel de MC primero se identifican los errores por CG y CP + AM y posteriormente de CP vs. AM. Para entrenar al clasificador de CG se ocupan 2000 datos aleatorios por clase para el entrenamiento, núcleo gaussiano y las 16 características: R_{1M} a X_{4DS} (ver tabla B.3).Para medir el desempeño se prueban con datos no vistos 572 para CG y 1965 para CP + AM, en la tabla 5.14 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.13 se muestra la clase de CG (azul) vs. CP+ AM(rojo) del problema binario a resolver en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

Para entrenar al clasificador de CP vs AM se ocupan 600 datos aleatorios por clase para el entrenamiento, núcleo gaussiano y las 16 características: R_{1M} a X_{4DS} (ver tabla B.3).Para medir el desempeño se prueban con datos no vistos 2461 para CP y 286 para AM, en la tabla 5.15 se resumen los parámetros utilizados de entrenamiento, las medidas de desempeño y matriz de confusión. En la figura 5.14 se muestra la clase de CP (azul) vs. AM(rojo) del problema binario a resolver en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

Los espectros detectados sin error se muestran en la figura 5.15 la clase Buenos (azul) y MC(rojo) detectados cómo espectros buenos en el espacio de características $S_{R_{DS}}$ vs. S_{X_M} .

5.3.1. Errores en las conexiones

Circuito Abierto

Parámetros		Datos		
Entrenamiento	$N = 2000$	Clase	Total	Prueba
Características	$R_{A_{DS}}, S_{RX_{DS}}$	CA	4768	2768
Núcleo	Lineal	Todos	16140	14140
No.Características	2	Total	20908	16908
C	8	<i>Desempeño</i>		
γ	2	Medida	Media	DS
		TA	99.92 %	0.11 %
		TVP	99.95 %	0.11 %
		TFN	0.05 %	0.11 %
		TFP	0.10 %	0.14 %
		TVN	99.90 %	0.14 %
		A_z	0.9997	0.0006
		<i>Matriz de Confusión</i>		
		Clase	CA	Todos
		CA'	2768	12
		Todos'	0	14128

Tabla 5.9: Parámetros y desempeño para Circuito Abierto

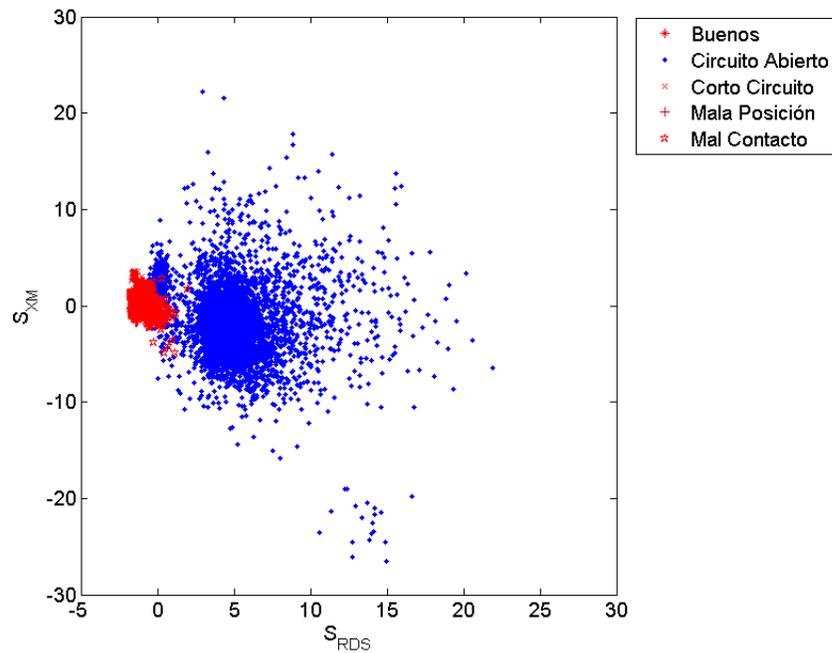


Figura 5.6: $S_{R_{DS}}$ vs. S_{X_M} de las clases CA, B, CC, MP y MC.

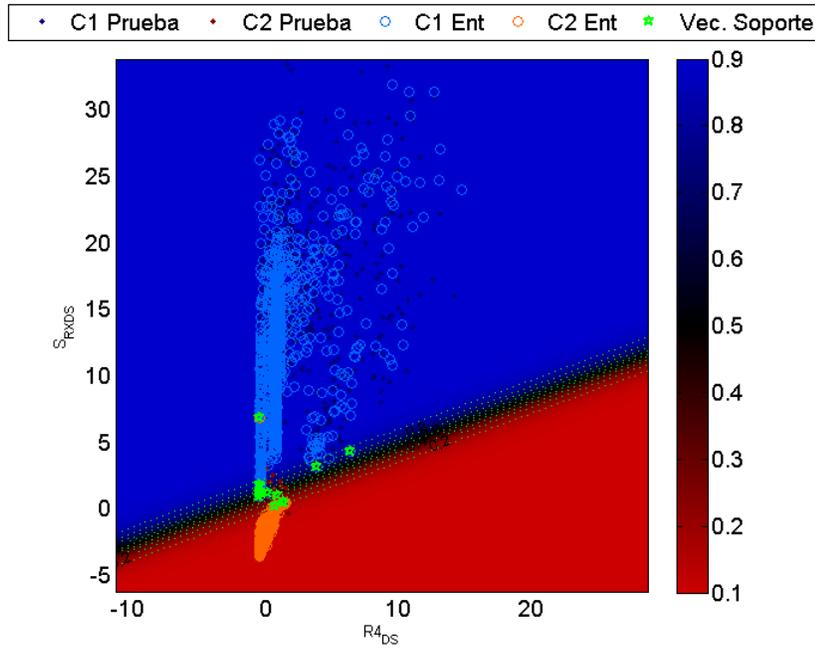


Figura 5.7: Clasificador lineal para CA

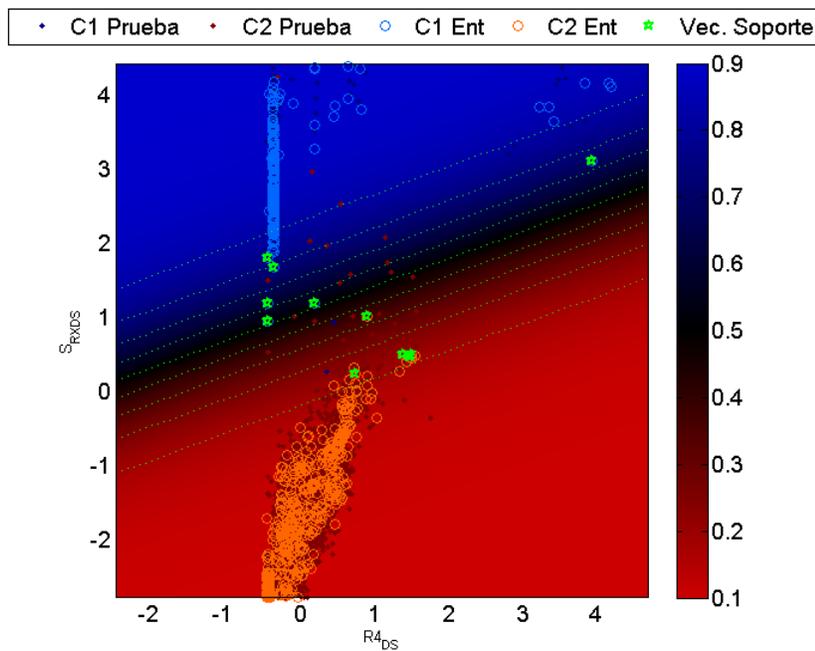
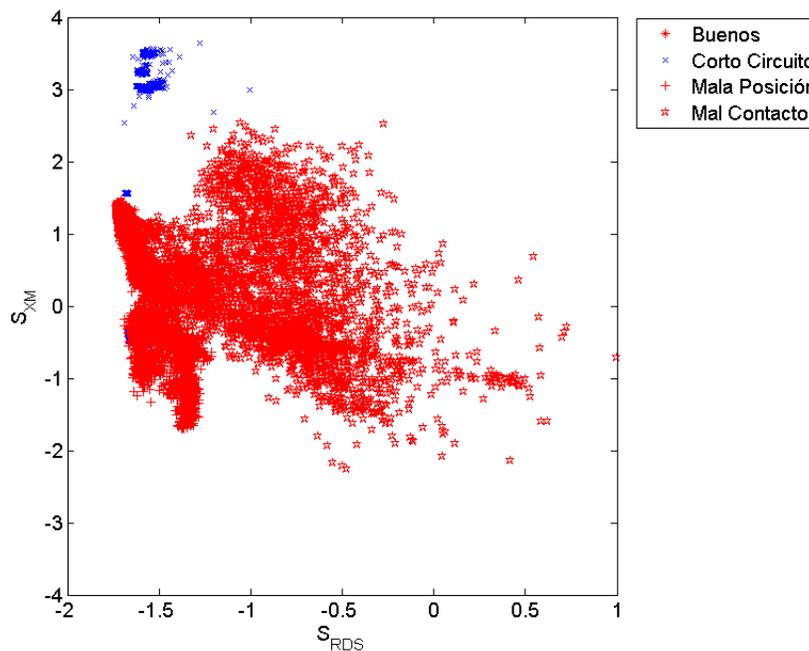


Figura 5.8: Clasificador lineal con un poco de zoom para CA

Corto Circuito

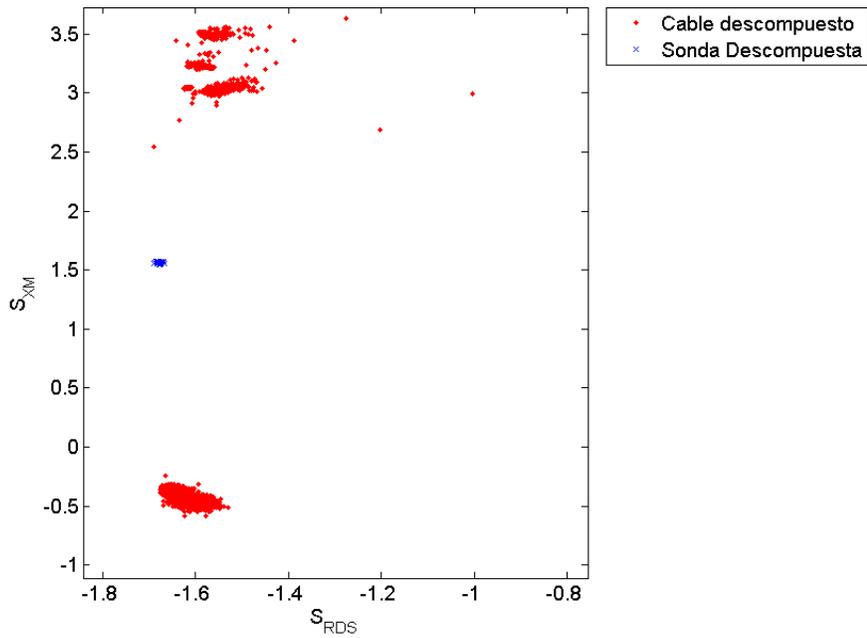
Parámetros		Datos		
Entrenamiento	$N = 2000$	Clase	Total	Prueba
Características	$S_{R_M}, S_{R_{X_M}}, S_{R_{X_{DS}}}$	CC	3797	1797
Núcleo	Gaussiano	Todos	12331	10331
No.Características	3	Total	16128	12128
C	8			
γ	2			
		Desempeño		
		Medida	Media	DS
		TA	100.00 %	0.00 %
		TVP	100.00 %	0.00 %
		TFN	0.00 %	0.00 %
		TFP	0.00 %	0.00 %
		TVN	100.00 %	0.00 %
		A_z	1.0000	0.0000
		Matriz de Confusión		
		Clase	CC	Todos
		CC'	1797	0
		Todos'	0	10331

Tabla 5.10: Parámetros y desempeño para Corto Circuito

Figura 5.9: $S_{R_{DS}}$ vs. S_{X_M} de las clases CC, B, MP y MC.

Sonda Descompuesta

<i>Parámetros</i>		<i>Matriz de Confusión</i>		
Característica	S_{X_M}	Clase	SoD	CD
Umbral	$1 \leq S_{X_M} \leq 2$	SoD'	24	0
		CD'	0	3773

Tabla 5.11: *Parámetros y desempeño para Sonda Descompuesta*Figura 5.10: S_{RDS} vs. S_{X_M} de las clases SoD y CD.

5.3.2. Errores en la Posición

<i>Parámetros</i>		<i>Datos</i>		
Entrenamiento	2000	Clase	N	Prueba
Características	$R4_M, X2_M, X4_M,$	<i>MP</i>	3232	1232
	$X2_{DS}, S_{RM}, S_{RX_M},$	<i>Todos</i>	9099	7099
	$S_{RX_{DS}}$	Total	12331	8331
Núcleo	Gaussiano			
No.Características	7	<i>Desempeño</i>		
C	8	Medida	Media	DS
γ	2	<i>TA</i>	99.95 %	0.07 %
		<i>TVP</i>	100.00 %	0.00 %
		<i>TFN</i>	0.00 %	0.00 %
		<i>TFP</i>	0.10 %	0.14 %
		<i>TVN</i>	99.90 %	0.14 %
		A_z	0.9998	0.0006
<i>Matriz de Confusión</i>				
		Clase	MP	Todos
		<i>MP'</i>	1232	1
		<i>Todos'</i>	0	7098

Tabla 5.12: Parámetros y desempeño para Mala Posición

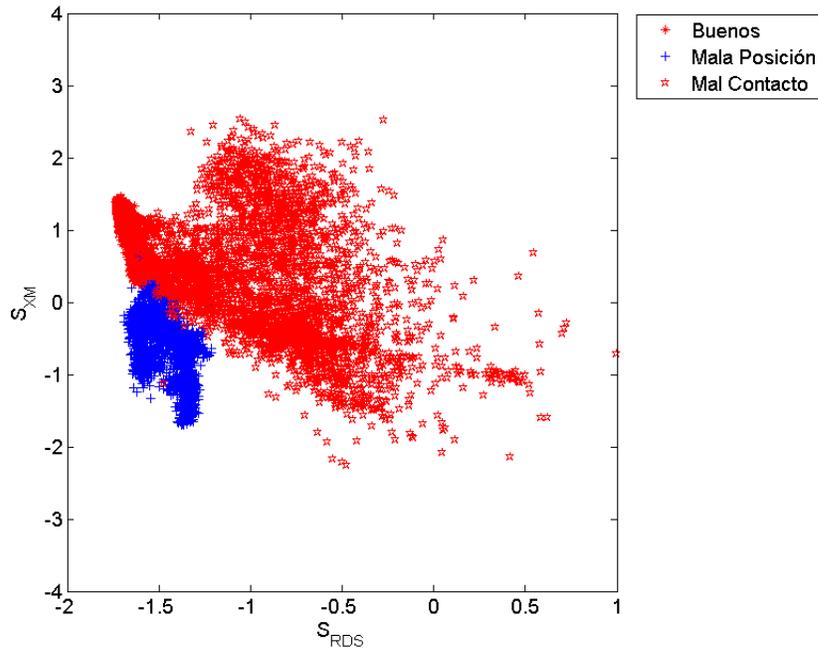


Figura 5.11: $S_{R_{DS}}$ vs. S_{X_M} de las clases *MP*, *B*, y *MC*.

5.3.3. Errores en el Contacto

<i>Parámetros</i>		<i>Datos</i>		
Entrenamiento	2000	Clase	N	Prueba
Características	$S_{RM}, S_{RDS}, S_{XDS},$ S_{RXM}, S_{RXDS}	<i>MC</i>	6598	4598
Núcleo	Gaussiano	<i>B</i>	2500	500
No. Características	5	Total	9098	5098
<i>C</i>	8	<i>Desempeño</i>		
γ	2	Medida	Media	DS
		<i>TA</i>	98.18 %	0.40 %
		<i>TVP</i>	98.30 %	0.50 %
		<i>TFN</i>	1.70 %	0.50 %
		<i>TFP</i>	1.94 %	0.61 %
		<i>TVN</i>	98.06 %	0.61 %
		<i>Az</i>	0.9931	0.0005
<i>Matriz de Confusión</i>				
Clase	MC	Buenos		
<i>MC'</i>	4525	12		
<i>Buenos'</i>	73	488		

Tabla 5.13: Parámetros y desempeño para Mal Contacto

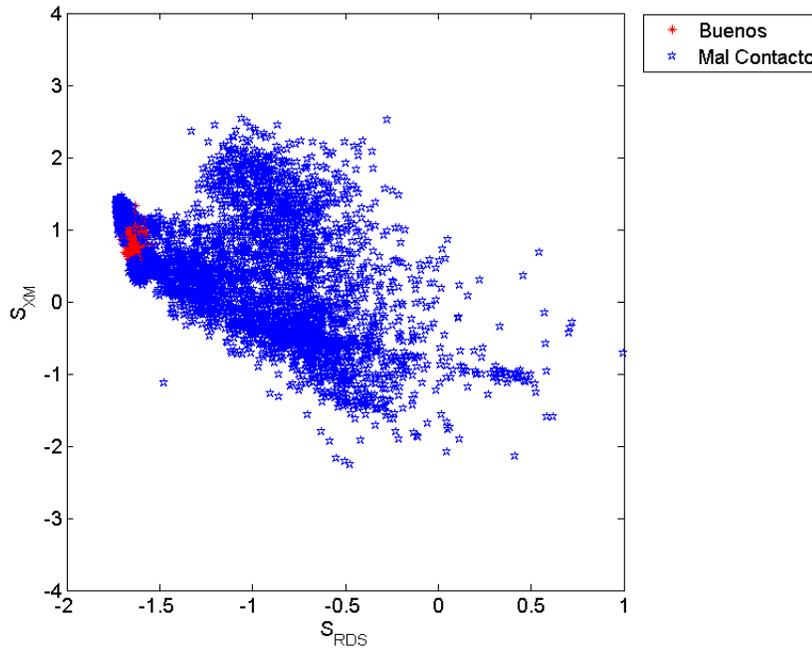


Figura 5.12: S_{RDS} vs. S_{XM} de las clases *MC* y *B*.

Errores por Contenido Gástrico

Parámetros		Datos		
Entrenamiento	2000	Clase	N	Prueba
Características	$R1_M$ a $X4_{DS}$	CG	2572	572
Núcleo	Gaussiano	Todos	3965	1965
No.Características	16	Total	6537	2537
C	8			
γ	2			
		Desempeño		
Medida	Media	DS		
TA	99.28 %	0.21 %		
TVP	98.96 %	0.32 %		
TFN	1.04 %	0.32 %		
TFP	0.40 %	0.38 %		
TVN	99.60 %	0.38 %		
A_z	0.997	0.002		
		Matriz de Confusión		
Clase	CG	Todos		
CG'	569	21		
Todos'	3	1944		

Tabla 5.14: Parámetros y desempeño para Contenido Gástrico

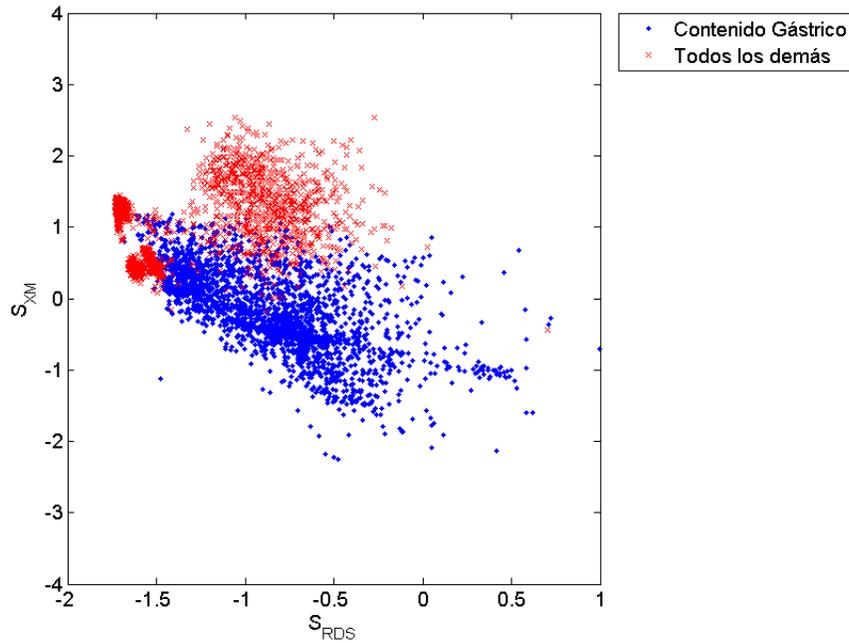


Figura 5.13: $S_{R_{DS}}$ vs. S_{X_M} de las clases CG y Todos.

Errores por Contacto Pobre

<i>Parámetros</i>		<i>Datos</i>	
Datos Entrenamiento	600	Clase	N Prueba
Características	$R1_M$ a $X4_{DS}$	CP	3061 2461
Núcleo	Gaussiano	AM	886 286
No. Características	16	Total	3947 2747
C	8		
γ	2		

<i>Desempeño</i>		
Medida	Media	DS
TA	99.67%	0.35 %
TVP	99.50%	0.45 %
TFN	0.50 %	0.45 %
TFP	0.17 %	0.37 %
TVN	99.83%	0.37 %
Az	0.998	0.002

<i>Matriz de Confusión</i>		
Clase	CP	AM
CP'	2454	6
AM'	7	280

Tabla 5.15: *Parámetros y desempeño para Contacto Pobre*

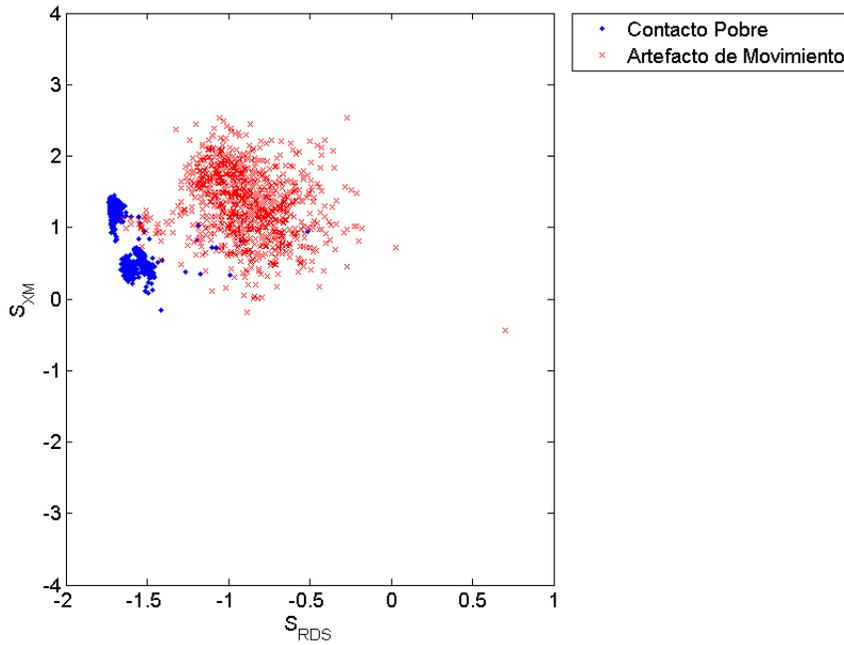


Figura 5.14: $S_{R_{DS}}$ vs. S_{X_M} de las clases CP y AF.

5.3.4. Espectros Buenos

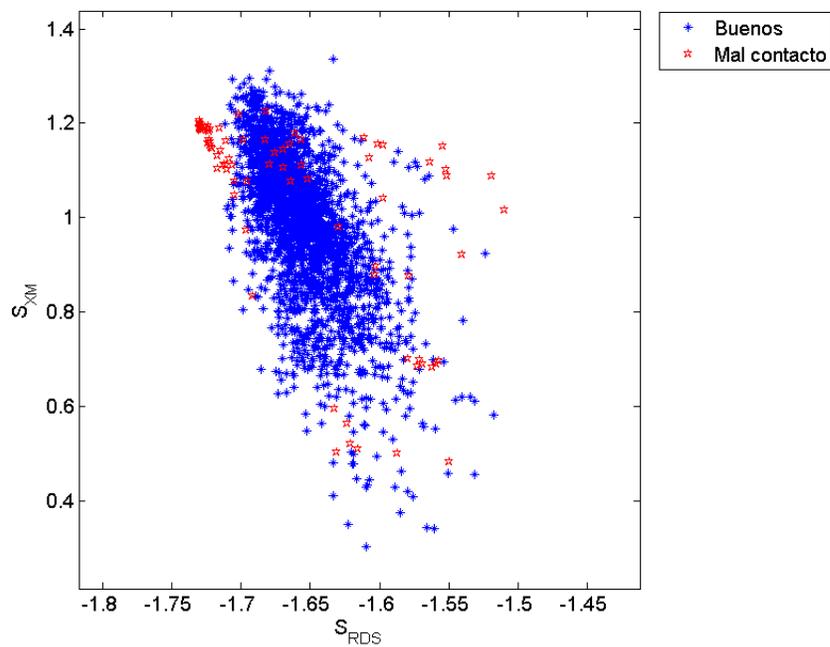


Figura 5.15: S_{RDS} vs. S_{XM} de las clases Buenos y MC

5.4. Validación con otras bases de datos

Una vez que se diseñó el ADE siendo validado en cada nivel con datos no vistos del mismo tipo de error, hay que comprobar que otra base de datos que ha sido clasificada sólo de espectros buenos no vistos ni para entrenar ni para probar, verificar que efectivamente no se estén identificando otro tipo de errores. Recordando que para esta prueba se tienen dos bases de datos con espectros buenos una del INCICH y otra del Multicéntrico. En la figura 5.16 se muestra la comparación de estas dos bases de datos, el estudio realizado en INCICH que cuenta con 2488 espectros (rombos azules) y la del estudio multicéntrico que cuenta con 11,145 espectros (asteriscos verdes), de ésta se toma un subconjunto de 2500 datos (círculos rojos) que se ocuparon para el diseño del algoritmo de detección de errores y que se excluyen para la realización de esta prueba. Cada espectro de la base de datos del INCICH es acondicionado para que pase a través del ADE teniendo una certeza de 91.6% y un 8.4% de error el cual 0.2% representa a CG y 8.2% a CP (ver Fig. 5.17), los círculos rojos representan el 8.4% de los espectros buenos clasificados como erróneos. Posteriormente se realiza nuevamente el acondicionamiento de cada espectro de la base de datos del multicéntrico para que pase a través del ADE teniendo como resultado una certeza de 97.7% y un error de 2.3% el cual el 0.1% representa a CG y 2.2% representa a CP) (ver Fig. 5.18), los círculos rojos representan el 2.3% de los espectros buenos clasificados como erróneos.

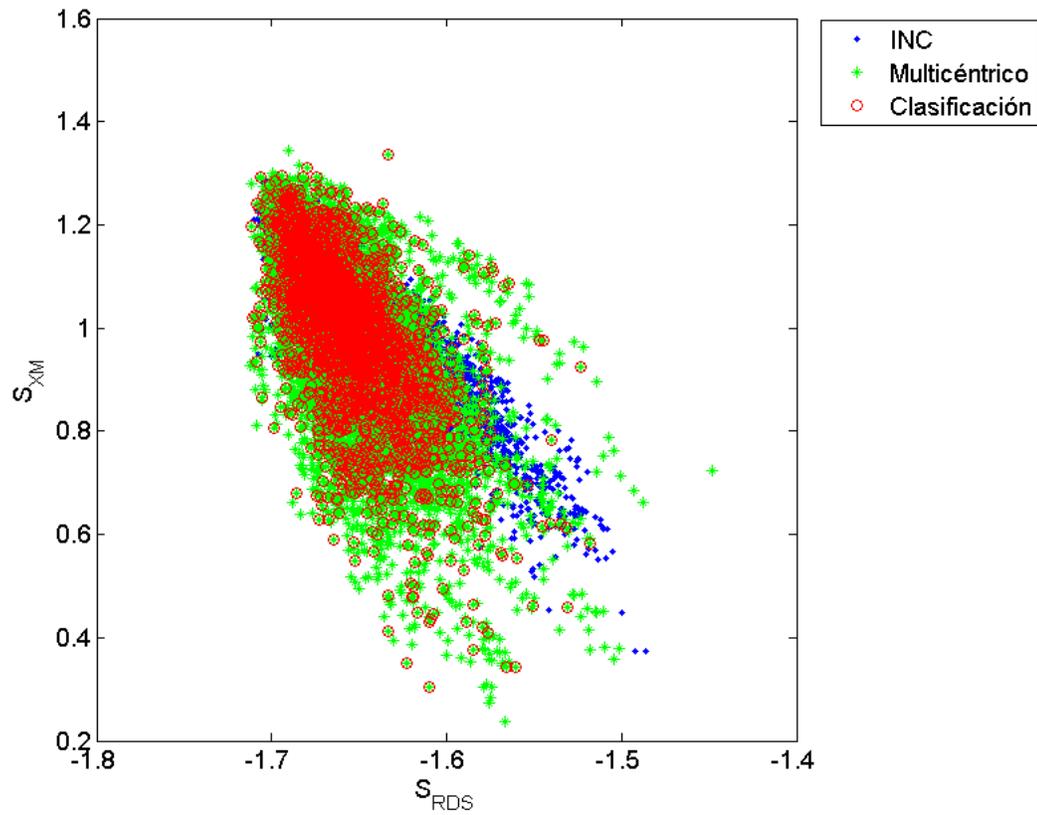


Figura 5.16: Comparación de espectros buenos de dos diferentes bases de datos.

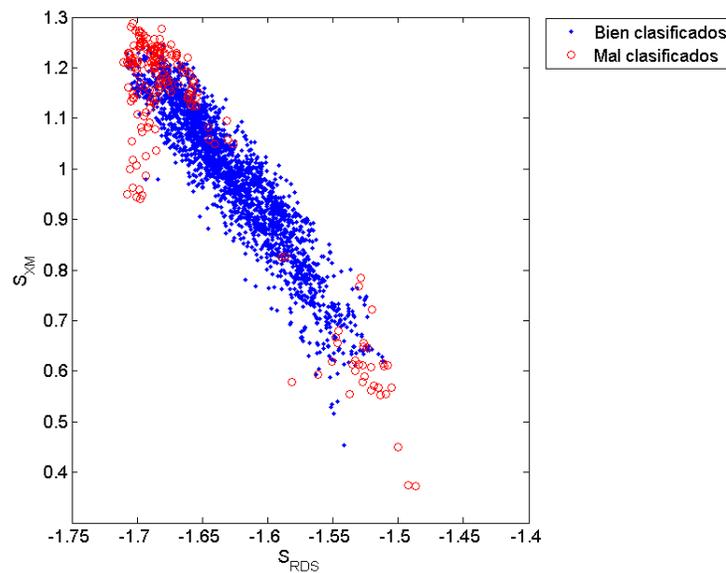


Figura 5.17: Comparación con espectros de INC en S_{RDS} vs. S_{XM} .

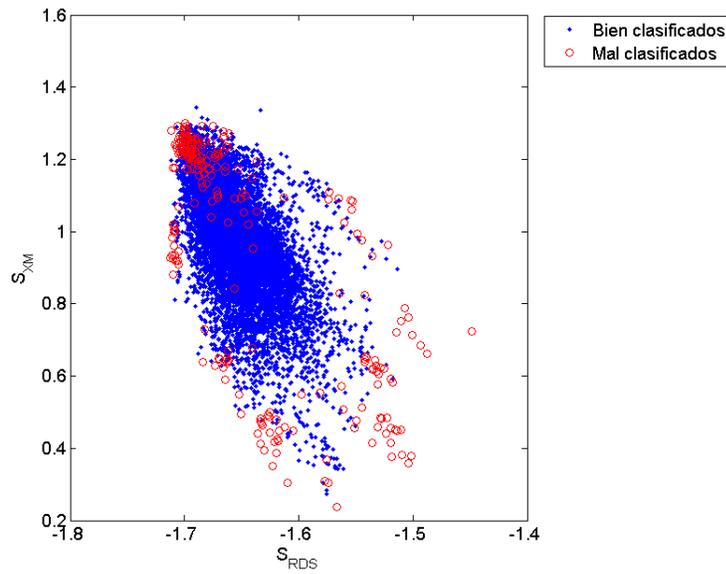


Figura 5.18: Comparación con espectros de Multicéntrico en S_{RDs} vs. S_{XM} .

5.5. Resultados Globales

Estableciendo la base de datos m_3 con ayuda del análisis de conglomerados y gráficas de siluetas se pudo dar inicio al análisis de selección de características del cual se obtuvieron las características de mayor relevancia en cada nivel (tabla 5.8), dando inicio a un proceso iterativo hasta encontrar el clasificador con el mejor desempeño ocupando el menor número de características posibles. Se resuelve el primer nivel y se establecen los cuatro clasificadores que lo conforman. Establecidos los clasificadores generales se resuelven los clasificadores particulares de los errores de CA y MC. En las figuras 5.19 y 5.20 se muestran a manera de resumen los desempeños de cada clasificador en los dos niveles así como el diagrama de flujo del algoritmo propuesto. En el apéndice D se muestran los códigos generados para el diseño de clasificadores y la implementación del algoritmo de detección de errores. En cuanto a la validación con otras bases de datos se ha obtenido una certeza mayor al 90 % y los errores detectados corresponden a la clase de errores por contacto.

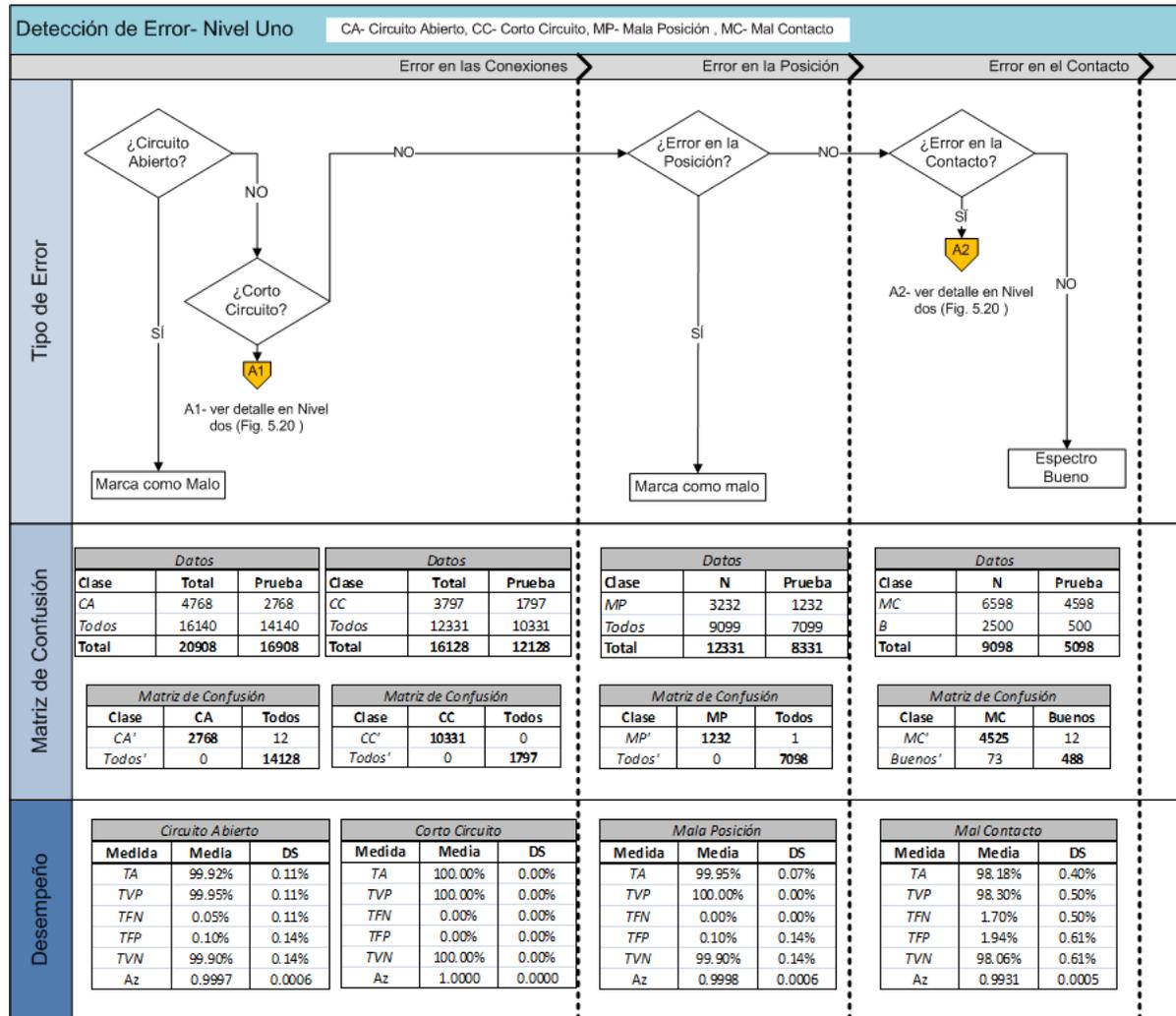


Figura 5.19: Desempeños de nivel uno. TA-Tasa de aciertos, TVP-Tasa de Verdaderos Positivos o sensibilidad, TFN- Tasa de falsos negativos, TFP- Tasa de Falsos Positivos o falsa alarma, TVN-Tasa de verdaderos negativos, Az- Área bajo la curva CRO

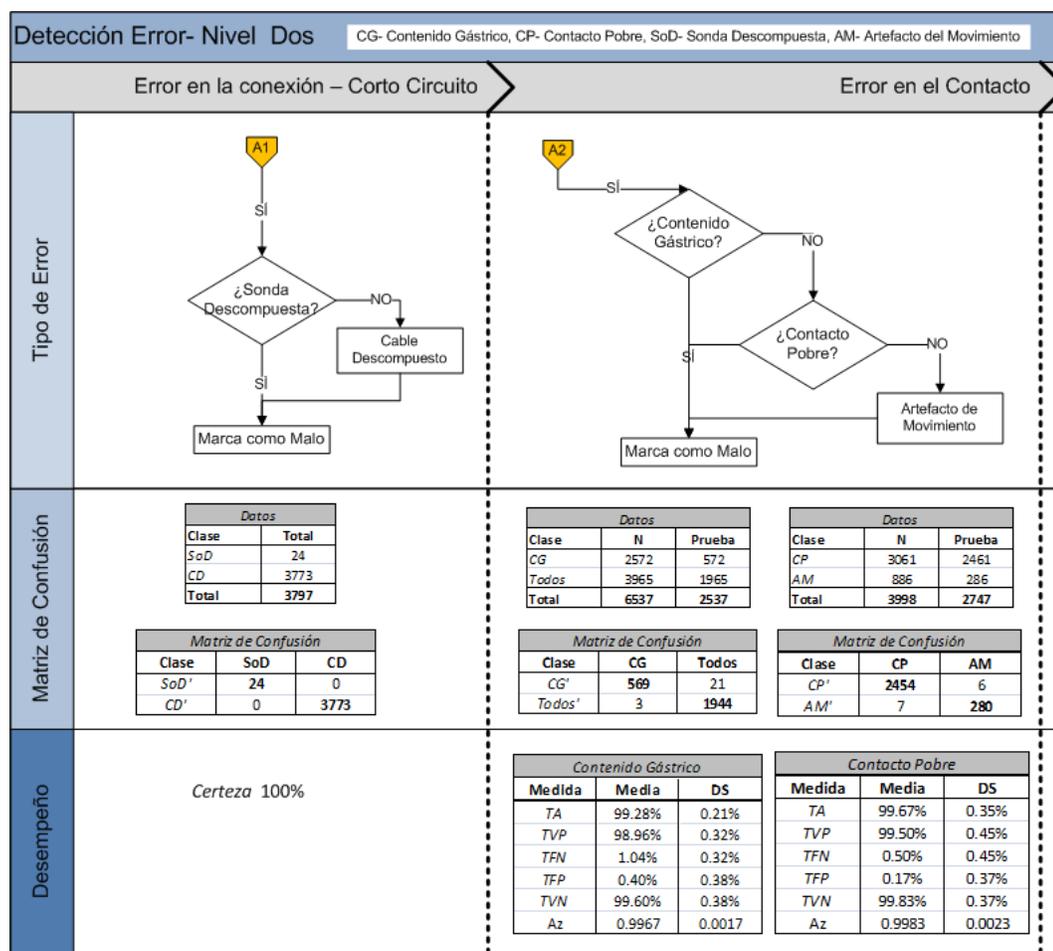


Figura 5.20: Desempeños de nivel dos. TA-Tasa de aciertos, TVP-Tasa de Verdaderos Positivos o sensibilidad, TFN- Tasa de falsos negativos, TFP- Tasa de Falsos Positivos o falsa alarma, TVN- Tasa de verdaderos negativos, Az- Área bajo la curva CRO

Capítulo 6

Discusión

En esta sección se discuten cada uno de los resultados presentados en el capítulo 5.1 de una forma detallada resaltando su relevancia y los hallazgos encontrados en cada caso.

6.1. Selección de Clases

Al realizar el barrido en un rango de k que va de 10 a 20 grupos se obtuvo que la mejor partición fue $k = 14$ por tener el mayor valor de silueta promedio de 0.61 (ver figura 5.2). En la gráfica de silueta asociada a esta partición (ver figura 5.3) se observa que hay conglomerados con datos mal clasificados ya que tienen valores negativos. Para saber cómo se están distribuyendo los conglomerados en las clasificaciones dadas por personal experto, se realizaron varios ensayos con $k = 14$ para poder determinar cuáles son las características que mejor separan naturalmente los datos y poder determinar una mejor clasificación de los datos. Se buscó que los espectros buenos no compartan conglomerado con algún error, es decir que tengan la mayoría de sus datos concentrados en un solo conglomerado o que sea mínimo su traslape con algún otro. Para observar la mejor distribución no existe un método en el cuál basarse pero las herramientas que se utilizaron son de gran ayuda para identificar la forma natural de agrupamiento de las clases y poder de esta manera elegir la mejor distribución de datos. A continuación presento un análisis detallado de cada ensayo esto con la finalidad de resaltar las diferencias que se encontraron entre ellos y cómo es que se pudo lograr formar nuevas matrices M_3 y m_3 partiendo sólo de analizar como los datos se comportan con diferentes características y bandas.

Analizando los resultados del ensayo 1 (tabla 5.1) se observa que para la mayoría de los conglomerados se obtiene una buena clasificación con respecto a la etiqueta previa. Esto quiere decir que los datos tienen podrían potencialmente con estas etiquetas ser bien clasificados, aunque hay algunos errores que se dispersan por varios conglomerados. Es importante ver el traslape que existen entre errores pero siempre se tiene que considerar

el número de datos que se tienen ya que las prevalencias no son las mismas, sólo hay que tener reservas cuando se observan los 100 % de que número de datos estamos hablando por ejemplo en el error *e4*. Pero en este ensayo se tienen contempladas todas las variables con todas sus características, los ensayos 2 y 3 muestran los resultados uno para A y F y el otro para R y X. Comparando estos dos resultados observamos que se tiene un porcentaje de más de 90 % para los espectros buenos pero para A y F aumenta el porcentaje de traslape con varios errores por ejemplo un 88.7 % de traslape con el error *e13*. Mientras que para R y X el traslape con *e12* se mantiene como en el ensayo 1.

Los resultados anteriores fueron para cada tipo de error en específico sin embargo la propuesta inicial de esta tesis fue identificar por 5 clases más generales como se muestran en m_2 , y los ensayos que se realizaron para esta matriz fueron los ensayos 3 bis a 6. En cualquiera de estos ensayos se observan que al unir los diversos tipos de error en *C1* están completamente dispersos entre todos los conglomerados lo que nos hizo regresar a observar las características que se tenían de cada error de esta clase, al hacer esto nos dimos cuenta de que algunos espectros tienen características muy diferentes aún estando dentro de la misma clase, por lo que se decidió separar los espectros de *C1*, que eran de conexión a dos clases circuito abierto y corto circuito esto por su morfología y similitudes de características. Se observó que los espectros de los errores *e9* y *e10* realmente eran muy diferentes que los de *e8* de la clase *C2* dado que realmente no presentaban un error en la posición sino de contacto por tratarse de contenido gástrico y este error cuando puede llegar a suceder es por que la sonda está en el estómago y está bien colocada, es decir que no está en el esófago boca y por este motivo se reasignaron a la clase de mal contacto, finalmente *C4* que solo representa al error *e15* fue tomado como parte de la clase mal contacto *C3* ya que si existen altas variaciones es por que la sonda no está haciendo buen contacto con la mucosa gástrica.

Los resultados que hicieron que se modificara la dimensionalidad de la base de datos fue la comparación del ensayo 1 con el 3 ya que los porcentajes de agrupación del ensayo 1 con las variables A, F, R y X son similares a los del ensayo 3 que ocupa las variables R y X lo que permite reducir el número de características a la mitad teniendo resultados muy similares y sobre todo que los espectros buenos tienen un mínimo traslape con otros errores. Este resultado nos sugiere que en estas variables son las adecuadas para generar una nueva clasificación en donde se ocupen valores de R y X solamente. La comparación de los ensayos 3bis a 6 fueron los resultados que hicieron que cambiarán las clases ya que se observó que la clase errores de conexión se distribuye en todos los conglomerados lo que hace notar que unirlos en una clase no hace sentido por lo que se observó a detalle el tipo de características y se encuentra que la diferencia radica en las altas y bajas variaciones por lo que se decide la separación de estos errores haciendo dos clases independientes pero conservando la misma prioridad. Con estas consideraciones se generaron las bases de datos M_3 y m_3 dando inicio al proceso de selección de características ya con bases de

datos redimensionadas y redistribuidas. El detalle de la conformación de la matriz M_3 y m_3 con todos estos cambios antes mencionados se encuentra detallado en el apéndice B. De manera general en este análisis es importante conocer la naturaleza de cada tipo de error y a través de las herramientas de análisis de conglomerados poder reclasificar de una manera más objetiva las clases de la base de datos.

De manera puntual observando y haciendo comparaciones entre los resultados de los ensayos tenemos:

- En el ensayo 3 se pueden agrupar los errores de una manera similar ocupando solamente R y X, que ocupando A, F, R y X, y esto nos reduce la dimensionalidad de 32 a 16 características.
- Hay espectros como: $e0, e2, e4, e5, e10$, que se agrupan naturalmente en un conglomerado casi en un 100 % en los ensayos 1, 2 y 3.
- En la clasificación por prioridad se ve que la clase **C1** está completamente dispersa entre todos los tipos de errores, por lo que se identificaron espectros con base en características similares (media, DS) entre los errores dando origen a una nueva base de datos reclasificada M_3 y m_3 .
- Los espectros buenos (**C0**) o $e0$ en la mayoría de los ensayos se agrupan naturalmente en más de 90 %, lo que nos indica que las características propuestas podemos esperar tener una buena clasificación.

6.2. Selección de características

Una vez definida la base de datos con las mejores agrupaciones naturales de los tipos de errores (M_3 y m_3), se encuentran las características más relevantes ocupando el algoritmo SFS. Primero se aplica para el esquema general (nivel 1, m_3) y posteriormente los casos particulares (nivel 2, M_3). Para el caso general como particular la búsqueda de las características más relevantes de cada clase se hace tomando un subconjunto de 1000 muestras aleatorias. La tabla 5.8 muestra las características relevantes en cada clase para la base de datos m_3 y se puede observar que el número de características aumenta conforme la prioridad aumenta. Hay características como S_{RXDS} que son identificadas como relevantes y compartidas en cada clase, así como la media y/o DS de la resistencia en la banda cuatro y la media y/o DS de la reactancia en la banda dos. Se observa también que las sumas de resistencia y reactancia con media y/o DS se identifican como relevantes. En todas las clases están presentes las sumas de resistencia y/o reactancia ya sea simples y/o combinadas.

Para el nivel 2 que son los casos particulares a diferencia del nivel general, no fue posible identificar características particulares por lo que en cada caso se ocuparon las

primeras 16 características de la tabla B.3. Cabe mencionar que el caso particular para sonda descompuesta (SoD) que fue resuelto por umbral y no fue necesario hacer una selección de características para esta clase. La identificación de características se vuelve un punto relevante para obtener el mejor desempeño de cada clasificador y se busca obtener con un menor número de características el mejor desempeño, estos algoritmos dan una buena idea para identificar de una manera sistemática las características relevantes para la identificación de las clases.

6.3. Clasificación

Para el diseño de los clasificadores se parte de las características encontradas como las mejores y se van haciendo pruebas de desempeño, validación cruzada para encontrar los mejores parámetros (costo y γ) para cada nivel, primero de una manera general (nivel 1, m_3) y después de una forma particular (nivel 2, M_3) (ver fig. 4.8); en cada caso se ocupan 2000 datos aleatorios para entrenamiento de los clasificadores y el resto para hacer las pruebas de desempeño.

Partiendo de una base de datos de 20908 espectros la primera clase a identificar es CA ($n=4768$), el clasificador para identificar ésta tiene como parámetros un núcleo lineal y las características: $R4_{DS}$ y $S_{RX_{DS}}$ obteniendo una tasa de aciertos del 99.92% y $A_z = 0,99 \pm 0,0006$ (ver tabla 5.9). Es posible mostrar gráficamente la probabilidad posterior por ser un caso bidimensional como se puede observar en las figuras 5.7 y 5.8. Para resolver el siguiente nivel se quitan los datos identificados como CA quedando una base de datos con 16128 espectros. Para identificar la clase CC ($n=3797$), el clasificador queda conformado por un núcleo gaussiano y las características S_{RM} , S_{RX_M} , $S_{RX_{DS}}$ obteniendo una tasa de aciertos de 100% y $A_z = 1$ (ver tabla 5.10) lo que indica una clasificación perfecta. Para resolver el siguiente nivel se quitan los datos identificados como CC quedando una base de datos de 12331. En el siguiente nivel se identifica la clase MP ($n=3232$), el clasificador para ésta queda conformado por un núcleo gaussiano y las características: $R4_{DS}$, $X2_M$, $X4_M$, $X2_{DS}$, S_{X_M} , S_{RX_M} , $S_{RX_{DS}}$, obteniendo una tasa de aciertos de 99.95% y $A_z = 0,99 \pm 0,0006$ (ver tabla 5.12). Para resolver el último nivel se quitan los datos identificados como MP quedando una base de datos de 9098 espectros y se procede a identificar la clase MC ($n=6598$). El clasificador para identificar MC queda conformado por un núcleo gaussiano y las características S_{RM} , $S_{R_{DS}}$, $S_{X_{DS}}$, S_{RX_M} , $S_{RX_{DS}}$ obteniendo una tasa de aciertos del 98.18% y $A_z = 0,99 \pm 0,0005$ (ver tabla 5.13), con esto queda resuelto el problema de identificación de errores ya que la clase que separa MC son los espectros buenos.

Una vez identificadas las clases del nivel uno, se resuelven los casos particulares en el

nivel dos. Se identifica el subnivel de CC que se compone de las clases SoD y CD. Partiendo de la clase identificada como CC ($n=3797$) la clase SoD y CD se separan a través del método de umbral con la característica S_{X_M} (ver tabla 5.11) ya que las clases son perfectamente separables cómo se muestra en la figura 5.10, se obtiene una certeza de 100 %. La clase MC se conforma por: CG, CP y AM, partiendo del subconjunto identificado como MC ($n=6537$) se identifica la clase CG, se entrena el clasificador con 2000 datos aleatorios con un núcleo gaussiano con 16 características, obteniendo una tasa de aciertos de 99.28 % y $A_z = 0,99 \pm 0,002$ (ver tabla 5.14). Para resolver el siguiente nivel se quitan los espectros identificados como CG quedando una base de datos de 3947 espectros, el clasificador para CP es entrenado con 600 datos aleatorios para cada clase, núcleo gaussiano y 16 características obteniendo una tasa de aciertos de 99.67 % y $A_z = 0,99$. Se puede observar que a medida que la prioridad aumenta el número de características aumenta, sobre todo para las clases: CG y CP que ocupan 16 características.

Es importante mencionar que las características relevantes encontradas para el nivel uno ocupando los algoritmos SFS comparten en cada clase ciertas características como: las bandas 2 y 4 en R y X (media \pm DS), sumas de todas las bandas en sus diferentes combinaciones(media \pm DS). Sin embargo para el nivel dos no se pudieron encontrar características relevantes y se tuvieron que ocupar las 16 características obteniendo buenos resultados de clasificación. Para cada clasificador se obtuvieron características diferentes pero prevalecieron las bandas uno y cuatro así como las sumas de medias y desviaciones. Una herramienta que también ayudó fue la representación gráfica de cada tipo de error en sus bandas.

Se puede observar que en cada clasificador se obtiene una tasa de aciertos mayor al 95 % y una $A_z = 0,99$, el núcleo que predomina es el gaussiano y que en cada clase haciendo la prueba con datos no vistos hay buenos resultados de clasificación. Es importante señalar que en el diseño de cada clasificador se siguió la misma metodología propuesta desde el inicio, es importante seguir estos pasos para obtener las medidas de desempeño de cada clasificador y así saber realmente el rango de error que puede tener de cada uno. Para validar el ADE se hace una prueba retrospectiva en donde se implementa un algoritmo de identificación según su prioridad con base a los clasificadores propuestos. Esta prueba se realiza con dos bases de datos no vistos para el entrenamiento de los clasificadores, esto para no sesgar el resultado con datos que se usaron para el diseño y validación de los clasificadores, para la base de datos del multicéntrico se obtiene un error de 2.3 % y para la del INCICH un error del 8.4 %. Los espectros buenos que se identificaron con error fueron para ambas bases de datos de la clase MC que representan errores en el contacto; esto quiere decir que no se dispararon errores por mala conexión ni posición. De hecho este es un error más o menos esperado ya que los espectros fueron clasificados por personal experto pero sin ninguna ayuda de algoritmos objetivos que eviten sesgar la clasificación, este resultado se puede interpretar desde dos puntos de vista uno el que el clasificador se equivocó o que el personal experto pasará espectros malos como buenos ya que la frontera

de decisión no se podía determinar con tan solo mirar el espectro. Pero por cualquiera de los dos motivos se tiene más del 90 % de buena clasificación lo que hace muy robusta la clasificación.

Capítulo 7

Conclusiones

Fue posible desarrollar y validar el ADE para la identificación de espectros buenos y espectros con error, haciendo un pre-procesamiento de datos para la reducción de la dimensionalidad de la base de datos, la re-clasificación y redimensionamiento de la base de datos con base a los resultados obtenidos del análisis de conglomerados, una selección de características relevantes con el algoritmo SFS para el diseño de clasificadores y el uso de curvas CRO como medidas de desempeño de cada clasificador propuesto como solución.

Es posible clasificar los diversos tipos de errores de la medición de impedancia gástrica en humanos agrupando las mediciones de R y X en cuatro bandas ($\text{media} \pm \text{DS}$), así como las suma de las bandas y sus combinaciones. Verificándolo en el análisis de conglomerados y la naturaleza del error donde se observó que cada tipo de error tiene características particulares y son estas diferencias las que permiten identificar la clase a la que pertenece cada tipo de error.

Es posible la implementación del ADE siguiendo un esquema jerárquico donde se establece que los errores tienen prioridades según el tipo de falla que se presente. Cada clasificador diseñado en el ADE es un clasificador binario implementando con MSV o método de umbral, con características específicas en cada nivel obtenidas de ocupar el algoritmo SFS. El mejor clasificador para nivel se elige a través de las medidas de desempeño ocupado validación cruzada, obteniendo porcentajes de certeza mayores del 95 % con una SD máxima del 0.4 % y $A_z > 0,98$ con una SD máxima de 0.0023.

Se validó el desempeño con otras bases de datos de espectros buenos no vistos ni usados para entrenar a los clasificadores obteniendo mas del 90 % de tasa de aciertos con un error para la base del multicéntrico de 8.4 % y 2.3 % de la base del INCICH. Sin identificar ningún error por conexión o posición, los errores que se identificaron fueron errores de la clase de mal contacto.

7.1. Trabajo futuro

Una de las siguientes etapas del desarrollo del EITG es la implementación de este algoritmo en una versión más reciente del EITG para su uso clínico. Por lo que un trabajo futuro es la validación in-vitro e in vivo de este algoritmo ya que hasta ahora sólo se ha validado retrospectivamente y con una versión anterior del EITG. Es de esperarse que algún tipo de error no este considerado en las bases de datos que se ocuparon para esta tesis. A medida que se tenga un mayor número de espectros y nuevos errores encontrados se podrá mejorar la identificación de errores por lo se tendrá que ver qué características tienen los nuevos errores y ubicarlos dentro de la jerarquía propuesta o en su defecto hacer una nueva clase y hacer nuevamente la metodología propuesta e implementar un algoritmo mejorado. Al tener nuevos datos se puede ver incrementada la n de cada tipo de error y poder reforzar los clasificadores con un mayor número de espectros.

Bibliografía

- [1] R. G. Fiddian-Green, P. M. Amelin, J. B. Herrmann, E. Arous, B. S. Cutler, M. Schiedler, H. B. Wheeler, and S. Baker, "Prediction of the development of sigmoid ischemia on the day of aortic operations. indirect measurements of intramural ph in the colon.," *Arch Surg*, vol. 121, pp. 654–660, Jun 1986.
- [2] M. G. Schiedler, B. S. Cutler, and R. G. Fiddian-Green, "Sigmoid intramural ph for prediction of ischemic colitis during aortic surgery. a comparison with risk factors and inferior mesenteric artery stump pressures.," *Archives of surgery (Chicago, Ill. : 1960)*, vol. 122, pp. 881–886, August 1987.
- [3] G. R. Doglio, J. F. Pusajo, M. A. Egurrola, G. C. Bonfigli, C. Parra, L. Vetere, M. S. Hernandez, S. Fernandez, F. Palizas, and G. Gutierrez, "Gastric mucosal ph as a prognostic index of mortality in critically ill patients.," *Crit Care Med*, vol. 19, pp. 1037–1040, Aug 1991.
- [4] G. Gutierrez, H. Bismar, D. R. Dantzker, and N. Silva, "Comparison of gastric intramucosal ph with measures of oxygen transport and consumption in critically ill patients.," *Crit Care Med*, vol. 20, pp. 451–457, Apr 1992.
- [5] G. Gutierrez, F. Palizas, G. Doglio, N. Wainsztein, A. Gallesio, J. Pacin, A. Dubin, E. Schiavi, M. Jorge, and J. Pusajo, "Gastric intramucosal ph as a therapeutic index of tissue oxygenation in critically ill patients.," *Lancet*, vol. 339, pp. 195–199, Jan 1992.
- [6] S. M. Hollenberg, T. S. Ahrens, D. Annane, M. E. Astiz, D. B. Chalfin, J. F. Dasta, S. O. Heard, C. Martin, L. M. Napolitano, G. M. Susla, R. Totaro, J.-L. L. Vincent, and S. Zanotti-Cavazzoni, "Practice parameters for hemodynamic support of sepsis in adult patients: 2004 update.," *Critical care medicine*, vol. 32, pp. 1928–1948, September 2004.
- [7] W. C. Shoemaker, P. L. Appel, H. B. Kram, K. Waxman, and T. S. Lee, "Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients.," *Chest*, vol. 94, pp. 1176–1186, December 1988.

-
- [8] C. D. Gomersall, G. M. Joynt, R. C. Freebairn, V. Hung, T. A. Buckley, and T. E. Oh, "Resuscitation of critically ill patients based on the results of gastric tonometry: a prospective, randomized, controlled trial.," *Critical care medicine*, vol. 28, pp. 607–614, March 2000.
- [9] P. R. Miller, E. H. Kincaid, J. W. Meredith, and M. C. Chang, "Threshold values of intramucosal ph and mucosal-arterial co2 gap during shock resuscitation.," *J Trauma*, vol. 45, pp. 868–872, Nov 1998.
- [10] S. O. Heard, C. M. Helsmoortel, J. C. Kent, A. Shahnarian, and M. P. Fink, "Gastric tonometry in healthy volunteers: effect of ranitidine on calculated intramural ph.," *Critical care medicine*, vol. 19, pp. 271–274, February 1991.
- [11] R. G. Fiddian-Green and S. Baker, "Predictive value of the stomach wall ph for complications after cardiac operations: comparison with other monitoring.," *Crit Care Med*, vol. 15, pp. 153–156, Feb 1987.
- [12] T. Gys, A. Hubens, H. Neels, L. F. Lauwers, and R. Peeters, "Prognostic value of gastric intramural ph in surgical intensive care patients.," *Critical care medicine*, vol. 16, pp. 1222–1224, December 1988.
- [13] P. E. Marik, "Sublingual capnography: a clinical validation study.," *Chest*, vol. 120, pp. 923–927, Sep 2001.
- [14] H. P. Povoas, M. H. Weil, W. Tang, B. Moran, T. Kamohara, and J. Bisera, "Comparisons between sublingual and gastric tonometry during hemorrhagic shock," *Chest*, vol. 118, pp. 1127–32, Oct 2000.
- [15] E. C. Rackow, P. O'Neil, M. E. Astiz, and C. M. Carpati, "Sublingual capnometry and indexes of tissue perfusion in patients with circulatory failure.," *Chest*, vol. 120, pp. 1633–1638, November 2001.
- [16] A. H. Kyle, C. T. Chan, and A. I. Minchinton, "Characterization of three-dimensional tissue cultures using electrical impedance spectroscopy," *Biophys J*, vol. 76, pp. 2640–8, May 1999. Journal Article.
- [17] E. Gersing, "Impedance spectroscopy on living tissue for determination of the state of organs," *Bioelectrochem Bioenerg*, vol. 45, pp. 145–49, 1998.
- [18] S. Kun and R. A. Peura, "Tissue ischemia detection using impedance spectroscopy," in *Proc. 16th Annual International Conference of the IEEE Engineering Advances: New Opportunities for Biomedical Engineers Engineering in Medicine and Biology Society*, pp. 868–869, 3–6 Nov. 1994.
- [19] M. Kun, Steven (Worcester, M. Peura, Robert A. (Princeton, and M. Ristic, Borislav (Worcester, "Impedance spectroscopy system for ischemia monitoring and detection," September 1998.

- [20] O. Casas, R. Bragos, P. J. Riu, J. Rosell, M. Tresanchez, M. Warren, A. Rodriguez-Sinovas, A. Carreno, and J. Cinca, "In vivo and in situ ischemic tissue characterization using electrical impedance spectroscopy," *Ann N Y Acad Sci*, vol. 873, pp. 51–8, Apr 20 1999.
- [21] A. Ivorra, M. Genescà, A. Sola, L. Palacios, R. Villa, G. Hotter, and J. Aguiló, "Bioimpedance dispersion width as a parameter to monitor living tissues.," *Physiol Meas*, vol. 26, pp. S165–S173, Apr 2005.
- [22] J. Songer, S. Kun, and S. Makarov, "Tissue ischemia monitoring using impedance spectroscopy: evaluation of neural networks for ischemia estimation," pp. 15–16, August 2002.
- [23] S. Othman and E. Sacristan, "Complex impedance spectrometer using parallel demodulation and digital conversion," *US Patent No. 6,970,738 B1*, 2005.
- [24] E. Sacristan, "Impedance spectroscopy system and catheter for ischemic mucosal damage monitoring in hollow viscous organs," *US Patent No. 6,882,879 B2*, 2005.
- [25] E. Sacristan, "Impedance spectroscopy method for monitoring ischemic mucosal damage in hollow viscous organs," *US Patent No. 6,965,795 B2.*, 2005.
- [26] C. A. Gonzalez, C. Villanueva, S. Othman, R. Narvaez, and E. Sacristan, "Impedance spectroscopy for monitoring ischemic injury in the intestinal mucosa," *Physiol Meas*, vol. 24, pp. 277–89, May 2003. Journal Article Research Support, Non-U.S. Gov't.
- [27] Y. Sánchez, "Modelo bioeléctrico del tejido intestinal sometido a espectroscopia de impedancia.," Master's thesis, Universidad Autónoma Metropolitana - Iztapalapa, Diciembre 2003.
- [28] S. Othman, *Caracterizacion in-vivo de la espectrometria de impedancia como tecnica de monitoreo de daño isquemico tisular*. PhD thesis, Posgrado en Ingeniería Biomédica. Universidad Autónoma Metropolitana Unidad Iztapalapa., 2003.
- [29] E. Sacristan, R. Baheza, A. Rodriguez, M. Does, B. Damon, and J. Gore, "Dual frequency coil and mris method for the characterization of ischemic injury in the intestinal wall," *25th Annual IEEE-EMBS Conference, Cancun Mexico*, p. 4, 2003.
- [30] C. A. González Díaz, *Validación de la Terapia Guiada por Espectroscopía de Impedancia Electrica Gástrica en un Modelo Experimental de Choque Séptico Inducido*. PhD thesis, Universidad Autónoma Metropolitana - Iztapalapa, 2003a.
- [31] N. Beltrán, U. Cerón, G. Sánchez, M. Remolina, and E. Sacristán, "Estudio multicéntrico de espectroscopia de impedancia gástrica en pacientes en estado crítico.," *Rev Asoc Mex Med Crit Ter Int*, vol. 21 (1), pp. 21–25, 2007.
- [32] R. G. Fiddian-Green, "Role of the gut in shock and resuscitation," *Bailliere's Clinical Anaesthesiology*, vol. 5, no. 1, pp. 75–99, 1991. Cited By (since 1996): 4.

- [33] P. E. Marik, "Sublingual capnometry: a non-invasive measure of microcirculatory dysfunction and tissue hypoxia.," *Physiol Meas*, vol. 27, pp. R37–R47, Jul 2006.
- [34] C. Ince and M. Sinaasappel, "Microcirculatory oxygenation and shunting in sepsis and shock.," *Crit Care Med*, vol. 27, pp. 1369–1377, Jul 1999.
- [35] U. Haglund, G. B. Bulkley, and D. N. Granger, "On the pathophysiology of intestinal ischemic injury. clinical review," *Acta Chir Scand*, vol. 153, no. 5-6, pp. 321–4, 1987.
- [36] K. R. Foster and H. P. Schwan, "Dielectric properties of tissues and biological materials: a critical review.," *Critical Reviews in Biomedical Engineering*, vol. 17, no. 1, pp. 25–104, 1989. Cited By (since 1996): 401.
- [37] J. J. Ackmann and M. A. Seitz, "Methods of complex impedance measurements in biologic tissue.," *Critical Reviews in Biomedical Engineering*, vol. 11, no. 4, pp. 281–311, 1984. Cited By (since 1996): 82.
- [38] H. P. Schwan, "Electrical properties of tissue and cell suspensions.," *Advances in Biological and Medical Physics*, vol. 5, pp. 147–209, 1957. Cited By (since 1996): 303.
- [39] A. Despopoulos and S. Silbernagl, *Color Atlas of Physiology (Thieme Flexibook)*. Thieme Medical Publishers, 4th ed., January 1991.
- [40] H. Schwan, "The bioimpedance field: some historical observations. in gersing, e. and schaefer, m. (eds.)," *IX International Conference on Electrical Bio- Impedance*, vol. 1, pp. 1–4, 1995.
- [41] D. W. Kim, "Detection of physiological events by impedance.," *Yonsei medical journal*, vol. 30, no. 1, pp. 1–11, 1989.
- [42] B. Rigaud, J. P. Morucci, and N. Chauveau, "Bioelectrical impedance techniques in medicine. part i: Bioimpedance measurement. second section: impedance spectrometry.," *Critical reviews in biomedical engineering*, vol. 24, no. 4-6, pp. 257–351, 1996.
- [43] S. Grimnes and O. G. Martinsen, "Comments on algorithm for tissue ischemia estimation based on electrical impedance spectroscopy.," *IEEE transactions on bio-medical engineering*, vol. 54, February 2007.
- [44] B. Ristic, S. Kun, and R. A. Peura, "Muscle tissue ischemia monitoring using impedance spectroscopy: quantitative results of animal studies," in *Proc. 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 5, pp. 2108–2111, 30 Oct.–2 Nov. 1997.
- [45] B. Rigaud, L. Hamzaoui, N. Chauveau, M. Granie, J. P. Scotto Di Rinaldi, and J. P. Morucci, "Tissue characterization by impedance: a multifrequency approach," *Physiol Meas*, vol. 15 Suppl 2a, pp. A13–20, May 1994.

-
- [46] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2 ed., November 2000.
- [47] E. J. Ciaccio, S. M. Dunn, and M. Akay, “Biosignal pattern recognition and interpretation systems,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, pp. 89–95, Sept 1993.
- [48] E. J. Ciaccio, S. M. Dunn, and M. Akay, “Biosignal pattern recognition and interpretation systems. 3. methods of classification,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 13, pp. 129–135, Feb 1994.
- [49] E. J. Ciaccio, S. M. Dunn, and M. Akay, “Biosignal pattern recognition and interpretation systems. 2. methods for feature extraction and selection,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, pp. 106–113, Dec 1993.
- [50] P. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.
- [51] D. Zongker and A. Jain, “Algorithms for feature selection: An evaluation,” in *Proc. 13th International Conference on Pattern Recognition*, vol. 2, pp. 18–22 vol.2, 1996.
- [52] A. Jain and D. Zongker, “Feature selection: evaluation, application, and small sample performance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153–158, Feb 1997.
- [53] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler, “Yale: Rapid prototyping for complex data mining tasks,” in *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (L. Ungar, M. Craven, D. Gunopulos, and T. Eliassi-Rad, eds.), (New York, NY, USA), pp. 935–940, ACM, August 2006.
- [54] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [55] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1st ed. 1995. corr. 2nd printing ed., December 1995.
- [56] H. W. Kuhn and A. W. Tucker, “Nonlinear programming,” in *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), pp. 481–492, University of California Press, Berkeley, CA, USA, 1951.
- [57] C. Chih-Chung and L. Chih-Jen, “Libsvm: a library for support vector machines.” 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [58] C. Chih-Chung and L. Chih-Jen, “Libsvm: a library for support vector machines - implementation details,” February 2009. <http://www.csie.ntu.edu.tw/~cjlin>.

-
- [59] H. Chih-Wei, C. Chih-Chung, and L. Chih-Jen, “A practical guide to support vector classification,” October 2008. <http://www.csie.ntu.edu.tw/~cjlin>.
- [60] T. Fawcett, “An introduction to roc analysis,” *Pattern Recogn. Lett.*, vol. 27, pp. 861–874, June 2006.

Apéndice A

Tipos de Errores

En este apartado se muestra la morfología de cada tipo de errores graficado en amplitud, fase, resistencia y reactancia tanto datos crudos como promediados \pm DS como se pueden observar en las figuras A.1 a A.16.

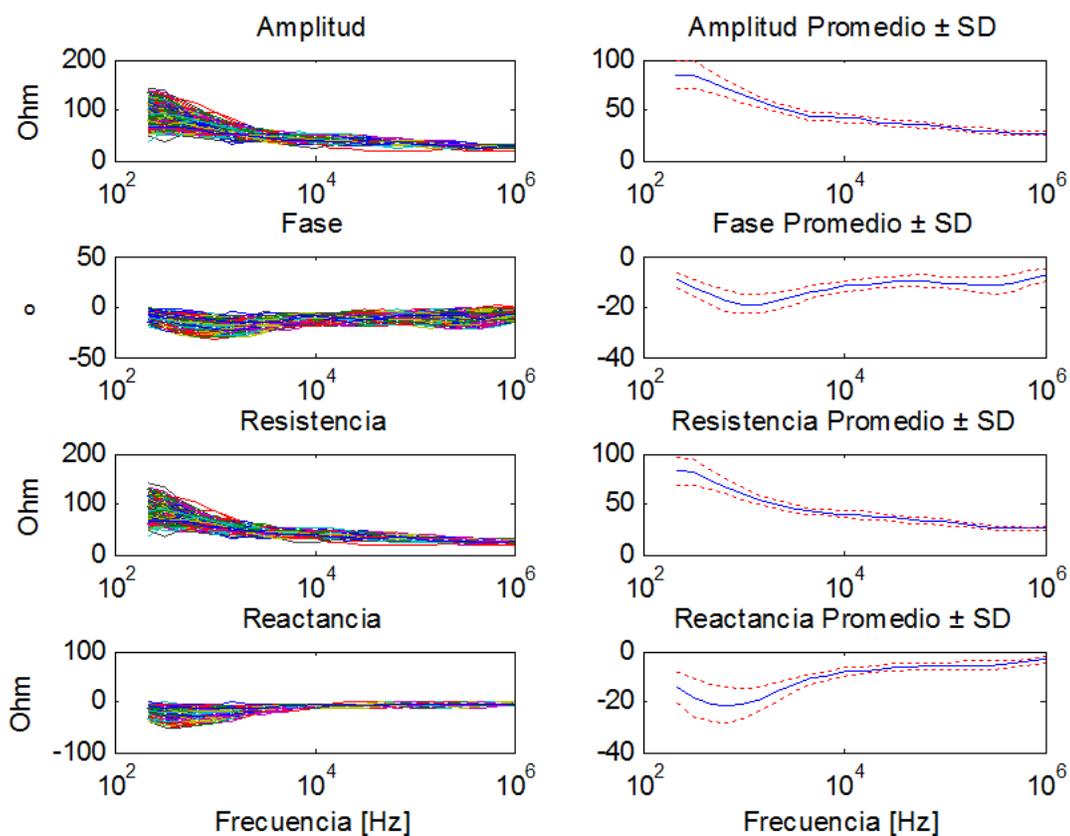


Figura A.1: *Espectros sin error*

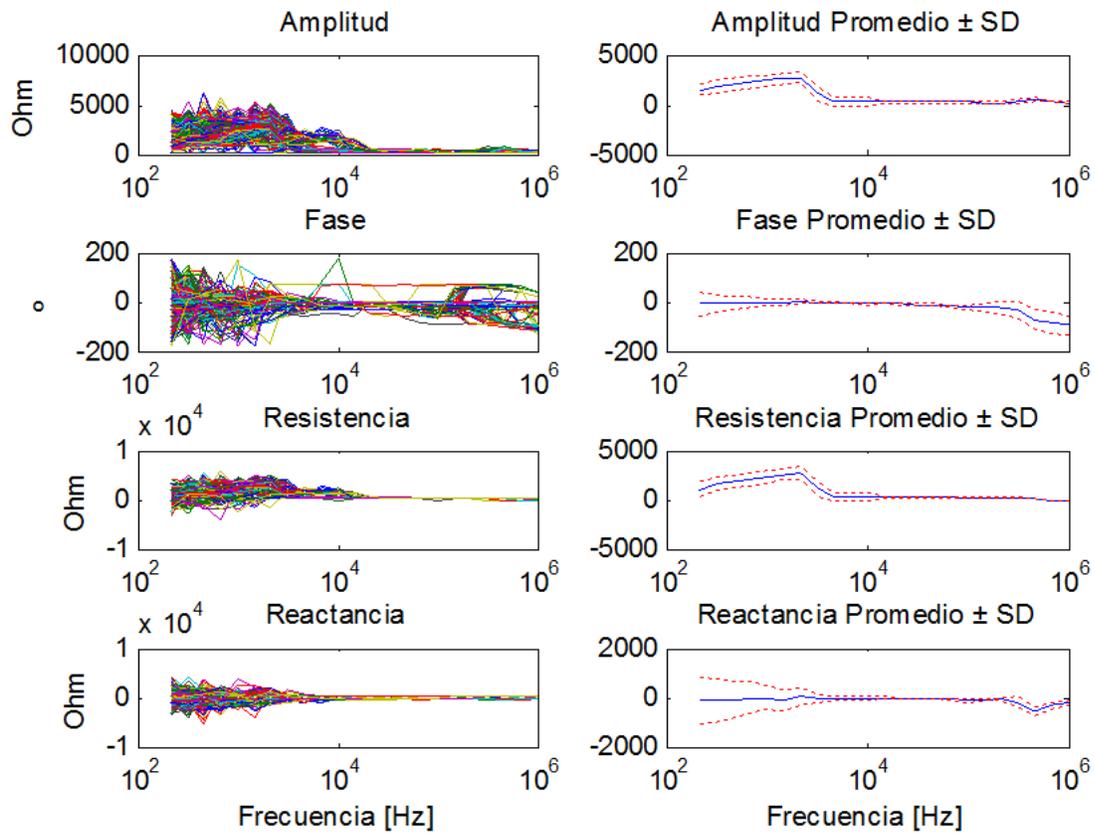


Figura A.2: *Cable Desconectado*

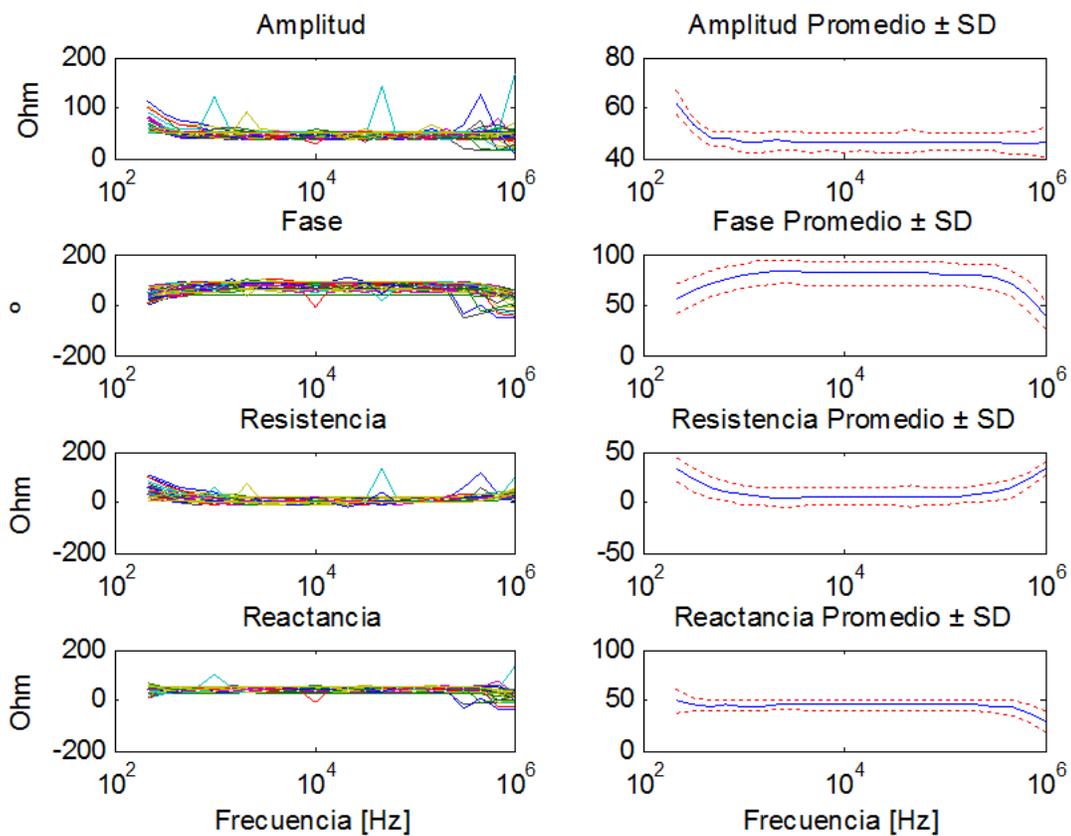


Figura A.3: Cable Descompuesto 1

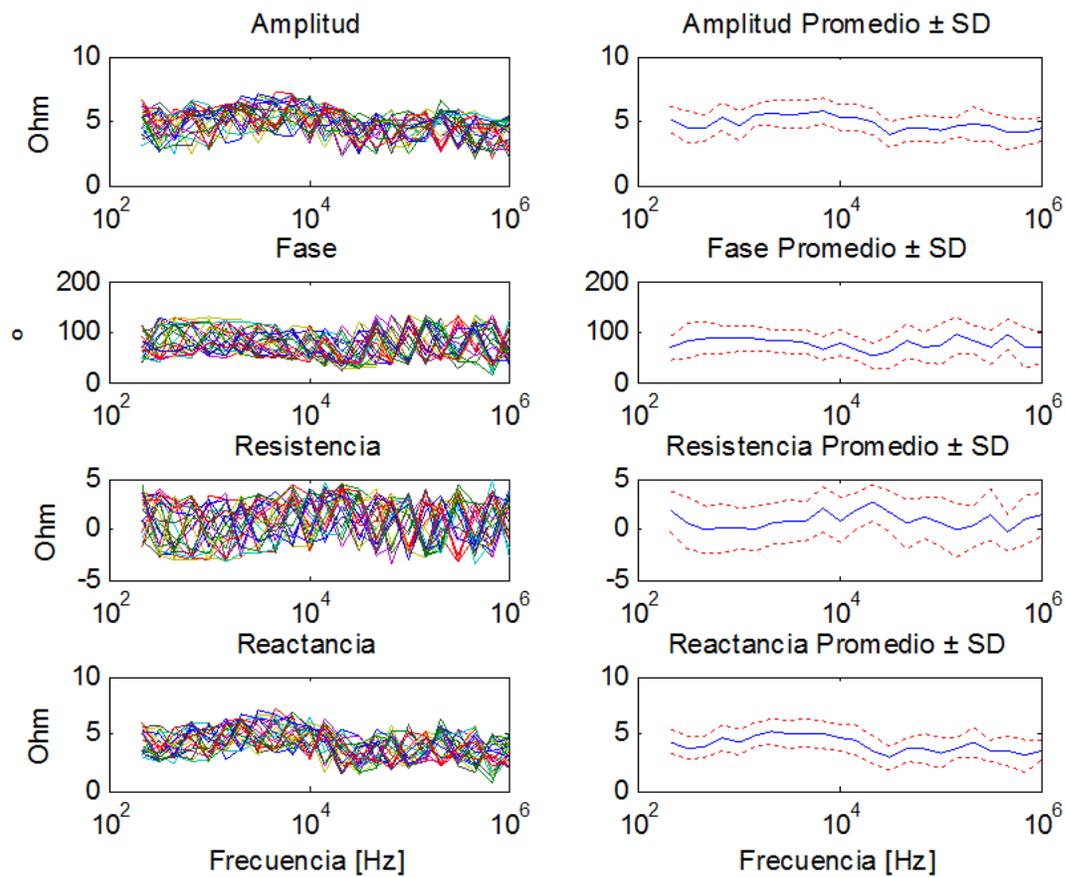


Figura A.4: Sonda Descompuesta

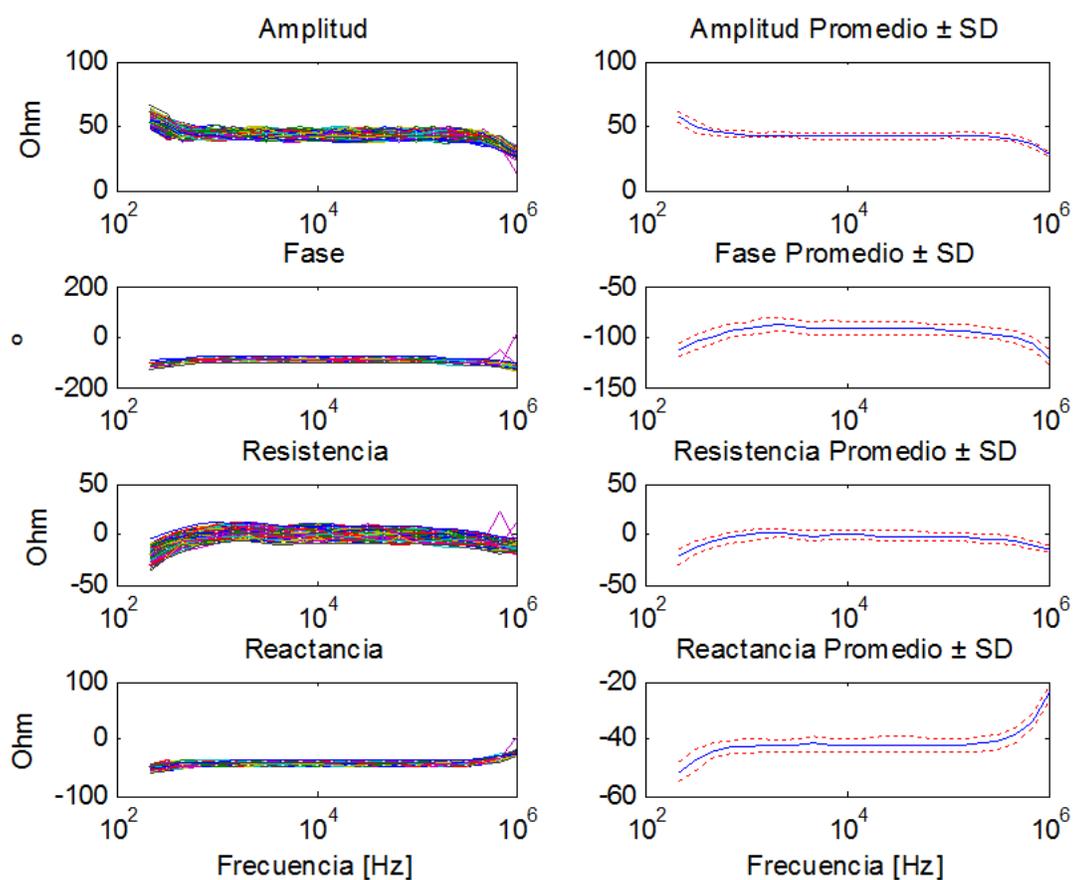


Figura A.5: Cable Descompuesto 2

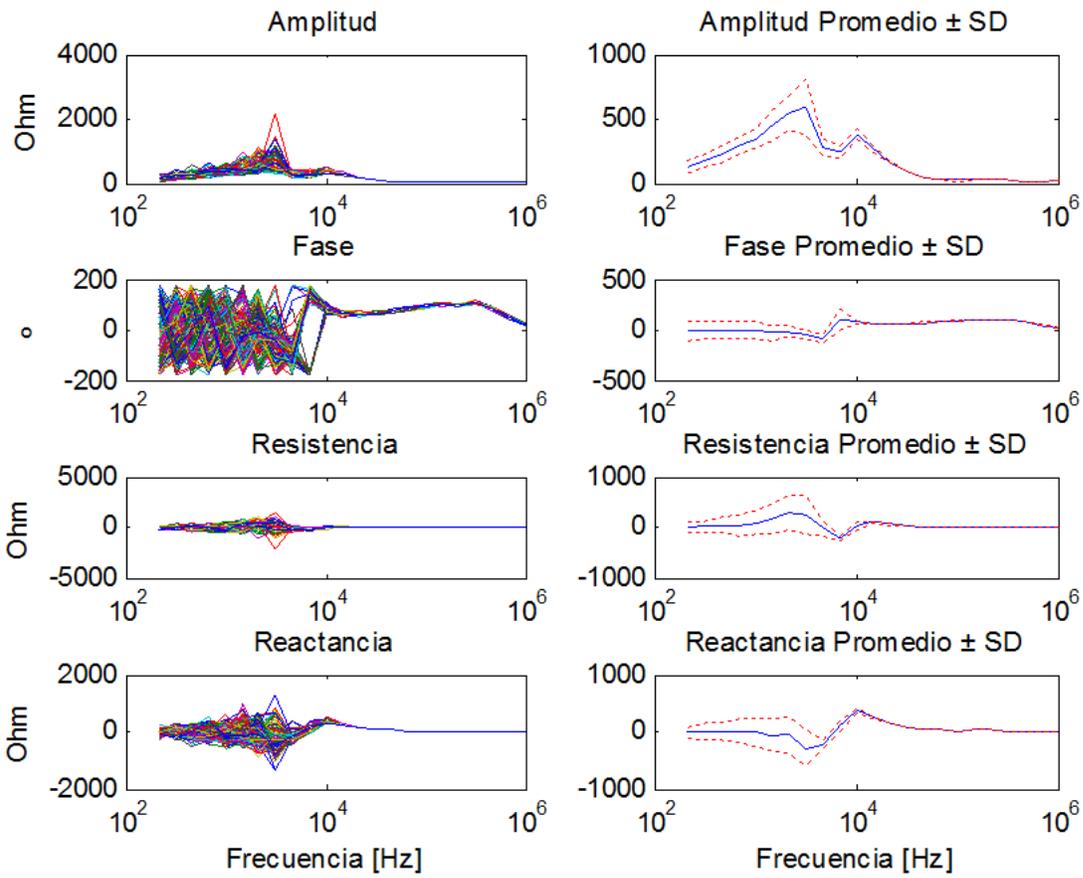


Figura A.6: Cable Descompuesto 3

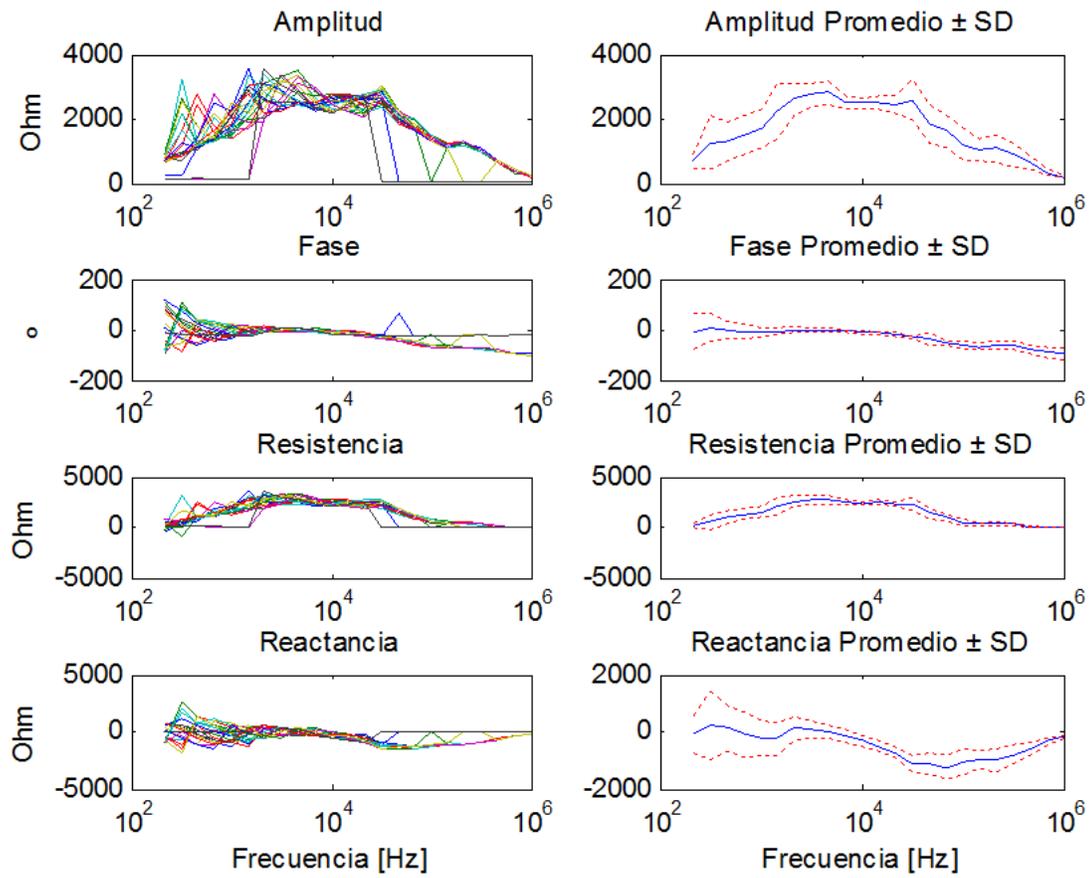


Figura A.7: Cable Descompuesto 4

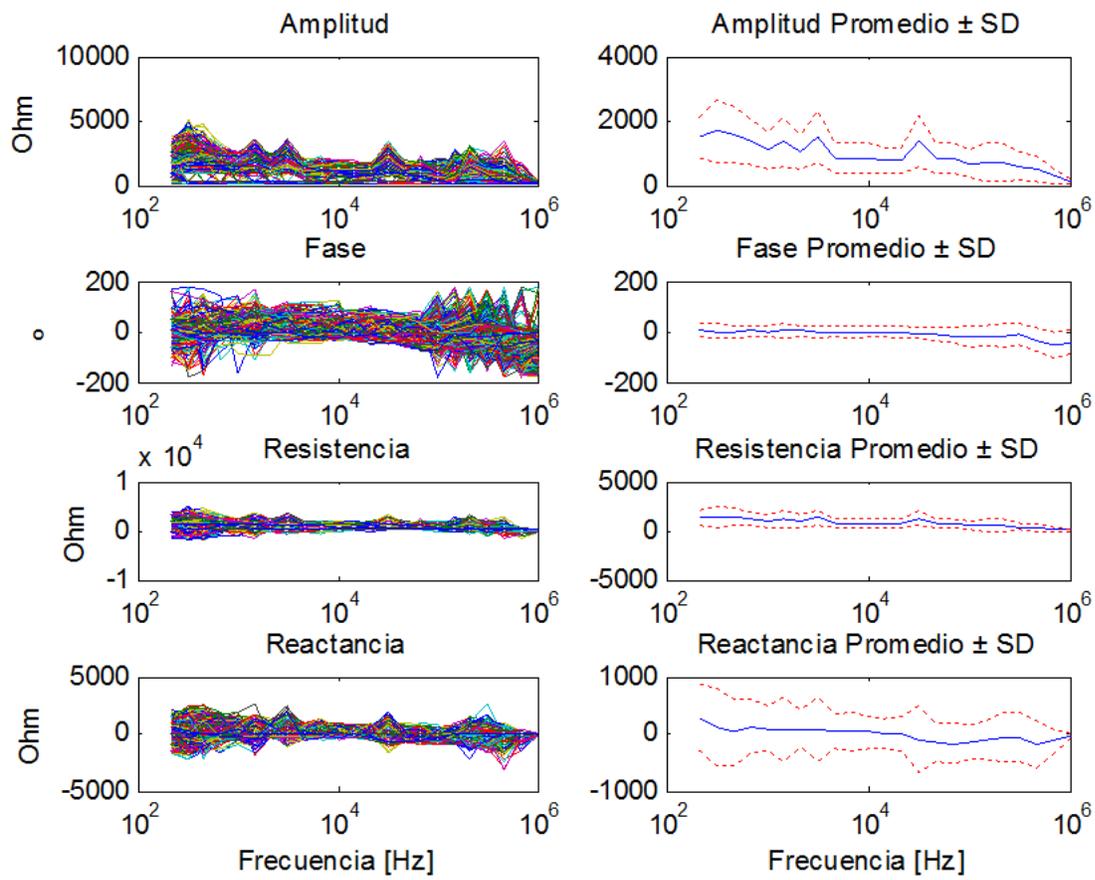


Figura A.8: Cable Descompuesto 5

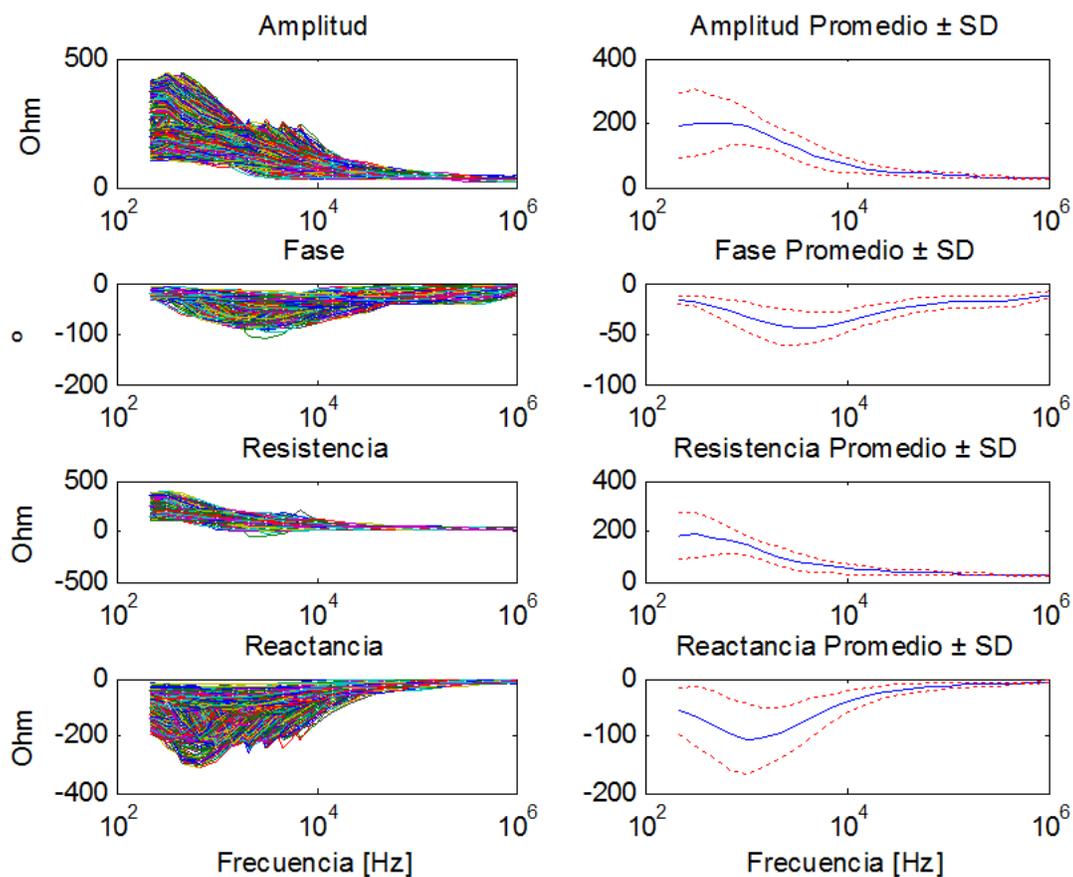


Figura A.9: Boca-Esófago

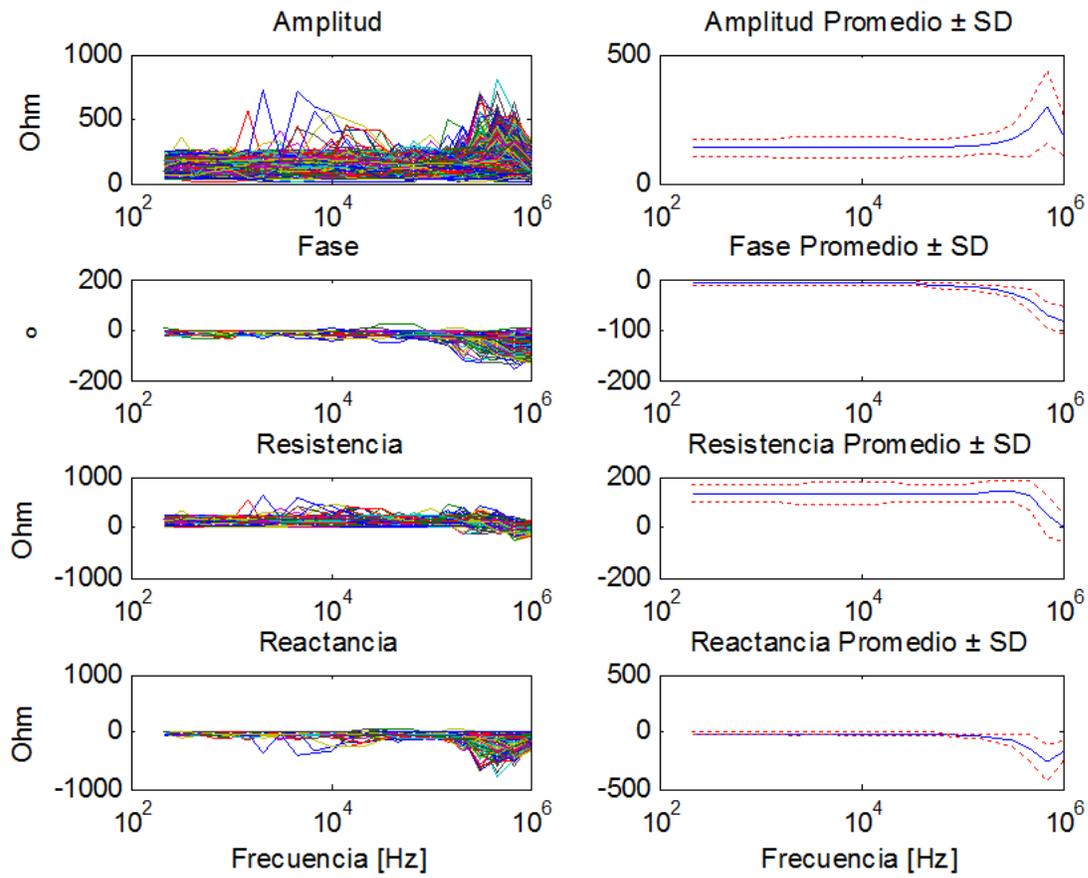


Figura A.10: *Contenido Gástrico 1*

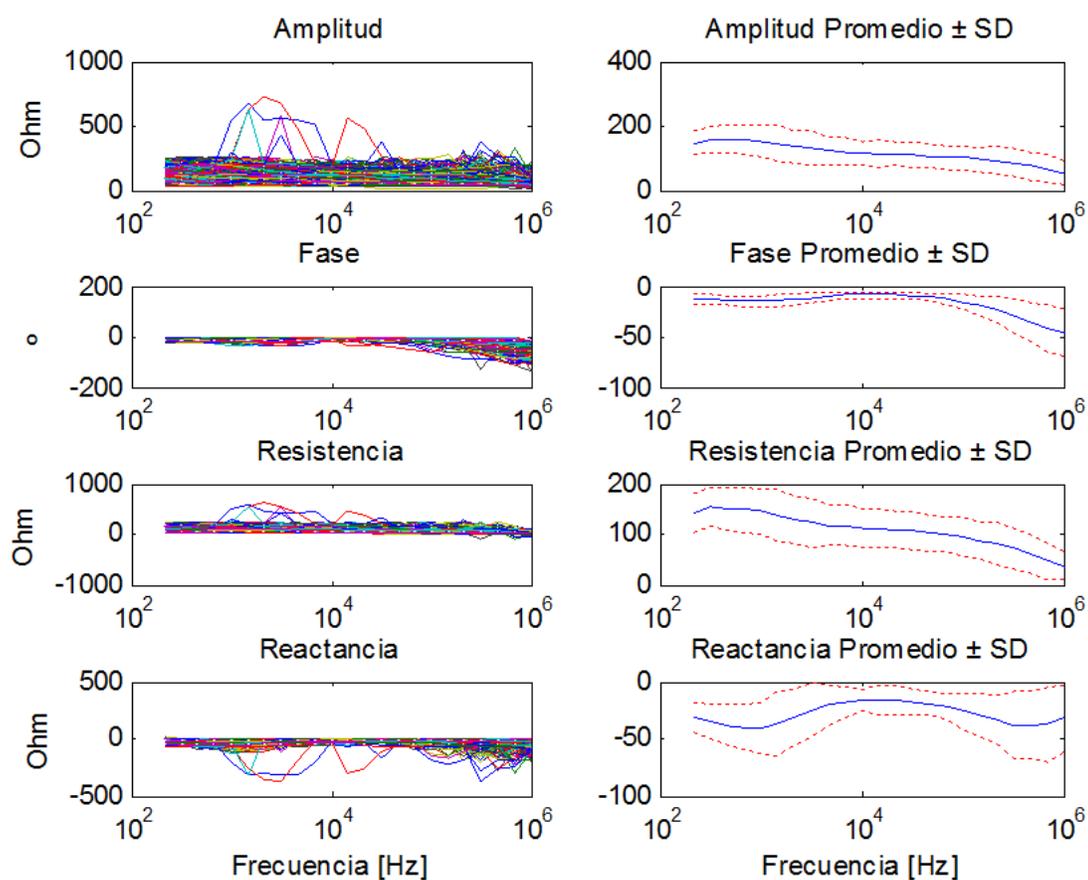


Figura A.11: *Contenido Gástrico 2*

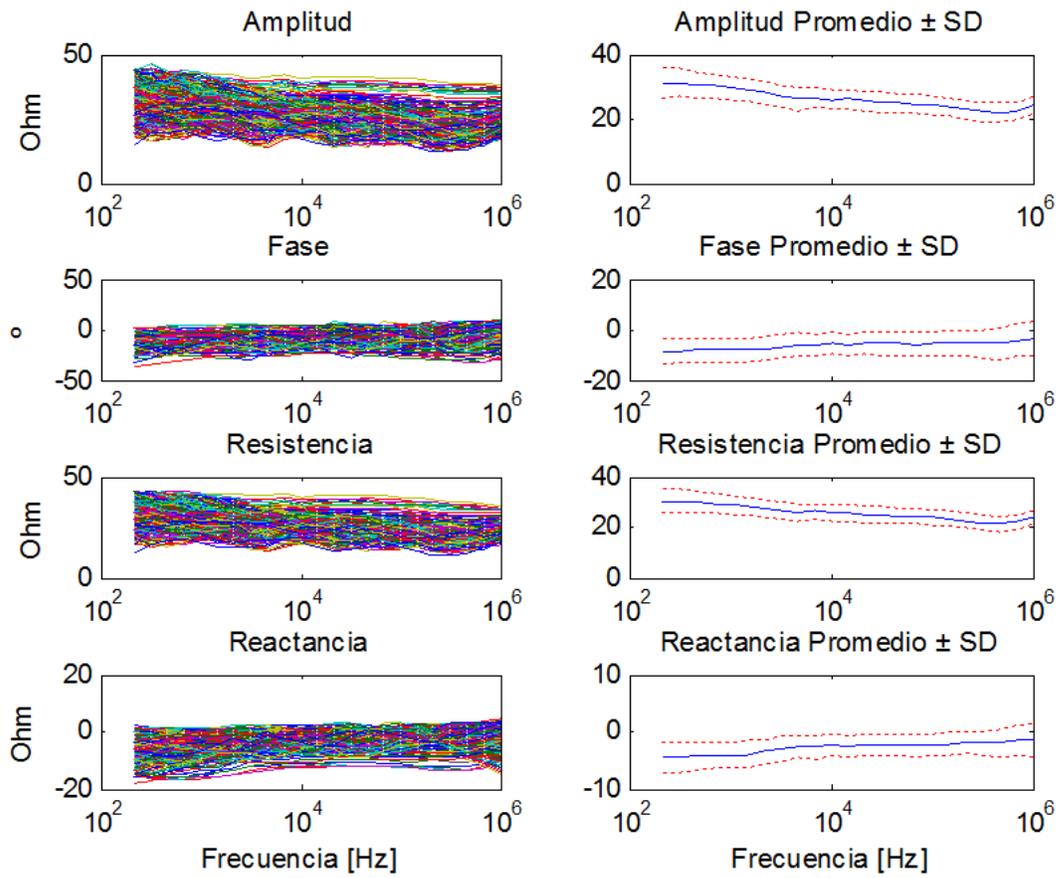


Figura A.12: Plano

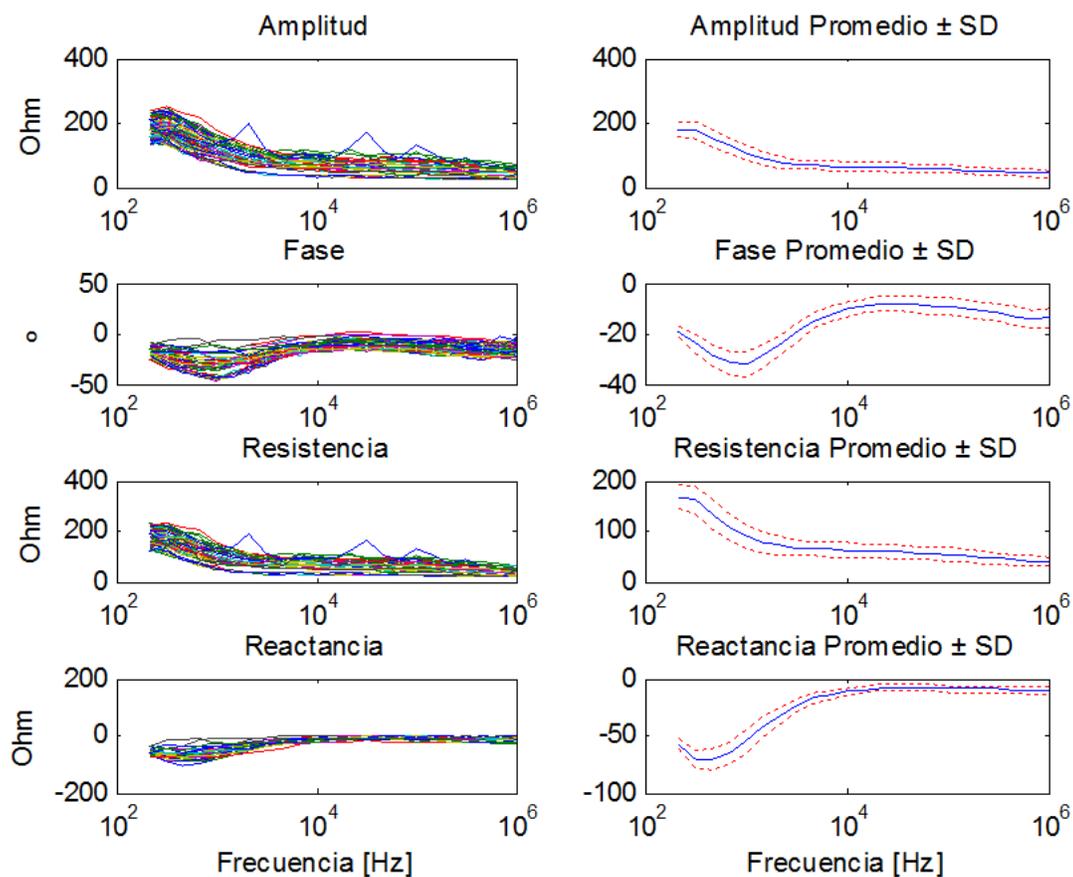


Figura A.13: *Alto*

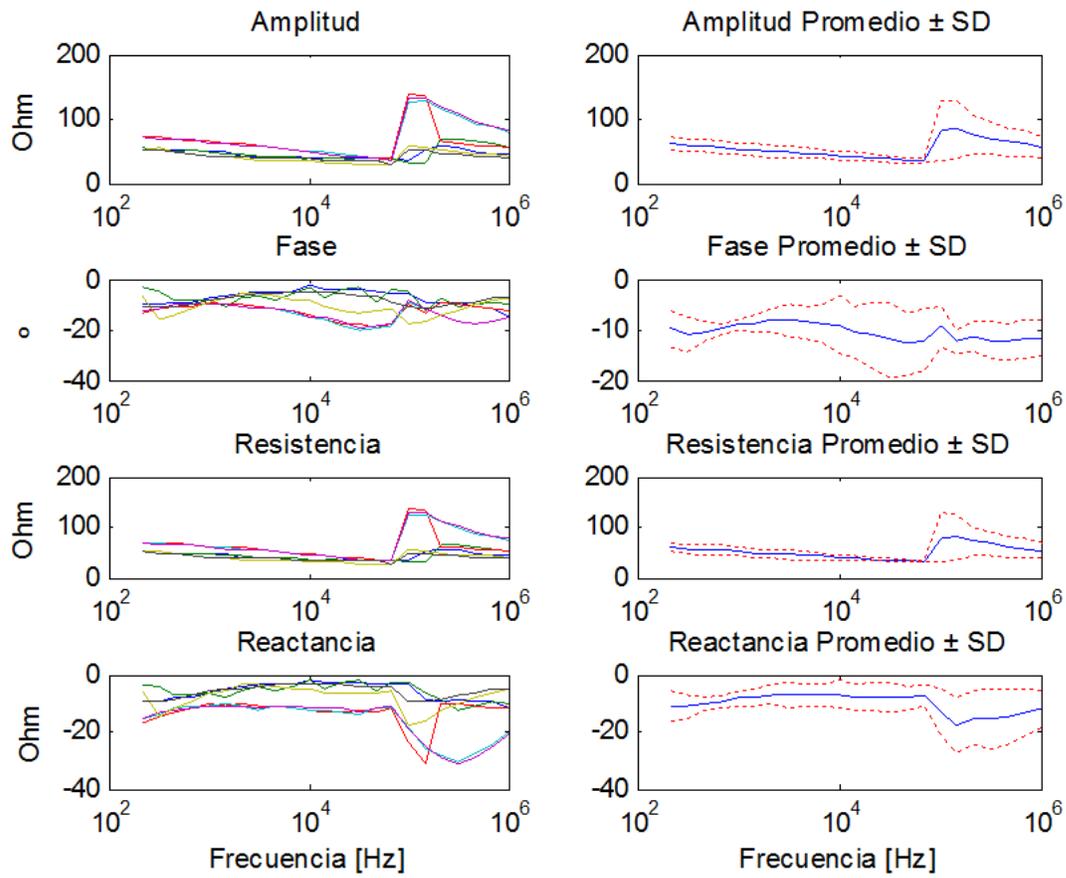


Figura A.14: Aire

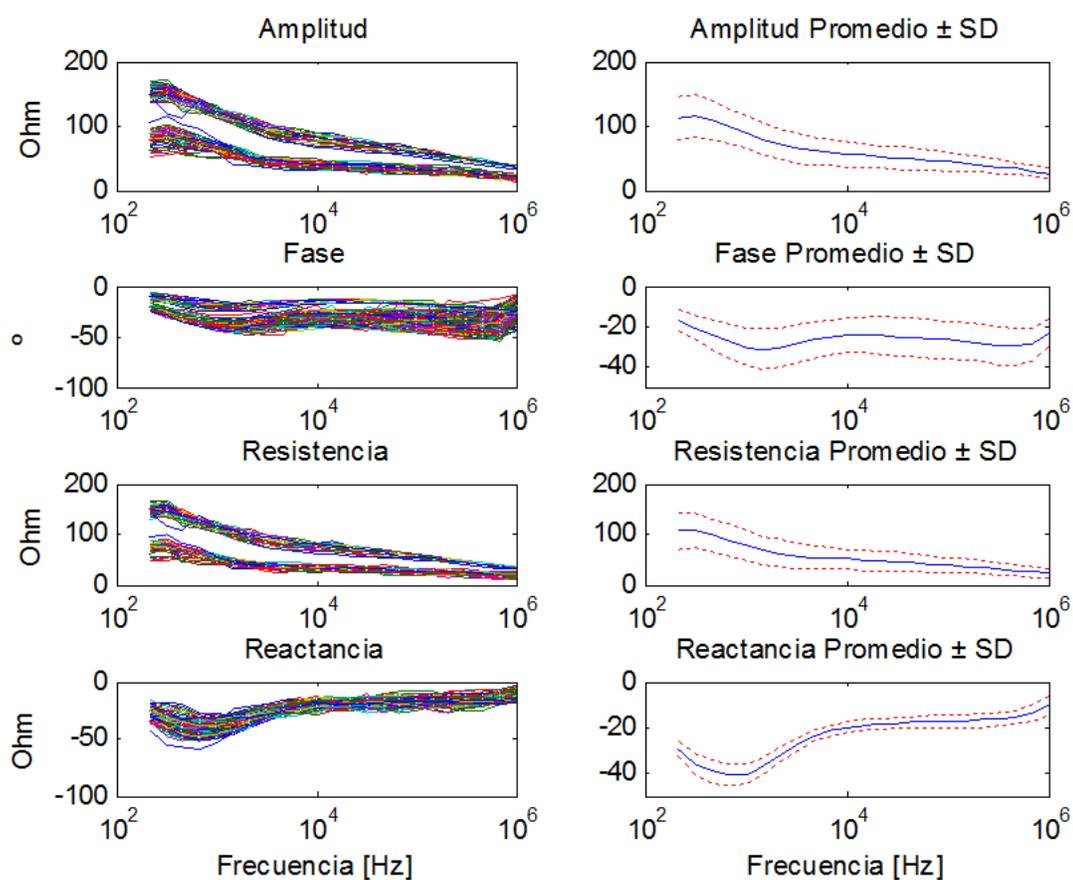


Figura A.15: *Offset*

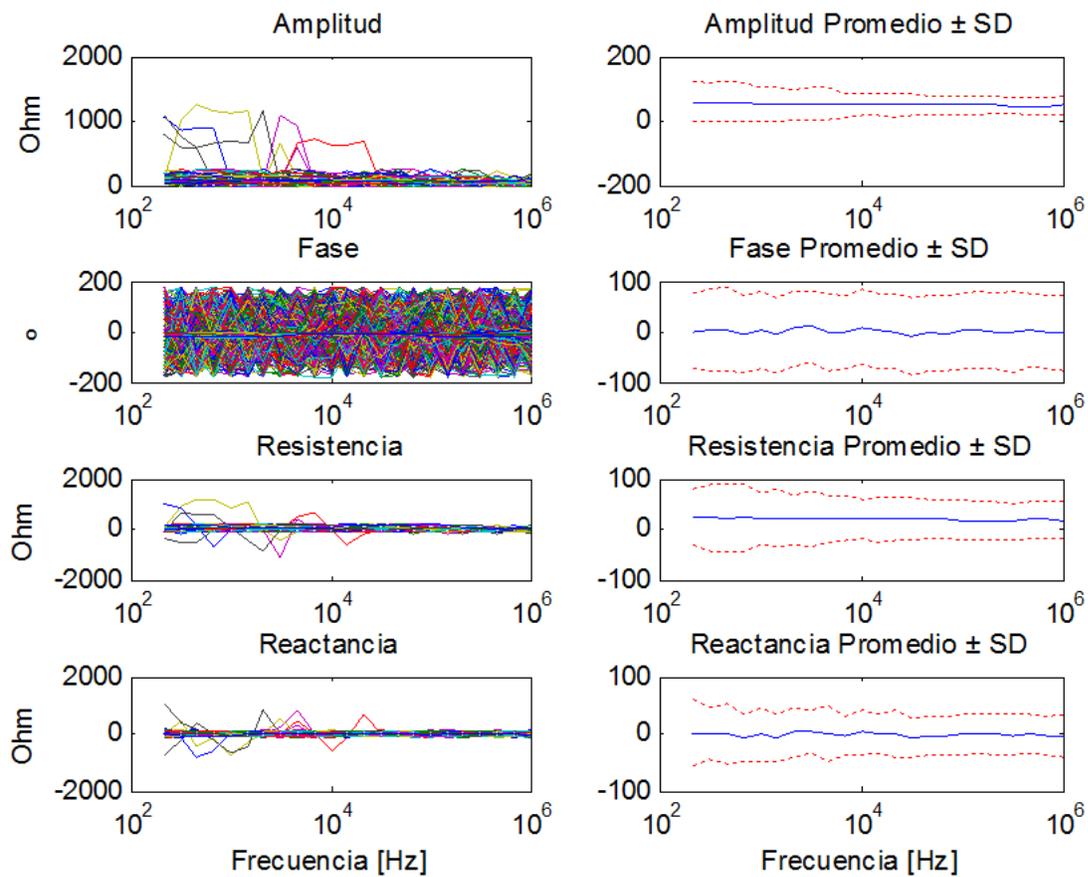


Figura A.16: *Altas Variaciones*

Apéndice B

Conformación de M_3 y m_3

En este apartado se explicará a detalle cómo fue creada la matriz M_3 y m_3 . Ésta la configuración que se ocupó para la clasificación y en fue la base de solución para la implementación del algoritmo de detección de errores. Con base a las características inherentes de cada espectro y a los hallazgos encontrados en el análisis de conglomerados, se asignaron nuevas etiquetas a los espectros con la finalidad de encontrar un mejor espacio para la clasificación.

Recordando que M_1 (20908 x 92) está clasificada en 16 tipos de espectros y 4 tipos de clases 4.1 y ésta fue redimensionada en cuanto al número de características y reclasificada con la fusión de algunos espectros dando origen a M_2 4.4 (20908 x 32) con 14 tipos de espectros y 4 tipos de clases. Las consideraciones que se tomaron en cuenta para las nuevas clases por cada tipo de espectro fue la siguiente:

- Se ocupan 16 características (R y X) de M_2 pero con la clasificación de los 16 tipos de espectros de M_1 .
- Se observa el comportamiento en las mediciones de media y DS en cada banda de cada tipo de espectro (Apéndice A).
- Se observa el mayor porcentaje de la agrupación natural y sus traslapes con otros errores en resistencia y reactancia en todas las bandas de la tabla 5.3.
- Se agregan 6 características: 4 por la suma de promedios y DS en todas las bandas y 2 por la suma de las sumas obtenidas de las bandas en media y DS.

Una vez que se identificaron las características de cada tipo de espectro se conformo la siguiente clasificación:

- *Circuito Abierto (CA)*
Se conforma por los errores: 1,5,6,7 por presentar DS en el rango 172 - 1128 principalmente en las bandas uno y dos. Pero sigue perteneciendo a la clase de errores

por conexión asociados a cable desconectado o descompuesto que puede implicar un circuito abierto en las conexiones. Se asigna prioridad uno.

- *Corto Circuito (CC)*

Se conforma por los errores: 2, 3 y 4 por presentar DS en el rango 1 - 10 y por los promedios de resistencia fueron menores a los 12 Ω . Se le asigna prioridad 2.

- *Mala Posición (MP)*

Se conforma por el error 8, se reasignan los errores 9 y 10 ya que no representan un error en la colocación de la sonda sino más bien en el contacto con la mucosa gástrica. Se le asigna prioridad 3.

- *Mal Contacto (MC)*

Se conforma por los errores 9-15 por que estos representan todos los errores debido al mal contacto de la sonda con la mucosa gástrica. Se elimina la clase *Movimiento* y es fusionada en esta clase. Se le asigna prioridad 4.

En la tabla B.1 la clase A representa la clasificación de M_1 y clase B es la nueva clasificación propuesta que conforman a M_3 con las consideraciones mencionadas previamente. En la tabla B.2 vemos la agrupación por las nuevas clases propuestas. Por lo tanto la matriz M_3 tiene una dimensión de 20908 datos x 22 características que se muestran en la tabla B.3, con 5 clases y 16 tipos de espectros.

No. Error	Descripción	Clase A	Clase B	Prioridad	N
0	Buenos	Bueno	Bueno	-	2500
1	Cable desconectado	Conexiones	Circuito Abierto	1	3443
5	Cable descompuesto 3	Conexiones	Circuito Abierto	1	386
6	Cable descompuesto 4	Conexiones	Circuito Abierto	1	21
7	Cable descompuesto 5	Conexiones	Circuito Abierto	1	918
2	Cable descompuesto 1	Conexiones	Corto Circuito	2	958
3	Cable descompuesto 2	Conexiones	Corto Circuito	2	2815
4	Sonda descompuesta	Conexiones	Corto Circuito	2	24
8	Boca Esófago	Posición	Mala Posición	3	3232
9	Contenido Gástrico 1	Posición	Mal Contacto	4	1897
10	Contenido Gástrico 2	Posición	Mal Contacto	4	698
11	Plano	Contacto	Mal Contacto	4	2500
12	Alto	Contacto	Mal Contacto	4	254
13	Aire	Contacto	Mal Contacto	4	7
14	Offset	Contacto	Mal Contacto	4	344
15	Altas Variaciones	Movimiento	Mal Contacto	4	911

Tabla B.1: Base de Datos reclasificada y redimensionada (M_3)

Clase	Tipo	N
0	Bueno (B)	2500
1	Circuito Abierto (CA)	4768
2	Corto Circuito (CC)	3797
3	Mala Posición (MP)	3232
4	Mal Contacto (MC)	6611
	Total	20908

Tabla B.2: Base de Datos reclasificada y redimensionada (m_3)

Código	Característica
$R1_M$	Promedio de Resistencia en la banda 1
$R2_M$	Promedio de Resistencia en la banda 2
$R3_M$	Promedio de Resistencia en la banda 3
$R4_M$	Promedio de Resistencia en la banda 4
$R1_{DS}$	DS de Resistencia en la banda 1
$R2_{DS}$	DS de Resistencia en la banda 2
$R3_{DS}$	DS de Resistencia en la banda 3
$R4_{DS}$	DS de Resistencia en la banda 4
$X1_M$	Promedio de Reactancia en la banda 1
$X2_M$	Promedio de Reactancia en la banda 2
$X3_M$	Promedio de Reactancia en la banda 3
$X4_M$	Promedio de Reactancia en la banda 4
$X1_{DS}$	DS de Reactancia en la banda 1
$X2_{DS}$	DS de Reactancia en la banda 2
$X3_{DS}$	DS de Reactancia en la banda 3
$X4_{DS}$	DS de Reactancia en la banda 4
S_{RM}	Suma de los promedios de Resistencia en todas las bandas
S_{XM}	Suma de los promedios de Reactancia en todas las bandas
S_{RDS}	Suma de las DS de Resistencia en todas las bandas
S_{XDS}	Suma de las DS de Reactancia en todas las bandas
S_{RXM}	Suma de S_{RM} y S_{XM}
S_{RXDS}	Suma de S_{RDS} y S_{XDS}

Tabla B.3: Características de Base de datos (M_3 y m_3)

Apéndice C

Características de Espectros por Banda

En las figuras C.1 a C.4 se muestran los promedios y DS vs. el número de muestras por tipo de error en cada banda por variable. En las tablas C.1 a C.16 se muestran los promedios \pm DS por banda en amplitud, fase, resistencia y reactancia. En las figuras C.5 a C.8 se muestran combinaciones de las variables.

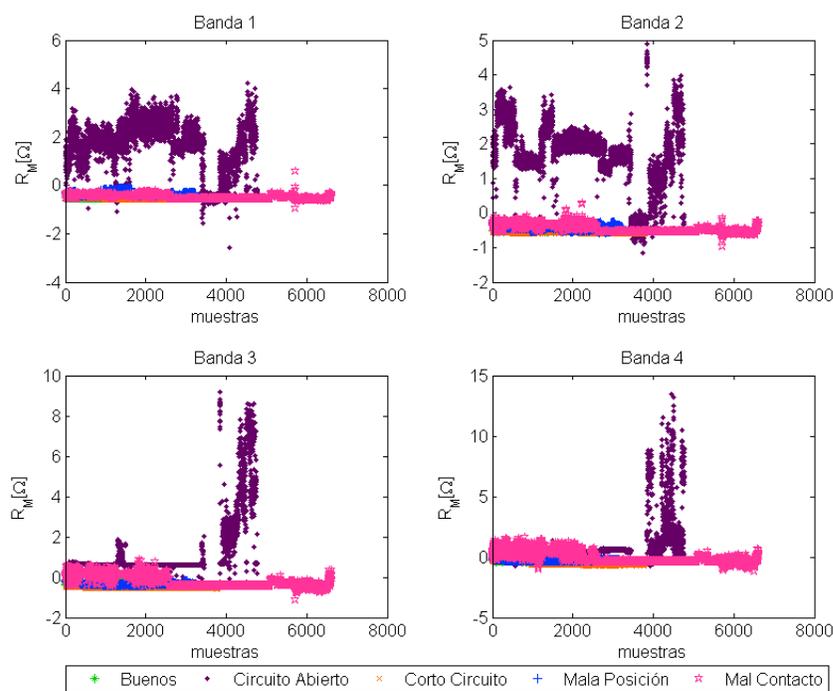


Figura C.1: R_M vs. muestras por banda

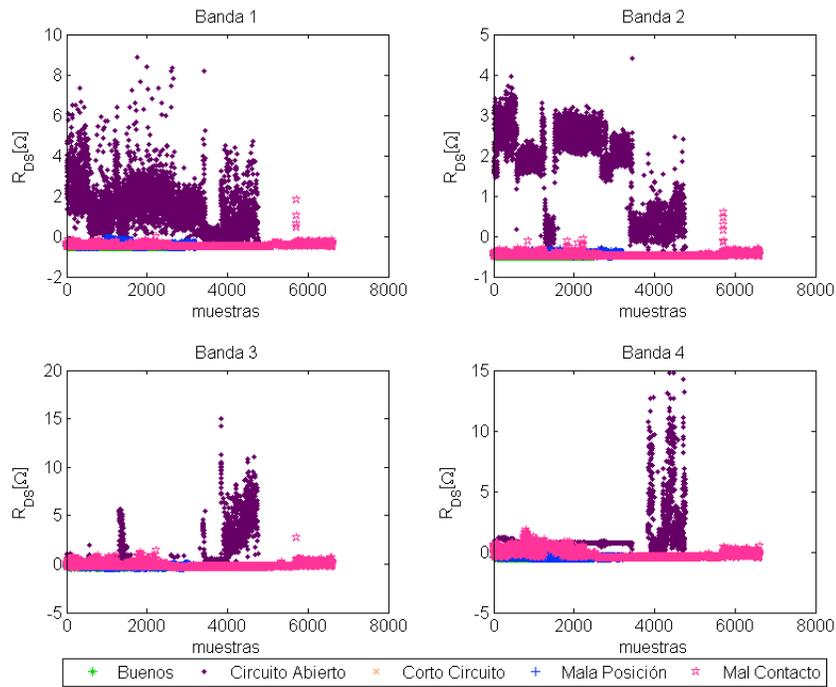


Figura C.2: R_{DS} vs. muestras por banda

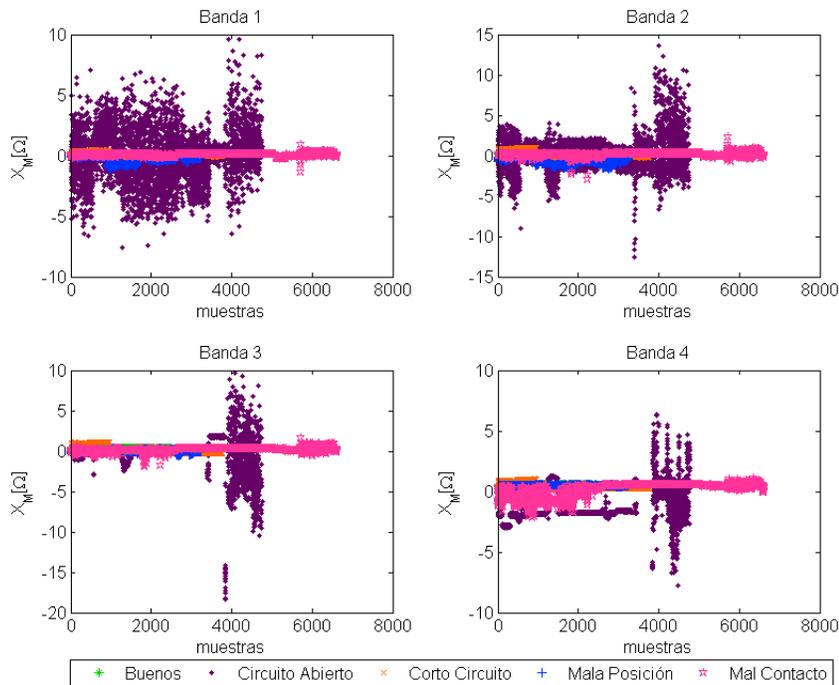


Figura C.3: X_M vs. muestras por banda

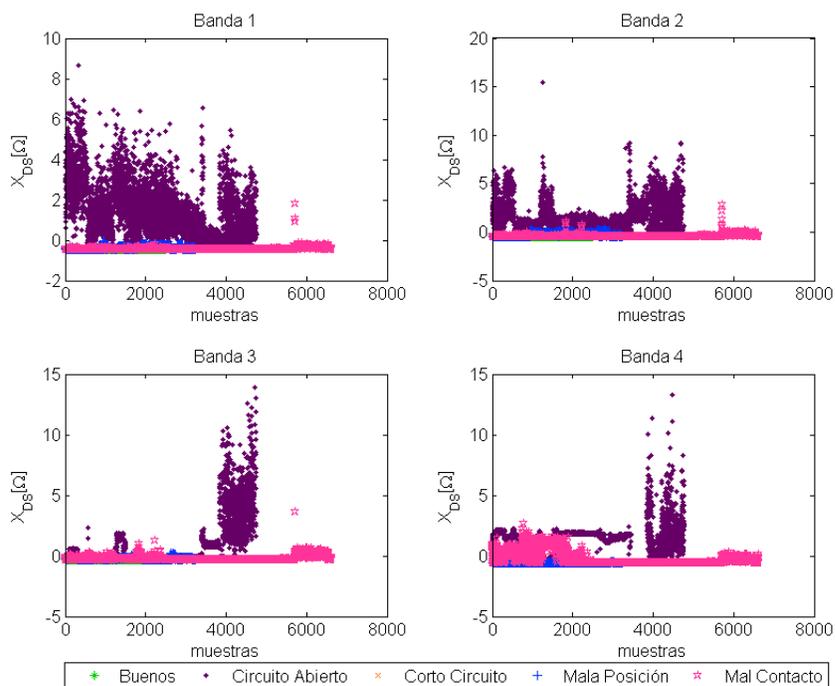


Figura C.4: X_{DS} vs. muestras por banda

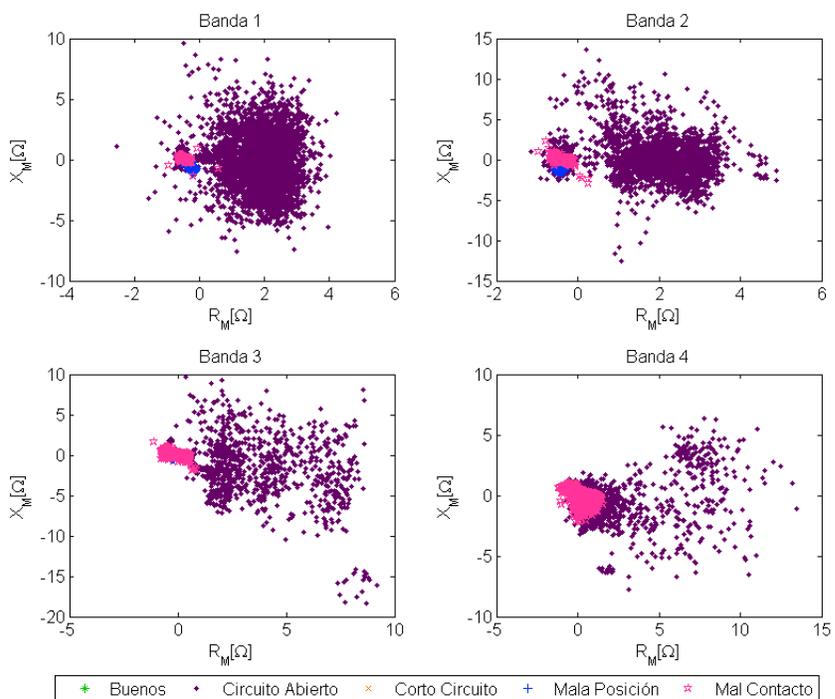


Figura C.5: R_M vs X_M

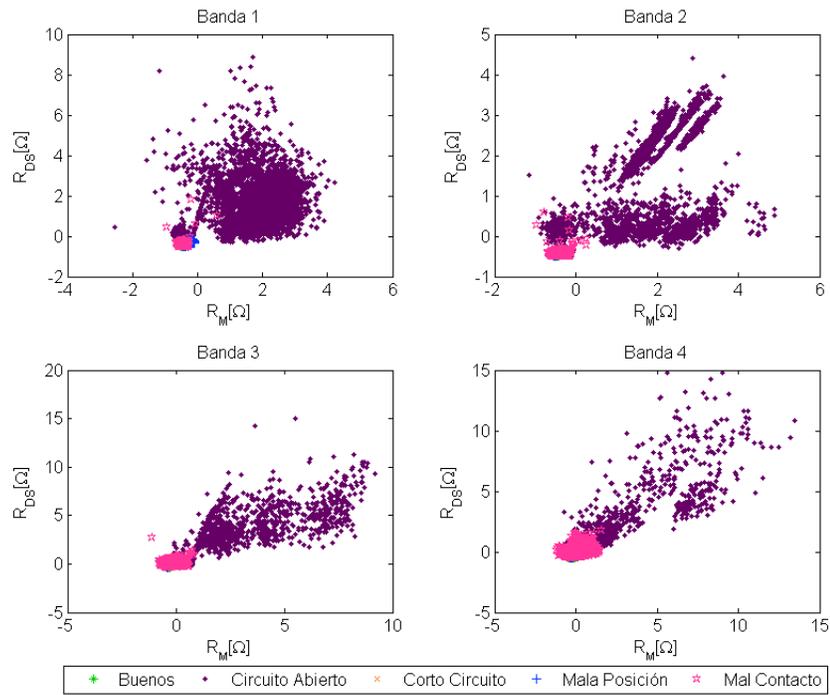


Figura C.6: $R_M vs R_{DS}$

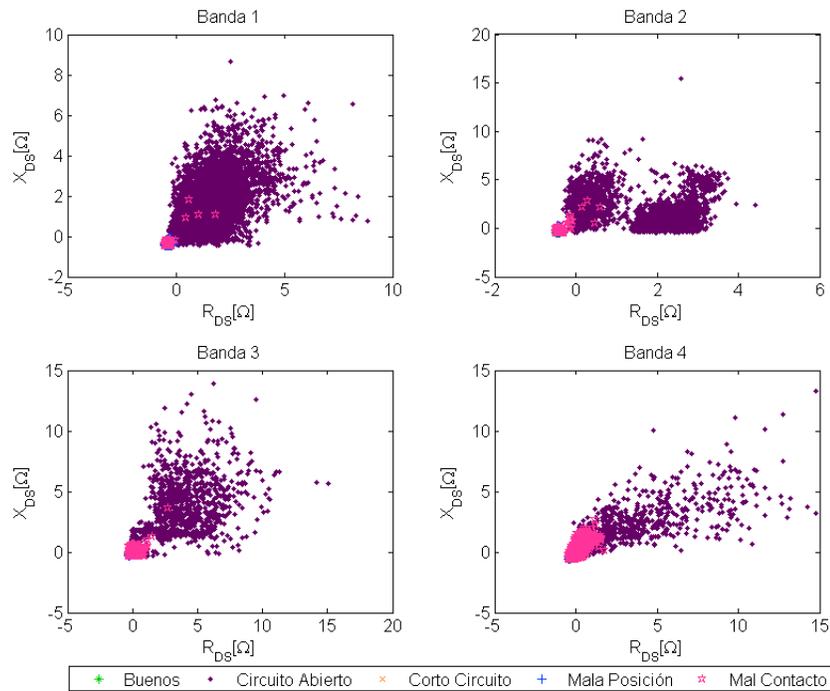


Figura C.7: $R_{DS} vs X_{DS}$

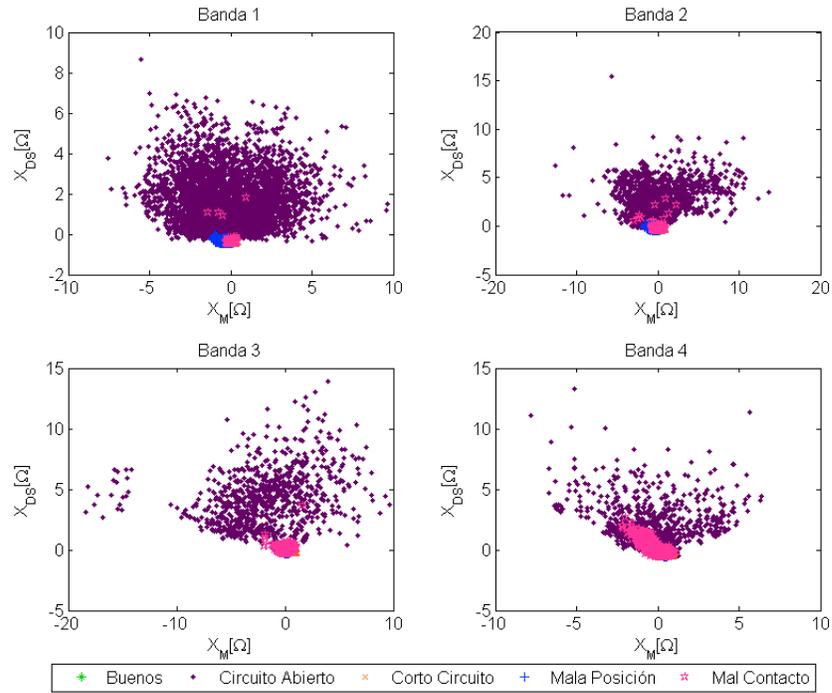


Figura C.8: X_M vs. X_{DS}

No. Error	Medida	Banda	Media	DS
0	<i>Amplitud</i>	1	76.25	10.03
0	[Ω]	2	47.79	6.71
0		3	36.40	3.08
0		4	28.00	2.39
0	<i>Fase</i>	1	-14.45	3.95
0	[°]	2	-15.22	3.04
0		3	-10.22	1.60
0		4	-10.03	2.56
0	<i>Resistencia</i>	1	73.55	10.76
0	[Ω]	2	45.92	5.77
0		3	35.79	3.00
0		4	27.52	2.31
0	<i>Reactancia</i>	1	-18.97	3.64
0	[Ω]	2	-12.79	4.24
0		3	-6.46	1.16
0		4	-4.91	1.34

Tabla C.1: Espectros sin error $N = 2500$.

No. Error	Medida	Banda	Media	DS
1	<i>Amplitud</i> [Ω]	1	2009.29	569.02
1		2	1279.42	1236.41
1		3	240.29	24.76
1		4	299.81	158.72
1	<i>Fase</i> [$^\circ$]	1	-3.11	23.54
1		2	-2.58	5.00
1		3	-12.27	6.59
1		4	-53.41	35.04
1	<i>Resistencia</i> [Ω]	1	1803.58	679.61
1		2	1264.63	1228.63
1		3	231.61	25.71
1		4	88.14	110.37
1	<i>Reactancia</i> [Ω]	1	-80.65	624.97
1		2	-32.34	189.53
1		3	-51.87	29.36
1		4	-235.64	196.56

Tabla C.2: Cable Desconectado $N = 3443$.

No. Error	Medida	Banda	Media	DS
2	<i>Amplitud</i> [Ω]	1	51.38	6.74
2		2	46.36	1.15
2		3	46.36	1.07
2		4	46.19	1.61
2	<i>Fase</i> [$^\circ$]	1	69.01	9.60
2		2	81.93	2.79
2		3	81.07	2.45
2		4	67.41	16.15
2	<i>Resistencia</i> [Ω]	1	18.11	10.29
2		2	5.76	2.14
2		3	6.50	1.89
2		4	16.04	10.86
2	<i>Reactancia</i> [Ω]	1	45.92	3.89
2		2	44.98	1.15
2		3	44.94	1.05
2		4	40.65	6.43

Tabla C.3: Cable Descompuesto 1 $N = 958$.

No. Error	Medida	Banda	Media	DS
3	<i>Amplitud</i> [Ω]	1	47.58	6.33
3		2	42.42	1.83
3		3	42.33	1.91
3		4	38.18	6.00
3	<i>Fase</i> [$^{\circ}$]	1	-100.47	9.05
3		2	-89.89	3.06
3		3	-91.91	2.84
3		4	-102.33	10.58
3	<i>Resistencia</i> [Ω]	1	-9.27	8.89
3		2	0.19	2.32
3		3	-1.30	2.18
3		4	-7.09	4.74
3	<i>Reactancia</i> [Ω]	1	-45.83	4.49
3		2	-42.15	1.79
3		3	-42.06	1.86
3		4	-36.88	7.37

Tabla C.4: Cable Descompuesto 2 $N = 2815$.

No. Error	Medida	Banda	Media	DS
4	<i>Amplitud</i> [Ω]	1	4.79	1.08
4		2	5.49	1.00
4		3	4.50	1.13
4		4	4.36	1.12
4	<i>Fase</i> [$^{\circ}$]	1	82.60	26.57
4		2	79.33	23.78
4		3	68.28	33.54
4		4	80.41	35.85
4	<i>Resistencia</i> [Ω]	1	0.56	2.10
4		2	0.89	2.11
4		3	1.47	2.29
4		4	0.66	2.45
4	<i>Reactancia</i> [Ω]	1	4.15	1.07
4		2	4.93	1.15
4		3	3.60	1.35
4		4	3.60	1.25

Tabla C.5: Sonda Descompuesta $N = 24$.

No. Error	Medida	Banda	Media	DS
5	<i>Amplitud</i> [Ω]	1	230.77	106.00
5		2	412.82	172.03
5		3	102.05	98.76
5		4	20.20	6.78
5	<i>Fase</i> [°]	1	-2.10	92.94
5		2	3.04	100.60
5		3	74.95	12.83
5		4	79.06	35.28
5	<i>Resistencia</i> [Ω]	1	47.63	172.08
5		2	97.52	307.33
5		3	38.53	45.41
5		4	1.32	9.32
5	<i>Reactancia</i> [Ω]	1	1.71	181.31
5		2	-24.61	328.24
5		3	93.46	88.45
5		4	17.44	9.04

Tabla C.6: Cable Descompuesto 3 N = 386.

No. Error	Medida	Banda	Media	DS
6	<i>Amplitud</i> [Ω]	1	1284.73	559.02
6		2	2560.37	474.66
6		3	2028.86	626.33
6		4	689.43	389.44
6	<i>Fase</i> [°]	1	-1.13	42.28
6		2	-2.15	9.11
6		3	-32.75	19.92
6		4	-73.40	13.94
6	<i>Resistencia</i> [Ω]	1	951.47	690.42
6		2	2530.00	482.16
6		3	1673.79	844.13
6		4	194.10	205.48
6	<i>Reactancia</i> [Ω]	1	-19.85	752.58
6		2	-84.26	384.07
6		3	-973.46	379.89
6		4	-649.70	353.40

Tabla C.7: Cable Descompuesto 4 N = 21.

No. Error	Medida	Banda	Media	DS
7	<i>Amplitud</i> [Ω]	1	1430.87	372.87
7		2	1058.86	324.09
7		3	873.65	319.37
7		4	482.12	291.57
7	<i>Fase</i> [$^\circ$]	1	3.84	16.22
7		2	3.56	15.81
7		3	-7.42	20.89
7		4	-26.63	35.93
7	<i>Resistencia</i> [Ω]	1	1279.79	403.51
7		2	966.05	325.59
7		3	780.75	320.29
7		4	337.66	312.97
7	<i>Reactancia</i> [Ω]	1	123.94	376.05
7		2	55.21	294.27
7		3	-97.28	284.50
7		4	-94.11	227.58

Tabla C.8: Cable Descompuesto 5 $N = 918$.

No. Error	Medida	Banda	Media	DS
8	<i>Amplitud</i> [Ω]	1	195.01	31.81
8		2	113.77	41.53
8		3	46.10	8.11
8		4	30.32	3.33
8	<i>Fase</i> [$^\circ$]	1	-23.37	7.21
8		2	-41.03	6.61
8		3	-24.39	5.64
8		4	-15.76	3.26
8	<i>Resistencia</i> [Ω]	1	172.72	29.50
8		2	81.38	28.44
8		3	40.90	5.23
8		4	28.95	2.85
8	<i>Reactancia</i> [Ω]	1	-81.47	23.88
8		2	-73.17	29.64
8		3	-19.88	7.49
8		4	-8.38	2.37

Tabla C.9: Boca-Esófago 5 $N = 3232$.

No. Error	Medida	Banda	Media	DS
9	<i>Amplitud</i>	1	136.91	11.24
9	[Ω]	2	136.97	13.13
9		3	138.96	16.84
9		4	192.96	85.09
9	<i>Fase</i>	1	-6.39	2.07
9	[$^\circ$]	2	-5.84	1.80
9		3	-8.99	3.42
9		4	-42.15	29.49
9	<i>Resistencia</i>	1	135.61	10.93
9	[Ω]	2	135.89	12.84
9		3	136.45	16.73
9		4	100.21	73.77
9	<i>Reactancia</i>	1	-14.55	5.86
9	[Ω]	2	-13.88	5.33
9		3	-22.53	9.69
9		4	-125.85	108.42

Tabla C.10: Contenido Gástrico 1 $N = 1897$.

No. Error	Medida	Banda	Media	DS
10	<i>Amplitud</i>	1	153.86	24.93
10	[Ω]	2	127.10	19.22
10		3	106.63	15.63
10		4	77.39	24.14
10	<i>Fase</i>	1	-13.65	2.75
10	[$^\circ$]	2	-10.57	3.22
10		3	-10.49	3.37
10		4	-32.06	11.86
10	<i>Resistencia</i>	1	148.72	23.74
10	[Ω]	2	123.98	17.66
10		3	104.39	15.95
10		4	65.18	24.41
10	<i>Reactancia</i>	1	-36.75	10.54
10	[Ω]	2	-24.76	10.24
10		3	-19.40	6.21
10		4	-34.59	14.52

Tabla C.11: Contenido Gástrico 2 $N = 698$.

No. Error	Medida	Banda	Media	DS
11	<i>Amplitud</i> [Ω]	1	30.36	1.48
11		2	27.13	1.66
11		3	25.05	1.30
11		4	22.81	1.52
11	<i>Fase</i> [\circ]	1	-8.10	1.81
11		2	-6.50	2.24
11		3	-5.44	2.36
11		4	-4.73	2.91
11	<i>Resistencia</i> [Ω]	1	29.96	1.51
11		2	26.87	1.65
11		3	24.86	1.34
11		4	22.61	1.55
11	<i>Reactancia</i> [Ω]	1	-4.28	0.91
11		2	-3.06	1.08
11		3	-2.34	0.96
11		4	-1.86	1.12

Tabla C.12: Plano $N = 2500$.

No. Error	Medida	Banda	Media	DS
12	<i>Amplitud</i> [Ω]	1	148.66	32.33
12		2	73.02	9.99
12		3	59.55	3.56
12		4	46.91	4.80
12	<i>Fase</i> [\circ]	1	-26.74	5.63
12		2	-18.33	7.49
12		3	-8.44	1.15
12		4	-12.01	2.49
12	<i>Resistencia</i> [Ω]	1	133.41	34.35
12		2	68.51	6.21
12		3	58.86	3.55
12		4	45.84	4.99
12	<i>Reactancia</i> [Ω]	1	-63.57	9.06
12		2	-23.35	12.28
12		3	-8.42	1.15
12		4	-9.49	1.51

Tabla C.13: Alto $N = 254$.

No. Error	Medida	Banda	Media	DS
13	<i>Amplitud</i>	1	57.59	3.50
13	$[\Omega]$	2	46.47	3.55
13		3	44.98	19.14
13		4	68.15	13.91
13	<i>Fase</i>	1	-9.89	1.75
13	$[\circ]$	2	-8.52	1.59
13		3	-11.18	2.77
13		4	-11.83	2.20
13	<i>Resistencia</i>	1	56.66	3.30
13	$[\Omega]$	2	45.83	3.60
13		3	43.95	19.06
13		4	66.44	13.68
13	<i>Reactancia</i>	1	-10.04	2.11
13	$[\Omega]$	2	-7.21	0.97
13		3	-8.62	2.75
13		4	-14.72	3.53

Tabla C.14: Aire $N = 7$.

No. Error	Medida	Banda	Media	DS
14	<i>Amplitud</i>	1	104.30	11.03
14	$[\Omega]$	2	65.26	9.46
14		3	48.74	4.24
14		4	34.47	5.93
14	<i>Fase</i>	1	-23.61	5.39
14	$[\circ]$	2	-27.43	3.43
14		3	-24.99	3.15
14		4	-27.56	5.57
14	<i>Resistencia</i>	1	96.10	13.32
14	$[\Omega]$	2	58.39	7.15
14		3	44.58	4.37
14		4	30.63	5.32
14	<i>Reactancia</i>	1	-37.46	5.28
14	$[\Omega]$	2	-27.04	6.85
14		3	-18.11	1.88
14		4	-14.99	3.08

Tabla C.15: Offset $N = 344$.

No. Error	Medida	Banda	Media	DS
15	<i>Amplitud</i> [Ω]	1	56.37	24.32
15		2	50.45	24.72
15		3	48.53	22.83
15		4	46.76	21.87
15	<i>Fase</i> [$^\circ$]	1	1.54	70.74
15		2	3.60	65.40
15		3	-1.09	65.26
15		4	1.78	65.37
15	<i>Resistencia</i> [Ω]	1	23.65	34.79
15		2	21.54	32.05
15		3	20.02	29.69
15		4	18.19	29.95
15	<i>Reactancia</i> [Ω]	1	-1.53	35.53
15		2	-0.34	32.50
15		3	-2.65	30.99
15		4	-1.61	30.50

Tabla C.16: *Altas Variaciones* $N = 911$.

Apéndice D

Códigos

En esta sección se presentan los códigos generales para la implementación en el diseño de los clasificadores (*función clasificador*) esto código se utiliza para generar cada uno de los clasificadores que se implementan para el algoritmo de identificación de errores (*función identifica*).

```
function clasificador (clase, N, carac, archivo, nucleo)
%*****
% Función que encuentra un clasificador con base a la clase, N y
% características del archivo seleccionado.
% clase- numero de clase a la cual se quiere separar.
% N- Número de datos de entrenamiento a los clasificadores.
% nucleo- tipo de kernel 't 0' Lineal , '-t 2' Gaussiano
%*****
load(archivo)
%Selección de clases binarias.
[data, t, c1, c2, i_c1, i_c2]=selectClase(data,clase, N);

% Selección de conjunto de características.
s=selectCarac(c1, c2, data, carac); disp(s.mens);

% Selección de mejores hiperparámetros.
[valores, palabra]=selectParam(nucleo, t, s.carac_train);

% Encuentra mejor clasificador con base a la validación cruzada.
[desem, modeloVc, modeloT]=valCruzada(5,t, s.carac_train, palabra);

% Validación con clasificador con todos los datos
disp('Modelo con todos los datos de entrenamiento')
[t1, acc1, F1] = svmpredict(data(:,1), s.carac_prueba, modeloT, '-b 1 ');
```

```
%Validación con mejor clasificador de validación cruzada
disp('Mejor Modelo con validación cruzada')
[t2, acc2, F2] = svmpredict(data(:,1), s.carac_prueba, modeloVc, '-b 1 ');

% Cálculo de incidencia de errores
[errores1, errores1a, data1a, data2a]=detectError(t1,data);
[errores2, errores2a, data1b, data2b]=detectError(t2,data);

DatosRC1=[data1b; c2(:, 1:33)];
DatosRC2=[data2b; c1(:, 1:33)];

%Gráfica probabilidad posterior si es de 2 características.

if s.numCarac==2
    graficaPP(modeloVc,s.etiq_carac_train,s.carac_train, s.carac_prueba,
i_c1, i_c2, s.r_carac1, s.r_carac2);
end
% Estructura para guardado de datos.
p=struct('N', N, 'Carac', carac , 'palabra' , palabra , 'desem',desem, 'modeloVc',
modeloVc, 'modeloT', modeloT,'Datos', data, 'DatosResto', data1b, 'DatosRC1',
DatosRC1, 'DatosRC2', DatosRC2, 'erroresMVC', errores2, 'erroresMT', errores1,
'erroresMVc1' ,errores2a, 'erroresMT1',errores1a, 'EtiqPred', t2,'ExactitudModVc',
acc2(1:2)', 'ExactitudModT', acc1(1:2)', 'ProbPost',F2, 'mens', s.mens);
r=p(1);
```

```

function errores=identifica(archivo)
%*****
% Función que identifica errores por conexión, posición y contacto.
% Los datos ya deben venir escalados y calaculadas sus características.
%*****
load(archivo)
%Identificación de características
todas=data(:,6:21);
SRM=data(:,22);
SXM=data(:,23);
SRSD=data(:,24);
SXSD=data(:,25);
SRX_M=data(:,26);
SRX_SD= data(:,27);
[m n]=size(data);
e=zeros(m,1);
t=zeros(m,1);
for i=1:m
e(i)=0;
tic
%Identificación de errores
    if (circuitoAbierto(data(i,13),SRSD(i)) > 0); %Circuito Abierto
        e(i)=e(i)+1;
    end;
    if (e(i) == 0)
        if (cortoCircuito(SRM(i),SRX_M(i), SRX_SD(i)) > 0); %Corto Circuito
            e(i)=e(i)+2;
            if (e(i)==2)
                if (sondaDes(SXM(i)) >0); % Sonda Descompuesta
                    e(i)=e(i)+3-2;
                else
                    e(i)=e(i)+4-2; %Cable descompuesto
                end;
            end;
        end
    end
end

    if (e(i) == 0)
        if (malaPosicion(data(i,9), data(i,15), data(i,17), data(i,19),
            SRM(i),SRX_M(i), SRX_SD(i)) > 0); % Mala Posición
            e(i)=e(i)+5;
        end
    end
end

```

```
end
if (e(i) == 0)
    if (malContacto(SRM(i), SRSD(i), SXSD(i), SRX_M(i),
        SRX_SD(i)) > 0); % Mal Contacto
        e(i)=e(i)+6;
        if (e(i)==6)
            if (contGas(todas(i,:))>0); % Contenido Gástrico
                e(i)=e(i)+7-6;
            elseif (contPob(todas(i,:))>0); % Contacto Pobre
                e(i)=e(i)+8-6;
            else
                e(i)=e(i)+9-6; %Artefacto de movimiento
            end;
        end
    end
end
t(i)=toc;
errores=e;
mens1=num2str(i);
mens2=num2str(m);
mens=['Espectros calculados: ',mens1,'/',mens2];
disp(mens);
end

save errv5 errores data etiq t
```

Universidad Autónoma Metropolitana-Iztapalapa

Detección de Errores en la Medición de Espectroscopía de Impedancia Tisular Gástrica

Tesis que presenta para obtener el grado de:

Maestría en Ciencias (Ingeniería Biomédica)

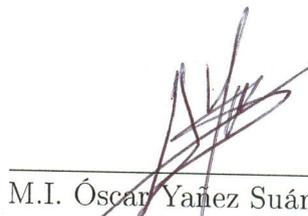
I.B. María de Montserrat Godínez García

Mayo 2010

Asesores:



Dr. Emilio Sacristán Rock
Universidad Autónoma Metropolitana



M.I. Óscar Yañez Suárez
Universidad Autónoma Metropolitana

Sinodales:



Dr. César Antonio González Díaz
Escuela Militar de Graduados de Sanidad



Dr. John Charles Goddard Close
Universidad Autónoma Metropolitana

M.I. Óscar Yañez Suárez
Universidad Autónoma Metropolitana