



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA**

**APLICACIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS PARA LA
DETECCIÓN DE HUSOS SIGMA EN SEÑALES DE SUEÑO**

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA**

ANDRÉS SEGUNDO MARINKOVIC GÓMEZ

**SANTIAGO DE CHILE
ENERO 2012**



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**APLICACIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS PARA LA
DETECCIÓN DE HUSOS SIGMA EN SEÑALES DE SUEÑO**

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA**

ANDRÉS SEGUNDO MARINKOVIC GÓMEZ

**PROFESOR GUÍA:
PABLO ANTONIO ESTÉVEZ VALENCIA**

**MIEMBROS DE LA COMISIÓN:
HÉCTOR MILER AGUSTO ALEGRÍA
CLAUDIO MARCELO HELD BARRANDEGUY**

**SANTIAGO DE CHILE
ENERO 2012**

APLICACIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS PARA LA DETECCIÓN DE HUSOS SIGMA EN SEÑALES DE SUEÑO

El estudio de señales de sueño es de gran importancia para evaluar la calidad del sueño de los individuos y para investigar las funciones que éste realiza. En ese sentido, resulta interesante desarrollar métodos automáticos que analicen las señales de sueño, ya que en la actualidad dicho análisis es desarrollado generalmente por un experto, lo que resulta tedioso, debido a la longitud de los registros.

Dentro del análisis que se puede realizar a las señales de sueño, está la detección de husos sigma (HS), que corresponden a un patrón particular que aparece en el electro encéfalo grama durante las fases 2, 3 y 4 de estado conocido como sueño quieto. El principal objetivo de este trabajo es desarrollar un sistema capaz de detectar automáticamente candidatos de HS en señales de sueño. Se desea que el sistema sea comparable con la salida de los módulos 1 y 2 de un sistema implementado por Leonardo Causa en el Departamento de Ingeniería Eléctrica, los cuales utilizan la potencia espectral de tres bandas de frecuencia, la transformada de Hilbert-Huang, la descomposición modal empírica y criterios difusos para llevar a cabo la detección.

El sistema planteado en este trabajo utiliza una novedosa herramienta llamada gráficos de recurrencia. Dicha herramienta permite construir, a partir de un trozo de la señal del sueño, un gráfico en dos dimensiones, el cual describe la periodicidad de dicho trozo. Luego, mediante mediciones a este gráfico, es posible extraer características numéricas que describen ciertas propiedades de aquel trozo de la señal. Además de las características deducidas del gráfico se agregan otras, obtenidas directamente del segmento de señal y relacionadas con otros atributos de los HS, como su duración y magnitud. Todos estos atributos se combinan mediante una red neuronal, la cual determina si el trozo sobre el cual se calcularon los atributos corresponde o no a un huso sigma. Así, se plantea el problema de detección de HS como un problema de clasificación.

Se dividen los datos en tres conjuntos: entrenamiento, validación y prueba. El conjunto de entrenamiento y validación es utilizado en principio para desarrollar el método planteado, determinado cuáles de sus variantes entregan mejores resultados. Posteriormente, los parámetros de este sistema se ajustan utilizando los mismos conjuntos de validación y entrenamiento, evaluando el desempeño del sistema frente a diferentes combinaciones de parámetros. Luego, estos mismos conjuntos de entrenamiento y validación son utilizados para entrenar el modelo.

Posteriormente se utiliza el conjunto de prueba para comparar el desempeño del sistema implementado en este trabajo con el desempeño de los módulos 1 y 2 del sistema original. Los resultados sobre el conjunto de prueba, el cual incluye cuatro pacientes, muestran un desempeño similar entre ambos sistemas. El sistema planteado obtiene una tasa de verdaderos positivos (TPR) de 80,8% y una tasa de falsas detecciones (FDR) de 81,9%. El sistema original obtiene una TPR de 79,5% y una FDR de 85,7%. Se observa que una sección en la mitad de la curva ROC (acrónimo de *Receiver Operator Characteristic*), calculada para el sistema implementado, es estrictamente mejor que los módulos 1 y 2 del sistema original, para los datos utilizados. Es posible que un híbrido de ambos métodos de origen a un método de mejor desempeño, ya que la información aportada por cada método parece ser complementaria.

Se define, por ende, un método apropiado de detección de HS. Algunas de las características introducidas por este método resultan bastante adecuadas para este problema, por lo que podrían ser usadas y mejoradas en futuras investigaciones.

Agradecimientos

Terminar de escribir este informe representa el final de un largo camino que he recorrido, un camino lleno de alegrías pero también lleno de esfuerzo y contratiempos. Recorrer este camino habría sido imposible sin el apoyo de un gran número de personas, las cuales intentaré mencionar a continuación. Espero no olvidar a ninguna de las personas que me ayudaron durante mi carrera de ingeniería. Al escribir esto, empiezo a cerrar un largo proceso que duró 7 años. Sin embargo, más que el fin de una etapa, es el comienzo de una nueva etapa en mi vida. Espero que esta nueva etapa me traiga tantas satisfacciones como la etapa que hoy dejo detrás.

Debo comenzar agradeciendo a mi familia. A mi padre, Boris, que desde pequeño me transmitió sus conocimientos y despertó en mí la curiosidad que me llevaría a estudiar ingeniería. A mi madre, Teresa, quien inculcó en mí disciplina y responsabilidad, dos virtudes sin las cuales jamás estaría donde estoy hoy en día. A mi hermano, Billo, quien ha sido un excelente compañero durante toda mi vida, siempre motivándome con su humor e inmensa energía. A mi abuelo, Jonás, quien ha sido un modelo a seguir y siempre me ha apoyado, especialmente durante la realización de este trabajo. Finalmente, no puedo dejar de mencionar a Eleonor, quien siempre ha estado para mí. Sin su entrega y cariño, jamás habría logrado ninguna de las metas que he alcanzado en mi vida.

Desde luego no puedo dejar de agradecer a Daniela, quien ha sido una excelente polola, amiga y compañera, y me ha brindado inmenso amor y apoyo estos últimos años. Sin su motivación y cariño, no habría podido terminar este trabajo.

También me gustaría agradecer a mis amigos del colegio: Daniel, Mark, Joni, JJ, Nacho, Neme, Matías y Alfonso, quienes siempre se han mantenido cercanos a mí, tanto en los días del colegio como durante toda mi carrera universitaria.

A mis compañeros de primer año: Alida, Diego, Tami, Nico, Panchito, Gabriel, Sebastián y Catalán, quienes hicieron muy ameno mi primer año en la universidad y han sido excelentes amigos desde entonces. Tampoco puedo dejar de mencionar a Vasco, quien me ayudó mucho en primer año y fue sido un gran amigo durante toda la carrera.

A mis compañeros y amigos de eléctrica: Fernando, Chino, Romina, Cristóbal y Herrman, quienes siempre me entregaron su simpatía y ayuda. También quisiera agradecer a otros amigos que hice en la universidad: Marrano, Omar, Gonzalo y Hernán, grandes mentes y sobre todo grandes personas.

No puedo dejar de agradecer a Leonardo Causa, quien me ayudo mucho durante la realización de este trabajo. Gracias a su excelente disposición pude

obtener y comprender la base de datos usada en esta memoria. También me gustaría agradecer a Pablo Huijse y a Jorge Vergara, quienes me ayudaron mucho durante las primeras fases de este proyecto.

En esta ocasión, quisiera agradecer a todos los profesores que tantos conocimientos me entregaron durante estos últimos 7 años. En especial, me gustaría agradecer a Sergio Rica, Éric Tanter y Pablo Estévez, grandes maestros y excelentes personas. Gracias al profesor Estévez por su ayuda durante la realización de este trabajo. También quisiera dar las gracias a los profesores Claudio Held y Hector Augusto, cuyos comentarios fueron muy importantes para hacer de éste un mejor trabajo.

Este trabajo fue posible gracias al financiamiento del proyecto Fondecyt Regular 1110701.

Índice de Contenidos

1. Introducción.....	1
1.1. Antecedentes.....	1
1.2. Objetivos.....	4
1.3. Estructura de Memoria.....	5
2. Revisión Bibliográfica	7
2.1. Clasificación.....	7
2.1.1. Generalidades.....	7
2.1.2. Característica Operativa del Receptor	10
2.1.3. Máquina de Soporte Vectorial	12
2.1.4. Redes Neuronales.....	15
2.2. Gráficos de Recurrencia	26
2.2.1. Motivación	26
2.2.2. Definición de Gráficos de Recurrencia	26
2.2.3. Método de Retraso de Tiempo.....	27
2.2.4. Distancias.....	28
2.2.5. Ejemplos de Gráficos de Recurrencia.....	32
2.2.6. Medidas sobre Gráficos de Recurrencia	35
2.3. Señales de Sueño	43
2.3.1. Generalidades.....	43
2.3.2. Estado del Arte.....	46
2.3.3. Sistema de Detección de HS Implementado en la Universidad de Chile	49
2.3.4. Descomposición Modal Empírica	56
2.4. Otros Métodos Usados en el Trabajo	61
2.4.1. Información Mutua.....	61
2.4.2. Filtros de Frecuencia.....	61
3. Descripción del Sistema Propuesto.....	66
3.1. Introducción	66
3.2. Algoritmo Propuesto	67

3.2.1. Pre-procesamiento de la Señal	67
3.2.2. Construcción de RP's y Extracción de Características.....	68
3.2.3. Clasificación y Obtención de Detecciones	80
3.3. Parámetros Principales.....	84
3.3.1. Parámetros de pre-procesamiento de la señal.....	85
3.3.2. Parámetros de Extracción de Características y Construcción de RPs .	85
3.3.3. Parámetros de Clasificación.....	85
3.3.4. Resumen de Parámetros Utilizados	86
4. Resultados	87
4.1. Metodología de Pruebas Realizadas	87
4.1.1. Criterio para Determinar Concordancia entre Sistema y Experto.....	87
4.1.2. Medidas de Desempeño	89
4.2. Base de Datos	89
4.3. Pruebas Preliminares.....	91
4.3.1. Metodología	91
4.3.2. Pruebas sobre Variantes.....	92
4.3.3. Pruebas sobre Parámetros	99
4.3.4. Análisis de Resultados.....	104
4.4. Prueba de Ajuste de Parámetros.....	108
4.4.1. Metodología	108
4.4.2. Resultados	110
4.4.3. Análisis de Resultados.....	112
4.5. Prueba Final	113
4.5.1. Metodología	113
4.5.2. Resultados	115
4.5.3. Análisis de Resultados.....	120
5. Conclusiones.....	123
5.1. Método.....	123
5.2. Resultados.....	125
5.3. Objetivos.....	126
5.4. Trabajo Futuro	127
6. Bibliografía	129
7. Anexos	134
7.1. Variantes del Método.....	134

7.1.1. Variantes de Pre-Procesamiento de la Señal.....	134
7.1.2. Variantes en Construcción de RP's y Extracción de Características..	136

Índice de Figuras

Figura 1: Tabla de Contingencia.	10
Figura 2: Ejemplo de Curva ROC.....	11
Figura 3: Red Neuronal MLP Genérica.	18
Figura 4: Red Neuronal MLP con Sesgo.....	21
Figura 5: Ejemplo de RP para Ruido Blanco Gaussiano.....	33
Figura 6: Ejemplo de RP para Señal Sinusoidal de Periodo 0.6 [s].....	34
Figura 7: Ejemplo de RP para 2 Señales Sinusoidales Concatenadas, la Primera de Periodo 0.6 [s] y la Segunda de Periodo 0.3 [s].	35
Figura 8: Ejemplo de Ventana de Theiler con $tw = 2$	36
Figura 9: Comparación de Distribuciones con Asimetría positiva y Negativa.....	38
Figura 10: Comparación de Curtosis para Distintas Distribuciones de Probabilidad. Aquellas distribuciones más empinadas tienen valores mayores de curtosis. Los valores de curtosis para las distribuciones en la curva son: Distribución de Laplace: 6; Distribución Secante Hiperbólica: 5; Distribución Logística: 4,2; Distribución Normal: 3; Distribución Coseno Elevado: 2,4; Distribución Semicírculo de Winger: 2; Distribución Uniforme: 1,8.....	39
Figura 11: Registros de Sueño en donde aparecen Husos Sigma.....	45
Figura 12: Ejemplo de Componente de un SSC1.....	53
Figura 13: Ejemplo de Algoritmo EMD Sobre Suma de Sinusoides de Distinta Frecuencia.....	58
Figura 14: EMD de una Señal de Tres Componentes.....	60
Figura 15: <i>Ripple</i> y Atenuación en un Filtro Pasa Bajos.	63
Figura 16: Función de Transferencia de 4 Filtros de Frecuencia de Distintos Tipos.	64
Figura 17: Diagrama del Funcionamiento General del Sistema Implementado.....	66
Figura 18: Diagrama del Funcionamiento del Sistema Propuesto.....	67
Figura 19: Ejemplo de Filtrado en Banda de Frecuencia de Señal Canal 3 de EEG para Etapa SQ2.....	68
Figura 20: Contrastación de Medidas de RP con Marcas del Experto.	70
Figura 21: Proceso de Cálculo de Envolvente de Señal Filtrada.....	71
Figura 22: Proceso de Cálculo de Ancho Peak de la Envolvente de Señal de EEG.	76
Figura 23: Filtros Gaussianos Usados en Convolución.....	78
Figura 24: Ejemplo de Determinación de Clases para Vector Asociado a Cada Ventana.....	80
Figura 25: Diagrama de la Red Neuronal Usada para Clasificar Husos Sigma. ...	82
Figura 26: Salida Interpolada de Red Neuronal Para un Ejemplo de SQ2.....	84
Figura 27: Ejemplos de Criterio Usado para Medir Desempeño del Sistema.....	88
Figura 28: Distribución de m Según Método de Falsos Vecinos.....	102
Figura 29: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $uh = -0.225$ y Salida de módulo 2, para Paciente 4. ...	116

Figura 30: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $uh = -0.225$ y Salida de módulo 2, para Paciente 5. ...	117
Figura 31: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $uh = -0.225$ y Salida de módulo 2, para Paciente 6. ...	118
Figura 32: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $uh = -0.225$ y Salida de módulo 2, para Paciente 7. ...	119
Figura 36: Curva ROC para Sistema Basado en Redes Neuronales sobre CP. .	120
Figura 37: Ejemplo de Descomposición EMD.	135
Figura 38: Contratación de Medidas de RP con Marcas del Experto.	138
Figura 39: Ejemplo de RP Calculado para un Instante de Tiempo en Donde Existe un HS con Poco Ruido.	140

Índice de Tablas

Tabla 1: Ejemplo de Valores de <i>fbj</i> para 7 Muestras y el Valor de <i>anchj</i> Deducido a Partir de <i>fb</i>	75
Tabla 2: Características Usadas por Clasificador.....	79
Tabla 3: Resumen de Parámetros del Sistema.....	86
Tabla 4: Duración de Registros para los 3 Pacientes y para Diferentes Fases de Sueño.....	90
Tabla 5: Inicio y Fin de Trozo de Tiempo de cada Registros Usados para Conjuntos de Entrenamiento y Validación para Pruebas Preliminares.	91
Tabla 6: Valores Óptimos Obtenidos para cada Distancia.	93
Tabla 7: Comparación de Información Mutua entre <i>Lmax</i> , <i>Lσ</i> y <i>Lmean</i> y Marcas de HS de Experto.....	94
Tabla 8: TP y FP Obtenidos sobre CP para Distintas Distancias Usadas para Construcción de RP's.....	95
Tabla 9: Comparación de Información Mutua entre Cada Medidas Deducidas del RP y Marcas de HS de Experto.....	96
Tabla 10: Comparación de Clasificaciones sobre CP entre un Sistema que Usa las 9 Características de RP y un Sistema que Usa solo las 3 con Mayor IM.	97
Tabla 11: Comparación de Información Mutua entre las Potencias de Bandas de Frecuencia del EEG y Marcas de HS de Experto.....	98
Tabla 12: Comparación de Clasificaciones sobre CV entre Sistema que Usa las 3 Características Basadas en Potencias de Bandas de Frecuencia del EEG y un Sistema que Usa solo las con Mayor IM.	98
Tabla 13: Comparación de IM entre <i>anch</i> y Marcas de HS, y entre <i>fb</i> y Marcas de HS, para distintas formas de calcular <i>upeaks</i>	99
Tabla 14: Comparación de Clasificaciones sobre CV entre 3 Sistemas Usando Distintos Métodos de Cálculo de <i>upeaks</i>	99
Tabla 15: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	100
Tabla 16: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	101
Tabla 17: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	101
Tabla 18: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	102
Tabla 19: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	103
Tabla 20: IM entre Características Deducidas de RPs (<i>Lmax</i> , <i>Lmedio</i> y <i>Ldesv</i>) y Marcas de HS de Experto.	103
Tabla 21: Valores de Parámetros Principales del Sistema para Cada Combinación de Parámetros Usada.	109
Tabla 22: Inicio y Fin de Trozo de Tiempo para los 5 Subconjunto Usados en Validación Cruzada.	110

Tabla 23: TP, FP, TPR y FDR Obtenidos para Prueba de Ajuste de Parámetros.	111
Tabla 24: Inicio y Fin de Trozo de Tiempo de cada Registro Usado para los Conjuntos de Entrenamiento, Validación y Prueba. El guión indica que dicho registro no fue usado para conformar el conjunto respectivo. Los límites de inicio y fin son inclusivos	114
Tabla 25: Comparación de TP, FP, FDR y TPR entre M2 y el Clasificador Basado en NN con $uh = -0.225$	115

Abreviaturas Usadas en Este Trabajo

E: Entropía

ECG: Electro-cardiograma

EEG: Electro-encefalograma

EMD: Descomposición modal empírica (sigla proveniente del inglés *Empirical Mode Decomposition*)

EOG: Electro-óculograma

FDR: Tasa de falsas detecciones (sigla proveniente del inglés *False Detection Rate*)

FFT: Transformada rápida de Fourier (sigla proveniente del inglés *Fast Fourier Transform*)

FN: Falsos negativos (sigla proveniente del inglés *False Negative*)

FNR: Tasa de falsos negativos (sigla proveniente del inglés *False Negative Rate*)

FP: Falsos positivos (sigla proveniente del inglés *False Positive*)

FPR: Tasa de falsos positivos (sigla proveniente del inglés *False Positive Rate*)

HHT: Transformada de Hilbert-Huang (sigla proveniente del inglés *Hilbert-Huang Transform*)

HS: Huso sigma

IM: Información mutua

LOI: Línea de identidad, presente en todo RP (sigla proveniente del inglés *Line Of Identity*)

M2: Sistema compuesto por los módulos 1 y 2 del trabajo de Leonardo Causa.

ML: Aprendizaje de máquinas (sigla proveniente del inglés *Machine Learning*)

MLP: Perceptrón Multicapa (sigla proveniente del inglés *Multi-layer perceptrón*)

NN: Red Neuronal (sigla proveniente del inglés *Neural Network*)

REM: Estado del sueño caracterizado por la presencia de movimientos oculares rápidos (sigla proveniente del inglés *Rapid eye movements*).

NREM: Estado del sueño caracterizado por la ausencia de oculares rápidos (sigla proveniente del inglés *Non-rapid eye movements*).

ROC: Característica operativa del receptor, curva utilizada para comparar sistemas de clasificación (sigla proveniente del inglés *Receiver Operating Characteristic*)

RP: Gráfico de recurrencia (sigla proveniente del inglés *Recurrence Plot*)

RR: Razón de recurrencia (sigla proveniente del inglés *Recurrence Rate*)

SFT: Transformada corta de Fourier (sigla proveniente del inglés *Short-time Fourier Transform*)

SP: Sueño paradójico

SQ: Sueño quieto

SQ1: Etapa 1 de sueño quieto

SQ2: Etapa 2 de sueño quieto

SQ3: Etapa 3 de sueño quieto

SQ4: Etapa 4 de sueño quieto

SSC1: Candidato a HS a la salida del módulo 1 (sigla proveniente del inglés *Sleep Spindle Candidate*)

SSC2: Candidato a HS a la salida del módulo 2 (sigla proveniente del inglés *Sleep Spindle Candidate*)

SVM: Máquina de soporte vectorial 2 (sigla proveniente del inglés *Support Vector Machine*)

SWS: Etapa del sueño compuesta por SQ3 y SQ4 (sigla proveniente del inglés *Slow Wave Sleep*)

TCC: Tasa de clasificaciones correctas

TN: Verdaderos negativos (sigla proveniente del inglés *True Negative*)

TNR: Tasa de verdaderos negativos (sigla proveniente del inglés *True Negative Rate*)

TP: Verdaderos positivos (sigla proveniente del inglés *True Positive*)

TPR: Tasa de verdaderos positivos (sigla proveniente del inglés *True Positive Rate*)

VI: Vigilia

1. Introducción

1.1. Antecedentes

El sueño es una parte fundamental de la vida del ser humano. Además de ocupar, en condiciones normales, un tercio del día de una persona adulta, es esencial para que durante los otros dos tercios del día el individuo pueda desempeñar sus labores de forma adecuada. El sueño no solo realiza funciones de descanso y restauradoras, sino que además es fundamental para la consolidación del aprendizaje ocurrido durante el día, según lo demuestran estudios recientes [55] [58].

Se sabe que el sueño no es un proceso constante, si no que se compone de diferentes estados, caracterizadas por distintos fenómenos neurológicos, y que desempeñan distintas funciones para el organismo [39]. Uno de estos estados, el sueño quieto (SQ), se divide en cuatro etapas indicadas por los números 1, 2, 3 y 4. Las etapas asociadas a un número de mayor valor corresponden a estados de sueño más profundo.

Existen muchos exámenes que se realizan para medir y diagnosticar la calidad del sueño de un ser humano. Entre los más importantes están el electroencefalograma (EEG), el electro-óculograma (EOG) y el electro-cardiograma (ECG). Diferentes patrones en el EOG y en el EEG permiten determinar el estado del sueño en que se encuentra el individuo, y si dichos estados ocurren con la frecuencia y en la proporción adecuada. Esto permite que expertos diagnostiquen la calidad del sueño de una persona, y si este proceso está realizando adecuadamente las funciones que debería.

Uno de los patrones que se detectan en las señales del sueño es el huso sigma (HS), el cual aparece en algunos canales del EEG. El HS corresponde a un tren de ondas de frecuencias entre 10 y 16 [Hz], repentina y de duración entre 0,5 y 2 [s], con amplitudes que primero crecen y luego decrecen. Dicho patrón aparece principalmente en la etapa SQ2, aunque también puede aparecer en las etapas SQ3 y SQ4. Además de su importancia para la detección la etapa 2 del sueño quieto, los husos sigma promueven los procesos necesarios para la formación de memoria [41] [49] [50]. Por otra parte, distorsiones en este patrón han sido asociadas a diferentes patologías neurológicas [37] [13] [8] [47] [25]. Es por ello que su detección es de vital importancia tanto para el estudio del sueño en condiciones normales, como para el diagnóstico de patologías.

Actualmente, el análisis de los estados y etapas del sueño y de los patrones en las mediciones que se realizan sobre un paciente durmiendo, es realizado, generalmente, en forma visual por un experto. Este proceso resulta lento y engorroso, debido a la gran cantidad de datos producidos por un solo paciente

durante únicamente una noche. Este problema se acentúa si se considera que cada registro de sueño está compuesto por decenas de señales fisiológicas [7]. Otro inconveniente del análisis realizado por un profesional es que este proceso es subjetivo, dependiendo tanto de características personales como del estado de ánimo en que el experto se encuentre al momento de realizar el estudio. Por esto, resulta de interés crear métodos automáticos de detección de patrones en señales de sueño. Además del potencial que éstos tienen para ser más rápidos, corresponden a una medición objetiva que producirá siempre los mismos resultados frente a la misma entrada.

Para enfrentar el problema de la detección automática de patrones en señales de sueño, y en particular de husos sigma, se han propuesto diferentes métodos computacionales. Un grupo de herramientas que resultan útiles para este tipo de problema son aquellas conocidas como aprendizaje de máquinas (o ML, sigla del inglés *Machine Learning*). En forma general, el aprendizaje de máquinas es un conjunto de técnicas estadísticas para aprender u obtener información de ciertos datos y luego utilizarla para analizar nuevos datos que se presenten. Normalmente, estos métodos crean una estructura o un modelo a partir de los datos actuales, la cual luego utilizan sobre datos futuros, con diversos propósitos. ML es una de ramas más exitosas de la inteligencia artificial en la actualidad.

En el pasado el aprendizaje de máquinas se ha sido utilizado en forma exitosa para detectar patrones en señales de sueño, en particular los husos sigma, como se puede ver en [3], [11], [15], [19], [22], [23], [44], [46], [54] y [60]. Uno de los métodos más exitosos desarrollados hasta ahora es el presentado en [6]. Este método, desarrollado en el Departamento de Ingeniería Eléctrica de la Universidad de Chile, se basa en diversas técnicas para llevar a cabo la detección de HS en EEG de niños de 10 años. En él, se procesan estas señales originales provenientes del EEG con la técnica de descomposición modal empírica [20] (o EMD, del inglés *Empirical Mode Decomposition*) y luego con la transformada de Hilbert-Huang [20] (o HHT, sigla proveniente del inglés *Hilbert-Huang Transform*). Sobre la señal que resulta de utilizar estas dos técnicas se aplican criterios de lógica difusa para seleccionar candidatos. Estos candidatos son filtrados posteriormente, utilizando criterios de frecuencia, simetría, amplitud y duración. A pesar de que este método es muy exitoso, ciertas componentes pueden ser mejoradas. Específicamente, los dos primeros módulos, donde se seleccionan candidatos a HS, entregan demasiados falsos positivos (o FP, del inglés *False Positive*), es decir, demasiados candidatos que no son realmente HS según el experto. En este trabajo se reemplazan estos dos módulos por un sistema que utiliza otras técnicas de ML. Dichas técnicas corresponden a gráficos de recurrencia (o RP, sigla proveniente del inglés *Recurrence Plots*) [31] y a redes neuronales (o NN, sigla proveniente del inglés *Neural Networks*) [2].

Un RP es una visualización de la periodicidad de un sistema dinámico. En otras palabras, es una función cuyo dominio es bidimensional, y donde cada punto del gráfico está asociado a dos instantes de tiempo. La función indica si en ambos instantes de tiempo el sistema dinámico se encuentra o no en estados similares,

bajo un cierto criterio. Debido a su carácter bidimensional, es posible graficar dicha función, generándose así una visualización de las recurrencias presentes en el sistema. Dado que un HS posee un gran número de componentes con frecuencias en una banda bien específica, corresponde a un patrón con un grado de periodicidad intrínseca. Por esto, este trabajo se basa en la idea de que los RP podrían resultar útiles a la hora de detectar estos patrones. Hasta donde se sabe, ésta es la primera vez que se usan RP para la detección de HS, lo que corresponde, por lo tanto, a un aporte original de este trabajo.

Por definición, un HS es un patrón con ciertas características de duración, amplitud, frecuencia y simetría. Por lo tanto, no resulta adecuado detectarlos utilizando únicamente los RP, ya que éstos están más enfocados en la morfología y en el espectro de frecuencias del evento. Debido a esto, este sistema considera otras mediciones, realizadas sobre la señal del EEG, para determinar la presencia de candidatos a HS en distintos instantes de tiempo. Dichas mediciones estaban enfocadas en otras características distintas a las que miden los RP, como, por ejemplo, la amplitud o la duración del posible HS. Mientras algunas de estas mediciones existían en otros sistemas desarrollados previamente, otras fueron desarrolladas específicamente para este trabajo, y por ende, corresponden a un segundo aporte original de esta memoria.

Considerando que existen muchas mediciones determinadas a partir de la señal, resulta necesario unificarlas al realizar la detección final. Esto se abordó como un problema de clasificación [12]. Inicialmente, a partir de un segmento de registro se deducen ciertas características numéricas, ya sea midiéndolas directamente del trozo de registro, o del RP calculado a partir del segmento. Luego, dichas características se agrupan en un vector y se hacen pasar por un clasificador, el cual entrega una salida binaria indicando si existe o no presencia de HS en el trozo original. Para realizar esta labor, el clasificador se entrena previamente con trozos de registro para los cuales la presencia o ausencia de HS es conocida.

El clasificador usado para detectar HS en los segmentos del EEG corresponde a una red neuronal de tipo perceptrón multicapa (o MLP, sigla proveniente del inglés *Multi-layer perceptron*). En términos básicos, una red neuronal MLP es una estructura computacional que permite encontrar una función de entrada-salida. Dicha función es capaz de adaptarse a partir de ejemplos presentados, aprendiendo la salida deseada para cada uno de estos ejemplos. Luego, es posible utilizar la red MLP sobre nuevos ejemplos de salida desconocida, para predecir esta salida. A pesar de que existen otros clasificadores, se optó por una red neuronal debido a su versatilidad y a que requieren poco parámetros para funcionar correctamente [18]. El clasificador desarrollado en este trabajo sirve para detectar HS en forma automática, sobre nuevos datos que reciba en su entrada.

1.2. Objetivos

El objetivo principal de este trabajo es desarrollar un detector basado en gráficos de recurrencia y redes neuronales que sea mejor que los módulos 1 y 2 del sistema desarrollado por Leonardo Causa en [6]. Dicha comparación, para poder ser válida, debe ser llevada a cabo utilizando los mismos registros de sueño como entrada de ambos sistemas. Al igual que el detector de [6], se espera que el clasificador sea capaz de aprender a partir de una muestra de pacientes, y luego generalizar a nuevos pacientes. Se espera que este detector obtenga mejores tasas de falsos positivos y de verdaderos positivos (o TP, del inglés *True Positive*) que los módulos 1 y 2 del sistema original.

El primer objetivo secundario es estudiar el aporte de los RP al problema de detección de patrones HS. Se busca ver las ventajas y desventajas de este método, y determinar si corresponde a un método adecuado para tal labor. Idealmente, un sistema de detección de HS basado en RP's debería tener un desempeño comparable a otros métodos del estado del arte que resuelvan dicho problema.

Un segundo objetivo secundario consiste en determinar si el uso de NN es adecuado para la detección de husos sigma. Esto implica también estudiar las ventajas y desventajas de plantear la detección de estos patrones como un problema de clasificación.

Finalmente, el último objetivo secundario es investigar y probar nuevas características para su uso en la clasificación de HS. Aunque existen actualmente un gran número de sistemas que miden diferentes propiedades de las señales de sueño, muchos de ellos no consideran la definición completa de HS, la cual incluye frecuencia, morfología, amplitud y duración. Así, muchos detectores existentes actualmente se basan sólo en algunas de estas características, dejando de lado las otras. Esto introduce errores en las detecciones, ya que al ignorar ciertos criterios, es posible que se detecten fragmentos que claramente no son HS. A futuro, las características desarrolladas en este trabajo pueden ser usadas en nuevos sistemas de detección de HS.

A continuación se listan algunos objetivos de menor importancia asociados a este trabajo:

- Entender las propiedades principales de los husos sigma, en el marco de su detección automática.
- Interiorizarse con las distintas materias abarcadas en este estudio, entre ellas: las señales de sueño, los RP y las redes neuronales, mediante su aplicación en esta memoria.

- Entender el estado del arte relativo a la detección automática de HS, considerando la posibilidad de este tema continúe en un trabajo futuro.
- Familiarizarse con las técnicas de desarrollo de una memoria y una publicación científica, considerando posibles desarrollos futuros.

1.3. Estructura de Memoria

A continuación se detalla la estructura de este informe.

En el capítulo 2 se hace una revisión bibliográfica. El objetivo de este capítulo es ser una referencia para los tópicos que aparecen en el resto del texto. En él se explican en detalle todas las temáticas específicas necesarias para comprender el sistema implementado, en particular, los temas relacionados con RP, NN y HS. Además de temas específicos relacionados directamente con el sistema desarrollado, se presentan temas más generales que corresponden al marco teórico del trabajo. Entre estos temas se incluyen el estado del arte de sistemas de detección de HS, generalidades sobre las señales de sueño, técnicas usadas por el sistema desarrollado por Leonardo Causa [6] y la definición de clasificadores en general.

En el capítulo 3 se explica en detalle el funcionamiento del sistema implementado. Se explican paso a paso los procesos que realiza el sistema desde la entrada, que corresponde a una señal de un EEG, hasta su salida, que corresponde a marcas indicando el inicio y el fin de cada HS.

En el capítulo 4 se muestran y analizan los resultados de las distintas pruebas realizadas con el sistema. Primero se presentan las pruebas realizadas para ajustar los parámetros de este algoritmo, de tal forma de adaptarlos al problema particular que se está resolviendo. Luego, se presenta la prueba principal de este trabajo, en donde se que compara el desempeño del detector de HS con el de el método desarrollado en [6], utilizando el mismo conjunto de datos en ambos sistemas.

En el capítulo 5 se presentan las conclusiones de este trabajo. Se analizan los resultados obtenidos y su implicancia en trabajos futuros. Además se examina el cumplimiento de los objetivos de esta memoria. Finalmente, se da una reseña sobre las posibles formas de extender este trabajo a futuro.

En el capítulo 6 se listan las referencias utilizadas en este informe.

En el capítulo 7 se presentan los anexos de este trabajo. Se explican algunas variantes del sistema, consideradas durante su desarrollo, pero

posteriormente eliminadas del sistema final presentado en el capítulo 3. Además, se muestran las pruebas que se realizaron para comparar estas diferentes variantes y así determinar dicho sistema final.

2. Revisión Bibliográfica

2.1. Clasificación

2.1.1. Generalidades

El problema de clasificación ha sido ampliamente estudiado en áreas relacionadas con inteligencia artificial, debido a sus numerosas aplicaciones prácticas. En forma bastante general, este problema consiste en que, dados ciertos objetos identificados por vectores de características, se determina a qué clase pertenecen dichos objetos. Estas clases corresponden a una partición finita y fija del espacio de características.

Se llama \vec{x} al vector de dimensión N en el espacio de características, y F a la función de clasificación. El dominio y recorrido de esta función estarán dados por [12]:

$$F: \mathbb{R}^N \rightarrow \{1, \dots, C\}. \quad (1)$$

Así, el clasificador es una función definida desde un espacio de dimensión N a un espacio finito de clases, definidas desde la 1 hasta la C .

Normalmente las funciones de clasificación se aprenden mediante entrenamiento, a partir de vectores cuya clase se conoce. Así, por ejemplo, se puede tener una base de datos de características, con sus respectivas clases, y usarlas para entrenar algún sistema. El objetivo de dicho entrenamiento será generalizar la clasificación, es decir, que el sistema sea capaz de clasificar correctamente frente a nuevos vectores de características que se le presenten.

Un caso particular, que corresponde a lo que sucede en esta memoria, es el caso en que sólo existen dos clases, es decir $C = 2$. En este caso, el clasificador sólo debe determinar si el vector de características pertenece a la clase 1, que llamaremos clase objetivo o si no pertenece a dicha clase, y por ende debe pertenecer a la clase 2. Se conoce como un positivo cuando el sistema determina que un vector pertenece a la clase objetivo y como un negativo cuando el sistema determina que el vector no pertenece a dicha clase.

Independiente de la salida del clasificador, cada objeto del mundo real que de origen a un vector de características debe estar asociado a una de las dos clases. Esta asociación, conocida como estándar dorado, está dada por algún método mejor de clasificación, ya sea otro clasificador que existía previamente o, en la mayoría de los casos, por un experto humano. Dicha asociación es

considerada perfecta, y el clasificador desarrollado será una aproximación de ella. En base a esto, para el problema de 2 clases existen 4 posibles casos al comparar dicho estándar dorado con la salida del clasificador:

- $F(\vec{x}) = 1$ y \vec{x} realmente pertenece a la clase 1 según el estándar dorado. Este caso se conoce como verdadero positivo (o TP, sigla del inglés *True Positive*).
- $F(\vec{x}) = 2$ y \vec{x} realmente pertenece a la clase 1 según el estándar dorado. Este caso se conoce como falso negativo (o FN, sigla del inglés *False Negative*).
- $F(\vec{x}) = 1$ y \vec{x} realmente pertenece a la clase 2 según el estándar dorado. Este caso se conoce como falso positivo (o FP, sigla del inglés *False Positive*).
- $F(\vec{x}) = 2$ y \vec{x} realmente pertenece a la clase 2 según el estándar dorado. Este caso se conoce como verdadero negativo (o TN, sigla del inglés *True Negative*).

Estas mediciones se pueden convertir a tasas, para obtener números entre 0 y 1. Para ello, se consideran positivos, o P, a aquellos objetos que según el estándar dorado son positivos. Éstos corresponden a la suma de los TP y los FN. Asimismo, se consideran negativos, o N, a aquellos objetos que según el estándar dorado son negativos. Estos corresponden a la suma de los FP y los TN. Con esto se obtienen las tasas, dividiendo por N o por P, según las siguientes fórmulas [12]:

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}, \quad (2)$$

$$FNR = \frac{FN}{P} = \frac{FN}{TP + FN}, \quad (3)$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}, \quad (4)$$

$$TNR = \frac{TN}{N} = \frac{TN}{FP + TN}. \quad (5)$$

En las fórmulas anteriores a las 4 siglas originales (TP, FN, FP y TN) se le ha agregado la letra R, proveniente de la palabra en inglés *rate* (que significa tasa) para denotar la tasa en que ocurren estos eventos. Así, por ejemplo, la tasa de

verdaderos positivos se denota TPR, sigla proveniente del inglés *True Positive Rate*. Esto es análogo para las otras tres tasas.

Nótese que, por definición, $TPR + FNR = 1$ y $FPR + TNR = 1$. La TPR es conocida también como la sensibilidad del sistema, mientras que la TNR es también conocida como la especificidad. Existen dos tasas más que suelen usarse en la literatura. Una de ellas es la tasa de falsos descubrimientos (o FDR, proveniente del inglés *False Discovery Rate*), dada por la siguiente ecuación [12]:

$$FDR = \frac{FP}{FP + TP}. \quad (6)$$

Nótese que la FDR es diferente a la FPR, ya que aparecen diferentes términos en su denominador. Mientras la primera considera el porcentaje de falsos positivos de entre los positivos según el sistema, la segunda considera el porcentaje de falsos positivos de entre los negativos para el estándar dorado. Otra tasa utilizada es la tasa de clasificaciones correctas (TCC), que corresponde a la tasa total de aciertos del sistema. Está dada por la siguiente ecuación:

$$TCC = \frac{TN + TP}{P + N}. \quad (7)$$

Un clasificador exitoso debería intentar maximizar tanto la TPR como la TNR y minimizar tanto la FPR como la FNR. Debido a la restricción sobre la suma de estas tasas, basta, por ejemplo, maximizar la TPR y minimizar la FPR para cumplir con los cuatro objetivos antes mencionados.

Un error muy común es pensar, por ejemplo, que un buen clasificador es aquel que tiene una alta TPR, sin importar la FPR. Sin embargo, realizar un clasificador con $TPR = 1$ es bastante sencillo, basta que $F(\cdot)$ sea la función constante, que a cualquier vector de entrada le asocie el valor 1. Desgraciadamente en este caso se debe cumplir que $FPR = 1$, pues necesariamente cualquier vector de la clase 2 será considerado por el clasificador como perteneciente a la clase 1. Por ello, la FPR debe minimizarse al mismo tiempo que la TPR es maximizada.

Una forma de graficar las cuatro tasas que se han descrito anteriormente se conoce como tabla de contingencia. En ella, las tasas se ordenan según se muestra en la Figura 1 [12].

		Estándar Dorado	
		P	N
Clasificador	P'	TP	FP
	N'	FN	TN

Figura 1: Tabla de Contingencia.

P' y N' corresponden a positivos y negativos, respectivamente, según el sistema. P y N corresponden a positivos y negativos, respectivamente, según el estándar dorado. El cruce entre estas posibilidades da origen a los TP, FP, FN y TN.

2.1.2. Característica Operativa del Receptor

La característica operativa del receptor, más conocida como curva ROC [12] (acrónimo de *Receiver Operating Characteristic* en inglés) es un gráfico que permite comparar distintos sistemas de clasificación.

Es común, aunque no corresponde a una regla general, que los clasificadores binarios posean en su salida un umbral o frontera. En estos casos, dicho clasificador se compone, normalmente, de una función continua la cual es discretizada mediante la aplicación de esta frontera. Así, si frente a un vector la salida de la función es inferior al valor del umbral, se considera que dicho vector pertenece a la clase negativa, y si, por el contrario, si esta salida supera dicho umbral, se considera al vector como miembro de la clase positiva. Por ende, esta frontera discrimina entre ambas clases, y al irse variando se privilegia una por sobre la otra. De esta forma, para valores extremos de este umbral, el clasificador le asignará la misma clase a todos los elementos del espacio de características, ya sea la clase positiva (valores extremadamente bajos del umbral) o la clase negativa (valores extremadamente altos del umbral). La curva ROC representa el comportamiento del sistema frente a las variaciones de dicha frontera, o en

general, de cualquier parámetro de este sistema. Se define como la tasa de verdaderos positivos, TPR, en función de la tasa de falsos positivos, FPR.

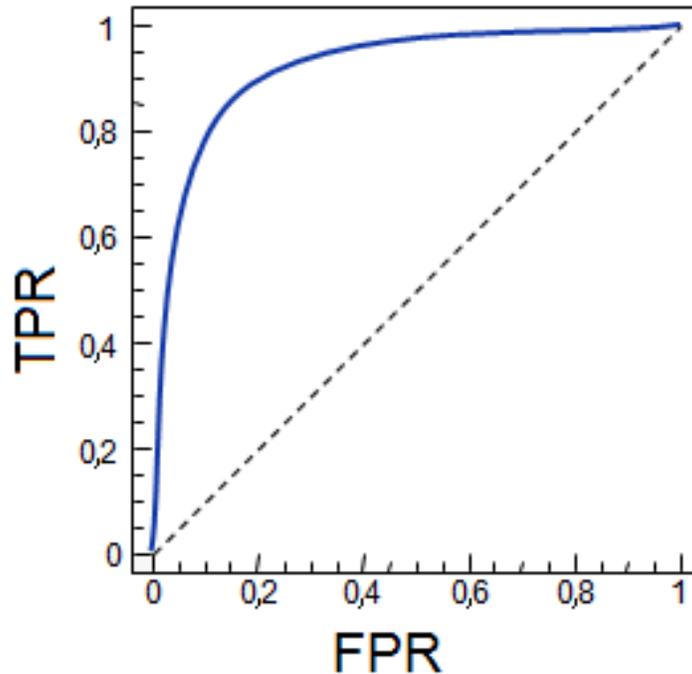


Figura 2: Ejemplo de Curva ROC.

La abscisa corresponde a la tasa de falsos positivos (FPR) y la ordenada a la tasa de verdaderos positivos (TPR).

Una característica muy interesante, que se puede apreciar de la Figura 2, es que la curva es creciente. La razón de esto es que si se quiere aumentar la TPR, es necesario modificar el umbral de tal manera que el sistema tienda a clasificar más elementos como de la clase positiva. Sin embargo, esto hará que algunos elementos que para el estándar eran de la clase negativa y estaban bien clasificados, pasen a ser mal clasificados como de la clase positiva por el sistema. Con esto, se aumentará necesariamente el FPR.

Nótese que los extremos de la curva ROC están en los puntos (0,0) y (1,1). El primer punto, (0,0), se obtiene para el caso en que el clasificador considere a todos los elementos como negativos. En este caso, es imposible que existan falsos positivos y falsos negativos. El segundo punto, (1,1), se obtiene para el caso en que el clasificador considera a todos los elementos como positivos. En este caso, todos los puntos que realmente pertenecen a la clase positiva se clasifican correctamente, pero todos los puntos de la clase negativa se clasifican incorrectamente. Dentro del espacio de la curva ROC, el punto de mayor interés es el (0,1). Este punto correspondería a un sistema perfecto, en donde todos los elementos se clasifican correctamente.

La curva ROC usualmente va por sobre la línea que representa la identidad $TPR = FPR$. Cualquier sistema que vaya por debajo de dicha línea, es peor que un

sistema que asigna las clases en forma aleatoria, y por lo tanto no es aceptable. Supóngase un clasificador que con probabilidad p le asigna la clase positiva a un vector cualquiera de entrada, por lo que con probabilidad $1 - p$ le asigna la clase negativa. A dicho clasificador se le presentan N_n vectores negativos y N_p vectores positivos, según el estándar dorado. En promedio, suponiendo N_n y N_p bastante grandes, habrá pN_n falsos positivos y pN_p verdaderos positivos, ya que, independiente del estándar dorado, un vector cualquiera será clasificado como positivo con probabilidad p . Utilizando las ecuaciones (2) y (3) se puede apreciar que en este caso $FPR = \frac{pN_n}{N_n} = p = \frac{pN_p}{N_p} = TPR$. Así, variando el parámetro p es posible construir una curva ROC que coincida con la línea $FPR = TPR$. Esta línea corresponde a un clasificador aleatorio, que no utiliza ninguna información del vector de características que se le presenta. Por ende, cualquier sistema que sí utilice dicho vector debe, por lo menos, estar sobre esta línea.

2.1.3. Máquina de Soporte Vectorial

Una máquina de soporte vectorial (o SVM, sigla del inglés *Support Vector Machine*) es uno de los métodos más usados de clasificación. El método SVM básico [43] intenta encontrar el hiperplano, en el espacio de características, que permita separar en forma óptima entre dos clases asignadas a dichos vectores. Se busca que el hiperplano separador (representado por (\vec{w}, b)) satisfaga las siguientes condiciones:

$$\vec{w} \cdot \vec{x}_k + b \geq +1 \text{ Si } y_k = +1 \text{ y} \quad (8)$$

$$\vec{w} \cdot \vec{x}_k + b \leq -1 \text{ Si } y_k = -1, \quad (9)$$

donde \vec{x}_k corresponde un vector de características utilizado en el entrenamiento, e y_k corresponde al objetivo (dado por el estándar dorado) para dicho vector. Estas condiciones deben satisfacerse para todos los vectores \vec{x}_k que conformen el conjunto de entrenamiento.

En términos intuitivos, el hiperplano separador es tal que, si se traza el vector normal a dicho hiperplano, en uno de los sentidos de esta dirección normal quedan todos los vectores de una clase, y en el otro sentido, quedan todos los vectores de la otra clase. Por ejemplo, en un espacio de tres dimensiones, el hiperplano es simplemente un plano que separa dos áreas del espacio 3D, hacia uno de sus lados solo hay vectores de una clase y hacia el otro solo hay vectores de la otra clase. Nótese que es posible que este hiperplano no exista, ya que se puede dar el caso, según como se asignen las clases a los vectores de características, en que ningún plano deje a todos los vectores de una de las clases hacia uno de sus lados y todos los vectores de la otra clase hacia el otro lado. En este caso se dice que el problema de clasificación no es linealmente separable.

La separación óptima será aquella que permita el máximo margen entre las clases. Para ello, se debe escoger el hiperplano cuya distancia hacia cualquiera de las dos clases sea máxima. Esto quiere decir que la distancia entre dicho plano separador y el vector más cercano a él de entre los de la clase positiva (los cuales se encontrarán todos hacia un lado de este hiperplano) se maximiza en conjunto con la distancia entre aquél plano separador y el vector más cercano a él de entre los de la clase negativa (los cuales se encontrarán hacia el otro lado del hiperplano). Se puede demostrar que el máximo margen estará dado por la solución al siguiente problema de optimización [53]:

$$\min_{\vec{w}} \frac{1}{2} \vec{w} \cdot \vec{w}, \quad (10)$$

sujeto a:

$$y_k(\vec{w} \cdot \vec{x}_k + b) \geq +1, \quad (11)$$

para todos los ejemplos. Este problema de optimización se puede resolver directamente o mediante el problema dual. La resolución de este problema, entregará el vector \vec{w}^* y la constante b^* óptimos. Con esto, la función utilizada por el clasificador es:

$$F(x) = \text{signo}(\vec{w}^* \cdot \vec{x} + b^*). \quad (12)$$

Se puede demostrar que, en general, la solución al problema de optimización planteado por SVM está dada por la siguiente ecuación [53]:

$$\vec{w} = \sum_{k=1}^l v_k \vec{x}_k, \quad (13)$$

donde los \vec{x}_k corresponden a los vectores de soporte, y los v_k son constantes escalares que corresponden a la solución al problema de optimización. Los vectores de soporte son los vectores más cercanos al hiperplano separador, y son finalmente los únicos que determinan el límite entre ambas clases.

Para mejorar el algoritmo de clasificación mediante SVM es posible transformar el espacio de características en otro, usualmente de mayor dimensión. Para ello se introduce una función φ , la cual se aplica a los vectores de características para obtener nuevos vectores, y luego se resuelve el problema de optimización en este nuevo espacio [53]. La razón por la que se realiza esto es que generalmente un problema de clasificación será más sencillo de separar en un espacio de alta dimensionalidad. Además, es posible que un problema que no es linealmente separable en un espacio, sí lo sea en el nuevo espacio de mayor dimensión. Con esto, el nuevo clasificador será de la forma:

$$F(x) = \text{signo}(\vec{w} \cdot \varphi(\vec{x}) + b). \quad (14)$$

En general la representación mostrada en la ecuación (14) no es conveniente, debido a que $\varphi(\vec{x})$ puede ser un vector de dimensión demasiado alta (incluso infinita), lo que hace que el cálculo de $F(x)$ sea muy costoso (o simplemente imposible). Debido a esto, resulta conveniente introducir una función de kernel. La función de kernel, $k(\cdot, \cdot)$, cumple con la siguiente propiedad:

$$k(\vec{x}, \vec{y}) = \varphi(\vec{x}) \cdot \varphi(\vec{y}), \quad (15)$$

para todos los vectores \vec{x} e \vec{y} del espacio de menor dimensión. Debido a que \vec{w} se puede representar una como suma ponderada de vectores de soporte (ver ecuación (13)), la función de clasificación queda como:

$$F(\vec{x}) = \text{signo} \left(\sum_{k=1}^l v_k \varphi(\vec{x}_k) \varphi(\vec{x}) + b \right). \quad (16)$$

Así, para efectos de la clasificación no es necesario calcular los vectores $\varphi(\vec{x})$ del espacio de alta dimensión, sino que simplemente se requiere calcular el producto interno entre dichos vectores. Por ello, si se logra introducir una función de kernel que cumpla con la propiedad de la ecuación (15), se evita calcular cada una de las dimensiones de $\varphi(\vec{x})$, las cuales pueden ser infinitas, y todo este cálculo se reduce a evaluar el kernel $k(\cdot, \cdot)$. Esto se resume en la siguiente ecuación para el clasificador:

$$F(\vec{x}) = \text{signo} \left(\sum_{k=1}^l v_k \varphi(\vec{x}_k) \varphi(\vec{x}) + b \right) = \text{signo} \left(\sum_{k=1}^l v_k k(\vec{x}, \vec{x}_k) + b \right). \quad (17)$$

Como se observa en la ecuación anterior, la nueva representación de la función del clasificador ya no utiliza a la función φ , sino que solamente a la función de kernel (la cual tiene como dominio un espacio de dimensión uno). Esto también es cierto para la resolución del problema de optimización, en donde, si se resuelve el problema dual, la solución quedará expresada en términos de la función de kernel, y no directamente de los vectores en el espacio de alta dimensión. Con esto, ya no se requiere trabajar directamente en el espacio de salida de la función φ , pero se siguen teniendo todas las ventajas de trabajar en dicho espacio.

Existen diversas funciones de kernel utilizadas, siendo las más usuales el kernel polinomial:

$$k(\vec{x}, \vec{y}) = (\gamma \vec{x} \cdot \vec{y} + c)^d, \quad (18)$$

el kernel de función de base radial (o kernel gaussianp):

$$k(\vec{x}, \vec{y}) = e^{(-\gamma \|\vec{x} - \vec{y}\|^2)}, \quad (19)$$

y el kernel sigmoidal:

$$k(\vec{x}, \vec{y}) = \tanh(\kappa(\vec{x} \cdot \vec{y}) + \theta). \quad (20)$$

Como ya se dijo, es posible que el problema de clasificación no sea linealmente separable, incluso después de aplicar la función φ . En estos casos, se modifica ligeramente el problema de optimización, para encontrar el hiperplano que logre separar las clases de la mejor manera posible, reduciendo al máximo los errores de clasificación, que corresponden a elementos que aparecen en el lado del hiperplano que no corresponde a su clase. Para ello, el nuevo problema de optimización ahora será [53]:

$$\min_{\vec{w}, \xi_k} \frac{1}{2} \vec{w} \cdot \vec{w} + \frac{C}{n} \sum_{k=1}^n \xi_k, \quad (21)$$

sujeto a:

$$y_k(\vec{w} \cdot \vec{x}_k + b) \geq 1 - \xi_k. \quad (22)$$

Como se puede ver, se introducen las variables ξ_k , una para cada uno de los n vectores de características con que se entrena la SVM, con el propósito agregar holgura a la solución, permitiendo que algunas clasificaciones sean incorrectas. Este algoritmo es conocido como C-SVM. Notar que el parámetro C determina la importancia de los errores de clasificación en la optimización. Mientras mayor es C , más importancia se dará a estos errores, y por lo tanto habrá menos holgura. El caso extremo corresponde a C tendiendo a infinito, en donde se regresa al algoritmo original SVM, donde los errores de clasificación no son tolerados.

2.1.4. Redes Neuronales

Las redes neuronales (o NN, acrónimo de *Neural Networks*, en inglés) corresponden a una de las técnicas de inteligencia artificial más usadas y más exitosas. Consisten en una serie de elementos, o neuronas, conectados entre sí mediante pesos sinápticos. Es decir, cada neurona tendrá como entrada la suma ponderada de otros elementos, y a su vez su salida será ponderada para luego ingresar a otras neuronas. Modificando los pesos sinápticos entre las neuronas, es posible modificar el comportamiento de la red, según el objetivo deseado. Con esto, las redes neuronales pueden ser usadas para aproximar todo tipo de funciones, con cualquier grado de exactitud. Dentro de sus ventajas se encuentra

su gran capacidad para adaptarse a situaciones desconocidas o cambios en las condiciones del problema, la posibilidad de generalizar resultados obtenidos a partir del entrenamiento y su alta tolerancia a la falla (debido a que la información es almacenada en forma redundante). Son muy útiles para generar un modelo de un sistema del cual se posee poca información o no se desea comprender en gran detalle su funcionamiento interno. Esto se debe a que se puede crear una red capaz de solucionar un problema sin la necesidad de explicar cómo se hace, sino que simplemente conociendo las salidas deseadas para las entradas con las que se entrena dicha red. Por otro lado, una desventaja de las redes neuronales es que éstas derivan un modelo de caja negra, y por lo tanto es difícil estudiar su estructura para comprender cómo ésta soluciona el problema objetivo. Los primeros modelos de NN están basados en el trabajo de McCulloch y Pitts [35].

2.1.4.1. Definiciones Básicas

En términos generales, una red neuronal [18] se compone de varias capas de neuronas unidas por pesos sinápticos. La entrada de cada capa de neuronas puede provenir ya sea de estímulos externos o de salidas de otras capas de neuronas. Existen distintas formas de modelar el flujo de señales entre una capa y otra, siendo las más comunes:

- **Redes Neuronales Pre-alimentadas:** Las neuronas se organizan en capas sucesivas. Las conexiones de salida de una capa solo se realizan con las neuronas de la capa próxima, y las conexiones de entrada de una capa solo se realizan con las conexiones de la capa anterior. Estas redes también son conocidas como perceptrón multicapa o redes MLP (acrónimo de *Multi-layer Perceptron*, en inglés).
- **Redes Recurrentes:** Al igual que las redes pre-alimentadas, sus neuronas se organizan en capas sucesivas. Sin embargo, en estas redes es posible que las conexiones de salida de una capa vayan a la entrada de neuronas en una capa anterior o incluso en la misma capa. En otras palabras, en estas redes existe al menos un camino cíclico formado por las conexiones entre las neuronas de la red. Esto le da la capacidad a las redes de tener estados en función a los datos ingresados, en donde el procesamiento de nuevos datos depende de dichos estados y por ende de datos ya vistos. Debido a esto, estas redes poseen memoria. En general las redes recurrentes son más poderosas que las redes MLP, pero son más difíciles de modelar y de entrenar.

Las redes neuronales se pueden usar para diferentes propósitos, entre ellos la clasificación. Dependiendo de este propósito y de la estructura que posea la red neuronal, existen diferentes formas de entrenarla. Normalmente esta forma estará basada en uno de los siguientes esquemas:

- **Aprendizaje Supervisado:** Este es el esquema más usado para problemas de clasificación. Requiere de una base de datos con vectores, los cuales corresponden a la entrada de la red neuronal, y además requiere, para cada uno de estos vectores, valores deseados para la salida. Consiste en ir presentando los vectores en la entrada en forma sucesiva e ir calculando el error o la diferencia entre la salida obtenida por la red y la salida deseada. A partir de este error, se ajustan las conexiones sinápticas de las neuronas, con el objetivo de reducirlo en cada iteración.
- **Aprendizaje no Supervisado:** En este caso, los datos de entrada no están asociados a una salida deseada para la red. Ésta debe, por sí sola, obtener información de estos datos, determinando relaciones y estructuras relevantes a partir de ellos. En general este esquema no es muy utilizado en problemas de clasificación, pero puede servir para etapas de pre-procesamiento.
- **Aprendizaje Reforzado:** En este esquema no se cuenta con salidas deseadas para la red. Sin embargo, el sistema realiza acciones y es capaz de analizar el desempeño de dichas acciones. Este desempeño es dado por el entorno, en base a acciones pasadas realizadas por la NN. Así, la red es capaz de entrenarse a partir de su interacción con el entorno, aunque no posea una salida objetivo para cada entrada.

En este trabajo se utilizarán redes neuronales pre-alimentadas, entrenadas mediante aprendizaje supervisado. A continuación se explica en mayor detalle el funcionamiento de estas redes y la forma en que serán entrenadas.

2.1.4.2. Definiciones redes MLP

En la Figura 3 se observa una red de neuronas, o nodos, ordenadas en capas de izquierda a derecha. Dicha red es una red neuronal MLP [2] que contiene varias capas de neuronas, en donde las salidas de las neuronas de una capa se conectan únicamente con todas las neuronas de la capa siguiente, y las entradas de las neuronas de una capa se conectan únicamente con todas las neuronas de la capa previa.

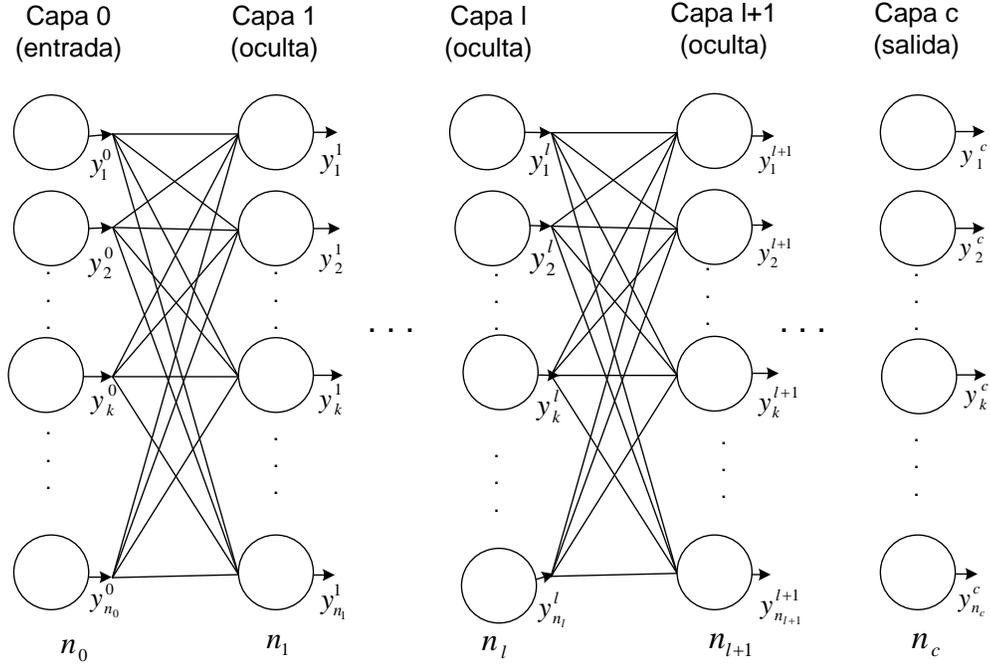


Figura 3: Red Neuronal MLP Genérica.

Cada neurona se conecta únicamente con las neuronas de la capa siguiente, y lo hace con todas ellas. Contiene $c + 1$ capas en total, donde una capa l cualquiera tiene n_l neuronas. La salida de una neurona k cualquiera, perteneciente a una capa l cualquiera, se denota y_k^l .

Se puede observar en la Figura 3 que cada neurona tiene una salida identificada con la letra y . El subíndice de dicha salida está asociado al número de la neurona en esa capa, mientras que el superíndice está asociado al número de la capa. Así, por ejemplo, la salida y_k^l corresponde a la salida de la neurona k de la capa l . Sin pérdida de generalidad, se ha supuesto que la red tiene $c + 1$ capas de neuronas, desde la capa 0, que corresponde a la capa de entrada, hasta la capa c , que corresponde a la capa de salida. Las otras capas, de la 1 a la $c - 1$, se conocen como capas ocultas. Es posible que c sea igual a 1, en cuyo caso no existirán capas ocultas. Para una capa l cualquiera existen n_l neuronas, numeradas desde la 1 hasta la n_l .

La ecuación para la salida de una neurona genérica (neurona k de la capa l) será la siguiente [2]:

$$y_k^l = \varphi(s_k^l), \quad (23)$$

donde

$$s_k^l = \sum_{j=1}^{n_{l-1}} w_{kj}^l y_j^{l-1}. \quad (24)$$

En la ecuación (24), el término w_{kj}^l corresponde al peso de la unión entre la neurona k de la capa l y la neurona j de la capa $l - 1$. φ es una función de activación que acota la suma de entradas provenientes de las capas anteriores. Para satisfacer esto, φ debe ser una función de saturación o de umbral, de tal forma que cuando s tome valores extremos, $\varphi(s)$ tome un valor relativamente constante, que corresponde a la cota impuesta sobre s . Las funciones más usadas para este propósito son la función logística y la función tangente hiperbólica, cuyas fórmulas mostramos a continuación. La fórmula de la función logística es:

$$y = \varphi(s) = \frac{1}{1 + e^{-as}}. \quad (25)$$

Esta función cumple con la siguiente propiedad:

$$y \in [0, 1], \quad (26)$$

y su derivada está dada por la siguiente ecuación:

$$\varphi'(s) = \varphi(s)(1 - \varphi(s)). \quad (27)$$

La fórmula de la función tangente hiperbólica es:

$$y = \varphi(s) = \tanh(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}. \quad (28)$$

Esta función cumple con la siguiente propiedad:

$$y \in [-1, 1], \quad (29)$$

y su derivada está dada por la siguiente ecuación:

$$\varphi'(s) = (1 + \varphi(s))(1 - \varphi(s)) = 1 - \varphi(s)^2. \quad (30)$$

Como se observa, ambas funciones tienen una salida acotada, la logística toma valores entre 0 y 1, y la tangente hiperbólica entre -1 y 1.

Las ecuaciones (23) y (24) son válidas para todo valor de l entre 1 y c . Nótese que no son válidas para $l = 0$, pues no existen neuronas antes de la capa 0. En este caso, las salidas de las neuronas en dicha capa se obtienen directamente de las entradas, según:

$$y_k^0 = x_k, \quad (31)$$

donde x_k corresponde a la entrada k de la red neuronal.

2.1.4.3. Sesgo

Normalmente una red neuronal funcionará mejor si se introduce un sesgo en la suma de sus entradas [18], es decir, si se agrega una constante a la suma s de cada neurona. De esta forma, aunque todas las salidas de las neuronas de una capa una sean nulas, cuando se calcule s en alguna neurona en la capa siguiente, esta suma puede tomar un valor distinto de cero. Una forma de introducir este sesgo es agregar una nueva neurona a cada capa, salvo a la capa de salida. La salida y de esta nueva neurona es siempre constante e igual a 1. Así, el sesgo en la suma de cada neurona está especificado por el peso de la unión entre dicha neurona y la neurona de salida constante de la capa anterior. Con esto, se agrega un sesgo a cada suma sin modificar las fórmulas básicas ya mostradas. Esto permite, además, ajustar dicho sesgo de la misma manera que se entrena un peso cualquiera de la red, lo que a su vez simplifica la modelación del algoritmo de entrenamiento.

La Figura 4 muestra una red neuronal MLP con sesgo. En dicha figura se puede apreciar que en cada capa, salvo la última, existe una neurona identificada por un 1 en su interior. Estas neuronas no tienen entradas, y su salida siempre es constante e igual a uno, es decir:

$$y_0^l = 1 \quad \forall l \in \{0, \dots, c - 1\}. \quad (32)$$

Con esto, todas las capas ocultas y la capa de entrada tienen una neurona más, que corresponde a la neurona de salida constante. Finalmente, las fórmulas (23) y (24), para las salidas de las capas ocultas y de entrada se modifica ligeramente, quedando de la siguiente forma:

$$y_k^l = \begin{cases} \varphi(s_k^l) & \text{si } k \neq 0 \\ 1 & \text{si } k = 0 \end{cases}, \quad (33)$$

donde:

$$s_k^l = \sum_{j=0}^{n_{l-1}} w_{kj}^l y_j^{l-1}. \quad (34)$$

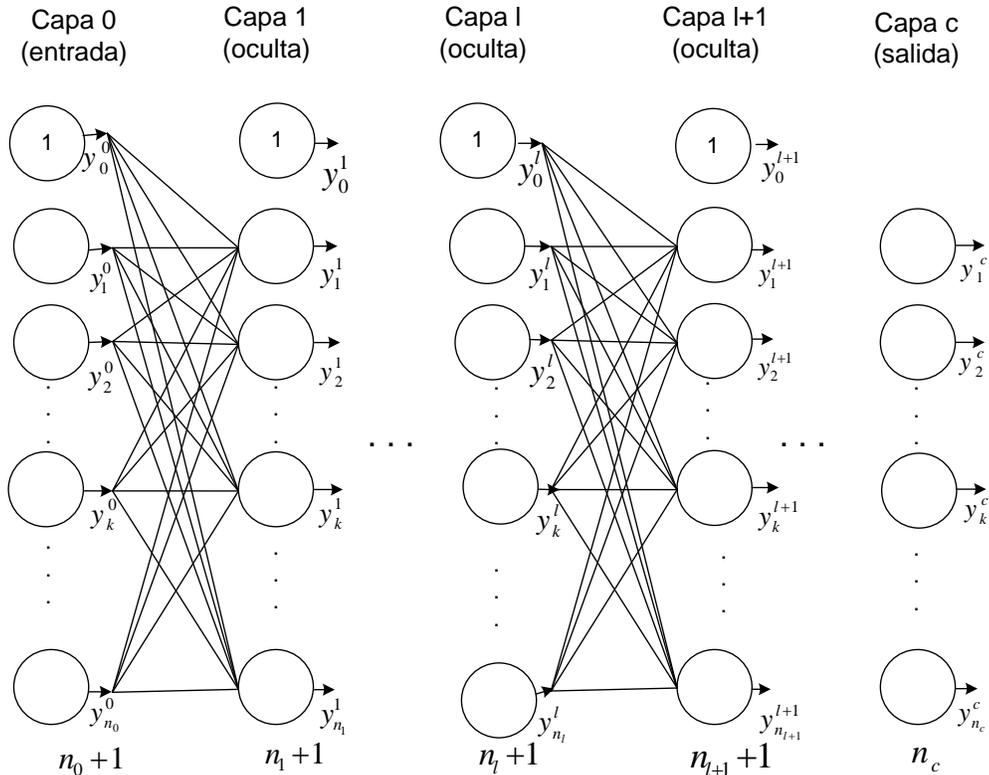


Figura 4: Red Neuronal MLP con Sesgo.

Contiene $c + 1$ capas en total, donde una capa l cualquiera tiene $n_l + 1$ neuronas (de la 0 a la n_l), salvo la capa c que no contiene neurona 0, y solo tiene n_c neuronas. La salida de una neurona k cualquiera, perteneciente a una capa l cualquiera, se denota y_k^l . Las salidas de las neuronas 0 de cada capa, denotadas por y_0^k , son constantes e igual a 1.

2.1.4.4. Entrenamiento Mediante Retropropagación

Como se explicó en la Figura 3, toda red MLP posee una capa de salida con n_c neuronas. Cada una de estas neuronas corresponde a una salida de la red. El entrenamiento mediante retropropagación [57] permite ajustar los pesos w_{kj}^l de las interconexiones entre las neuronas, de tal manera que las salidas la red se asemejen a las deseadas. Así, mediante entrenamiento supervisado, es posible modificar los pesos de la red para resolver el problema específico que se tenga.

La red comienza con todos sus pesos iniciados aleatoriamente, y posteriormente éstos se van ajustando en el proceso de entrenamiento. Durante esta fase de adiestramiento se tiene un cierto número de vectores de características de salida conocida, llamado conjunto de entrenamiento. Los elementos de dicho conjunto son presentados a la red uno por uno, registrando la diferencia entre la salida obtenida y la salida deseada. Dicha diferencia es conocida como el error. Este error es minimizado mediante la resolución de un problema de optimización, donde las variables de optimización son los valores de

los pesos de la red. Una forma de realizar dicha minimización es mediante el algoritmo de retropropagación del error, el cual se explica a continuación.

La regla de retropropagación funciona de la siguiente manera. Suponiendo que el algoritmo de entrenamiento se encuentra en la iteración t , se alimenta la red con un vector de características $\vec{x}(t)$, de dimensión n_0 . Dicho vector de entrada estará asociado a un vector de salida deseado, $\vec{d}(t)$, de dimensión n_c . Al alimentar la red, se producirá en la capa de salida un vector $\vec{y}^c(t)$, cuyas componentes serán las salidas escalares de las n_c neuronas en esa capa. Dicho de otra forma, para una neurona de salida k cualquiera, se tendrá una salida $y_k^c(t)$ y un objetivo $d_k(t)$, siendo estas las componentes escalares de los vectores $\vec{y}^c(t)$ y $\vec{d}(t)$, respectivamente. Con esto, el error en la neurona de salida k estará dado por la siguiente expresión:

$$\varepsilon_k^2(t) = (d_k(t) - y_k^c(t))^2 = (d_k(t) - \varphi(s_k^c(t)))^2 = \left(d_k(t) - \varphi \left(\sum_{j=0}^{n_{c-1}} w_{kj}^c(t) \cdot y_j^{c-1}(t) \right) \right)^2. \quad (35)$$

El error total a minimizar se define como:

$$\varepsilon_T(t) = \sum_{k=1}^{n_c} \varepsilon_k^2(t). \quad (36)$$

El error total, $\varepsilon_T(t)$, es minimizado mediante el algoritmo de gradiente descendiente. Para ello, se consideran como las variables de la minimización a todos los pesos $w_{kj}^l(t)$, de todas las capas de la red. En primer lugar, se calculan las derivadas parciales del error total con respecto a los pesos que van a neuronas en la de salida, $w_{kj}^c(t)$, las cuales están dadas por:

$$\frac{\partial \varepsilon_T(t)}{\partial w_{kj}^c(t)} = \frac{\partial \varepsilon_k^2(t)}{\partial w_{kj}^c(t)} = 2\varepsilon_k(t) \frac{\partial \varepsilon_k(t)}{\partial w_{kj}^c(t)} = 2\varepsilon_k(t) \left\{ -\varphi'(s_k^c(t)) y_j^{c-1}(t) \right\}. \quad (37)$$

Posteriormente, cada uno de estos pesos se actualiza según las siguientes ecuaciones:

$$\Delta w_{kj}^c(t) = -\mu \frac{\partial \varepsilon_k^2(t)}{\partial w_{kj}^c(t)} = 2\mu \varepsilon_k(t) \varphi'(s_k^c(t)) y_j^{c-1}(t), \quad (38)$$

y

$$w_{kj}^c(t+1) = w_{kj}^c(t) + \Delta w_{kj}^c(t) = w_{kj}^c(t) + 2\mu \varepsilon_k(t) \varphi'(s_k^c(t)) y_j^{c-1}(t). \quad (39)$$

En las dos ecuaciones anteriores se utiliza un parámetro μ , que se conoce como tasa de aprendizaje. Dicho parámetro multiplica al gradiente en la actualización de los pesos y debe ser ajustado para permitir controlar la estabilidad y la velocidad

de convergencia del algoritmo. Aumentarlo hace que el método se acerque más rápido a la solución, pero lo vuelve más inestable, pudiendo éste no converger para valores muy altos de μ . Debido a esto, es importante fijarlo correctamente, sin que tome valores demasiado altos o demasiado bajos.

Definiendo como delta a la siguiente expresión:

$$\delta_k^c(t) = 2\varepsilon_k(t)\varphi'(s_k^c(t)), \quad (40)$$

se tiene que:

$$\Delta w_{kj}^c(t) = \mu \delta_k^c(t) y_j^{c-1}(t). \quad (41)$$

La ecuación anterior se conoce como la regla delta. Nótese que:

$$\delta_k^c(t) = 2\varepsilon_k(t)\varphi'(s_k^c(t)) = -\frac{\partial \varepsilon_T(t)}{\partial s_k^c(t)}. \quad (42)$$

Con esta igualdad, es posible extender la regla a las capas ocultas y a la capa de entrada, según las siguientes fórmulas:

$$\Delta w_{kj}^l(t) = \mu \delta_k^l(t) y_j^{l-1}(t) \quad (43)$$

y

$$\delta_k^l(t) = -\frac{\partial \varepsilon_T(t)}{\partial s_k^l(t)} = -\sum_{j=1}^{n_{l+1}} \frac{\partial \varepsilon_T(t)}{\partial s_j^{l+1}(t)} \cdot \frac{\partial s_j^{l+1}(t)}{\partial s_k^l(t)}. \quad (44)$$

Nótese que:

$$\frac{\partial \varepsilon_T(t)}{\partial s_j^{l+1}(t)} = \delta_j^{l+1}(t) \quad (45)$$

y

$$\frac{\partial s_j^{l+1}(t)}{\partial s_k^l(t)} = \frac{\partial (\sum_{p=1}^{n_l} w_{jp}^{l+1}(t) y_p^l(t))}{\partial s_k^l(t)} = \frac{\partial (\sum_{p=1}^{n_l} w_{jp}^{l+1}(t) \varphi(s_p^l(t)))}{\partial s_k^l(t)} = w_{jk}^{l+1}(t) \varphi'(s_k^l(t)), \quad (46)$$

de donde concluimos finalmente que el δ de cualquier neurona entre las capas 1 y $c - 1$ se puede expresar como:

$$\delta_k^l(t) = -\varphi'(s_k^l(t)) \sum_{j=1}^{n_{l+1}} \delta_j^{l+1}(t) w_{jk}^{l+1}(t). \quad (47)$$

En resumen, en la ecuación (43) se muestra que la regla de actualización del método del gradiente conjugado para un peso $w_{kj}^l(t)$ cualquiera de la red. Dicha expresión depende solamente del parámetro μ , el cual es conocido; de la salida de una neurona de la capa anterior, $y_j^{l-1}(t)$, la cual también es conocida; y del valor $\delta_k^l(t)$, el cual debe ser calculado. En la ecuación (47) se muestra una forma de calcular este valor para todas las neuronas de las capa ocultas o la de entrada (l entre 1 y $c - 1$). Nótese que el cálculo de $\delta_k^l(t)$ requiere conocer todos los $\delta_j^{l+1}(t)$ de la capa siguiente. Por lo tanto, la ecuación (47) resulta ser una ecuación de recurrencia, en donde el caso base corresponde a la capa de salida. En esta capa es posible calcular los $\delta_k^c(t)$ directamente, utilizando la ecuación (40). Así, para calcular los valores $\Delta w_{kj}^l(t)$, necesarios para actualizar los pesos sinápticos en la iteración t , se comienza alimentando la red con la entrada $\overrightarrow{x(t)}$, y se calculan las salidas de todas las neuronas. Luego, se actualizan todos los pesos de la capa de salida y después se va retrocediendo en las capas, calculando los $\delta_k^l(t)$ y los $\Delta w_{kj}^l(t)$ en forma recursiva. Debido a esta segunda parte, en donde se calculan los valores $\Delta w_{kj}^l(t)$ desde la salida hacia la entrada, este método recibe el nombre de algoritmo de retropropagación. Una vez que se han calculado todos los $\Delta w_{kj}^l(t)$, éstos se utilizan para modificar los pesos de la red, y se pasa a la iteración $t + 1$, en donde se utiliza un ejemplo distinto, que corresponde a $\overrightarrow{x(t + 1)}$. Así, se implementa el método de gradiente descendente para minimizar el error en función de los pesos de la NN, pero de una forma eficiente que permite ahorrar muchos cálculos a la hora de evaluar las derivadas parciales.

Existen dos variantes de este algoritmo, el aprendizaje en grupo (más conocido por su palabra en inglés, *batch*) y el aprendizaje estocástico. En el aprendizaje *batch*, luego de terminar la iteración t y calcular los $\Delta w_{kj}^l(t)$, éstos no se aplican directamente a los $w_{kj}^l(t)$, si no que se guardan y se pasa a la iteración $t + 1$. Posteriormente, cuando ya se ha acumulado un cierto número de Δw_{kj}^l , para distintos valores de t , estos se aplican en conjunto sobre los pesos de la red. En el aprendizaje estocástico, en cambio, cada vez que se calculan los $\Delta w_{kj}^l(t)$, estos se aplican inmediatamente a los pesos de la red, y luego se pasa a la iteración siguiente. El aprendizaje estocástico es usualmente más rápido que el aprendizaje *batch*, obtiene mejores soluciones, y se puede usar en línea, es decir, en una red que se va entrenando al mismo tiempo que se utiliza para calcular salidas desconocidas. Por su parte, para el aprendizaje *batch* es más fácil modelar su comportamiento teóricamente, y muchas técnicas de aceleración del algoritmo de retropropagación, por ejemplos métodos de segundo orden, solo se aplican a esta variante.

En el algoritmo anterior, se introdujo un parámetro μ , conocido como tasa de aprendizaje. Se vio que este parámetro hace más rápida la convergencia del algoritmo de optimización, pero puede causar que el método se vuelva inestable y

no converja. Una forma de aumentar la velocidad de convergencia del método y al mismo tiempo disminuir su inestabilidad, es introducir una tasa de momentum, denotada por α . Con esta tasa, las ecuaciones (41) y (43) se modifican, quedando de la siguiente forma:

$$\Delta w_{kj}^c(t) = \mu \delta_k^c(t) y_j^{c-1}(t) + \alpha \Delta w_{kj}^c(t-1) \quad (48)$$

y

$$\Delta w_{kj}^l(t) = \mu \delta_k^l(t) y_j^{l-1}(t) + \alpha \Delta w_{kj}^l(t-1). \quad (49)$$

Como se puede observar, se agrega un término en donde aparece el Δw de la iteración anterior ($t-1$), de tal manera de agregar inercia a las variaciones de los pesos w .

El algoritmo de retropropagación se beneficia en el caso en que las funciones utilizadas sean sigmoideas, tales como las funciones mostradas en las ecuaciones (25) y (28). Esto se debe a que la derivada de estas funciones, las cuales se muestran en las ecuaciones (27) y (30), se puede expresar como un polinomio de la función original. Así, en estos casos, las ecuaciones para los δ , es decir, las ecuaciones (40) y (47), se pueden volver a escribir en la forma que se muestra a continuación. Para el caso de la función logística, estas ecuaciones se reemplazan por las siguientes:

$$\delta_k^c(t) = 2 \cdot \varepsilon_k(t) \cdot y_k^c(t) \cdot (1 - y_k^c(t)) \quad (50)$$

y

$$\delta_k^l(t) = y_k^l(t) \cdot (1 - y_k^l(t)) \sum_{j=1}^{n_{l+1}} \delta_j^{l+1}(t) \cdot w_{jk}^{l+1}(t). \quad (51)$$

Para la tangente hiperbólica, las ecuaciones (40) y (47) quedan de la siguiente forma:

$$\delta_k^c(t) = 2 \cdot \varepsilon_k(t) \cdot (1 + y_k^c(t)) \cdot (1 - y_k^c(t)) \quad (52)$$

y

$$\delta_k^l(t) = (1 + y_k^l(t)) \cdot (1 - y_k^l(t)) \sum_{j=1}^{n_{l+1}} \delta_j^{l+1}(t) \cdot w_{jk}^{l+1}(t). \quad (53)$$

Como se puede apreciar, para estas dos funciones sigmoideas los δ ya no dependen de la derivada de φ evaluada en la suma ponderada s , si no que

directamente de la salida y de la neurona. Esto simplifica los cálculos que realiza el método, pues ahora sólo es necesario que el algoritmo recuerde las salidas y de las neuronas, calculadas durante la evaluación inicial de la salida de red frente al nuevo ejemplo, sin la necesidad de almacenar en memoria o volver a computar las sumas s durante la retro propagación.

2.2. Gráficos de Recurrencia

2.2.1. Motivación

Los gráficos de recurrencia (o RP, sigla del inglés *Recurrence Plot*), corresponden a gráficos 2D construidos a partir de los estados de un sistema dinámico, con el objetivo de visualizar las recurrencias presente en dicho sistema [31]. En términos sencillos, los RP permiten observar gráficamente la similitud de los estados en distintos instantes de tiempo, para determinar cuándo y cómo estos estados vuelven a ser similares.

Los RP se componen de dos ejes iguales de tiempo y marcan, en cada coordenada (t_1, t_2) , si es que el estado del sistema dinámico en t_1 es similar o no al estado en t_2 . En caso de que sea similar, se pinta de negro el punto (t_1, t_2) , y en caso contrario, se deja en blanco. Cómo se definen los estados del sistema, y qué se entiende por dos estados similares, depende del problema particular.

En general, los RP son usados en sistemas en que se quiere estudiar alguna recurrencia o periodicidad, pudiendo ésta ser bastante compleja. Debido a esto, son usados en este trabajo para detectar HS, ya que éstos, por definición, corresponden a una onda cuyos componentes de frecuencia se encuentra principalmente en un rango bastante acotado. Por lo tanto, su periodo es relativamente constante, por lo que los HS corresponden a patrones recurrentes o con cierto grado de periodicidad. Otras aplicaciones en donde los RP han sido utilizados incluyen la astrofísica [1] [27] [59], geofísica [30] [32] [33], ingeniería [9] [36] y biología [14] [29], todas aplicaciones en donde aparecen fenómenos periódicos de interés.

2.2.2. Definición de Gráficos de Recurrencia

En términos formales, un RP se construye a partir de N vectores de estado de un sistema. Estos estados, se miden en distintos instantes de tiempo, los cuales denotaremos mediante la siguiente sucesión: $\{t_k\}_{k=0}^N$. Normalmente, estos tiempos están separados por distancias iguales, aunque esto no es una condición obligatoria. Para cada tiempo t_k , existe un vector que representa el estado del sistema en dicho momento, el cual denotaremos \vec{x}_k , y todos los estados se denotarán mediante la sucesión: $\{\vec{x}_k\}_{k=0}^N$. Dados estos estados, se define la matriz

de recurrencia como una matriz de tamaño $N \times N$ cuyos componentes, $R_{k,q}$, se determinan según la siguiente ecuación:

$$R_{k,q} = \begin{cases} 1 & \text{si } \vec{x}_k \approx \vec{x}_q \\ 0 & \text{si } \vec{x}_k \not\approx \vec{x}_q \end{cases}, \quad (54)$$

en donde $\vec{x}_k \approx \vec{x}_q$ quiere decir que los estados \vec{x}_k y \vec{x}_q son similares. En la mayoría de los casos, dicha similitud estará dada por una distancia $d(\cdot, \cdot)$. Así, se dirá que dos estados son similares si se cumple que $d(\vec{x}_k, \vec{x}_q) < \varepsilon$. Con esto, se redefinen los elementos $R_{k,q}(\varepsilon)$ de la matriz de recurrencia según la siguiente ecuación:

$$R_{k,q}(\varepsilon) = \begin{cases} 1 & \text{si } d(\vec{x}_k, \vec{x}_q) < \varepsilon \\ 0 & \text{si } d(\vec{x}_k, \vec{x}_q) \geq \varepsilon \end{cases}. \quad (55)$$

Dada la matriz de recurrencia, el RP es directamente una visualización de dicha matriz, en donde los píxeles de color negro corresponden a $R_{k,q} = 1$, y los de color blanco corresponden a $R_{k,q} = 0$. Cada elemento $R_{k,q}$ de la matriz de recurrencia está asociado a los instantes t_k y t_q , por lo que los ejes de esta matriz se pueden considerar como ejes de tiempo. Por convención, estos ejes crecen hacia la derecha y hacia arriba. Nótese que, debido a que $d(\vec{x}_k, \vec{x}_k) = 0$ para cualquier valor de k , siempre se tiene que $R_{k,k} = 1$, lo que implica que existe una línea negra en la diagonal principal de cualquier RP. Esta línea es conocida como Línea de Identidad (o LOI, proveniente del inglés *Line Of Identity*).

Nótese que el parámetro ε influye en gran medida en el RP obtenido, y por ende determinar un valor correcto para dicho parámetro es de gran importancia. En general, un ε demasiado pequeño causa que no existan puntos recurrentes (es decir, puntos donde $R_{k,q} = 1$), mientras que un ε demasiado grande causa que muchos puntos sean recurrentes, induciendo a conclusiones erróneas sobre la dinámica del sistema.

2.2.3. Método de Retraso de Tiempo

En situaciones experimentales, es común que no se puedan medir todas las componentes del vector \vec{x}'_k que define completamente el estado del sistema dinámico en el instante de tiempo t_k . En la mayoría de los casos, se tienen mediciones de una única variable de este sistema. Dichas mediciones serán denotadas por la siguiente sucesión: $\{u_k\}_{k=0}^N$. Una forma muy común de reconstruir los vectores de estado en su espacio original, a partir de estas mediciones escalares, es el método de retraso de tiempo [31] (más conocido como *time delay method*). Cada vector se reconstruye mediante la siguiente fórmula:

$$\vec{x}_k = \sum_{l=1}^m u_{k+\tau(l-1)} \vec{e}_l, \quad (56)$$

donde los vectores $\{\vec{e}_l\}_{l=1}^m$ corresponden a una base ortogonal del espacio de estados, τ a un parámetro conocido como retraso y m a la dimensión de dicho espacio. Con esto, a partir la serie de tiempo se obtendrán los vectores $\{\vec{x}_k\}_{k=0}^N$, con los cuales se puede construir el RP según el procedimiento ya explicado. Nótese que estos vectores reconstruidos no son exactamente los vectores $\{\vec{x}'_k\}_{k=0}^N$ originales que representaban la trayectoria del sistema dinámico. Sin embargo, el teorema de Takens [42][51] asegura que bajo ciertas condiciones del valor de m , ambas trayectorias pueden considerarse como una representación del mismo sistema dinámico, pero en sistemas de coordenadas distintos. Así, las propiedades de recurrencia que se descubran para una trayectoria, siguen siendo válidas para la otra.

En general, si solo se poseen las mediciones escalares u_k , el verdadero valor de la dimensión del espacio de estados del sistema dinámico es desconocido. Por ello, m debe ser fijado como un parámetro, conocido como dimensión de imbibición (o *embedding dimension* en inglés). Para que el teorema de Takens sea válido m debe ser superior a un cierto valor especificado en dicho teorema. Se han desarrollado métodos para estimar el menor valor de m que satisface esta condición, permitiendo tener la dimensión imbibición más pequeña, y por lo tanto los vectores reconstruidos más sencillos, donde siga siendo válido el teorema de Takens. En este trabajo se utiliza el método de los falsos vecinos para fijar el parámetro m , cuya explicación aparece en [24]. También resulta fundamental fijar correctamente el parámetro τ o retraso. Ambos parámetros, τ y m , influyen sustancialmente en la calidad de los RP construidos, por lo que deben ser fijados correctamente.

2.2.4. Distancias

En la ecuación (55) se observa que para calcular un RP es necesario definir una distancia entre vectores de estado del sistema dinámico. Esta distancia se puede escoger de distintas maneras, lo que dará origen a distintos tipos de RP. A continuación, se detallan las distancias más comunes a la hora de construir gráficos de recurrencia. Además, se detallan aquellas distancias que no son normalmente utilizadas en la literatura, pero que fueron utilizadas en este trabajo.

2.2.4.1. Distancia Euclidiana

La distancia euclidiana es una de las más utilizadas para calcular RP, debido a su simpleza. Se define según la siguiente ecuación:

$$d_e(\vec{x}_k, \vec{x}_q) = \|\vec{x}_k - \vec{x}_q\|_2 = \sqrt{\sum_{i=1}^m (x_k^i - x_q^i)^2}. \quad (57)$$

Para este trabajo, se modifica ligeramente esta distancia. La razón de esto es que las señales utilizadas en esta memoria tienen amplitudes que se mueven en rangos muy grandes. A causa de esto, las distancias entre los elementos en ciertas zonas de la señal son varios órdenes de magnitud mayores que las distancias en otras zonas. A causa de esto, los RP en distintas zonas de la señal no eran comparables, haciendo imposible procesarlos de la misma manera para detectar HS. Debido a ello, se modifica la distancia euclidiana, dividiéndola por la suma de la norma de ambos vectores. Gracias a esta normalización, se logra que el tamaño de los vectores influya poco en el valor de esta distancia, lo que permite resultados invariantes a la magnitud de la señal. La fórmula para esta distancia normalizada es:

$$d_{en}(\vec{x}_k, \vec{x}_q) = \frac{2\|\vec{x}_k - \vec{x}_q\|_2}{\|\vec{x}_k\| + \|\vec{x}_q\|} = \frac{2}{\|\vec{x}_k\| + \|\vec{x}_q\|} \sqrt{\sum_{i=1}^m (x_k^i - x_q^i)^2}. \quad (58)$$

2.2.4.2. Distancia Máxima

La distancia máxima es una distancia muy usada para construir RP. Se define mediante la siguiente fórmula:

$$d_{max}(\vec{x}_k, \vec{x}_q) = \|\vec{x}_k - \vec{x}_q\|_{max} = \max\{|x_k^1 - x_q^1|, |x_k^2 - x_q^2|, \dots, |x_k^m - x_q^m|\}. \quad (59)$$

En este trabajo no se utiliza esta distancia, pues, según estudios preliminares, no era adecuada para el tipo de problemas abordados en este caso.

2.2.4.3. Distancias a Partir de Correlación

La correlación es una medida de similitud entre vectores que alcanza su valor máximo de 1 cuando ambos vectores son paralelos. Se define una distancia a partir de la correlación como la resta entre este valor máximo posible y dicha correlación, es decir:

$$d_{correl}(\vec{x}_k, \vec{x}_q) = 1 - \frac{\sum_{i=1}^m (x_k^i - \bar{x}_k)(x_q^i - \bar{x}_q)}{\sqrt{\sum_{i=1}^m (x_k^i - \bar{x}_k)^2} \sqrt{\sum_{i=1}^m (x_q^i - \bar{x}_q)^2}}. \quad (60)$$

Nótese que cuando la correlación es máxima, esta distancia es cero. Se puede demostrar que la correlación corresponde al ángulo entre los dos vectores. Debido a esto, su valor no depende de la magnitud de dicho vectores, lo que hace innecesario normalizar esta distancia.

Para esta distancia, se introdujo una modificación de tal manera de resaltar ciertas frecuencias específicas presentes en la señal del EEG. Esto se hizo mediante un kernel periódico, el cual se define mediante la siguiente fórmula:

$$k(t_k, t_q, \sigma_t, p) = G_{\sigma_t} \left(2 \sin \left(\frac{\pi}{p} (t_k - t_q) \right) \right), \quad (61)$$

donde:

$$G_{\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{x^2}{2\sigma} \right). \quad (62)$$

El parámetro p corresponde al periodo del kernel. El valor máximo de la expresión mostrada en la ecuación (61) se alcanza cuando:

$$|t_k - t_q| = zp \quad \forall z \in \mathbb{Z}, \quad (63)$$

es decir, cuando t_k y t_q se encuentran separados por un número entero de periodos. Debido a esto, al multiplicar la correlación en la expresión (60) por un kernel periódico, se obtienen menores distancias entre aquellos elementos que se encuentran separados temporalmente por un número entero de periodos p . En otras palabras, este kernel tiende a resaltar las recurrencias de periodo p que pudieran aparecer en el sistema dinámico. La fórmula para esta distancia es:

$$d_{correl-kp}(\vec{x}_k, \vec{x}_q, \sigma_t, p) = G_{\sigma_t}(0) - \left[\frac{\sum_{i=1}^m (x_k^i - \bar{x}_k) \cdot (x_q^i - \bar{x}_q)}{\sqrt{\sum_{i=1}^m (x_k^i - \bar{x}_k)^2} \cdot \sqrt{\sum_{i=1}^m (x_q^i - \bar{x}_q)^2}} \right] \left[G_{\sigma_t} \left(2 \sin \left(\frac{\pi}{p} \cdot (t_k - t_q) \right) \right) \right], \quad (64)$$

donde σ_t se conoce como el tamaño del kernel. En este caso, el nuevo valor máximo que puede tomar que la expresión que se resta en la ecuación (64) es $G_{\sigma_t}(0) = \frac{1}{\sqrt{2\pi}\sigma_t}$, por lo que se modifica dicha ecuación agregando este valor en la parte izquierda de la resta, de tal manera que la máxima similitud esté asociada a una distancia nula.

2.2.4.4. Distancias a Partir de Auto-Correntropía

Las tres distancias definidas a continuación se basan en la correntropía [17], que corresponde a una medida no lineal de similitud entre dos vectores. Se define como:

$$C(\vec{x}_k, \vec{x}_q, \sigma_d) = \frac{1}{m} \sum_{i=1}^m G_{\sigma_d}(x_k^i - x_q^i), \quad (65)$$

donde $G_{\sigma_d}(\cdot)$ es la misma función definida en (62). Se habla de auto-correntropía debido a que, en el caso de este trabajo, ambos vectores provienen de una misma señal. El parámetro σ_d se conoce como el tamaño del kernel gaussiano. A partir de esta medida de similitud, se define la distancia a partir de la auto-correntropía como:

$$d_{corren}(\vec{x}_k, \vec{x}_q, \sigma_d) = G_{\sigma_d}(0) - \frac{1}{m} \sum_{i=1}^m G_{\sigma_d}(x_k^i - x_q^i), \quad (66)$$

donde el valor máximo posible para el kernel, $G_{\sigma_d}(0)$, aparece a la izquierda de la resta, de tal forma de que la máxima similitud entre vectores esté asociada a una distancia nula.

Por la misma razón que se debió normalizar la distancia euclidiana, es necesario normalizar la distancia a partir de la auto-correntropía. Esto se hace mediante la siguiente fórmula:

$$d_{corren-n}(\vec{x}_k, \vec{x}_q, \sigma_d) = G_{\sigma_d}(0) - \frac{1}{m} \sum_{i=1}^m G_{\sigma_d} \left((x_k^i - x_q^i) \frac{2}{\|\vec{x}_k\| + \|\vec{x}_q\|} \right). \quad (67)$$

Al igual que con la distancia a partir de la correlación, resulta conveniente introducir un kernel periódico en la ecuación anterior. Esto da origen a la siguiente fórmula:

$$d_{corren-kp}(\vec{x}_k, \vec{x}_q, \sigma_d, \sigma_t, p) = G_{\sigma_d}(0)G_{\sigma_t}(0) - \left[\frac{1}{m} \sum_{i=1}^m G_{\sigma_d} \left((x_k^i - x_q^i) \frac{2}{\|\vec{x}_q\| + \|\vec{x}_k\|} \right) \right] \left[G_{\sigma_t} \left(2 \sin \left(\frac{\pi}{p} (t_k - t_q) \right) \right) \right], \quad (68)$$

En esta ecuación σ_d es el tamaño del kernel de la correntropía. Además σ_t y p son respectivamente el tamaño y el periodo del kernel periódico. El valor máximo posible para la expresión de la derecha de la resta es $G_{\sigma_d}(0)G_{\sigma_t}(0)$, por lo que este valor aparece al lado izquierdo de la resta, de tal manera que la máxima similitud esté asociada a la mínima distancia.

2.2.5. Ejemplos de Gráficos de Recurrencia

Debido a que los RP corresponden a una visualización de un sistema dinámico, es de suma importancia presentar ejemplos para demostrar cómo aparecen en ellos las periodicidades de los sistemas. A continuación, se presentan 3 ejemplos de RP.

2.2.5.1. Ruido Blanco Gaussiano

Para el caso de ruido blanco gaussiano, los puntos sobre el RP aparecen en forma aleatoria con cierta probabilidad, la cual depende de la distancia utilizada y del valor del parámetro ε . Así, en general, no se observa ningún patrón en dichos puntos, salvo la LOI, que aparece en todos los RP. En la Figura 5 se puede apreciar una señal de ruido blanco gaussiano y el RP obtenido a partir de dicha señal. Se observa que, además de la LOI, los demás puntos aparecen distribuidos de manera estocástica en todo el espacio bidimensional.

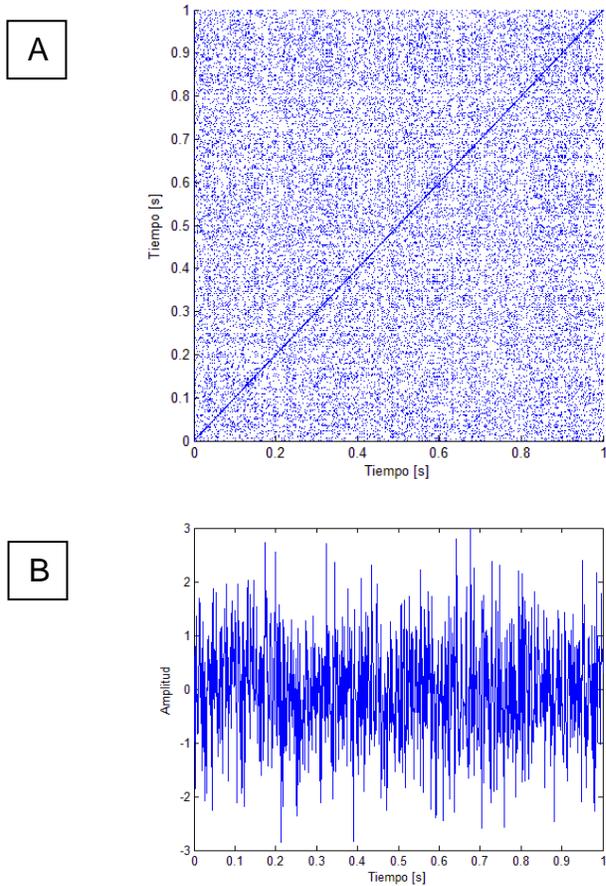
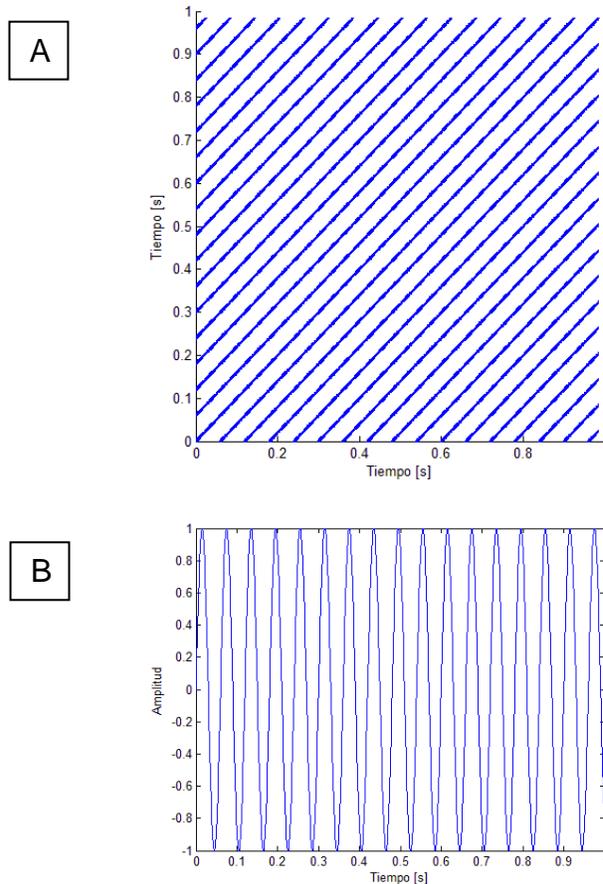


Figura 5: Ejemplo de RP para Ruido Blanco Gaussiano. A: RP obtenido para señal de ruido blanco gaussiano. B: Señal de ruido blanco gaussiano original.

2.2.5.2. Señal Sinusoidal

En la Figura 6B se puede apreciar una señal sinusoidal de periodo constante e igual a 0.6 [s]. En la Figura 6A se muestra el RP calculado a partir de dicha señal. Como se puede observar, la señal da origen a un RP que presenta largas líneas separadas por una distancia uniforme y paralelas a la LOI. Como es de esperar, la distancia entre líneas depende del periodo, y el hecho de que esta distancia sea constante se debe a que el periodo también lo es. Una línea representa la propiedad del sistema de que cada estado en un instante de tiempo es similar al estado que viene un número entero de periodos antes o después de aquel instante. Por ejemplo, las 2 líneas diagonales más cercanas a la LOI representan la similitud entre estados separados por ± 1 periodo. Así, líneas diagonales largas y continuas implican que existe una periodicidad que abarca un gran número de estados del sistema. Esta señal, de periodo constante, corresponde a un caso extremo, ya que la periodicidad de la señal es máxima, y

por lo tanto el largo de las líneas y la uniformidad con que éstas están distribuidas también los es.

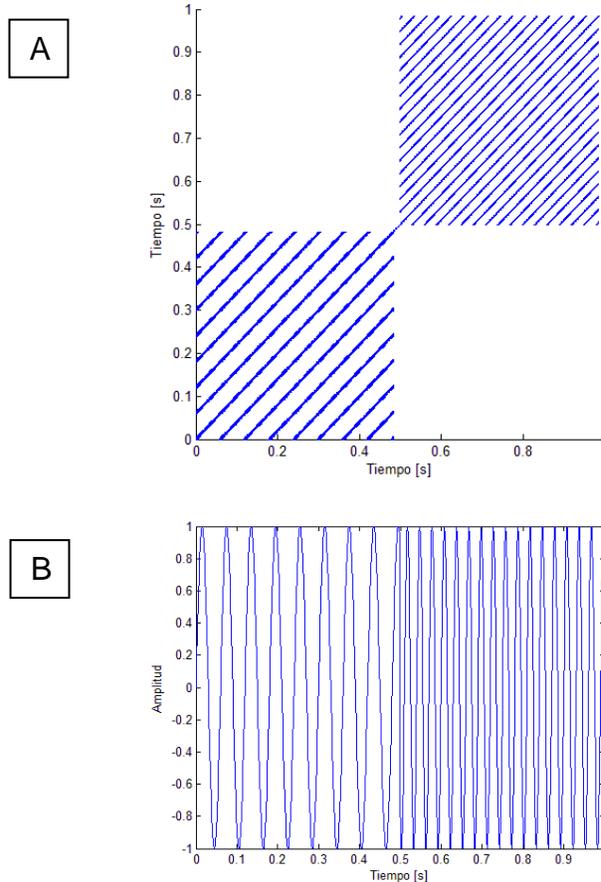


**Figura 6: Ejemplo de RP para Señal Sinusoidal de Periodo 0.6 [s].
A: RP obtenido para señal sinusoidal. B: Señal sinusoidal de periodo 0.6[s].**

2.2.5.3. Señales Sinusoidales de Distinta Frecuencia

En la Figura 7 se puede observar una señal que se compone por dos señales sinusoidales de distinta frecuencia, colocadas una después de la otra, y el RP calculado a partir de dicha señal. La primera mitad de la señal mostrada en la Figura 7B corresponde a una senoide de cierto periodo y la segunda mitad corresponde a otra senoide de un periodo distinto. Como se observa, el RP de la señal completa contiene a los dos RP de las señales individuales que componen la señal, colocados en los dos cuartos pertenecientes a la diagonal principal del RP. Notar que los otros dos cuartos, los anti diagonales, son completamente blancos, lo que indica que no existe ninguna periodicidad entre la primera parte de la señal y la segunda. Este ejemplo demuestra que señales que poseen cierto grado incompleto de periodicidad presentan líneas diagonales que no abarcan todo el RP. Así, en el RP de sistemas semi-periódicos aparecerán líneas paralelas

a la LOI, pero estas se verán interrumpidas en alguna parte. Otro aspecto interesante de este ejemplo, es que el RP permite visualizar en forma clara el periodo de las señales. Se puede apreciar en la señal original que la frecuencia de la segunda senoide es el doble de la primera, lo cual se observa claramente en el RP, donde las líneas del segundo sub-RP están separadas por la mitad de la distancia que separa las líneas del primer sub-RP.



**Figura 7: Ejemplo de RP para 2 Señales Sinusoidales Concatenadas, la Primera de Periodo 0.6 [s] y la Segunda de Periodo 0.3 [s].
A: RP obtenido para señales sinusoidales concatenadas. B: Señales sinusoidales concatenadas.**

2.2.6. Medidas sobre Gráficos de Recurrencia

Existen varias medidas sobre los RP [31], usadas para cuantificar ciertas características de estos gráficos. Es importante destacar que cuando se calculan estas medidas se deben ignorar los puntos del RP cercanos a la LOI, como se muestra en la Figura 8. Esto, debido a que los puntos de la diagonal principal aparecen en todos los RP y por ende no aportan información a las mediciones.

Por su parte, los puntos en diagonales cercanas a la LOI no corresponden realmente a puntos recurrentes, sino que a puntos cercanos entre sí temporalmente, y que por ende resultan similares debido a que el sistema no varía lo suficiente entre un instante y el siguiente. Así, los puntos de la LOI y cercanos a ella tienden a distorsionar las mediciones sobre los RP, causando que sistemas que no poseen recurrencias parezcan tenerlas. Para eliminar estos puntos, en todas las mediciones se ignoran los coeficientes i y j que satisfacen la siguiente desigualdad:

$$|i - j| < tw, \quad (69)$$

donde tw corresponde a un parámetro, conocido como el ancho de la ventana de Theiler [52].

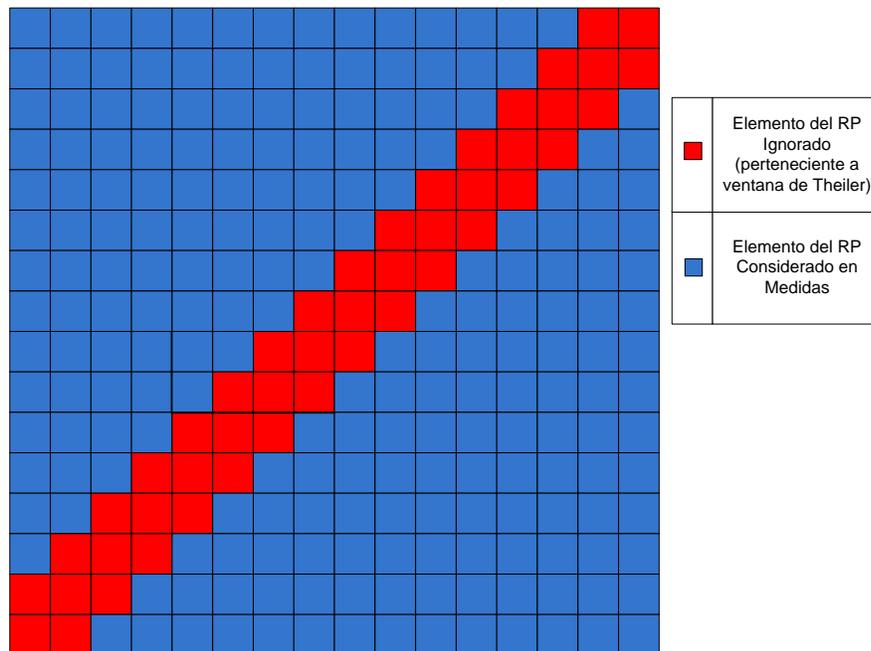


Figura 8: Ejemplo de Ventana de Theiler con $tw = 2$. El parámetro tw corresponde al ancho de la ventana. Los puntos en torno a la diagonal principal son ignorados en el cálculo de medidas.

A continuación se enumeran algunas de las principales medidas existentes, junto con una explicación de su propósito.

2.2.6.1. Tasa de Recurrencia

La tasa de recurrencia (o RR, sigla del inglés *Recurrence Rate*) mide la densidad de puntos de todo el RP. Se calcula mediante la siguiente fórmula:

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{k,q=1}^N R_{k,q}(\varepsilon). \quad (70)$$

Su propósito es medir en forma global la recurrencia total que existe en el sistema, independiente de las líneas u otros patrones que pudieran aparecer en el RP.

2.2.6.2. Histograma de Líneas Diagonales

Las líneas diagonales paralelas a la LOI indican la presencia de recurrencias en el sistema dinámico estudiado. Debido a esto, es importante estudiar la frecuencia con que ocurren los diferentes valores posibles para los largos de estas líneas, lo que se conoce como el histograma de líneas diagonales. Este histograma resume en la función $P(l)$, la cual, para un largo l cualquiera, indica el número de líneas diagonales, paralelas a la LOI, de longitud l . Los valores de esta función se calculan mediante la siguiente ecuación:

$$P(l) = \sum_{k,q=1}^N \left(1 - R_{k-1,q-1}(\varepsilon)\right) \left(1 - R_{k+l,q+l}(\varepsilon)\right) \prod_{i=0}^{l-1} R_{k+i,q+i}(\varepsilon). \quad (71)$$

A partir de este histograma es posible calcular distintas estadísticas de definen las propiedades de la distribución de largos. Además de la media y de la entropía de la distribución, las cuales se explican en las secciones 2.2.6.5 y 2.2.6.6, respectivamente, se calcula la desviación estándar, el coeficiente de asimetría de Fisher y la curtosis. El cálculo de estas variables se explica a continuación.

Para simplificar los cálculos, se toma la función $P(l)$ y se transforma en una muestra de largos llamada L . Dicha muestra contiene, para cada largo l , $P(l)$ elementos de valor l . Por ejemplo, supóngase que el histograma de largos está dado por los siguientes valores: $P(1) = 2, P(2) = 3, P(3) = 5$ y $P(l) = 0 \forall l \geq 4$. En este caso, se tiene $L = \{1, 1, 2, 2, 2, 3, 3, 3, 3, 3\}$. En otras palabras, el valor l se agrega a L un número $P(l)$ de veces. Se define como l_i al elemento i del conjunto L .

Usualmente, L no se define desde $l = 1$, si no que desde $l = L_{min}$, que corresponde a un parámetro determinado en función del problema en particular. Esto quiere decir que, al construir L , se ignoran los valores de $P(l)$ tales que $l < L_{min}$. Así, en el ejemplo anterior, si $L_{min} = 2$, se tiene $L = \{2, 2, 2, 3, 3, 3, 3, 3\}$.

La desviación estándar, L_{desv} , se define según:

$$L_{desv} = \left(\frac{1}{|L| - 1} \sum_{i=1}^{|L|} (l_i - \bar{l}) \right)^{\frac{1}{2}}, \quad (72)$$

donde $|L|$ es el tamaño del conjunto L y \bar{l} es la media de dicho conjunto.

El coeficiente de asimetría de Fisher (o *skewness* en inglés) se define mediante la siguiente ecuación:

$$L_{skew} = \frac{\frac{1}{n} \sum_{i=1}^{|L|} (l_i - \bar{l})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2} \right)^3}. \quad (73)$$

Este coeficiente mide el grado de asimetría de la distribución con respecto a la media. Es positivo si el histograma al lado derecho de la media tiene valores más dispersos que al lado izquierdo. Por el contrario, será negativo si el lado derecho de la media tiene valores más concentrados que el lado izquierdo. En la Figura 9 se muestra un ejemplo de este coeficiente. Una distribución perfectamente simétrica tiene un valor nulo del coeficiente de Fisher.

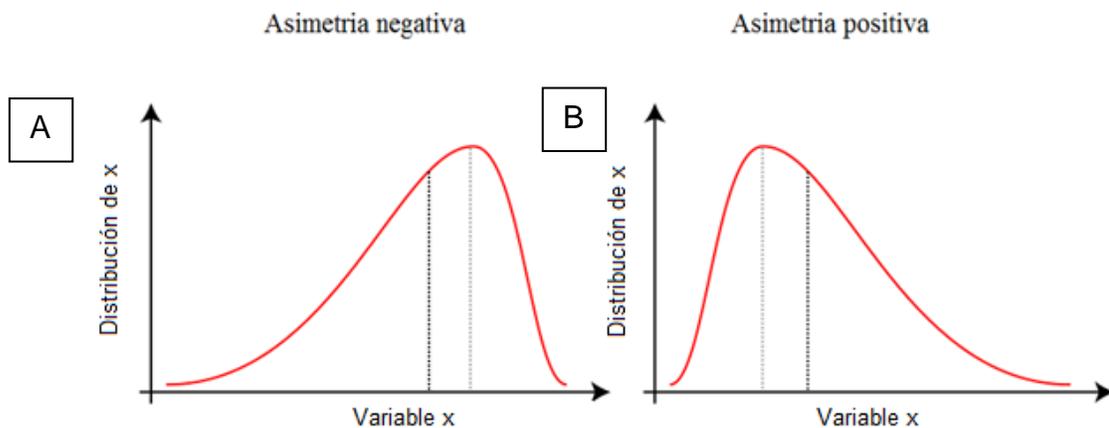


Figura 9: Comparación de Distribuciones con Asimetría positiva y Negativa. La línea punteada gris representa la moda de las distribuciones, y la línea punteada negra la media. A: Distribución con asimetría negativa. Los elementos a la izquierda de la media se encuentran más dispersos que los elementos a la derecha. B: Distribución con asimetría positiva. Los elementos a la derecha a de la media se encuentran más dispersos que los elementos a la izquierda.

La curtosis (o *kurtosis* en inglés) se define mediante la siguiente ecuación:

$$L_{kurt} = \frac{\frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^4}{\left(\frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2\right)^2} \quad (74)$$

Esta medida determina el grado de empinamiento de la distribución en torno a la media. Este concepto se puede apreciar en la Figura 10, donde se muestra una comparación de los valores de curtosis de distintas distribuciones de probabilidad comunes. Aquellas distribuciones cuyo máximo es más puntiagudo mayores presentan valores de curtosis.

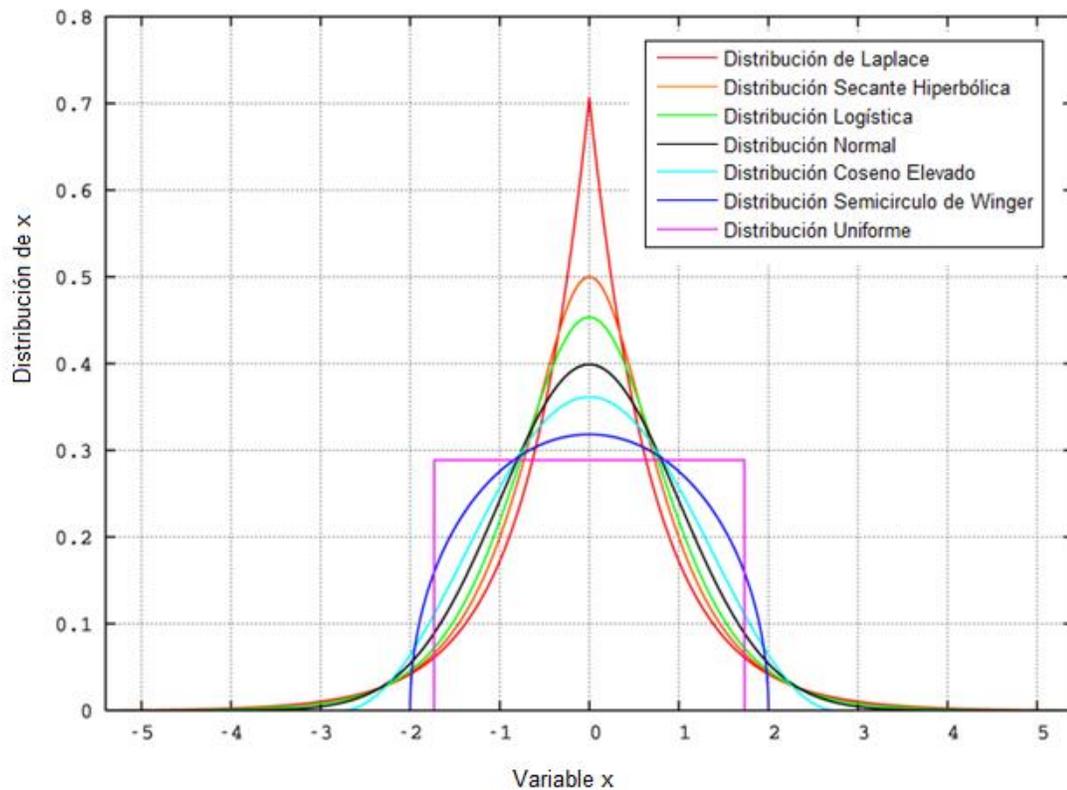


Figura 10: Comparación de Curtosis para Distintas Distribuciones de Probabilidad. Aquellas distribuciones más empinadas tienen valores mayores de curtosis. Los valores de curtosis para las distribuciones en la curva son: Distribución de Laplace: 6; Distribución Secante Hiperbólica: 5; Distribución Logística: 4,2; Distribución Normal: 3; Distribución Coseno Elevado: 2,4; Distribución Semicírculo de Winger: 2; Distribución Uniforme: 1,8.

2.2.6.3. Largo Máximo de Líneas Diagonales

Esta medida corresponde al tamaño máximo de las líneas diagonales en el RP. Se calcula mediante la siguiente fórmula:

$$L_{max} = \max(\{l : P(l) \neq 0\}). \quad (75)$$

Un valor alto de L_{max} indica que el sistema posee un alto grado de periodicidad, pues existe al menos una línea del RP que abarca varios estados, lo que quiere decir que existe al menos un periodo que domina la recurrencia de la trayectoria en varias partes. Por lo tanto, esta medida es muy apropiada para discriminar entre sistemas periódicos y aquellos que no lo son.

2.2.6.4. Determinismo o Predictibilidad del Sistema

Generalmente los RP obtenidos a partir de sistemas con comportamientos más estocásticos o caóticos muestran líneas diagonales más cortas. Una forma de cuantificar esto es medir la razón de puntos que pertenecen a líneas diagonales largas de entre todos los puntos del RP. Se entiende por líneas largas aquellas cuya longitud es mayor o igual a L_{min} , que es un parámetro que se fija en función de las características particulares del problema considerado. Esta razón se conoce como determinismo o predictibilidad, pues aquellos sistemas en donde es más alta deberían ser más periódicos, y por lo tanto más predecibles. Se calcula mediante la siguiente fórmula:

$$DET(L_{min}) = \frac{\sum_{l=L_{min}}^N l \cdot P(l)}{\sum_{l=1}^N l \cdot P(l)}. \quad (76)$$

Nótese que el denominador de esta división corresponde a la cantidad total de puntos en el RP o, equivalentemente, al RR , pues se considera que todos los puntos del gráfico pertenecen a una línea, aún cuando esta sea de largo uno.

2.2.6.5. Largo Promedio de Líneas Diagonales

El largo promedio de líneas diagonales está dado por la siguiente ecuación:

$$L_{medio} = \frac{\sum_{l=L_{min}}^N l \cdot P(l)}{\sum_{l=L_{min}}^N P(l)}. \quad (77)$$

Nótese que este promedio se calcula sobre las líneas consideradas largas, es decir, aquellas cuya longitud es mayor o igual a L_{min} , y no sobre todas las líneas del histograma definido por $P(l)$. Al igual que L_{max} , esta medida indica la periodicidad del sistema.

2.2.6.6. Entropía de Shannon

A partir del histograma de líneas diagonales, $P(l)$, es posible definir una probabilidad $p(l)$ para cada largo, dada por la siguiente ecuación:

$$\left(p(l) = \frac{P(l)}{N_l} \right), \quad (78)$$

donde

$$N_l = \sum_{l \geq L_{min}} P(l). \quad (79)$$

Esta probabilidad estará definida solo para las líneas de longitud mayor o igual al parámetro L_{min} . A partir de esta probabilidad, se puede calcular la entropía de Shannon de la distribución de líneas diagonales, mediante la siguiente fórmula:

$$E = - \sum_{l=L_{min}}^N p(l) \cdot \ln(p(l)). \quad (80)$$

La medida E está relacionada con la complejidad del sistema. Sistemas más complejos, con distintos grados de periodicidad, estarán asociados a valores altos de esta medición, mientras que sistemas poco complejos, como el ruido blanco gaussiano, por ejemplo, darán menores valores de E .

2.2.6.7. Medidas τ

Todas las medidas introducidas hasta ahora pueden ser definidas para una sola línea diagonal a una distancia τ de, y paralela a, la LOI. Dichas medidas son marcadas con un subíndice τ . A continuación mostramos las fórmulas de algunas de ellas:

$$RR_{\tau} = \frac{1}{N - \tau} \cdot \sum_{k=1}^{N-\tau} R_{k,k+\tau}, \quad (81)$$

$$DET_{\tau} = \frac{\sum_{l=L_{min}}^{N-\tau} l \cdot P_{\tau}(l)}{\sum_{l=1}^{N-\tau} l \cdot P_{\tau}(l)} \quad y \quad (82)$$

$$L_{\tau} = \frac{\sum_{l=L_{min}}^{N-\tau} l \cdot P_{\tau}(l)}{\sum_{l=L_{min}}^{N-\tau} P_{\tau}(l)}. \quad (83)$$

Estas variantes miden las mismas propiedades que sus mediciones originales, pero enfocadas en solo uno de los posibles periodos del sistema. Por ejemplo, RR_τ mide la tasa de recurrencia considerando únicamente los estados separados por el periodo τ . Así, si para un τ fijo RR_τ es alto, esto indica que muchos estados del sistema son similares a estados separados por una distancia τ , lo que indica la presencia de recurrencias de periodo τ .

2.2.6.8. Razón

La razón (o *ratio* en inglés) se define como la división entre el DET y el RR . Está basada en un estudio que mostró que dicho cociente puede servir para detectar transiciones en la dinámica del sistema [31]. Esto, debido a que durante ciertas transiciones el RR disminuía, mientras que el DET permanecía constante. Se calcula mediante la siguiente fórmula:

$$RA = N^2 \frac{\sum_{l=L_{min}}^N l \cdot P(l)}{(\sum_{l=1}^N P(l))^2}. \quad (84)$$

2.2.6.9. Histograma de Líneas Verticales

Así como las líneas diagonales del RP representan trayectorias periódicas del sistema dinámico, las líneas verticales del RP representan trayectorias estáticas, en las cuales la trayectoria se encuentra atascada y relativamente inmóvil. Resulta interesante, en algunos casos, estudiar en qué medida el sistema pasa por estos estados. En forma análoga al histograma de líneas diagonales, se define el histograma de líneas verticales. Dicho histograma corresponde a una función, $P(v)$, que para un valor de v cualquiera indica el número de líneas verticales de largo v que aparecen en el gráfico. Dicha función se calcula mediante la siguiente fórmula:

$$P(v) = \sum_{k,q=1}^N (1 - R_{k,q}(\epsilon)) (1 - R_{k,q+v}(\epsilon)) \prod_{i=0}^{v-1} R_{k,q+i}(\epsilon). \quad (85)$$

2.2.6.10. Laminaridad del Sistema

En forma análoga al determinismo, se llama laminaridad al cociente entre el número de puntos que pertenecen a líneas verticales largas y el total de puntos del RP. Una línea vertical larga se define como aquella que sea mayor o igual al parámetro V_{min} , el cual debe ser fijado en función del problema en particular. La laminaridad está dada por la siguiente fórmula:

$$LAM(V_{min}) = \frac{\sum_{v=V_{min}}^N vP(v)}{\sum_{v=1}^N v \cdot P(v)}. \quad (86)$$

Nótese que el denominador de esta razón corresponde a la cantidad total de puntos en el RP, pues todos los puntos pertenecen al menos a una línea vertical de largo uno. Esta medida muestra la presencia de estados laminares en el sistema, es decir, la presencia de estados para los que la trayectoria queda atrapada por un cierto lapso de tiempo. Así, esta medida decrecerá si el RP posee pocas estructuras verticales.

2.2.6.11. Tiempo de Captura

El largo promedio de líneas verticales, también conocido como tiempo de captura (o *trapping time* en inglés) está dado por la siguiente ecuación:

$$TT = \frac{\sum_{v=V_{min}}^N vP(v)}{\sum_{v=V_{min}}^N P(v)}. \quad (87)$$

Nótese que este promedio se calcula sobre las líneas consideradas largas, es decir, aquellas cuya longitud es mayor o igual a V_{min} , y no sobre todas las líneas del histograma definido por $P(v)$. Esta medida representa el tiempo promedio en que el sistema se encuentra atrapado en un estado constante.

2.3. Señales de Sueño

2.3.1. Generalidades

El estudio de patrones de sueño corresponde a un campo muy interesante para la ciencia moderna. La razón de dicho interés se debe a que el análisis de señales tomadas mientras una persona duerme puede ser usado, entre otras cosas, para comprender de mejor manera cómo funcionan los ciclos de sueño en los humanos, qué funciones cumplen, y también para estudiar anomalías de estos ciclos.

El sueño corresponde a un periodo de pérdida de conciencia y de una disminución de la respuesta a estímulos externos. Aunque usualmente se considera un periodo de descanso, es también una etapa de importante actividad cerebral, en donde muchas neuronas permanecen activas. Además de las neuronas responsables de funciones fisiológicas vitales, como las que controlan la

respiración, neuronas pertenecientes a otras áreas continúan trabajando, normalmente en forma muy sincrónica [48].

No se conoce completamente todas las funciones que desempeña el sueño. Claramente una de sus principales usos es proveer un descanso tanto físico como mental al cuerpo, pero diversas hipótesis postulan que ésta no sería su única función. Se cree que el sueño es también una etapa de mantenimiento de las estructuras sinápticas del cerebro, lo que favorece la memoria y los procesos de cambio en la estructura cerebral [38] [55] [58]. Así, además de ser un proceso de reposo, es una etapa de consolidación de lo aprendido durante el día, y de preparación para el día siguiente.

Dormir no es un proceso homogéneo, sino que uno descrito por diferentes estados. Dichos estados son: sueño paradójico (SP) y sueño quieto (SQ) [34] [39], los cuales presentan diferencias tanto fisiológicas como funcionales. Pueden distinguirse mediante diversas mediciones que se realizan a un paciente, siendo las más usadas para este propósito el electroencefalograma (EEG), el electrooculograma (EOG) y el electromiograma (EMG). Estas mediciones normalmente forman parte de un examen más general conocido como polisomnograma. Dicho examen mide varios indicadores del sueño de un paciente y lo registra en un polígrafo. Además de las tres mediciones ya mencionadas, este examen puede incluir: el electrocardiograma (ECG), el esfuerzo respiratorio abdominal (MV) y el movimiento del paciente (MC, sigla de Movimiento Corporal).

El sueño paradójico (SP), también conocido como sueño REM (acrónimo de *Rapid Eye Movements*, en inglés), se caracteriza por la presencia de movimientos oculares rápidos (MOR) en el EOG, por la ausencia de tono muscular y por ondas rápidas y de bajo voltaje en el EEG [39].

El sueño quieto (SQ), también conocido como sueño NREM (acrónimo de *Non Rapid Eye Movement*, en inglés) se caracteriza por ondas mayor amplitud en el EEG, la no ocurrencia de movimientos oculares rápidos, y frecuencias cardiacas y respiratorias generalmente más lentas que en SP. Antiguamente el SQ era dividido en 4 etapas, las cuales indicaban qué tan profundo era el sueño. Actualmente las etapas 3 y 4 se consideran una sola, conocida como sueño profundo (o SWS, sigla proveniente del inglés *Slow Wave Sleep*). Las dos primeras etapas de SQ se abrevian como SQ_n, donde n corresponde al número de la etapa. Así, por ejemplo, SQ₂ corresponde a la etapa 2 de SQ.

2.3.1.1. Husos Sigma

Un patrón característico de las señales del EEG, y que marca el inicio de la etapa 2 del sueño quieto, es el huso sigma (HS). La definición de HS varía según el autor, pero para los efectos de este estudio, se considera como un tren de ondas de frecuencias entre 10 y 16 [Hz], de amplitud ascendente y luego descendente, y de duración entre 0,5 y 2 [s] [39]. Normalmente, los HS aparecen

superpuestos a ondas de menor frecuencia y de mayor amplitud. La banda de frecuencias entre 10 y 16 [Hz] es conocida como banda sigma. Es importante destacar que los límites de esta banda sigma varían según los autores, y su tamaño es normalmente menor al que se utiliza en este trabajo. La razón por la que se considera este rango de frecuencias es debido a que los HS en niños son más difusos y pueden presentar frecuencias espectrales que se salen del rango normal para adultos. Dado que la base de datos que se utilizó en este estudio estaba compuesta por mediciones realizadas a niños de 10 años, resulta necesario considerar un rango espectral un poco mayor al que se usa normalmente. En la Figura 11 se muestran ejemplos de HS.

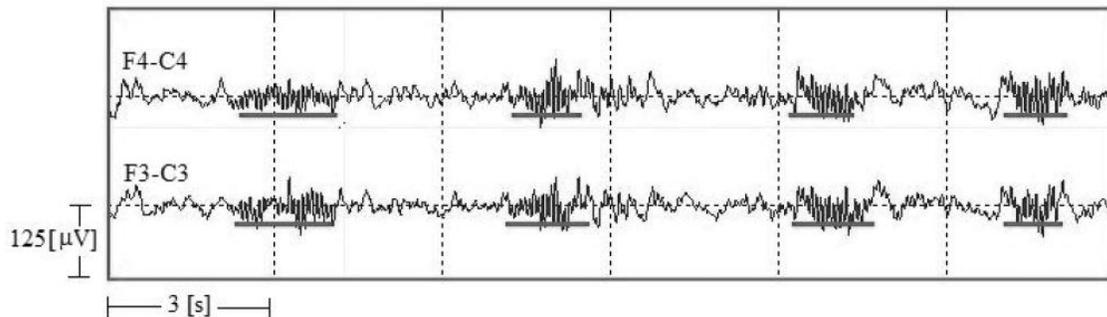


Figura 11: Registros de Sueño en donde aparecen Husos Sigma.
Las líneas horizontales marcan la detección según el experto. Las señales corresponden a las desviaciones fronto-centrales del EEG. F3-C3 es la diferencia de potencial entre los electrodos F3 y C3, y corresponde al canal izquierdo de derivación fronto-central. F4-C4 es la diferencia de potencial entre los electrodos F4 y C4, y corresponde al canal derecho de derivación fronto-central. Esta figura fue tomada de [7].

La detección y el estudio de los HS es de enorme interés actualmente. Se cree que los HS son indicadores de un desarrollo normal del cerebro y en general del sistema nervioso central. Normalmente alteraciones en estos patrones se relacionan a desviaciones en el funcionamiento y en el desarrollo neuronal [37]. Además, anomalías en la estructura de los HS han sido relacionados con diversas patologías neurológicas, tales como: retardo mental [13], sordera [21], autismo [28] e hiperkinesia [8].

Los HS son patrones característicos del sueño quieto, en particular de la etapa SQ2. Por ende, su detección es de interés para la determinación de las etapas por las que transita un individuo al dormir. Esto, a su vez, es de gran utilidad en estudios clínicos y fisiológicos. Algunas de las aplicaciones donde es interesante detectar esta etapa son: la evaluación del sistema nervioso central, la determinación de déficit nutricional y la evaluación del riesgo de muerte súbita en infantes [10].

Se cree que los HS están relacionados directamente con los procesos de memoria del cerebro [41] [50], en particular, que son los promovedores de las condiciones necesarias para la plasticidad cerebral requerida para la formación de memoria. Estudios han demostrado que la densidad de HS aumenta después de tareas de aprendizaje. Con esto, se ha respaldado el hecho que dichos patrones están críticamente involucrados en el proceso de consolidación de memoria [45] [56].

En la actualidad, el proceso de detección de HS se hace generalmente en forma visual por un experto. Esta es una tarea lenta, tediosa y propensa a errores. Dichos errores causan que existan diferencias entre distintos expertos, y también para el mismo experto, el que puede realizar detecciones distintas en momentos diferentes [10] [50]. Además, las mediciones son realizadas durante un lapso de varias horas, pues se efectúan durante todo el periodo de sueño normal del paciente. Se debe tener en cuenta, además, que en el registro de una noche aparecen aproximadamente 2000 eventos de HS [5]. Todo esto hace conveniente la creación de sistemas automáticos de detección de patrones de sueño, en particular de HS. No sólo corresponden a métodos estandarizados de detección, sino que además permiten aumentar la velocidad con que se hacen dichas detecciones.

2.3.2. Estado del Arte

Debido al gran interés que genera la detección automática de HS, existen diversos trabajos que han abordado este tema. En esta sección, se describen brevemente los avances más significativos existentes en el presente.

En [3] se muestra un sistema automático de detección de HS llamado método de ajuste individual (o IAM, sigla de *Individual Adjustment Method*, en inglés). Este método comienza filtrando pasa-banda trozos de la señal de EEG pertenecientes al estado SQ. Al resultado de este filtrado se aplica la transformada rápida de Fourier (o FFT, sigla de *Fast Fourier Transform*, en inglés) para obtener el espectro de frecuencias de zonas de la señal del EEG marcadas como SQ. Este espectro, que corresponde a una señal de amplitud en función de la frecuencia, es usado para definir las frecuencias de corte de un filtro pasa banda. Para ello, se calcula la segunda derivada del espectro y para la función resultante de este cálculo se detectan los cruces por cero. Cada par consecutivo de cruces por cero envuelve a un *peak* de la señal. De estos *peaks*, se selecciona aquel de amplitud más negativa, y los dos cruces por cero que lo envuelven se eligen como las frecuencias de corte de un filtro pasa banda. Este filtro es aplicado nuevamente a la señal de EEG. A esta señal filtrada se calcula su envolvente, y si dicha envolvente supera cierto umbral por más de 0,5 [s], se marca como un HS. Este método fue usado sobre segmentos de la señal de EEG de 12 individuos sanos. La duración total de estos registros era de aproximadamente 4 horas. Se obtuvo una TPR de 93% y una FDR de 58.4%.

En [22] se utilizan y comparan 4 métodos diferentes para detectar HS en adultos. El primer método usa un índice sigma basado en FFT. Este índice se calcula en ventanas de 1 [s] de duración, como la razón entre la amplitud espectral de la banda sigma (10,5 a 16 [Hz]) y la suma de las amplitudes espectrales de las siguientes dos bandas: 4 a 10 [Hz] y 20 a 40 [Hz]. En paralelo, se filtra la señal del EEG en la banda sigma y se determina la máxima amplitud de esta señal filtrada para cada ventana de 1 [s] de duración. Este valor es conocido como amplitud sigma. Luego, se determina un umbral óptimo de amplitud, el cual es aplicado sobre la amplitud sigma calculada en cada ventana para determinar la presencia de HS. En este proceso se considera, además, el valor del índice sigma en la ventana. Es decir, se marcan como HS aquellos que superen un umbral de amplitud sigma y un umbral de índice sigma. El segundo método solo usa el valor del índice sigma para realizar la detección. Sobre dicho valor, se aplica un umbral determinado en forma empírica para determinar la presencia de HS. El método tres utiliza un detector difuso de actividad sigma. Dicho detector utiliza tres características deducidas de la FFT: el cociente entre la amplitud espectral de la banda sigma y la de la banda theta (4 a 7 [Hz]), el cociente entre la amplitud espectral de la banda sigma y la de la banda alfa (7,5 a 10 [Hz]) y el cociente entre la amplitud espectral de la banda sigma y la de la banda gamma (20 a 40 [Hz]). El detector difuso combina estas tres características mediante 17 reglas, deduciendo un índice de presencia de HS. Para que exista un HS según este método, dicho índice debe superar cierto umbral. El cuarto método utiliza el mismo procedimiento del método 1 para calcular la amplitud sigma. Los trozos cuya amplitud sigma supera cierto umbral empírico por más de 0.5 [s] son considerados HS. Todos los métodos se probaron sobre 12 registros completos, cada uno de duración aproximada de 8 horas. Los mejores resultados, que son los que se muestran a continuación, fueron los obtenidos en SQ2. De todos los métodos, el mejor fue el método 1, el cual obtuvo una TPR de 70%, una TNR de 98,6% y una FDR de 32%. El método 4, que fue el peor de todos, obtuvo una TPR de 70%, una TNR de 97,7% y una FDR de 46%. En [23], que corresponde a un trabajo previo del mismo grupo que realizó [22], se determina un umbral óptimo para detectar HS basado en amplitud. La forma de determinar este umbral es la misma que se usa en el método 1 de [22].

En [11] se utiliza la transformada corta de Fourier (o SFT, sigla del inglés *Short-time Fourier Transform*), extracción de característica y gas neural para detectar HS en infantes. La SFT corresponde a la FFT aplicada en una ventana de la señal. En este trabajo se aplica la SFT sobre la señal del EEG, usando ventanas de Hamming de 0,5 [s], para determinar la potencia de 5 bandas: delta-rápida-theta (2 a 7 [Hz]), alfa (7 a 10 [Hz]), sigma (10 a 15 [Hz]), beta (12 a 30 [Hz]) y alta frecuencia (30 a 60 [Hz]). Estas características son usadas en un clasificador basado en redes neuronales, específicamente, basado en el algoritmo *Merge Neural Gas* (MNG). Este método fue utilizado para detectar HS en infantes. Se dividió el registro en dos conjuntos, uno de entrenamiento y uno de prueba. Se obtuvo una TPR de 62,7% y una TCC de 88,7%.

En [54], se aplica una red neuronal MLP para detectar HS sobre el EEG filtrado pasa-banda. La entrada de la NN eran valores de la señal misma, obtenidos al sub-muestrearla. La señal es filtrada en la banda sigma (10,5 a 16 [Hz]) y luego dividida en ventanas de 0,5 [s]. Cada ventana es dividida en 64 muestras, las que corresponden a la entrada de la NN. La salida de la NN era una salida binaria indicando la presencia o ausencia de HS en la ventana. La red se entrenó mediante el algoritmo de retropropagación. Se tomó el registro completo de un adulto y se dividió en tres partes iguales. Para cada uno de estos tres segmentos se seleccionó la etapa de SQ2 con mayor duración. Una de estas etapas fue utilizada como conjunto de entrenamiento, y las otras dos como conjunto de prueba. La duración total del conjunto de prueba era aproximadamente de 10 minutos. Se obtuvo una TPR de 80,2%, una TNR de 95% y una FDR de 7%.

En [15] se usa SFT para extraer características del EEG, y clasificadores basados en MLP y SVM para detectar los HS. A partir de SFT entrega el espectro de frecuencias de ventanas de la señal. Este espectro es muestreado para obtener 32 coeficientes, para frecuencias entre 2 y 64 [Hz], los cuales son usados como características para dos clasificadores distintos, uno basado en MLP y otro basado en SVM. Ambos sistemas fueron entrenados con segmentos de un registro clasificado como SQ2. Para el clasificador MLP, se utilizaron 1000 ejemplos de entrenamiento y 1142 de prueba, obteniéndose una TCC de 88,7%. Para el clasificador SVM, se utilizaron 1575 ejemplos de entrenamiento y 175 de prueba, obteniéndose una TCC de 95,4%.

En [46] y [60] se utiliza un método llamado *matching pursuit* (MP). MP es una forma de descomponer una señal utilizando una gran base de funciones. Es similar a la descomposición por serie de Fourier, donde una señal es descompuesta en una suma de funciones sinusoidales. En el caso de Fourier, las funciones sinusoidales representan un conjunto completo mínimo, es decir, es el más pequeño que permite representar todas las señales. En MP, se usa un conjunto mayor de funciones, el cual es completo pero redundante. Esto quiere decir que algunas señales de dicho conjunto pueden ser eliminadas y este seguirá siendo completo. Esto permite que la descomposición se adapte mejor a la señal por descomponer, pues al poseer una mayor base de datos de funciones, es más probable que existan funciones más similares a la señal original. En [60] se utiliza MP con funciones Gabor para descomponer la señal del EEG. Esto permite obtener una representación en el plano de Winger 3D de la función original. Este plano permite relacionar la frecuencia y el tiempo con la amplitud. Es una función de dos dimensiones, donde existe una amplitud (representada por el eje Z) asociada a cada par tiempo-frecuencia (representados en el eje X e Y respectivamente). Dentro de este gráfico, los *peaks* son conocidos como átomos. Luego de descomponer la señal y obtener su representación en el plano de Winger, se buscan átomos en la banda sigma (10 a 16 [Hz]) de este plano. Aquellos átomos que superan cierto umbral de amplitud y de duración son marcados como un HS. Se utilizaron trozos de la etapa SQ2 seleccionados de 9 polisomnograma de adultos. No se separó en conjunto de entrenamiento y prueba,

usándose los mismos registros para ambos propósitos. Se obtuvo una TPR de 90%. En [46] se utiliza MP para definir un umbral de amplitud utilizado para la detección de HS. Primero, se aplica MP con funciones Gabor para obtener una representación en plano Winger de la señal. Luego, se buscan átomos en el rango de frecuencias entre 11 y 15 [Hz], determinando la amplitud de cada uno. La detección de HS se hace mediante la definición de un umbral de amplitud, de tal manera que aquellos átomos cuya amplitud supera este umbral son considerados HS. Dicho umbral se varía, para construir una curva ROC del sistema. A partir de esta curva ROC, se determina el umbral óptimo. Se utilizaron trozos de de SQ2, SWS y SP, seleccionados de 9 polisomnograma de adultos. El registro duraba aproximadamente 6 horas, y contenía 725 HS. No se separó en conjunto de entrenamiento y prueba, usándose los mismos registros para ambos propósitos. Para todas las etapas juntas, se obtuvo una TPR de 88,4%, una TNR de 81,2% y una FDR de 86,9%.

En [19] se utilizan criterios de amplitud y duración para detectar HS en registros de siesta de lactantes. Este algoritmo se compone de 3 módulos. El módulo 1 realiza un filtrado en el rango de frecuencias de 10 a 15 [Hz]. Sobre esta señal filtrada aplica criterios difusos de amplitud, obteniendo la amplitud difusa, y luego aplica umbrales sobre esta amplitud de detección para generar un grupo de candidatos a HS. El módulo 2 funciona en paralelo al módulo 1. Este módulo toma la señal del EEG original (sin filtrar) y utiliza criterios de amplitud, frecuencia y simetría para generar un segundo grupo de candidatos. El módulo 3 combina ambos grupos, considerando como HS solo aquellos eventos que sean candidatos para ambos módulos. Se utilizó como conjunto de entrenamiento un registro de siesta con 358 HS, y como prueba dos registros completos de siesta. La duración total de estos dos registros era de 6 horas, y tenían 803 HS marcados por expertos. Se obtuvo una TPR de 87,7% con una FDR de 8,1%. Algo similar se hace en [44], en donde además se agregan técnicas de detección de artefactos. Este método consta de dos etapas aplicadas en paralelo. En la etapa 1, se filtra pasa-banda la señal del EEG y se aplican criterios de duración y amplitud para definir umbrales de detección. En la segunda etapa se busca detectar y eliminar el ruido generado por la contracción muscular y por la actividad alfa (7 a 10 [Hz]). Se consideran HS aquellos eventos que satisfacen los criterios de la etapa 1 y que no son eliminados por presencia de artefactos en la etapa 2. Se utilizaron 10 polisomnogramas de adultos sanos, sin separar conjunto de prueba y entrenamiento. Se obtuvo un TPR de 89,7% y un FDR de 6,5%.

2.3.3. Sistema de Detección de HS Implementado en la Universidad de Chile

En [6] y [7] se explica el sistema original sobre el cual este trabajo está basado. Dicho sistema proviene de un trabajo conjunto entre miembros del Instituto de Nutrición y Tecnología de los Alimentos (INTA) y del Laboratorio de Ingeniería Biomédica del Departamento de Ingeniería Eléctrica de la Universidad de Chile.

Uno de los grandes aportes de este estudio fue la base de datos recopilada, conformada por varias mediciones de polisomnograma, durante toda la noche, para 56 niños sanos de 10 años. Estas mediciones corresponden a una cantidad sustancial de datos que posteriormente fueron usados no solo para el sistema implementado en [6], sino que para otros estudios. En particular, de esta base de datos provienen los registros usados en este trabajo.

El sistema, que se describe en detalle a continuación, se basa en lógica difusa, descomposición modal empírica (EMD) y la transformada de Hilbert-Huang (HHT) [20], para la detección de HS.

2.3.3.1. Módulo 1

El módulo 1 tiene como propósito seleccionar ciertas zonas del EEG en donde podría haber efectivamente HS, es decir, sirve para detectar las etapas SQ2 y SWS [6]. Esto se hace debido a que el módulo 2 es una etapa bastante costosa computacionalmente, por lo que no conviene aplicarla a todas las señales.

Básicamente, este módulo divide el todo el registro fronto-central del EEG (ya sea F3-C3 o F4-C4) en ventanas de Hamming móviles de 2,56 [s] (512 muestras), con un traslape de 1,28 [s] (256 muestras). Para cada ventana, se calcula la potencia de las bandas delta (0,5 a 3 [Hz]), sigma (10 a 16 [Hz]), alta frecuencia (30 a 60 [Hz]) y banda fisiológica (0,5 a 60 [Hz]). Esta última potencia es usada para normalizar las 3 primeras. Así, para cada ventana de 2,56 [s] se calculan tres valores que corresponden a las potencias de la banda sigma, delta y alta frecuencia, cada una dividida por la potencia de la banda fisiológica. Luego, se toman ventanas de 30 [s], y a cada una de ellas se le calcula el promedio de las potencias normalizadas calculadas en todas las ventanas de 2,56 [s] contenidas dentro de los 30 [s] de la ventana. Con esto, se determinan tres potencias promedio: P_D para la banda delta, P_S para la banda sigma y P_{HF} para alta frecuencia, para cada ventana n de 30 [s]. Luego, se clasifica cada una de estas ventanas n según las siguientes reglas:

- Si $P_d(n) \geq \alpha_1$ y $P_{HF} \leq \alpha_3$ entonces la ventana n pertenece a una zona de interés.
- Si $P_s(n) \geq \alpha_2$ y $P_{HF} \leq \alpha_3$ entonces la ventana n pertenece a una zona de interés.
- Si no, la ventana n no pertenece a una zona de interés y no será usada por el módulo 2.

Los parámetros α_1 , α_2 y α_3 se ajustan en forma empírica durante la construcción del método, utilizando el conjunto de entrenamiento. Las reglas antes mostradas se basan en el hecho que los trozos de sueño clasificados como SQ2

normalmente tienen un alto P_S y los clasificados como SWS tienen un alto P_D , mientras que en vigilia existe un alto P_{HF} .

2.3.3.2. Módulo 2

Este módulo solo es aplicado a las zonas detectadas como zonas de interés por el módulo 1 [6]. A partir de dichas zonas, el módulo 2 genera los primeros candidatos a HS, conocidos como SSC1 (acrónimo de *Sleep Spindle Candidate 1*, en inglés).

La primera etapa de este módulo consiste en aplicar EMD y HHT [20] sobre los trozos de EEG definidos por el módulo 1. EMD (explicado en detalle en la sección 2.3.4) es utilizado para separar la señal en varias componentes. De ellas, solo se utilizan las primeras 3, pues empíricamente se comprobó que son suficientes para capturar los HS presentes en la señal. Dichas componentes son conocidas como IMF_1 , IMF_2 e IMF_3 . Debido a la naturaleza de EMD, es posible que la potencia en la banda sigma a veces sea mayor en una de las componentes y otras veces sea mayor en otra componente. Por ello, es necesario usar FFT para calcular la potencia de la banda sigma de las 3 componentes, en cada ventana, y elegir aquella componente donde dicha potencia se maximiza. Esta componente se denomina la componente primaria o IMF_P . Se calcula mediante la siguiente fórmula:

$$IMF_P = \max\{P_{SS IMF_1}, P_{SS IMF_2}, P_{SS IMF_3}\}, \quad (88)$$

donde $P_{SS IMF_i}$ es la potencia en la banda sigma de IMF_i , con $i \in \{1, 2, 3\}$. Las dos componentes de EMD que no corresponden a IMF_P , son llamadas componentes secundarias (IMF_{S1} e IMF_{S2}) y son descartadas si la potencia de su banda sigma no supera cierto umbral. Este criterio se muestra a en la siguiente ecuación:

$$IMF_{Sj} = \begin{cases} IMF_i & \text{si } P_{SS IMF_i} \geq \beta_1 \\ 0 & \text{en caso contrario} \end{cases}, \quad (89)$$

donde $j \in \{1, 2\}$, $i = 1, 2 \text{ ó } 3$ e $i \neq P$. El parámetro β_1 es determinado empíricamente.

Posteriormente, en este módulo se aplica HHT a IMF_P y a los IMF_{Sj} que hayan satisfecho el criterio de la ecuación (89). HHT es una descomposición que permite obtener simultáneamente la amplitud instantánea de una señal, $a(t)$, y su frecuencia instantánea, $\omega(t)$. Así, tras la aplicación de HHT en este módulo se obtiene al menos una función de amplitud instantánea: $a_{IMF(i)}(t)$ y su correspondiente frecuencia instantánea: $\omega_{IMF(i)}(t)$.

En [6] se considera que la amplitud de un HS debe estar entre 15 y 120 [μV], y que su frecuencia debe estar dentro de la banda sigma, vale decir, entre 10 y 16 [Hz]. A partir de estos criterios, se crearon conjuntos difusos para clasificar $a_{IMF(i)}(t)$ y $\omega_{IMF(i)}(t)$, y así determinar la amplitud instantánea difusa: $f a_{IMF(i)}(t)$, y la frecuencia instantánea difusa, $f \omega_{IMF(i)}(t)$. El conjunto difuso de amplitud está dado por:

$$\{0/10, 1/15, 1/120, 0/150\}. \quad (90)$$

El conjunto difuso de frecuencia está dado por:

$$\{0/9,5, 1/10, 1/16, 0/16,5\}. \quad (91)$$

Los conjuntos difusos están definidos como trapezoides cuyos 4 puntos característicos se muestran entre las llaves. Para cada punto, el numerador muestra el valor de la función de pertenencia asociado al valor de la variable mostrada en el denominador. Por ejemplo, para la frecuencia, cualquier valor entre 10 y 16 [Hz] tiene pertenencia 1, y cualquier valor que esté fuera del rango entre 9,5 y 16,5 [Hz] tiene pertenencia 0.

Luego de calcular, a partir de los conjuntos difusos, los valores difusos tanto para frecuencia como amplitud, dichos valores se multiplican para obtener el producto instantáneo $PI_{IMF(i)}(t)$. Este producto asegura que se cumpla tanto con el criterio de frecuencia como con el criterio de amplitud. Se calcula según la siguiente fórmula:

$$PI_{IMF(i)}(t) = f a_{IMF(i)}(t) \cdot f \omega_{IMF(i)}(t). \quad (92)$$

Luego, se toma el máximo para este producto entre la componente principal de EMD y las dos secundarias, según:

$$PI_{max}(t) = \max \{PI_{IMFP}(t), PI_{IMFS1}(t), PI_{IMFS2}(t)\}. \quad (93)$$

Si $PI_{max}(t)$ supera cierto umbral, se considera que en ese instante de tiempo existe un candidato a HS. Se determinó empíricamente que el valor de dicho umbral era 0,5.

Los candidatos a HS a la salida del módulo 2, conocidos como SSC1, se determinan a partir de un criterio de duración. Básicamente, aquellos elementos marcados como HS que se encuentran a menos de 0,2 [s] de distancia se unen, expandiendo su duración desde el comienzo del primero hasta el final del último. Dichos elementos corresponden a la salida del módulo 2, o sea, a los SSC1.

2.3.3.3. Módulo 3

El módulo 3 se creó para reducir el número de FP del sistema, minimizando la pérdida de TP [7]. Para ello, los segmentos de EEG que se marcaron como SSC1 por el módulo anterior se filtran pasa banda con frecuencias de corte 0,5 y 25 [Hz], para contribuir a la eliminación de artefactos y ruido. Luego, en cada SSC1 se buscan 3 *peaks* consecutivos, ya sea del tipo mínimo-máximo-mínimo o máximo-mínimo-máximo. Dichos *peaks* se encuentran observando 3 cambios consecutivos en la pendiente de la función. La pendiente de la función es determinada ajustando una recta mediante regresión lineal. Para esta regresión se utilizan 5 muestras consecutivas. Cada uno de estos conjuntos de 3 *peaks* consecutivos es denominado componente. Para cada componente, sus *peaks* son llamados: izquierdo (I), centro (C), y derecho (D). Cada uno de estos *peaks* posee una amplitud y un tiempo, lo que da origen a los siguientes tres pares: (A_I, t_I) , (A_C, t_C) y (A_D, t_D) . En la Figura 12 se muestra un ejemplo de una componente.

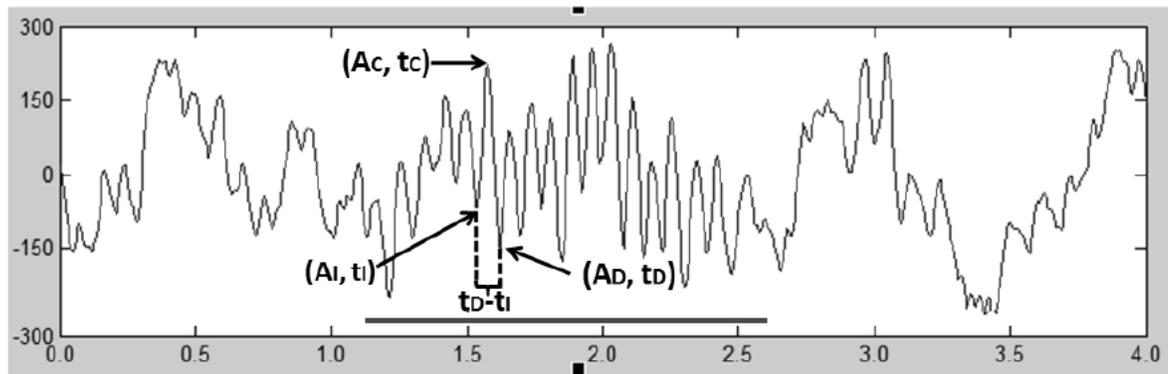


Figura 12: Ejemplo de Componente de un SSC1.

El HS está marcado con una línea horizontal. Cada uno de los 3 *peaks* de la componente aparece identificado con un par amplitud-tiempo. Esta figura fue tomada de [7].

Para cada componente se calculan tres características: frecuencia, amplitud y simetría. La frecuencia de cada componente del SSC1 se determina mediante la siguiente ecuación:

$$frec_{(k)} = \frac{1}{t_D(k) - t_I(k)}, \quad (94)$$

donde k toma valores entre 1 y $n_c(j)$, siendo $n_c(j)$ el número total de componentes del SSC1 número j . Así, k sirve como un identificador de las diferentes componentes de cada SSC1.

Una vez calculadas estas frecuencias, para cada SSC1 (identificado por la letra j), se calcula la desviación estándar y el valor promedio de las frecuencias de sus componentes. Dicho cálculo se realiza mediante las siguientes fórmulas:

$$\overline{frec_{SSc1(j)}} = \frac{1}{n_c(j)} \sum_{k=1}^{n_c(j)} frec_{(k)} \quad (95)$$

y

$$\sigma_{frec_{SSc1(j)}} = \sqrt{\frac{1}{n_c - 1} \sum_{k=1}^{n_c(j)} (frec_{(k)} - \overline{frec_{SSc1(j)}})^2}. \quad (96)$$

Se define la amplitud de cada componente mediante la siguiente fórmula:

$$Ampl_{(k)} = \min\{|A_{C(k)} - A_{I(k)}|, |A_{C(k)} - A_{D(k)}|\}, \quad (97)$$

donde k toma valores entre 1 y $n_c(j)$, siendo $n_c(j)$ el número total de componentes del SSC1 número j . Luego, el SSC1 asociado al número j se caracteriza mediante el promedio y la desviación estándar de los valores de la amplitud de sus componentes, dados por las siguientes ecuaciones:

$$\overline{Ampl_{SSc1(j)}} = \frac{1}{n_c(j)} \sum_{k=1}^{n_c(j)} Ampl_{(k)} \quad (98)$$

y

$$\sigma_{Ampl_{SSc1(j)}} = \sqrt{\frac{1}{n_c - 1} \sum_{k=1}^{n_c(j)} (Ampl_{(k)} - \overline{Ampl_{SSc1(j)}})^2}. \quad (99)$$

Finalmente, se realiza un cálculo de simetría entre las dos semiondas de cada componente, debido a que el ruido puede formar patrones que se encuentren dentro del rango sigma, y que podrían introducir falsos HS en la detección. Una forma de eliminar parcialmente esta confusión es exigir un grado de simetría en la componente. Se define una medida de simetría según la siguiente ecuación:

$$Sim_{(k)} = \frac{|\min(|t_{I(k)} - t_{C(k)}|, |t_{D(k)} - t_{C(k)}|)|}{|\max(|t_{I(k)} - t_{C(k)}|, |t_{D(k)} - t_{C(k)}|)|}, \quad (100)$$

donde k toma valores entre 1 y $n_c(j)$, siendo $n_c(j)$ el número total de componentes del SSC1 número j . Luego, el SSC1 asociado al número j se

caracteriza mediante el promedio y la desviación estándar de los valores de la simetría de sus componentes, dados por las siguientes ecuaciones:

$$\overline{Sim}_{SSC1(j)} = \frac{1}{n_c(j)} \sum_{k=1}^{n_c(j)} Sim_{(k)} \quad (101)$$

y

$$\sigma_{SimSSC1(j)} = \sqrt{\frac{1}{n_c - 1} \sum_{k=1}^{n_c(j)} (Sim_{(k)} - \overline{Sim}_{SSC1(j)})^2}. \quad (102)$$

Con esto, para cada SSC1 se tienen 6 características, las cuales corresponden a los promedios y las desviaciones de la simetría, la amplitud y la frecuencia de sus componentes. Sobre estos valores se aplican umbrales empíricos para filtrar los FP. Estos umbrales son ajustados utilizando un conjunto de entrenamiento. La razón por la que se aplican dichos umbrales es que se observó que para los TP los promedios de las mediciones de sus componentes están dentro de un cierto rango y las desviaciones estándar de dichas mediciones son mayoritariamente menores a las de los FP. Con estos umbrales se descartan algunos candidatos, filtrándose los SSC1 para así obtener los SSC2, que corresponden a la salida del módulo 3.

2.3.3.4. Módulo 4

Es posible que los HS aparezcan superpuestos a ondas más lentas y con mayores amplitudes. Esto causa que a veces estos patrones se encuentren distorsionados. Por esto, cuando el experto marca visualmente los HS, no se basa únicamente en la frecuencia para realizar su detección, sino que también en la duración y en la morfología de los eventos. Esto permite descartar elementos cuya frecuencia se encuentra en la banda sigma pero que no corresponden realmente a HS.

El módulo 4 intenta replicar el criterio del experto, filtrando y combinando los SSC2 en base a su duración [6]. En primer lugar, los SSC2 que están divididos por una distancia menor a 0,25 [s] se combinan, expandiendo su duración desde el comienzo del primero hasta el final del último. La amplitud de los nuevos segmentos se determina como el promedio ponderado de los segmentos unidos. Luego, segmentos de duración menor a 0,1 [s] son eliminados, y posteriormente segmentos a una distancia mayor a 0,5 [s] son combinados usando el mismo procedimiento ya mencionado.

En el módulo 1 se calculan, para la ventana n , tres promedios de potencias: $P_S(n)$, $P_D(n)$ y $P_{HF}(n)$. Los dos primeros son utilizados nuevamente para definir las siguientes reglas sobre cada ventana n :

- Si $P_D(n) \geq \delta_1$, se descartan segmentos SSC2 de largo menor a 0,4 [s] en la ventana n .
- En caso contrario, si $P_S(n) \geq \delta_2$, se descartan segmentos SSC2 de largo menor a 0,3 [s] en la ventana n .
- En caso contrario, se descartan segmentos SSC2 de largo menor a 0,5 [s] en la ventana n .

Los parámetros δ_1 y δ_2 se determinan empíricamente a partir del conjunto de entrenamiento.

2.3.4. Descomposición Modal Empírica

2.3.4.1. Definición

El método de Descomposición Modal Empírica (o EMD, sigla de *Empirical Mode Decomposition*) [20], es una forma de descomponer una señal en varias componentes, las cuales sumadas permiten recuperar la señal original. Es usada para separar partes más importantes de la señal de otras que pueden no ser útiles para el problema considerado. Para ello, este método extrae iterativamente las partes de mayor variación de la señal original, las cuales son consideradas las más importantes y corresponden a las primeras componentes de la descomposición. Debido a que este algoritmo se basa en la señal misma para determinar cuáles son las componentes que debe extraer, se adapta, de alguna manera, al problema particular en donde se aplique. Esto lo hace distinto de otros métodos de descomposición, como la transformada de Fourier, para los que la forma de separar la señal se determina mediante límites fijos. El objetivo principal de EMD es descomponer una señal no estacionaria en una suma de componentes de media cero. A continuación se explica en qué consiste el algoritmo de descomposición.

2.3.4.2. Conceptos Básicos

La idea de EMD es considerar las oscilaciones de las señales a nivel local. Si se considera una señal $x(t)$, que posee dos extremos del mismo tipo (mínimos por ejemplo) en los instantes t_1 y t_2 , se puede definir la parte de alta frecuencia, $d_1(t)|_{t_1}^{t_2}$, entre t_1 y t_2 . Esta componente también es conocida como el detalle de la función. Complementariamente, existe una parte de baja frecuencia $m_1(t)|_{t_1}^{t_2}$, también definida entre t_1 y t_2 , tal que $x(t)|_{t_1}^{t_2} = m_1(t)|_{t_1}^{t_2} + d_1(t)|_{t_1}^{t_2}$. Dicha componente es conocida como la tendencia local de la función (o *local trend*, en

inglés) [40]. Si esto se hace de manera similar para todos los extremos (máximos y mínimos) de la señal, y si se supone que existe una forma adecuada de combinar las distintas descomposiciones locales, se obtiene una descomposición en dos componentes de $x(t)$ para todos los instantes de tiempo en que esta señal está definida. Estas dos componentes son $d_1(t)$ y $m_1(t)$, ahora definidas para todo t , y corresponden a la unión de todas componentes $d_1(t)|_{t_1}^{t_2}$ y $m_1(t)|_{t_1}^{t_2}$, respectivamente. Posteriormente, se puede volver a aplicar este procedimiento para la parte de baja frecuencia, $m_1(t)$, obteniéndose dos descomposiciones de esta señal, una de alta frecuencia, $d_2(t)$, y una de baja frecuencia, $m_2(t)$. Este procedimiento se puede continuar, obteniéndose iterativa y empíricamente las componentes de mayor variación de las señales, que corresponderán a las componentes $d_i(t)$, resultantes de cada iteración i del algoritmo. A continuación, mostramos los pasos del algoritmo EMD [40]:

Dada una señal $x(t)$, el algoritmo es el siguiente:

1. Asignar a la variable $m_1(t)$ el valor de la señal $x(t)$.
2. Asignar a la variable i el valor 1.
3. Identificar los extremos (máximos y mínimos) de $m_i(t)$. Si no es posible realizar esto, al no existir cambios en la derivada de $m_i(t)$, terminar algoritmo.
4. Realizar una interpolación entre todos los mínimos de $m_i(t)$.
5. Asignar el resultado de la interpolación en paso 4 a la variable $e_{min}(t)$.
6. Realizar una interpolación entre todos los máximos de $m_i(t)$.
7. Asignar el resultado de la interpolación en paso 6 a la variable $e_{max}(t)$.
8. Asignar a la variable $d_i(t)$ la señal $m_i(t) - \frac{(e_{min}(t)+e_{max}(t))}{2}$.
9. Asignar a la variable m_{i+1} el valor $m_i(t) - d_i(t)$.
10. Asignar a la variable i el valor $i + 1$
11. Volver al paso 3.

En la Figura 13 se muestran los pasos de este algoritmo sobre una señal calculada como la suma de dos sinusoides de frecuencias distintas. La primera componente extraída de EMD ($d_1(t)$ en el algoritmo) se muestra en la Figura 13E. Esta componente corresponde a la senoide de mayor frecuencia usada para construir la función original. La señal resultante de restar $d_1(t)$ de la señal original se muestra en la Figura 13F. Corresponde a la senoide de menor frecuencia usada para construir la función original.

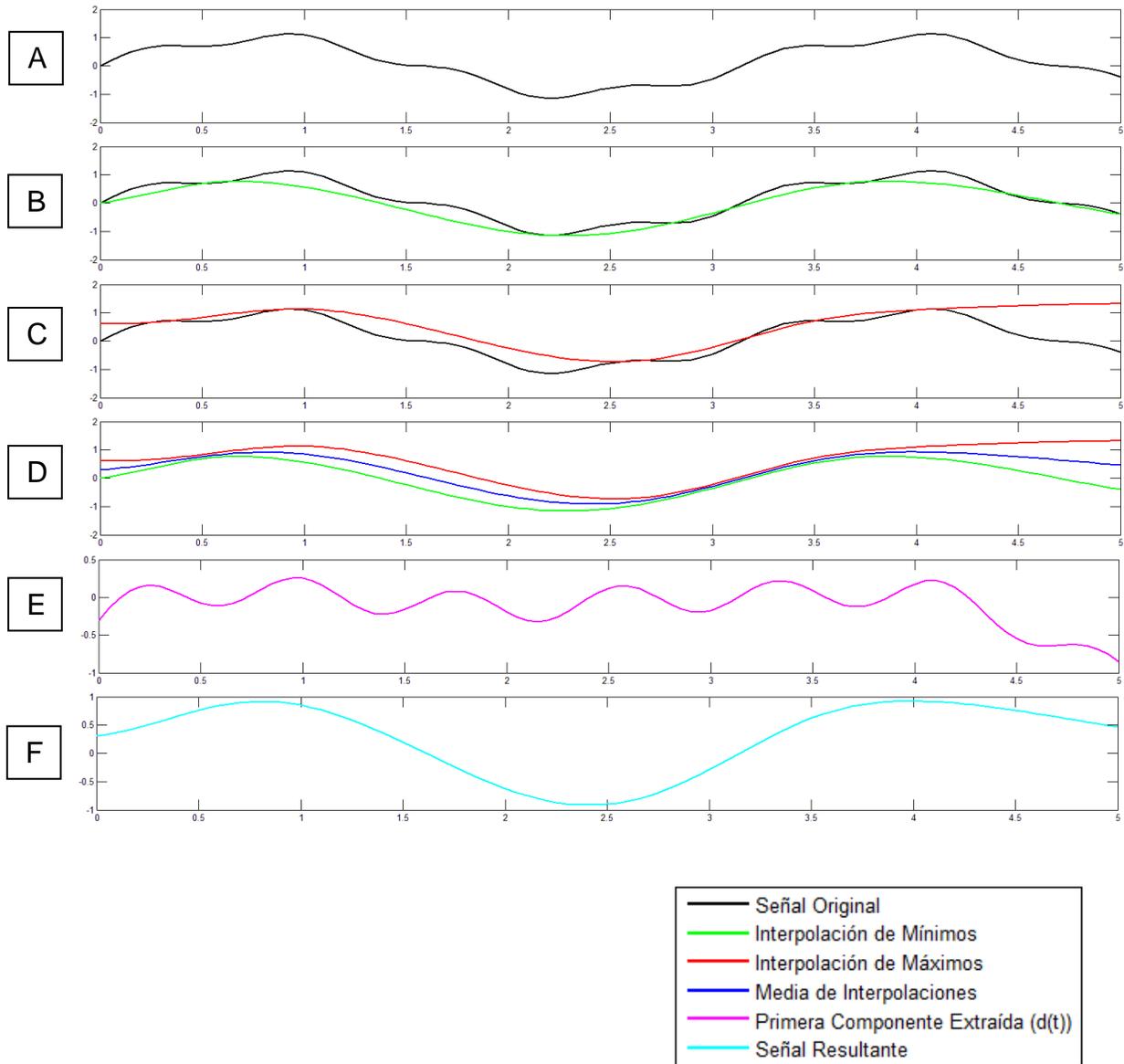


Figura 13: Ejemplo de Algoritmo EMD Sobre Suma de Sinusoides de Distinta Frecuencia.

A: Señal original de ejemplo. Está formada por una suma de dos sinusoides. Corresponde a la señal $m_1(t)$. **B:** Interpolación de los mínimos de la señal original. Corresponde a la señal $e_{min}(t)$ **C:** Interpolación de los máximos de la señal original. Corresponde a la señal $e_{max}(t)$ **D:** Media entre interpolación de mínimos e interpolación de máximos. **E:** Primera componente extraída de EMD, obtenida de restar la media entre las interpolaciones de la señal original. Corresponde a la señal $d_1(t)$. Para este ejemplo, se asemeja a la senoide de mayor frecuencia usada para formar la señal original. **F:** Resultante de de restar $d_1(t)$ a la señal original. Corresponde a la señal $m_2(t)$. Para este ejemplo, se asemeja a la senoide de menor frecuencia usada para formar la señal original.

Por construcción, el número de extremos (es decir, máximos o mínimos locales) de la función decrecen de una iteración a la próxima. Esto hace que luego de ciertas iteraciones, $m_i(t)$ no tenga extremos. Así, se garantiza que el algoritmo termina en un número N finito de iteraciones. La descomposición encontrada por el algoritmo está dada por las N funciones $d_i(t)$, calculadas cada una en una iteración i , y por la función de $m_{N+1}(t)$, calculada en la iteración final.

Es importante, para el correcto funcionamiento del algoritmo, que cada $d_i(t)$ calculado tenga media muy cercana a cero. En la Figura 13E se observa un ejemplo en donde esto no sucede. Debido a ello, esta primera componente de EMD no corresponde exactamente a la señal sinusoidal de mayor frecuencia utilizada para formar la señal original y la descomposición no es perfecta. En la práctica, para lograr que cada $d_i(t)$ tenga efectivamente media cercana a cero, el algoritmo original debe refinarse, agregándole un proceso de colado, o *sifting* en inglés, realizado sobre $d_i(t)$. Para ello, se define una sub-rutina llamada colar, dada por lo siguientes pasos:

Función colar($d_i(t)$):

1. Identificar los extremos (máximos y mínimos) de $d_i(t)$.
2. Realizar una interpolación entre todos los mínimos de $d_i(t)$.
3. Asignar el resultado de la interpolación en paso 2 a la variable $e_{min}^d(t)$.
4. Realizar una interpolación entre todos los máximos de $d_i(t)$.
5. Asignar el resultado de la interpolación en paso 4 a la variable $e_{max}^d(t)$.
6. Asignar a la variable $d_i(t)$ la señal $d_i(t) - \frac{(e_{min}^d(t) + e_{max}^d(t))}{2}$.
7. Si la media de $d_i(t)$ es cercana a cero, terminar subrutina, en caso contrario volver al paso 1.

Esta subrutina es agregada al algoritmo original, el cual queda de la siguiente forma:

1. Asignar a la variable $m_1(t)$ el valor de la señal $x(t)$.
2. Asignar a la variable i el valor 1.
3. Identificar los extremos (máximos y mínimos) de $m_i(t)$. Si no es posible realizar esto, terminar algoritmo.
4. Realizar una interpolación entre todos los mínimos de $m_i(t)$.
5. Asignar el resultado de la interpolación en paso 4 a la variable $e_{min}(t)$.
6. Realizar una interpolación entre todos los máximos de $m_i(t)$.
7. Asignar el resultado de la interpolación en paso 6 a la variable $e_{max}(t)$.
8. Asignar a la variable $d_i(t)$ la señal $m_i(t) - \frac{(e_{min}(t) + e_{max}(t))}{2}$.

9. Colar $d_i(t)$ (Ver subrutina colar).
10. Asignar a la variable m_{i+1} el valor $m_i(t) - d_i(t)$.
11. Asignar a la variable i el valor $i + 1$
12. Volver al paso 3.

2.3.4.3. Ejemplo

En la Figura 14 se muestra un ejemplo de EMD que muestra la utilidad de tiene este método. Se puede apreciar una señal original que corresponde a la suma de tres señales: una sinusoidal de periodo T sumada a dos señales triangulares, una de periodo mayor y otra de periodo menor a T. La descomposición EMD en tres componentes de la señal original permite recuperar perfectamente las tres señales que conformaban dicha señal. Cualquier descomposición basada las componentes armónicas de las señales obtendría una descomposición menos acertada. Esto, debido a que una descomposición basada únicamente en frecuencias no es capaz de discriminar completamente las señales, pues hay ciertas frecuencias presentes en dos o más de ellas.

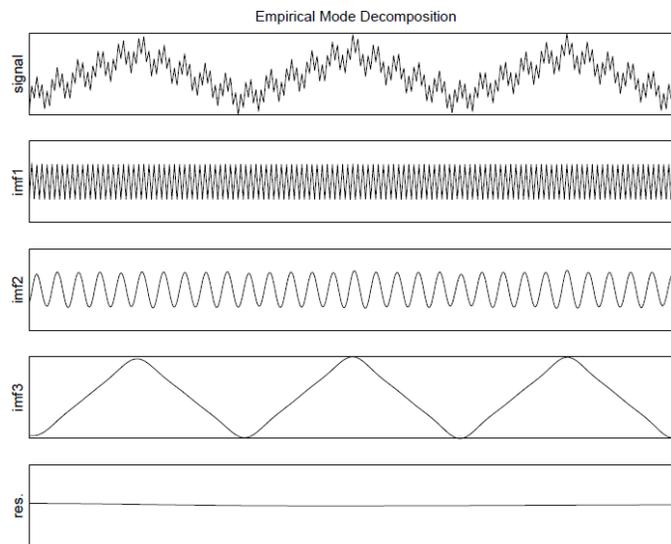


Figura 14: EMD de una Señal de Tres Componentes.

En el gráfico de más arriba se muestra una señal artificial construida con 3 señales: Una sinusoidal sumada a dos triangulares de mayor y menor frecuencia que la sinusoidal. Los siguientes 3 gráficos muestran la descomposición EMD, cuyas componentes corresponden a las señales originales. Esta figura fue tomada de [40].

2.4. Otros Métodos Usados en el Trabajo

A continuación, se muestran otros métodos utilizados durante la realización de este trabajo.

2.4.1. Información Mutua

La información mutua (IM) es un valor que mide la dependencia mutua de dos variables aleatorias X e Y . Corresponde a una medida proveniente del área de teoría de la información [16]. La información mutua se define por la siguiente ecuación [16]:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right), \quad (103)$$

donde X e Y corresponden a variables aleatorias, $p(x,y)$ es la distribución conjunta de X e Y , y $p(x)$ y $p(y)$ son las distribuciones marginales de X y de Y respectivamente.

En forma intuitiva, la información mutua indica cuánta información sobre una variable se tiene al conocer la otra variable. Así, por ejemplo, si X e Y son independientes, conocer una no entrega información sobre la otra, por lo que la IM entre ambas es igual a cero. Cualquier valor mayor a cero implica que conocer una variable entrega algo de información sobre la otra.

En este trabajo se utiliza información mutua para determinar qué características son relevantes para el sistema de clasificación y cuáles no lo son. Específicamente, se calcula la información mutua entre distintas medidas y el estándar dorado de clasificación (para el conjunto de entrenamiento). Con esto, se observa el aporte de cada entrada del clasificador, determinándose la relevancia de cada una de ellas. Esto permite tener un método rápido para comparar distintas características sin la necesidad de entrenar un clasificador completo.

Debido a que en la práctica es imposible conocer las probabilidades conjuntas y marginales de las variables aleatorias, estas deben ser estimadas. En particular, en este trabajo estas probabilidades se estimaban en el conjunto de entrenamiento, considerando las características utilizadas por sistema y el estándar dorado como un muestreo de diferentes variables aleatorias.

2.4.2. Filtros de Frecuencia

En forma general, un filtro es un sistema capaz de eliminar ciertas componentes de una señal. En particular, un filtro de frecuencia descarta frecuencias de las señales, con distintos propósitos. Uno de los propósitos más

usados es eliminar el ruido, pero existen otros, tales como, eliminar una cierta banda de frecuencias, o, por el contrario, resaltar únicamente una banda. En este estudio, se utilizan filtros de frecuencia para este último propósito. Existen diferentes tipos de filtros de frecuencia, con diferentes características y usos. Por razones que se entregan más adelante, en este trabajo se optó por utilizar filtros de Butterworth [4].

Por definición, todos los filtros, por muy complejos que sean, poseen dos tipos de zonas o bandas. Las primeras son conocidas como las bandas de paso (*passband* en inglés). En estas zonas la ganancia del filtro es alta, de tal forma que las frecuencias en dichas bandas son dejadas pasar por el filtro. Las otras zonas son conocidas como bandas de rechazo (*stopband* en inglés). En estas zonas la ganancia del filtro es baja, lo que implica que las frecuencias en estas zonas no atraviesan el filtro. En función a esto, se definen 4 tipos generales de filtros: el filtro pasa bajos, cuya banda de paso son todas las frecuencias menores a cierto valor, conocido como frecuencia de corte, y cuya banda de rechazo son todas aquellas frecuencias mayores a la frecuencia de corte; el filtro pasa altos, cuya banda de paso son todas las frecuencias mayores a la frecuencia de corte y cuya banda de rechazo son todas aquellas frecuencias menores a la frecuencia de corte; el filtro pasa banda, que posee dos frecuencias de corte, tal que la banda de paso está compuesta por los valores entre dichas frecuencias, y las bandas de rechazo se componen de todas aquellas frecuencias fuera de este rango; y el filtro rechaza banda, que posee dos frecuencias de corte, tal que la banda de rechazo está compuesta por los valores entre dichas frecuencias, y las bandas de paso se componen de todas aquellas frecuencias fuera de este rango. Normalmente existirá una pequeña zona en torno a las frecuencias de corte llamada banda de corte, en donde las frecuencias no pertenecen ni a la banda de paso ni a la banda de rechazo.

Dos características importantes de los filtros son el *ripple* y la atenuación. El *ripple* corresponde a la amplitud (en decibels) de las oscilaciones en la ganancia del filtro. Se puede calcular tanto en las bandas de paso como en las bandas de rechazo, aunque normalmente está asociado a las bandas de paso. La atenuación corresponde a la diferencia (en decibels) entre la ganancia en las bandas de paso y la ganancia en las bandas de rechazo. En la Figura 15 se muestran estas características.

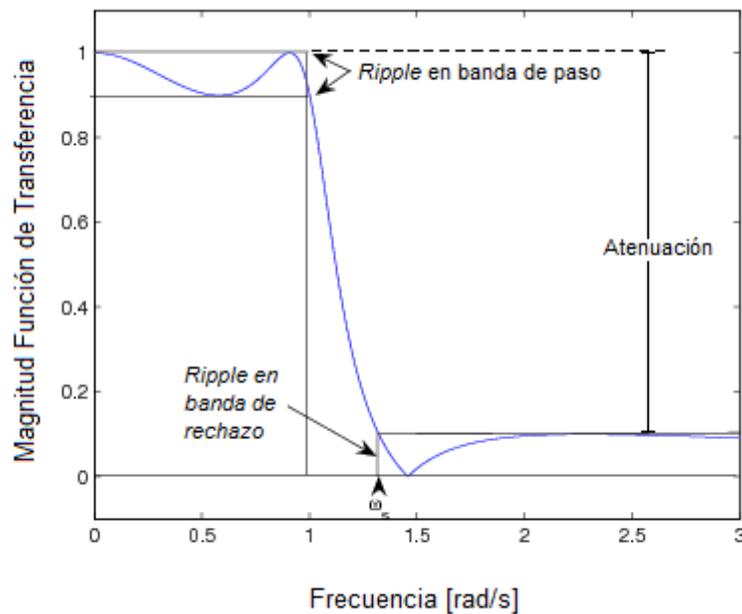


Figura 15: Ripple y Atenuación en un Filtro Pasa Bajos.

Ripple se define como la magnitud de las oscilaciones en la banda de paso o de rechazo. **Atenuación** se define como la diferencia entre las ganancias de la banda de paso y de rechazo.

Los filtros de Butterworth [4] son una familia de filtros ampliamente utilizados en la práctica, debido a que obtienen buenos desempeños. La función de transferencia de un filtro pasa bajos de Butterworth está dada por la siguiente ecuación:

$$|H(j\omega)|^2 = \frac{G_0^2}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}, \quad (104)$$

donde ω_c corresponde la frecuencia de corte del filtro, n corresponde al orden del filtro, y G_0 corresponde a la ganancia para $\omega = 0$. Un filtro pasa bajos se define como uno que solo deja pasar frecuencias inferiores a la frecuencia de corte.

El filtro pasa bajos mostrado en la ecuación (104) se puede modificar para construir un filtro pasa altos, y varios de estos filtros se pueden poner en serie para formar filtros pasa banda o rechaza banda, según sea necesario.

Los filtros de Butterworth poseen varias características que los colocan dentro de los mejores filtros de frecuencia que existen. Su principal ventaja, si se comparan con otros filtros como los de Chebyshev, es que su ganancia en la

banda paso es relativamente constante, es decir, tienen poco *ripple* en dicha banda. Esto también es cierto para la banda de rechazo. Esta característica los hace apropiados para casos en que se quiere que el filtro no distorsione la relación entre las frecuencias que pasan a través de él. Su desventaja, comparados con otros filtros, es que la banda de corte es más difusa, es decir, la ganancia que existe en la banda de paso se transforma más lentamente en la ganancia en la banda de rechazo. Esto hace que su frecuencia de corte no sea tan absoluta. En la Figura 16 se muestra una comparación entre un filtro de Butterworth y otros tipos de filtros.

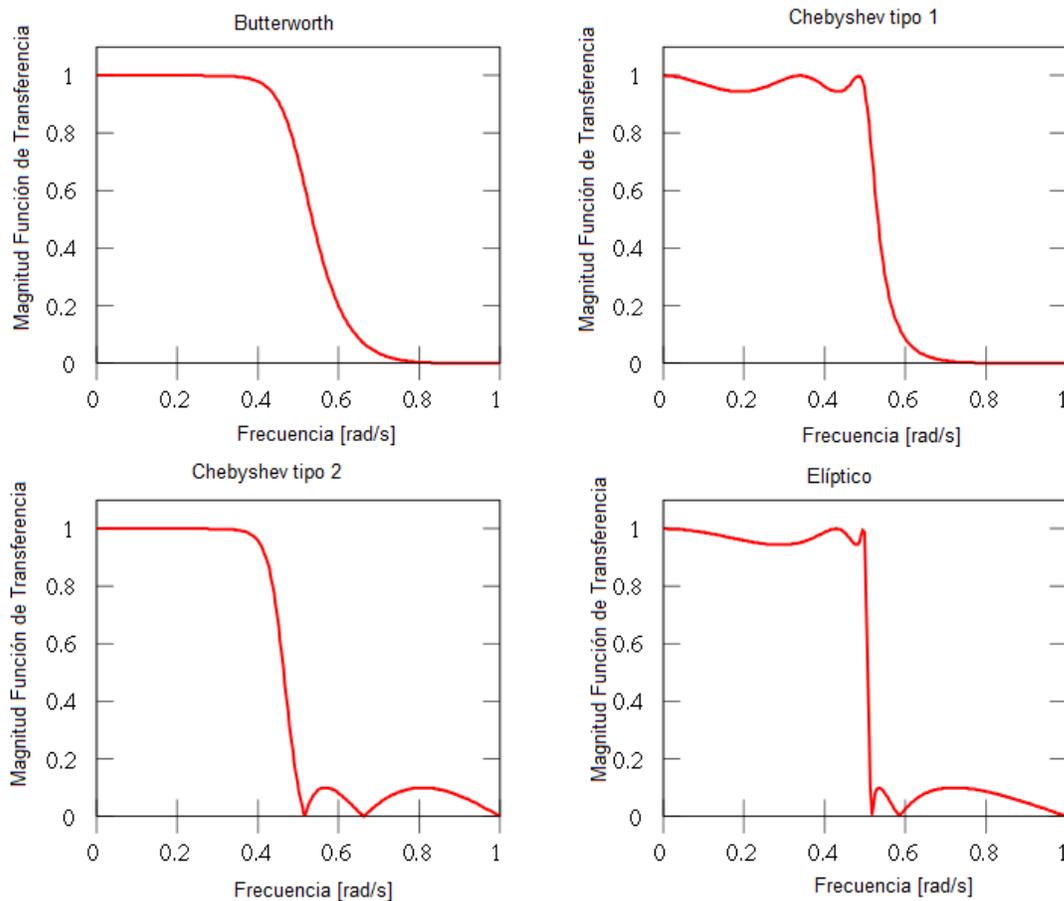


Figura 16: Función de Transferencia de 4 Filtros de Frecuencia de Distintos Tipos. El filtro de Butterworth es el que tiene menor *ripple*, pero su ganancia en la banda de paso se transforma en la ganancia en la banda de rechazo en forma más lenta que para otros filtros.

En este trabajo se utilizan filtros de Butterworth para filtrar las señales del EEG en bandas de frecuencia que resultan interesantes para la detección de HS. Se escogió este tipo de filtro sobre otros debido a que no distorsionan la componente

de la señal en la banda que se quiere dejar pasar. Esto permite que el filtro no introduzca artefactos a la señal, los cuales pueden aparecer como falsas detecciones. Además, permite que los verdaderos HS no se distorsionen, de tal forma que puedan ser detectados correctamente. Debido a que los límites de las bandas de frecuencia de estas señales fisiológicas, tales como la banda sigma, no son absolutos, si no que son más bien difusos, no es grave que la frecuencia de corte del filtro no esté bien definida, como sucede con los filtros de Butterworth.

3. Descripción del Sistema Propuesto

3.1. Introducción

El sistema implementado está diseñado para seleccionar candidatos a HS tal como lo hacen los módulos 1 y 2 del sistema explicado en la sección 2.3.3. En términos generales, la idea es utilizar los mismos registros el sistema en [7], es decir las señales F3-C3 o F4-C4 del EEG, y a partir de ellos deducir características con las cuales construir un clasificador capaz de detectar si el trozo de tiempo del cual se dedujeron dichas características corresponde o no a un HS.

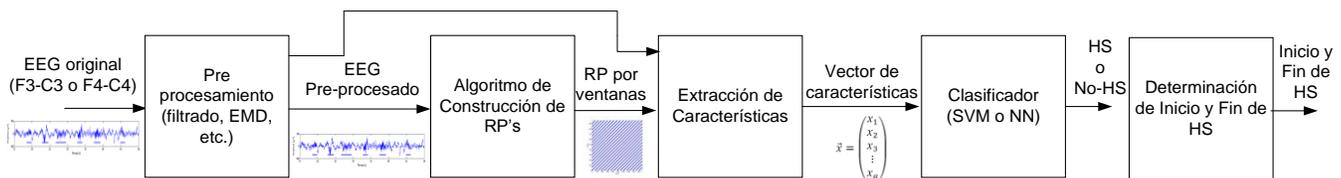


Figura 17: Diagrama del Funcionamiento General del Sistema Implementado.

La señal original (proveniente del EEG) es pre-procesada y luego se construye un RP a partir de trozos de ella. Para cada RP se calculan características, las cuales se hacen pasar por un clasificador, el cual determina si dicho trozo corresponde o no a un HS. A partir de la salida del clasificador se determina donde comienzan y terminan los HS, de forma similar a como lo haría el experto.

La Figura 17, ilustra un diagrama de bloques del sistema. Éste comienza con un pre-procesamiento del EEG. Este proceso se fue variando durante este trabajo, con el propósito de ver cuál era el método que más convenía. Se probaron distintas técnicas, tales como EMD y filtrado pasa banda de la señal, o la combinación de ambas técnicas. Luego, se dividen los registros en ventanas de cierto largo. El tamaño y el traslape de dichas ventanas son parámetros del sistema, y son discutidos en mayor detalle más adelante. En cada ventana se calcula uno o varios RP, cada uno a partir de una medida de distancia y de parámetros τ , m y ε . El valor de estos parámetros, cuántos grupos de ellos son usados y cuántas distancias se utilizan para construir los RP, también es discutido más adelante. Una vez calculados los RP, se extraen características a partir de ellos, tales como las mediciones explicadas en la sección 2.2.6. Además, se agregan otras características extraídas directamente de trozos de la señal pre-procesada. Con esto se forma un vector de características, el cual es alimentado a un clasificador binario, que detecta si cada trozo de registro corresponde o no a un HS. Este clasificador es una SVM o una NN. El enfoque de este trabajo son los RP, y no el clasificador en sí. Dicho clasificador es entrenado previamente usando parte de la base de datos para tal propósito. Finalmente, a partir de la salida del

clasificador se determina dónde comienzan y dónde terminan los HS detectados, lo que corresponde a la salida final del método.

3.2. Algoritmo Propuesto

En la Figura 18 se muestra la versión final del algoritmo implementado. Las etapas de este algoritmo se detallan en las siguientes secciones.

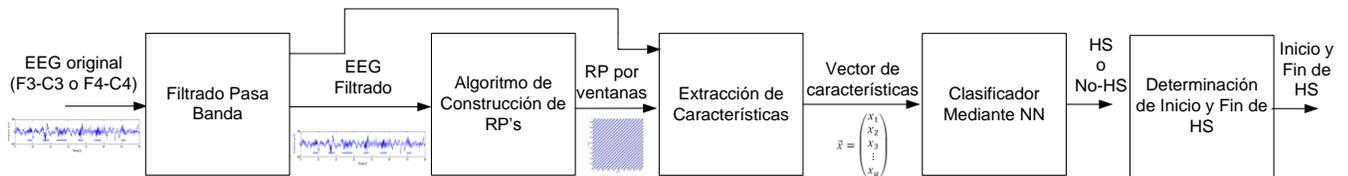


Figura 18: Diagrama del Funcionamiento del Sistema Propuesto.

La señal original (proveniente del EEG) es filtrada y luego se construye un RP a partir de trozos de ella. Para cada RP se calculan características, las cuales, junto con otras características deducidas del EEG filtrado, se hacen pasar por un clasificador, el cual determina si cada trozo corresponde o no a un HS. A partir de la salida del clasificador se determina donde comienzan y terminan los HS, de forma similar a como lo haría el experto.

3.2.1. Pre-procesamiento de la Señal

Debido a la presencia de ruido y artefactos en la señal original del EEG siempre es recomendable que ésta sea procesada antes de ingresar al sistema. Esto no solo permite un sistema más robusto, sino que además tiende a homogenizar las entradas al clasificador, haciéndolo más efectivo.

La señal del EEG es filtrada utilizando un filtro pasa banda. Las frecuencias de corte de dicho filtro, f_{cl} (frecuencia de corte inferior) y f_{ch} (frecuencia de corte superior), son parámetros del sistema, y se discuten en detalle en la sección 3.3. Dichos parámetros se escogen de manera que la banda sigma quede incluida en la banda definida por estas frecuencias. Esto quiere decir que se deben cumplir las siguientes condiciones:

$$f_{cl} \leq 10 \text{ [Hz]}, \quad (105)$$

$$f_{ch} \geq 16 \text{ [Hz]}. \quad (106)$$

Los límites mostrados en las ecuaciones (105) y (106) están basados en la banda sigma considerada en este trabajo, cuya frecuencia mínima es de 10 [Hz] y cuya frecuencia máxima es de 16 [Hz].

El objetivo del filtrado es resaltar los HS y hacerlos más homogéneos al momento de clasificarlos. Esto resulta principalmente importante a la hora de construir los RP, debido a que las estructuras que aparecen en ellos son muy sensibles al ruido. Por otra parte, un gran inconveniente de realizar este filtrado es que en el EEG casi siempre existen componentes de frecuencias en la banda sigma, incluso en aquellos trozos que no tienen presencia de HS. Debido a esto, muchos objetos que no son HS, parecen husos en la señal filtrada. Los verdaderos HS deben cumplir, además de criterios de frecuencia, con criterios de amplitud, morfología y duración. Por ello, hay que considerar en el sistema características que midan estas variables, para así diferenciar entre verdaderos HS y zonas con frecuencias en la banda sigma que no corresponden a HS. El filtro es un pasa banda de tipo Butterworth con un *ripple* de 3 [dB] y una atenuación de 10 [dB].

En la Figura 19 se ilustra un ejemplo del filtrado de un trozo de un registro del EEG. Como se observa, las zonas alta amplitud en la banda sigma tienden a estar asociadas a los HS marcados por el experto en la Figura 19A, aunque aparecen *peaks* en la señal filtrada (Figura 19B) que no corresponden a HS.

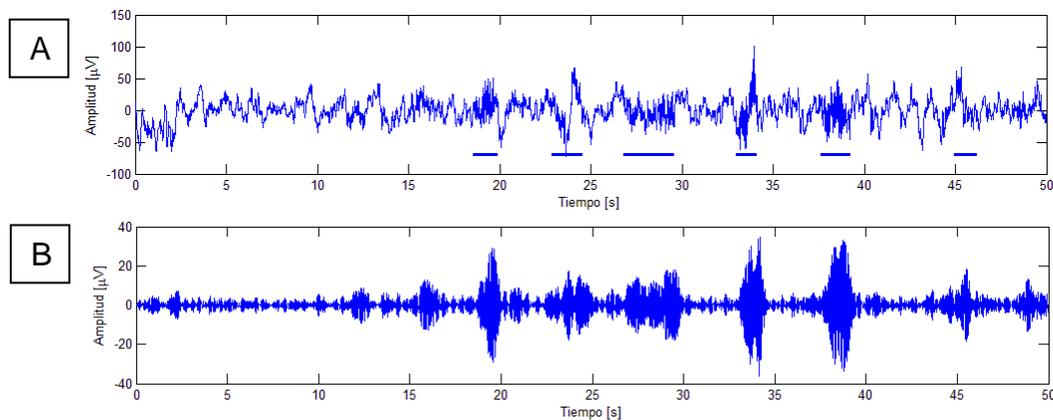


Figura 19: Ejemplo de Filtrado en Banda de Frecuencia de Señal Canal 3 de EEG para Etapa SQ2.

A: Señal original con HS marcados por experto. B: Señal filtrada entre f_{cl} y f_{ch} [Hz].

3.2.2. Construcción de RP's y Extracción de Características

3.2.2.1. Construcción de RP's

Se considera el ancho de los RP (conocido como a) como un parámetro del sistema. Al igual que otros parámetros del sistema, el valor de a es fijado mediante pruebas sobre el conjunto de entrenamiento.

Para construir los RP se toman ventanas de la señal filtrada cuyo tamaño está dado por la siguiente ecuación:

$$T = a + (m - 1) \cdot \tau, \quad (107)$$

donde m y τ son los parámetros de imbibición del espacio de estado explicados en 2.2.3. El valor para T mostrado en la ecuación permite determinar a vectores de tamaño m . A partir de estos vectores, se construye el RP de tamaño a . Nótese que si toma una ventana de menor tamaño, es imposible reconstruir los a vectores, pues para poder construir el vector número a , según el método de retraso de tiempo, se requieren $(m - 1)$ puntos subsecuentes a la muestra de índice a , separados entre sí por una distancia τ . Por ello, se requiere como mínimo una ventana de tamaño T . Estas ventanas son obtenidas para toda la señal filtrada, con un traslape tr entre ventanas, parámetro que también fue fijado mediante pruebas sobre el conjunto de entrenamiento.

A partir de cada ventana se extraen RPs, y a partir de dichos RPs se computan características numéricas. Además, se extraen otras características de la señal filtrada en esta ventana, no relacionadas con RPs. Todas estas características se combinan en un vector y se asocian al instante de tiempo correspondiente a la mitad de la ventana de la cual fueron deducidas.

Para calcular los RP se usan dos distancias alternativas, una es la correlación usando un kernel periódico ($d_{correl-kp}$) y la otra es la auto-correntropía normalizada ($d_{corren-n}$). Estas distancias se definen en las ecuaciones (64) y (67), respectivamente. Se escogieron estas distancias tras pruebas preliminares sobre el conjunto de entrenamiento. En la sección 4.3.2.1 se discuten dichas pruebas en mayor detalle. Los parámetros de esta distancias, vale decir, el periodo p y el sigma temporal σ_t de $d_{correl-kp}$, y el sigma espacial σ_d de $d_{corren-n}$, también fueron escogidos en pruebas preliminares (ver sección 4.3.3).

En la ecuación (55) se requiere de un parámetro ε para construir cada RP, que determina cuándo dos vectores son similares o no. Debido a que se construyen dos RP en este método, uno para cada distancia, se definen dos instancias de este parámetro distintas: ε_l para $d_{correl-kp}$ y ε_n para $d_{corren-n}$. Estos parámetros fueron escogidos mediante pruebas en el conjunto de entrenamiento, explicadas con mayor detalle en la sección 4.4.

3.2.2.2. Características a partir de RP's

A partir de los RP se obtienen mediciones numéricas que son utilizadas por el clasificador para determinar la presencia de HS. Estas mediciones son algunas de las explicadas en la sección 2.2.6. En pruebas preliminares se determinó que los mejores resultados se obtienen al usar las tres siguientes medidas: L_{max} , L_{medio} y L_{desv} . Las demás medidas, aunque estaban correlacionadas con las

marcas de los HS, no permitían obtener un mejor clasificador. Los detalles de esta prueba se encuentran en la sección 4.3.2.2. El sistema propuesto calcula estas tres mediciones a partir de los RPs de cada ventana, y las asigna al tiempo correspondiente a la mitad de cada ventana. En la Figura 20 se muestra como ejemplo una comparación entre los valores de dichas medidas y las marcas de HS realizadas por el experto, para un registro de la etapa SQ2. Se observa que existe una fuerte correlación entre las mediciones y la presencia de HS.

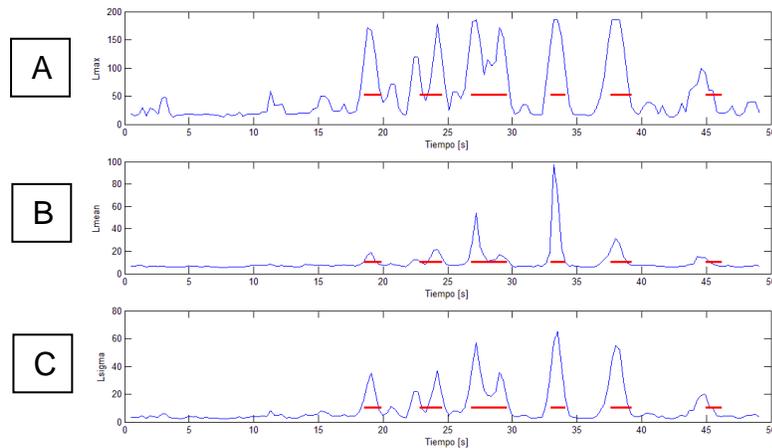


Figura 20: Contrastación de Medidas de RP con Marcas del Experto.
En todos los gráficos la marca del experto está representada como líneas horizontales rojas. A: Gráfico de L_{max} versus tiempo. B: Gráfico de L_{medio} versus tiempo. C: Gráfico de L_{desv} versus tiempo.

El cálculo de estas mediciones requiere de dos parámetros: el tamaño de la ventana de Theiler (t_w) y el largo mínimo de líneas diagonales considerado en el histograma o (L_{min}). En pruebas preliminares se observó que dichos parámetros tienen poca influencia en los resultados, por lo que no se ajustan exhaustivamente. En la sección 4.3.3.3 se detallan estas pruebas. Los valores finales de estos parámetros se especifican en la sección 3.3.

3.2.2.1. Otras características

Además de las características de RP ya explicadas, se utilizan otras características en el clasificador. Las características provenientes de los RPs están asociadas principalmente a la periodicidad del HS. Por ello, se agregaron características asociadas a la amplitud y a la duración de los eventos.

Amplitud de Envolvente

Esta característica corresponde a la envolvente de las componentes de la señal cuya frecuencia está en torno a la banda sigma. Se calcula como la interpolación lineal de los *peaks* del valor absoluto de la señal del EEG filtrada. En la Figura 21 se puede apreciar el proceso para obtener dicha señal. Primero, la señal original del EEG es filtrada pasa banda con frecuencias de corte f_{cl} y f_{ch} . Luego, se calcula el valor absoluto de la señal resultante, y sobre el resultado de esta operación se interpolan los máximos, obteniéndose la función envolvente, llamada *env*. Nótese que los *peaks* de la envolvente tienden a estar asociados a los HS marcados por el experto. Luego de calcularla, esta función se submuestra en los mismos instantes de tiempo a los que se asocian las características obtenidas de los RP (es decir, la mitad de la ventana usada para calcular el RP). De esta forma, se agrega una nueva característica al vector asociado a dichos instantes de tiempo.

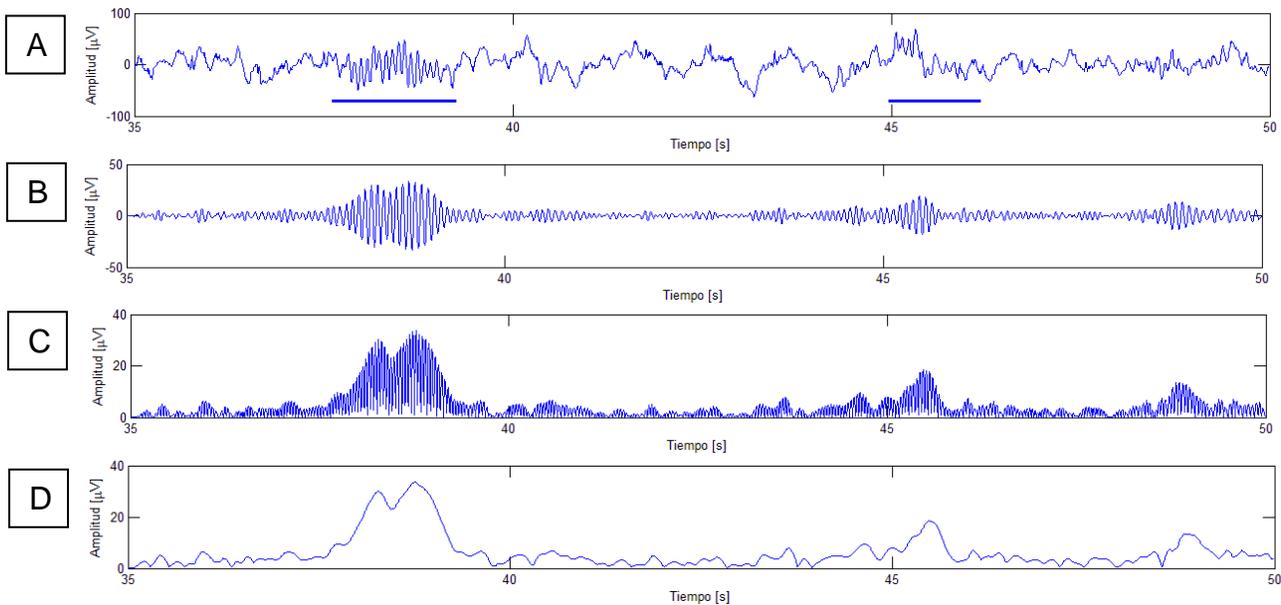


Figura 21: Proceso de Cálculo de Envolvente de Señal Filtrada.

A: Canal 3 de EEG para registro SQ2, con HS marcados por experto. **B:** Señal de A filtrada entre f_{cl} y f_{ch} . **C:** Valor absoluto de la señal en C. **D:** Envolvente de señal de C, calculada como la interpolación lineal de los *peaks* de C.

Potencias de Banda Sigma

Existe una fuerte presencia de frecuencias de la banda sigma en los HS. Debido a ello, es importante utilizar como característica del sistema clasificador la

potencia espectral, en la banda entre 10 y 16 [Hz]. A partir de dicha banda se deducen dos características, las que se explican a continuación.

La primera característica corresponde a la potencia espectral de la banda sigma para la señal sin filtrar, es decir, para la señal original obtenida del EEG. Para calcular dicha potencia, la señal original del EEG se divide en los mismos trozos de tamaño T utilizados para calcular los RP. Sobre estas ventanas se calcula la FFT (considerando una ventana rectangular). Posteriormente, se suma el cuadrado de los coeficientes de Fourier asociados a frecuencias dentro del intervalo definido por 10 y 16 [Hz]. Con esto, se obtiene la potencia para la banda sigma, P_σ . Análogamente, se calcula la potencia espectral de la banda fisiológica, es decir, entre 0,5 y 60 [Hz]. Esta potencia, llamada P_{BF} , es usada para normalizar la primera potencia calculada. Así, se obtiene la frecuencia en la banda sigma normalizada, según la siguiente ecuación:

$$P_\sigma^n = \frac{P_\sigma}{P_{BF}} \quad (108)$$

La segunda característica es la potencia en la banda sigma de la señal del EEG filtrada pasa banda entre las frecuencias f_{cl} y f_{ch} . Al igual que en el caso anterior, se calcula la FFT (con ventana rectangular) en cada ventana en las que fue dividida la señal, y posteriormente se suma el cuadrado de los coeficientes de Fourier asociados a las frecuencias dentro de la banda sigma. Con esto, se obtiene la potencia espectral de la señal filtrada, para la banda sigma, P_σ^f . Debido a que la señal ya estaba filtrada, no se normaliza por las potencias en la banda fisiológica.

Cada uno de estas 2 características se agrega al vector asociado con el instante de tiempo asignado a la mitad de la ventana en donde se realiza la respectiva medición.

Ancho de *Peak* de Señal de Envolvente

Además de la amplitud de la envolvente de la señal filtrada, es importante considerar el ancho de los *peaks* de esta envolvente. Esto, debido a que generalmente los HS corresponden a máximos de esta envolvente que se extienden por un cierto período de tiempo, mientras que el ruido y los artefactos están asociados a máximos menos prolongados.

Para estimar el ancho de los *peaks* se define primero un umbral, llamado u_{peaks} . Aquellas partes de la señal que superen dicho umbral se consideraran pertenecientes a un *peak*. El valor de u_{peaks} se calcula en el conjunto de entrenamiento. En primer lugar, se toman los H segmentos, de este conjunto, marcados como HS por el experto. Para el segmento i cualquiera, con $i \in \{1 \dots H\}$,

la marca del experto consiste en un tiempo de inicio y uno de fin del HS, definidos como t_{inic}^i y t_{fin}^i , respectivamente. Estos tiempos están definidos en segundos y se asocian a dos números enteros n_{inic}^i y n_{fin}^i , que son los índices de las muestras en los instantes t_{inic}^i y t_{fin}^i . Esto, considerando que la señal de envolvente se encuentra muestreada a una tasa $f_s = 200[Hz]$ y que valor del primer índice es 1. Así, por ejemplo, si $t_{inic}^i = 0$ [s] y $t_{fin}^i = 2$ [s], se tiene que $n_{inic}^i = 1$ y $n_{fin}^i = 401$. Con esto, el promedio de la amplitud de cada segmento i se calcula mediante la siguiente fórmula:

$$pe_i = \frac{\sum_{j=n_{inic}^i}^{n_{fin}^i} env_j}{n_{fin}^i - n_{inic}^i + 1}, \quad (109)$$

donde env_j corresponde al valor de la muestra j de la función envolvente (ver Figura 21D). Luego, el umbral se fija como:

$$u_{peaks} = \bar{pe} - 2 \cdot std_{pe}, \quad (110)$$

donde:

$$\bar{pe} = \frac{1}{H} \sum_{i=1}^H pe_i \quad (111)$$

y

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^H (pe_i - \bar{pe})^2}. \quad (112)$$

Se elige este valor para u_{peaks} de tal manera de abarcar prácticamente todos los HS marcados por el experto, es decir, se pretende que ningún *peak* asociado a los HS tenga una amplitud promedio (calculada mediante la ecuación (109)) menor al umbral. Se utilizaron, en las pruebas preliminares, otras formas de calcular este umbral, las cuales son discutidas en la sección 4.3.2.4. En estas pruebas se concluyó que los mejores resultados se obtienen utilizando el método definido por la ecuación (110).

A la señal de envolvente (Figura 21D) se le aplica el umbral u_{peaks} , creándose una función binaria que a cada instante de tiempo asocia un 1 si en ese tiempo dicha señal de envolvente es mayor a u_{peaks} , o un 0 en caso contrario. Para una muestra j cualquiera, dicha señal, llamada fb , se define como:

$$fb_j = \begin{cases} 1 & \text{si } env_j > u_{peaks} \\ 0 & \text{si } env_j \leq u_{peaks} \end{cases} \quad (113)$$

donde env_j corresponde al valor de la muestra j de la función envolvente (ver Figura 21D). Luego, se define la función de ancho de cada *peak*, llamada *anch*. El valor de $anch_j$ será 0 si $fb_j = 0$. En caso contrario, se tiene necesariamente que $fb_j = 1$, y el valor de $anch_j$ se calcula mediante el siguiente procedimiento:

1. Se busca el índice k más grande que cumpla $k < j$ y $fb_k = 0$. Si no existe, $k = 0$.
2. Se busca el índice q más pequeño que cumpla $q > j$ y $fb_q = 0$. Si no existe, q es igual al tamaño total de la muestra más 1.
3. Se asigna a $anch_j$ el valor $\frac{q-k-1}{F_s}$. Donde F_s es la frecuencia de muestreo. Para el método implementado en este trabajo, se tiene que $F_s = 200$ [Hz].

Por ejemplo, supóngase que fb está definido sobre 7 muestras y toma los valores mostrados en la Tabla 1. Se supone además, por simpleza, que $F_s = 1$ [Hz]. A continuación se calcula el valor de $anch_j$ para cada valor de j :

1. Para $j = 1$: $fb_1 = 0$, por lo que $anch_1 = 0$.
2. Para $j = 2$: $fb_2 = 1$. Se busca el índice k más grande que cumple $k < 2$ y $fb_k = 0$. Dicho índice corresponde a 1. También se busca el índice q más pequeño que cumple $q > 2$ y $fb_q = 0$. Dicho índice corresponde a 4.
 $anch_2 = 4 - 1 - 1 = 2$.
3. Para $j = 3$: $fb_3 = 1$. Se busca el índice k más grande que cumple $k < 3$ y $fb_k = 0$. Dicho índice corresponde a 1. También se busca el índice q más pequeño que cumple $q > 3$ y $fb_q = 0$. Dicho índice corresponde a 4.
 $anch_3 = 4 - 1 - 1 = 2$.
4. Para $j = 4$: $fb_4 = 0$, por lo que $anch_4 = 0$.
5. Para $j = 5$: $fb_5 = 1$. Se busca el índice k más grande que cumple $k < 5$ y $fb_k = 0$. Dicho índice corresponde a 4. También se busca el índice q más pequeño que cumple $q > 5$ y $fb_q = 0$. Dicho índice corresponde a 8.
 $anch_5 = 8 - 4 - 1 = 3$.
6. Para $j = 6$: $fb_6 = 1$. Se busca el índice k más grande que cumple $k < 6$ y $fb_k = 0$. Dicho índice corresponde a 4. También se busca el índice q más pequeño que cumple $q > 6$ y $fb_q = 0$. Dicho índice corresponde a 8.
 $anch_6 = 8 - 4 - 1 = 3$.
7. Para $j = 7$: $fb_7 = 1$. Se busca el índice k más grande que cumple $k < 7$ y $fb_k = 0$. Dicho índice corresponde a 4. También se busca el índice q más pequeño que cumple $q > 7$ y $fb_q = 0$. Dicho índice corresponde a 8.
 $anch_7 = 8 - 4 - 1 = 3$.
8. Para $j = 4$: $fb_8 = 0$, por lo que $anch_8 = 0$.

En la Tabla 1 se resumen los valores de $anch_j$.

Tabla 1: Ejemplo de Valores de fb_j para 7 Muestras y el Valor de $anch_j$ Deducido a Partir de fb .

j	fb_j	$anch_j$
1	0	0
2	1	2
3	1	2
4	0	0
5	1	3
6	1	3
7	1	3
8	0	0

Intuitivamente, este procedimiento asigna a cada j que cumple $fb_j = 1$, el ancho del conjunto de índices contiguos a j tal que fb vale 1 para todos esos índices. En la Figura 22 se ilustra este proceso. Como se observa, tras aplicar el umbral se obtiene una señal binaria (Figura 22C). Todos los puntos contiguos en donde esta señal binaria vale 1 son agrupados, y a cada uno de ellos se le asigna el ancho del grupo al que pertenecen (Figura 22D).

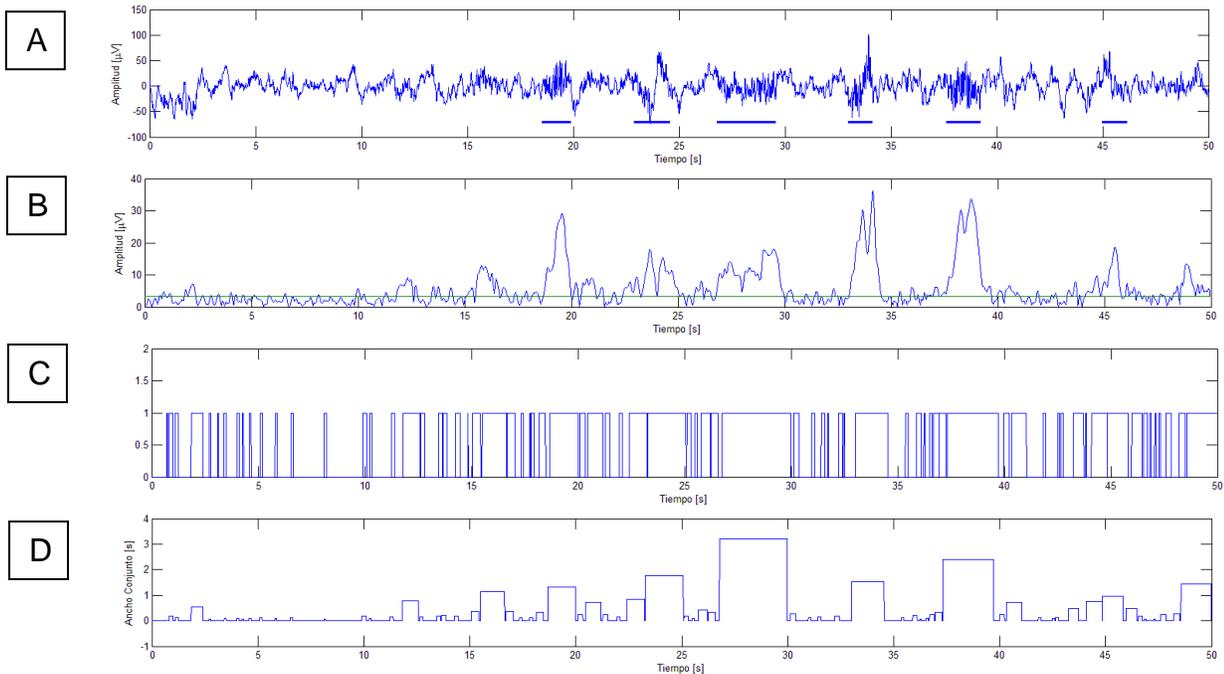


Figura 22: Proceso de Cálculo de Ancho Peak de la Envolvente de Señal de EEG.
A: Señal original con HS marcados por experto. B: Envolvente de señal de EEG, llamada env , con el respectivo umbral aplicado, u_{peaks} . C: Señal binaria, llamada fb . Toma el valor 1 si en el instante de tiempo j la función env_j supera u_{peaks} o 0 en caso contrario. D: Función de ancho de los conjuntos de 1's definidos por la señal C, llamada $anch$. A cada punto tal que $fb_j = 1$ se le asigna el ancho del conjunto de índices contiguos a j donde fb vale 1.

Tras calcular las funciones fb y $anch$ mostradas la Figura 22 C y D, éstas se sub-muestran en los instantes de tiempo a los que se asociaron las características obtenidas de los RP (es decir, la mitad de la ventana usada para calcular dichos RP), de tal forma de agregar dos nuevas características al vector asociado a dichos instantes de tiempo.

Envolvente Menos Media Local

En pruebas preliminares del sistema, se observó que era importante ver qué secciones de la señal envolvente tenían valores altos debido a que pertenecían a un *peak*, y cuáles pertenecían a una zona en dónde la amplitud de esta señal era alta en forma más extendida. La razón de esto es que normalmente los valores altos de amplitud ocasionados por HS tienden a ser locales y duran pocos segundos, mientras que el ruido y otros patrones pueden tener duraciones mayores. Esto es especialmente válido para vigilia, donde pueden aparecer señales de gran amplitud sin un patrón claro. Debido a esto, se utiliza como

característica una función que tiene valores altos para *peaks* cortos, como los asociados a los HS, pero valores bajos para *peaks* de mayor duración.

Se calculan tres funciones de media local de la señal de envolvente (*env*). Para calcular estas funciones se utilizan ventanas Gaussianas de duraciones de 2, 4 y 8 [s]. En general, una ventana Gaussiana, G_d , de duración d [s], tiene un número N de coeficientes dados por:

$$N = F_s d + 1, \quad (114)$$

donde F_s corresponde a la frecuencia de muestreo. Para este trabajo, se tiene $N = 200d + 1$. Los coeficientes de G_d se calculan mediante la siguiente ecuación:

$$G_d(i) = \frac{1}{M} e^{-\frac{1}{2} \left(\frac{5i}{N} \right)^2}, \quad (115)$$

donde:

$$i \in \left[-\frac{N-1}{2}, \frac{N-1}{2} \right] \quad (116)$$

y

$$M = \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} e^{-\frac{1}{2} \left(\frac{5i}{N} \right)^2}. \quad (117)$$

M es un factor de normalización que permite que:

$$\sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} G_d(i) = 1. \quad (118)$$

Las 3 funciones de media local se calculan como la convolución entre *env* y 3 funciones Gaussianas, según las siguientes fórmulas:

$$mm_2 = env * G_2, \quad (119)$$

$$mm_4 = env * G_4, \quad (120)$$

$$mm_8 = env * G_8. \quad (121)$$

En estas ecuaciones, mm_2 corresponde a la media local considerando una ventana de 2 [s], mm_4 corresponde a la media local considerando una ventana de 4 [s] y mm_8 corresponde a la media local considerando una ventana de 8 [s]. El ancho de los filtros Gaussianos está relacionado con la duración de un HS. Se decidió que uno de los anchos fuera la duración máxima de un HS, y que las otras dos fueran múltiplos de esta duración. En la Figura 23 se pueden observar los filtros Gaussianos utilizados.

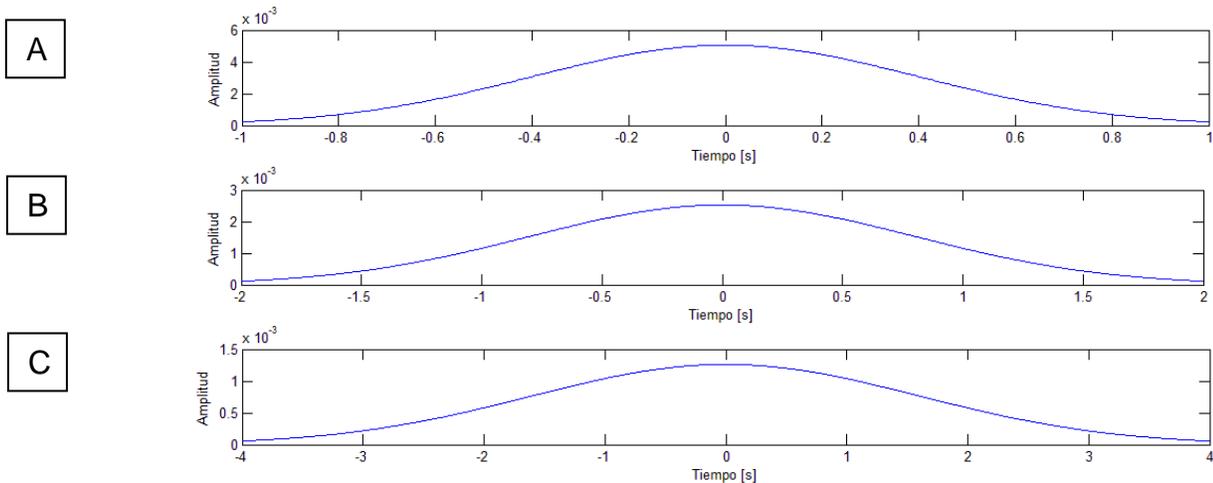


Figura 23: Filtros Gaussianos Usados en Convolución
A: Filtro Gaussiano de duración 2 [s], G_2 . B: Filtro Gaussiano de duración 4 [s], G_4 .
C: Filtro Gaussiano de duración 8 [s], G_8 .

A partir de las medias locales, se calculan las siguientes tres señales:

$$env_{mm2} = env - mm_2, \quad (122)$$

$$env_{mm4} = env - mm_4, \quad (123)$$

$$env_{mm8} = env - mm_8. \quad (124)$$

Estas señales corresponden la envolvente menos tres tipos distintos de media local y son usadas como características del sistema. Para ellos, se submuestrean en los instantes de tiempo a los que se asociaron las características obtenidas de los RP (es decir, la mitad de la ventana usada para calcular el RP), y se agregan al vector de características asociado a dichos instantes de tiempo.

3.2.2.2. Vector de Características Final

Considerando todas las características antes mencionadas, se construye a partir de ellas un vector que corresponde a la entrada del clasificador. Nótese que todas estas características están muestreadas la tasa de una muestra por cada ventana tomada de la señal pre-procesada. Así, resulta un vector de características para cada ventana, el cual se asocia al instante de tiempo correspondiente a la mitad de dicha ventana. En la Tabla 2 se resume el listado de características y se indica la propiedad de los HS que cada una de ellas busca medir.

Tabla 2: Características Usadas por Clasificador.

Número	Característica	Deducida de RP	Propiedad Medida
1	env (envolvente de EEG filtrado)	no	Amplitud
2	P_{σ}^n (potencia en banda sigma de EEG original)	no	Frecuencia
3	P_{σ}^n (potencia en banda sigma de EEG filtrado)	no	Frecuencia
4	env_{mm2} (envolvente de EEG menos media local (2 [s]))	no	Amplitud y Duración
5	env_{mm4} (envolvente de EEG menos media local (4 [s]))	no	Amplitud y Duración
6	env_{mm8} (envolvente de EEG menos media local (8 [s]))	no	Amplitud y Duración
7	L_{max} de RP para distancia de correlación	si	Morfología y Frecuencia
8	L_{medio} de RP para distancia de correlación	si	Morfología y Frecuencia
9	L_{desv} de RP para distancia de correlación	si	Morfología y Frecuencia
10	L_{max} de RP para distancia de correntropía	si	Morfología y Frecuencia
11	L_{medio} de RP para distancia de correntropía	si	Morfología y Frecuencia
12	L_{desv} de RP para distancia de correntropía	si	Morfología y Frecuencia
13	$anch$ (ancho de HS)	no	Amplitud y Duración
14	f_b (envolvente supera o no u_{peaks})	no	Amplitud y Duración

Como se puede apreciar en la Tabla 2, las características miden diferentes propiedades de la señal. Esto es importante, pues para clasificar correctamente HS, los expertos se basan en todos estos factores, y no sólo en uno de ellos. Nótese que solo 6 de las 14 características se deducen de RP's, siendo éstas las

únicas que miden la morfología del HS. Esta morfología es fundamental para clasificar correctamente dichos eventos.

3.2.3. Clasificación y Obtención de Detecciones

3.2.3.1. Entrenamiento del Clasificador

Una vez calculados los vectores de características se requiere entrenar un clasificador capaz de determinar si el tiempo en el cual se calculó cada vector está asociado a un HS o no. Se divide el espacio de características en dos clases: la clase 1 corresponde a vectores cuyo tiempo asociado (es decir, la mitad de la ventana en la cual fueron calculados) cae entre el inicio y el fin de un HS marcado por el experto. La clase 2 corresponde a vectores en donde dicha condición no se cumple, es decir, el centro de la ventana a partir de la cual se dedujo el vector de características cae fuera de un par inicio-fin marcado por un experto. En la Figura 24 se ilustra un ejemplo de cómo se llevó a cabo este proceso. En dicha figura se observan dos ventanas A y B de tamaño T , cuyo centro ha sido marcado por una línea vertical. En la parte inferior de la figura aparecen dos marcas de HS, dadas por el experto. El centro de la ventana A cae dentro de una marca del experto, y por lo tanto el vector calculado desde dicha ventana es considerado de la clase 1. Por otro lado, el centro de la ventana B cae fuera de cualquier HS, y por lo tanto el vector de características deducido desde esta ventana pertenece a la clase 2, a pesar de que parte de la ventana cae dentro de una marca del experto.



Figura 24: Ejemplo de Determinación de Clases para Vector Asociado a Cada Ventana.

El centro de la Ventana A cae entre las marcas de inicio y fin de un HS, por lo que el vector deducido a partir de esta ventana es considerado de la clase 1. El centro de la Ventana B cae fuera las marcas de inicio y fin de cualquier HS, por lo que el vector deducido a partir de esta ventana es considerado de la clase 2.

Bajo este criterio es posible asociar una clase a cada uno de los vectores de características calculados. Los trozos de registro fueron divididos en tres conjuntos: entrenamiento, validación y prueba. Los detalles de cómo se construyeron estos conjuntos se encuentran en la sección 4.2. Los conjuntos de entrenamiento y validación fueron usados en la construcción y selección del clasificador, mientras que el conjunto de prueba fue utilizado para evaluar la calidad de la clasificación. Una vez asociada una clase a cada vector de características, el problema quedó reducido a uno de clasificación binario, donde

los objetivos de la clasificación eran las clases ya conocidas para los conjuntos de entrenamiento y validación.

En las pruebas preliminares se utilizó como clasificador una SVM entrenada mediante el algoritmo C-SVM con kernel gaussiano. Sin embargo, este método no permitía construir una curva ROC en forma directa, pues sólo se podían variar los parámetros C de la optimización y γ del kernel gaussiano (ver ecuaciones (19) y (21)). Mover cualquiera de estos parámetros no permite obtener una curva clara en el espacio TPR-FPR, si no que más bien puntos aislados en dicho espacio. Por ello, se utiliza una NN con una capa oculta, cuya salida es discretizada mediante un umbral. Así, si se supera dicho umbral se considera que el vector a clasificar pertenece a la clase 1, y en caso contrario, a la clase 2. Modificando este umbral es posible construir una curva ROC.

La NN es entrenada mediante el algoritmo de retropropagación (gradiente descendiente) con momentum. Se ocupa una red de una capa oculta con sesgo, y como función de activación se utiliza la tangente hiperbólica, lo que implica que los valores de las salidas de las neuronas están entre -1 y 1. Los vectores de característica son normalizados para que sus componentes estén el rango [-1,1], y por ende puedan ser usadas como entradas a la NN. Para cada uno de los tres conjuntos en los que se dividen los datos (es decir: entrenamiento, validación y prueba) se toman todos los vectores pertenecientes cada conjunto, y para cada característica de estos vectores se busca su valor máximo y su valor mínimo. Luego, al valor máximo se le asigna un valor de 1, al valor mínimo un valor de -1, y todos los valores intermedios se interpolan linealmente entre dichos valores. En otras palabras, supóngase que para el conjunto C fijo (donde C puede ser el conjunto de validación, el de entrenamiento o el de prueba) el valor máximo de la característica k se denota C_k^M y el valor mínimo por C_k^m . Se llama \vec{v}_i al i -ésimo vector de características perteneciente a C , y $v_i(k)$ a la componente k de dicho vector. El valor normalizado de $v_i(k)$, llamado $\hat{v}_i(k)$, está dado por la siguiente ecuación:

$$\hat{v}_i(k) = 2 \cdot \frac{v_i(k) - C_k^m}{C_k^M - C_k^m} - 1 \quad (125)$$

Cada una de las componentes normalizadas de los vectores de características ingresa como una entrada a la red. Debido a esto, se dispone de 14 neuronas en la capa de entrada. La red neural tiene una única salida, y_1^c , con valores continuos entre -1 y 1, lo que implica una única neurona en la capa de salida. La salida de esta neurona indica el grado de certeza con el cual el trozo a partir del cual se dedujo el vector de características es o no un HS. Un valor 1 indica absoluta certeza de que el trozo es un HS, mientras que -1 indica absoluta certeza de que no es un HS. Hay N_o neuronas en la capa oculta. Este valor es un parámetro del sistema, y fue ajustado en pruebas preliminares (ver 4.3.3.4). Los parámetros del entrenamiento de esta red, es decir, la tasa de aprendizaje μ y la

tasa de momentum α , también fueron ajustados en pruebas preliminares. El diagrama general de esta red se muestra en la Figura 25.

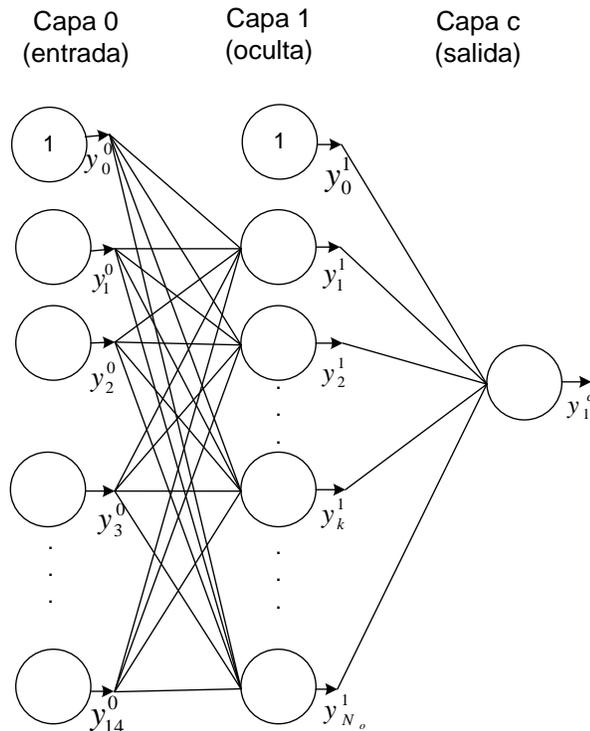


Figura 25: Diagrama de la Red Neuronal Usada para Clasificar Husos Sigma. Contiene 3 capas en total. La capa oculta tiene N_o neuronas, la capa de entrada 14 neuronas y la capa de salida una neurona. La salida de una neurona k cualquiera, perteneciente a una capa l cualquiera, se denota y_k^l .

Se debe tener en cuenta que, dadas las características del problema a resolver, la clasificación es bastante asimétrica, pues la mayoría de las ventanas obtenidas de la señal original están asociadas a la clase 2 (no HS). Esto, debido a que el porcentaje de tiempo en un EEG en donde ocurren husos sigma es bastante bajo. Si al momento de entrenar la red neuronal simplemente se minimiza el error cuadrático de clasificación, se tiende a crear una red cuyas clasificaciones están sesgadas hacia la clase 2. En pruebas preliminares se observó que la red tendía a tener pocos FP, a cambio de un gran número de FN. Una forma de resolver este problema es aumentar, el costo de los falsos negativos durante la optimización, para así reducir su número. Sin embargo, se optó por una solución más sencilla, en donde se aumenta artificialmente la cantidad de HS (marcados por el experto) en el conjunto de entrenamiento, de tal manera que durante el entrenamiento el número de positivos y de negativos fueran más parecidos. Para lograr esto, se toman todos los positivos reales del conjunto de entrenamiento, y se repiten un número n_p de veces, antes de presentarlos a la red. Con esto, se agranda el número de positivos en el conjunto de entrenamiento sin requerir de más muestras. El número n_p es un parámetro del sistema. Una alternativa a este método, que fue considerada pero luego descartada, era disminuir el número de

negativos del conjunto de entrenamiento, sacando algunas muestras. Esto, sin embargo, no resultó conveniente debido a que requería definir un criterio para seleccionar los negativos, lo cual podría resultar demasiado arbitrario. Si, en forma alternativa, se seleccionan aleatoriamente los elementos a eliminar, se agrega cierta varianza a los resultados del método, lo cual complica las comparaciones entre diferentes métodos y parámetros.

3.2.3.2. Determinación de Inicio y Fin de HS

Luego del entrenamiento, la red neuronal se utiliza para clasificar nuevas muestras, las cuales se extraen del conjunto de prueba. El procedimiento usado para determinar los vectores de características a partir de los nuevos registros es igual al usado durante el entrenamiento.

Los vectores normalizados a clasificar son pasados por la red neuronal, obteniéndose una salida para cada uno de ellos. Con esto, para cada ventana en las que ha sido dividida la función original, se tiene un valor entre -1 y 1 que indica el grado de pertenencia de dicha ventana a la clase de los HS. Dicho valor se asigna al centro de la ventana. Con esto se obtiene una función de salida, muestreada a la misma tasa a la que se muestrea durante la determinación de los vectores de características. Esta función se interpola linealmente a la misma tasa de muestreo de la señal original, es decir, 200 muestras por segundo. Así, se obtiene una función con una resolución bastante grande (0.005 [s]) cuya ordenada representa la certeza de que cada instante de tiempo corresponde a un HS.

La idea de la interpolación es que la determinación de los sectores marcados como HS sea más gradual. Originalmente se había considerado clasificar directamente a partir de la salida de la red neuronal, sin interpolar. Así, para aquellas ventanas clasificadas en la clase 1 por el sistema, se consideraba toda la duración de dicha ventana como parte de un HS. Esto no resultó ser óptimo, pues el inicio y fin de los HS estaba limitado únicamente al inicio y fin de las ventanas utilizadas, lo que empeoraba la resolución del sistema. Además, esto causaba que la elección de las posiciones de la ventana influyera enormemente en el desempeño del sistema. Por ello, se realiza una interpolación de las salidas de la red neuronal, y posteriormente la función resultante de esta interpolación es pasada por un umbral, llamado u_h , para determinar el inicio y fin de los HS. Así, todos los puntos de la función interpolada que superaran a u_h son considerados parte de un HS. Debido a que la función interpolada es suave, es sencillo encontrar el inicio y el fin de los HS, ya que las regiones definidas como por el umbral tienden a ser relativamente grandes, es decir, poseen varios puntos. El umbral u_h es determinado en el conjunto de entrenamiento.

El uso de interpolación resulta ser muy ventajoso, pues le agrega gradualidad a la clasificación. Debido a que el problema de determinación de clases es inherentemente discreto, es difícil de adaptar a series de tiempo, como sucede en este caso. Utilizar interpolación permite extender de forma adecuada la

clasificación a los instantes de tiempo donde no se calcula un vector de características. Así, por ejemplo, si para dos ventanas consecutivas se obtienen valores altos de la salida de la red neuronal, se considera (probablemente en forma acertada) que los puntos intermedios pertenecen a un HS. Alternativamente, si una ventana es clasificada como HS, y la siguiente no lo es, el punto en donde se considera que termina el evento marcado por la primera ventana queda determinado por las magnitudes de la salida de la red neuronal para ambas ventanas. Así, por ejemplo el HS tenderá a extenderse en la medida que las salidas para la primera ventana y la segunda ventana sean ambas altas (aunque en el caso de la segunda ventana no se supere el umbral) y tenderá a contraerse en la medida que disminuya cualquiera de las dos salidas. En este sentido, usar una NN es una buena idea, pues su salida es un número real entre -1 y 1 que permite realizar una interpolación. Si, en cambio, la salida fuese una clase, como ocurre con otros clasificadores, por ejemplo SVM, la interpolación, aunque posible, carecería de sentido. En la Figura 26 se ilustra la salida interpolada de la red neuronal. Se observa que el hecho de que esta salida pueda tener valores reales entre -1 y 1 permite encontrar en forma más precisa, mediante un umbral, dónde comienza y dónde terminan los HS.

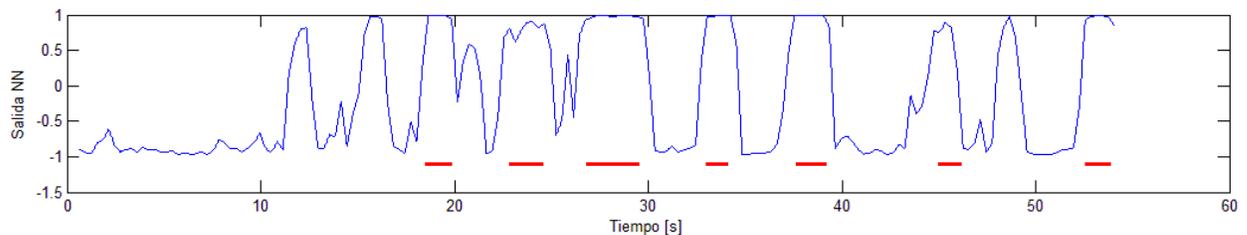


Figura 26: Salida Interpolada de Red Neuronal Para un Ejemplo de SQ2.
Para efectos comparativos, se han puesto las marcas de HS del experto como líneas rojas horizontales en la parte inferior del gráfico. Los dos *peaks* de la izquierda de la señal interpolada corresponden a FP.

3.3. Parámetros Principales

A continuación se detallan los parámetros principales del sistema y cómo éstos fueron ajustados. En este trabajo se utilizaron dos formas de ajustar los parámetros. La primera, aplicada a parámetros de menor importancia, consiste en dejar el resto de los parámetros constantes y variar solo el parámetro en cuestión, buscando el valor óptimo de dicho parámetro. Se utilizan dos métodos para encontrar este valor óptimo:

1. Calcular la IM entre las características del sistema que dependen del parámetro y las marcas de HS del experto, y ver para qué valor del parámetro se obtienen los mayores valores de IM.

2. Entrenar distintos clasificadores variando el parámetro, y ver para cuál se obtienen mejores clasificaciones.

Esto se realiza en un subconjunto del conjunto de entrenamiento, para que las pruebas tomaran menos tiempo. La segunda forma de ajustar los parámetros, que fue usada solo para parámetros más importantes, consiste en asignar a cada uno de estos parámetros un conjunto de valores posibles. Así, para cada parámetro importante se tiene un rango de valores. Luego, se hace un producto cruz entre estos rangos de valores para los parámetros, obteniéndose todas las combinaciones posibles de valores para los parámetros principales. Para cada una de estas combinaciones, se entrena el sistema y se comparaba el desempeño de la clasificación. Esta prueba es conocida como la prueba final de ajuste de parámetros. Los detalles de esta prueba se describen en 4.4.

3.3.1. Parámetros de pre-procesamiento de la señal

Los principales parámetros del pre-procesamiento de señal son las dos frecuencias de corte del filtro pasa banda, las cuales son llamadas f_{ch} y f_{cl} . Estas frecuencias son importantes, por lo que su valor fue ajustado en la prueba final de ajuste de parámetros.

3.3.2. Parámetros de Extracción de Características y Construcción de RPs

Los parámetros usados durante la construcción de RP's corresponden a: m y τ , usados en el método de retraso de tiempo (ver 2.2.3); a y tr , que corresponden al ancho de las ventanas utilizadas para construir los RP y al traslape entre ellas, respectivamente (ver 3.2.2.1); p y σ_t , que corresponden al periodo y al ancho del kernel temporal para $d_{correl-kp}$ (ver 2.2.4.3); σ_d que corresponde al parámetro de la función gaussiana usada para el cálculo de $d_{corren-n}$ (Ver 2.2.4.4); tw y L_{min} , que corresponden al ancho de la ventana de Theiler y al largo mínimo usado en las estadísticas del largo de las líneas en los RPs, respectivamente (ver 2.2.6); y ε_l y ε_n , que corresponden a los umbrales usados sobre las distancias entre los vectores, para construir los RPs. Los parámetros, ε_l , ε_n , a y tr , son considerados importantes, por lo que su valor fue ajustado en la prueba final de ajuste de parámetros. Los demás parámetros mencionados en esta sección fueron determinados mediante pruebas preliminares.

3.3.3. Parámetros de Clasificación

Los parámetros usados durante el entrenamiento son: la tasa de momentum α , el factor de aprendizaje μ , el número de neuronas en la capa oculta N_o y el número de veces que los positivos son repetidos en el conjunto de

entrenamiento, n_p . Ninguno de estos parámetros es considerado crítico, por lo que fueron ajustados mediante pruebas preliminares.

3.3.4. Resumen de Parámetros Utilizados

En la Tabla 3 se resumen todos los parámetros del sistema, mostrándose la parte del sistema donde éstos son usados, y la forma en que fueron ajustados.

Tabla 3: Resumen de Parámetros del Sistema.
Para cada parámetro, se explica en qué parte del sistema era usado, y la forma en que fue determinado (pruebas preliminares o en la prueba final de ajuste de parámetros).

Parámetro	Uso	Forma de Ajuste
m	Construcción características	Prueba Preliminar
τ	Construcción características	Prueba Preliminar
p	Construcción características	Prueba Preliminar
tw	Construcción características	Prueba Preliminar
L_{min}	Construcción características	Prueba Preliminar
σ_t	Construcción características	Prueba Preliminar
σ_d	Construcción características	Prueba Preliminar
ε_l	Construcción características	Ajuste final de parámetros
ε_n	Construcción características	Ajuste final de parámetros
tr	Construcción características	Ajuste final de parámetros
a	Construcción características	Ajuste final de parámetros
n_p	Entrenamiento clasificador	Prueba Preliminar
N_o	Entrenamiento clasificador	Prueba Preliminar
α	Entrenamiento clasificador	Prueba Preliminar
μ	Entrenamiento clasificador	Prueba Preliminar
f_{cl}	Pre-procesamiento	Ajuste final de parámetros
f_{ch}	Pre-procesamiento	Ajuste final de parámetros

En este resumen se excluye el parámetro u_h , ya que su valor se varía para construir la curva ROC. En las en que el desempeño del sistema es comparado con el desempeño del sistema implementado en [7], este parámetro es fijado en el conjunto de entrenamiento.

4. Resultados

4.1. Metodología de Pruebas Realizadas

4.1.1. Criterio para Determinar Concordancia entre Sistema y Experto

Para medir la concordancia entre el sistema y las marcas del experto, es necesario definir previamente un criterio que permita determinar qué se entiende por un FP, un TP y un FN. El criterio establecido en este trabajo es casi el mismo al usado en [6] y [7], y se describe a continuación. La única diferencia entre ambos es que en esta memoria no se utilizan ni definen los TN, por razones que se explican más adelante. En la Figura 27 se resume este criterio, el cual es explicado en detalle en las líneas siguientes.

Se considera verdadero positivo (TP) a todo trozo marcado por el sistema que coincida con más del 70% de una marca del experto, independiente de donde empiecen o donde terminen la marcas.

Un FN corresponde a una marca del experto que no coincide con ninguna marca del sistema. También se considera FN a cualquier marca del experto que cumple que 70% o menos de su largo coincide con una marca del sistema. Es decir, si no se cumple la condición para que la marca del experto sea TP, será considerado un FN.

Se considera FP a cualquier marca del sistema que no esté asociada a una marca del experto. También, si una marca del sistema supera por 0,5 [s] o más a la marca de un experto, hacia cualquiera de los lados, ese trozo es considerado FP. En este caso, es posible que, de todos modos, la marca del experto coincida en más de 70% con la marca del sistema. En esta situación se contabiliza un FP y un TP. Notar que es posible que la marca del sistema supere a la marca del experto por más de 0,5 [s] hacia ambos lados. En este caso, se contabilizan 2 FP (además de un TP si se cumple el criterio de traslape).

En este trabajo no se contabilizan los TN. La razón de esto es que, en contraste con un problema de clasificación convencional, en este caso los negativos no corresponden a un evento, sino más bien a la ausencia de eventos. Así, por ejemplo, es posible tener marcas de 3 HS de duración 1 [s] cada uno en una señal de 20 [s]. Es claro que los segundos donde están presentes los HS se pueden considerar como positivos, ya sea que estos estén marcados por el sistema o por el experto, pero no es claro qué considerar como negativo. Por

ejemplo, es difícil decidir si un trozo de 10 [s] sin marcas corresponde a un solo negativo, o a varios. Todas las otras mediciones (TP, FN y FP) están asociadas a una marca: los TPs provienen de una del sistema y del experto, los FNs provienen del experto y los FPs provienen de una del sistema. Sin embargo, los TN no provienen de ninguna marca lo que hace difícil definirlos. Debido a esto, no se contabilizaron los TN, en este criterio, y por lo tanto las medidas de desempeño usadas no dependen de esta medición. En [6] y [7] se definen los TN como aquellos segmentos del EEG con frecuencias en la banda sigma, pero que no cumplen con los criterios morfológicos necesarios para ser considerados HS. Esto requiere de un método para definir dichos segmentos, y además resulta demasiado arbitrario, debido a que los segmentos del EEG que no tienen frecuencias en la banda sigma también deberían ser considerados como TN. Por ello, no se usa este método, y las comparaciones con el método presentando en [6] se hacen exclusivamente en base a tasas que no incluyen el cálculo de TN.

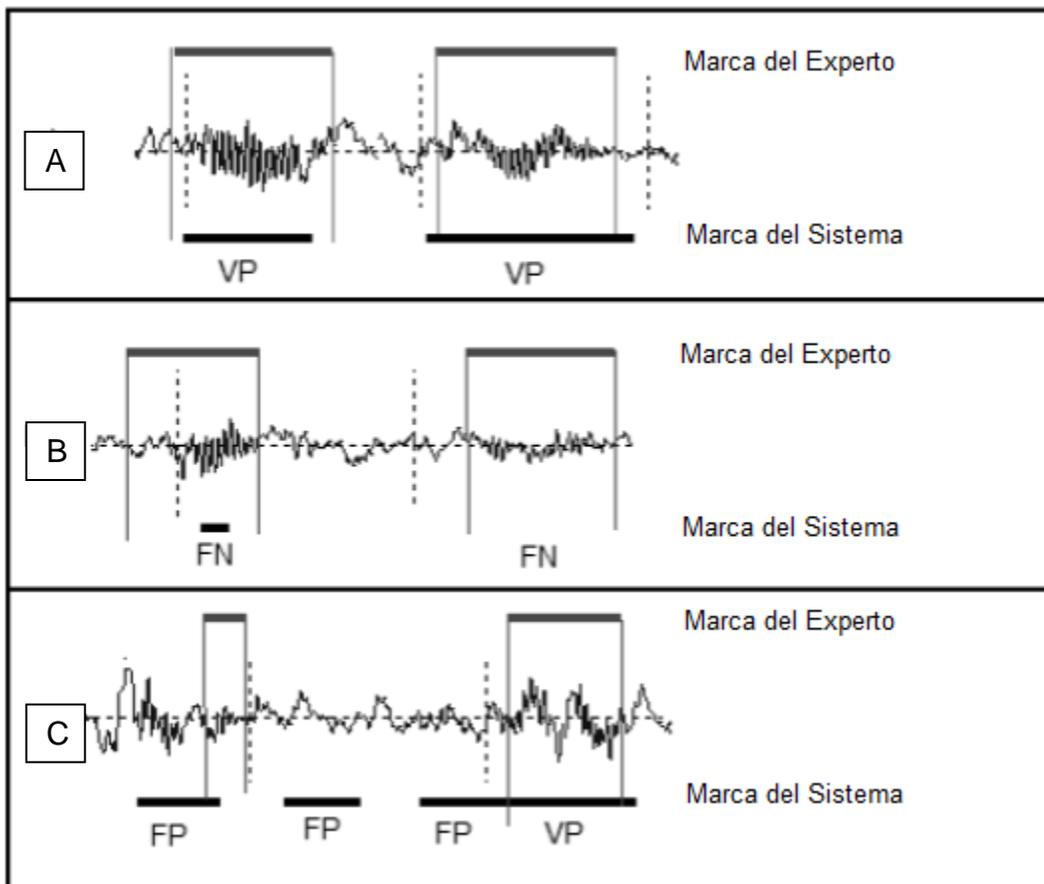


Figura 27: Ejemplos de Criterio Usado para Medir Desempeño del Sistema.
A: Ejemplos de TP: Debe existir una coincidencia mayor al 70% de la marca del experto. **B:** Ejemplos de FN: Se consideran casos en que el sistema no detecta el HS marcado por el experto, o donde la coincidencia entre la marca del sistema y la del experto es inferior al 70% de la marca del experto. **C:** Ejemplos de FP: Están dados por detecciones del sistema que no corresponden a marcas del experto, o

bien por sobre estimaciones de marcas del experto de más de 0,5 [s]. En este último caso la porción excedente se considera FP y el resto puede ser un TP, si se cumple el criterio establecido en A. Esta figura fue tomada de [7].

4.1.2. Medidas de Desempeño

Las dos medidas de desempeño usadas son la *TPR* y la *FDR* (ecuaciones (2) y (6), respectivamente). Debido a que no es posible calcular la *FPR* (ya que en este trabajo los TN no están definidos y por ende no pueden ser calculados), la curva ROC se calculó graficando *TPR* versus *FDR*. Aunque esta curva no corresponde a una curva ROC propiamente tal, ya que no utiliza la *FPR*, resulta ser un método adecuado de comparación, y es el mismo método que el usado en [6] y [7] para construir curvas ROC. Nótese que aunque en [6] y [7] es posible calcular la *TPR*, no se utiliza esta medición para la construcción de la curva ROC.

4.2. Base de Datos

La base de datos utilizada en este es un subconjunto de la base de datos utilizada en [6]. La base de datos original tenía mediciones del polisomnograma de 56 niños y niñas de 10 años de edad. Los registros tienen una duración de una noche, y contienen las siguientes señales: 5 canales para EEG mediante electrodos colocados según el sistema internacional 10-20, utilizándose las derivaciones F4-C4, C4-O2, F3-C3, C3-O1 y C4-C3; 2 canales para movimientos oculares, monitoreados por EOG; cuatro canales para EMG, monitoreándose el mentón, el diafragma y ambas piernas; un canal para ECG, utilizando la derivación V1; dos canales para actividad motora de las extremidades superior e inferior; un canal para el flujo de aire nasal; un canal para el esfuerzo respiratorio abdominal; un canal para el pulso; un canal para la temperatura corporal; un canal para la oximetría; un canal para la posición del paciente; y uno para registrar los ronquidos de individuo. Con esto se totalizaban 21 canales, los cuales fueron todos muestreados a 200 [Hz] [7]. Esta base de datos fue posteriormente analizada visualmente por expertos del INTA, marcando el principio y el fin de los HS mediante un software computacional. Con esto, se dispone de una base de aproximadamente 400 horas de grabación, en donde los eventos están marcados completamente. Esta base fue y podrá ser utilizada para muchos sistemas de detección automática de HS.

En este estudio se utiliza una base de datos que corresponde a un subconjunto de la base de datos original. Esto debido a que se trata de un estudio preliminar, lo que hizo necesario ejecutar muchas veces las simulaciones para ajustar el sistema. Estas pruebas habrían demorado demasiado tiempo para una base mayor, por lo que resulta más conveniente utilizar una base más pequeña a la utilizada en [6].

La base de datos se compone por algunos segundos de señales de 6 pacientes distintos. Para cada paciente, existían registros distintos, divididos según la etapa del sueño en la que se encontraba el individuo al tomar esos datos. Estos registros corresponden a: SP, SQ1, SQ2, SQ3, SQ4 y VI. Todos estos corresponden a estados o etapas de sueño, salvo VI que fue medido durante vigilia. No todos los pacientes tienen datos para todos estos registros. En la Tabla 4 se muestra la duración total de los registros que se tenían para cada paciente y para cada fase.

Tabla 4: Duración de Registros para los 3 Pacientes y para Diferentes Fases de Sueño.

Paciente	Registro	Duración [s]
1	SQ2	292
1	SQ3	487
1	SP	292
1	VI	779
2	SQ2	2347
2	SQ3	1702
2	SP	1124
2	VI	700
3	SP	638
3	SQ1	179
3	SQ2	638
3	SQ3	359
3	SQ4	638
3	VI	638
4	SP	638
4	SQ1	179
4	SQ2	638
4	SQ3	359
4	SQ4	638
4	VI	638
5	SP	359
5	SQ2	638
5	SQ3	638
5	SQ4	638
5	VI	638
6	SP	638
6	SQ1	359
6	SQ3	359
6	SQ4	179

La forma en que cada uno de estos conjuntos se dividió en CE, CP y CV depende de la prueba, por lo que se explica en la sección correspondiente.

4.3. Pruebas Preliminares

Debido a que inicialmente no se sabía qué variantes del método serían más exitosas que otras, fue necesario implementar una serie de pruebas preliminares para tener alguna base para tomar decisiones respecto a la forma de operar del sistema. Similarmente, era necesario tener algún método para ajustar los parámetros menos importantes del sistema en forma rápida, de tal forma de no entorpecer la prueba de ajuste de parámetros principales (en la Tabla 3 se muestra qué parámetros son importantes y cuáles no). Para todas las pruebas preliminares se siguió una metodología común, la cual se explica en detalle a continuación. Posteriormente, se muestran los resultados para cada prueba.

4.3.1. Metodología

A causa de que las pruebas preliminares requerían resultados rápidos, se tomó un grupo de muestras de la base de datos relativamente pequeño para llevarlas a cabo. Este grupo fue dividido en un conjunto de entrenamiento (CE), uno de validación (CV) y uno de prueba (CP). En la Tabla 5 se muestra cómo se eligieron los datos de cada uno de estos conjuntos. El CE fue utilizado para entrenar la NN. El CV fue utilizado durante el entrenamiento de la NN para evitar que ésta se sobre entrenara. Para ello, a medida que se entrenaba la NN, se medía su desempeño en CV, y cuando éste dejara de disminuir, se detenía el entrenamiento. El CP fue utilizado para comparar el desempeño de las distintas variantes del sistema.

Tabla 5: Inicio y Fin de Trozo de Tiempo de cada Registros Usados para Conjuntos de Entrenamiento y Validación para Pruebas Preliminares.

Paciente	Etapa	Inicio CE [s]	Fin CE [s]	Inicio CV [s]	Fin CV [s]	Inicio CP [s]	Fin CP [s]
1	SQ2	0	50	51	75	76	100
1	SQ3	0	50	51	75	76	100
1	VI	0	50	51	75	76	100
1	SP	0	50	51	75	76	100
2	SQ2	0	50	51	75	76	100
2	SQ3	0	50	51	75	76	100
2	VI	0	50	51	75	76	100
2	SP	0	50	51	75	76	100

En todas las pruebas preliminares se compararon dos o más sistemas diferentes, donde el funcionamiento de éstos estaba basado en las distintas

variantes del método o en diferentes valores de parámetros. Estos sistemas fueron comparados de dos formas distintas. El primero método de comparación fue tomar aquellas características que eran diferentes en los dos sistemas a comparar (ya sea porque se habían calculado mediante métodos o parámetros distintos) y calcular la IM entre cada una estas características y las marcas de HS realizadas por el experto. El segundo método consistió en medir el desempeño de la clasificación llevada a cabo por ambos sistemas. Para ello, se compararon los FP obtenidos por los distintos métodos en el conjunto de prueba, considerando, en cada caso, un umbral para la salida de la red neuronal tal que se obtuvieran los mismos TP que el módulo 2 del sistema de [6] y [7] sobre los mismos datos (CP). Al comparar los FP se realizaron, para cada sistema, tres entrenamientos diferentes de la NN, con lo que obtuvieron 3 desempeños distintos para cada clasificador. Esto, debido a que al comienzo del entrenamiento de la NN los pesos entre las neuronas se fijan aleatoriamente. Esto causa que el desempeño también sea aleatorio. Por ello, para efectos comparativos, es importante reducir esta aleatoriedad.

Se considera en informe que los resultados entregados por la IM son de carácter complementario, dándose más prioridad a los resultados obtenidos al medir los FP de la salida del sistema. Esto, debido a que la IM considera el aporte de una sola característica a la clasificación, y no el aporte que un grupo de características puede significar. Además no se considera la forma en que el clasificador combina las características. Así, es posible que tres características, a , b y c , tengan información mutua decreciente, siendo a la que tiene más información mutua y c la que tiene menos, pero que la combinación de a y c entregue más información que la combinación de a y b , debido a que a y b son redundantes. Por esto, se le da más prioridad a los resultados que consideran la salida del sistema (FP obtenidos por distintos sistemas), pues reflejan mejor el aporte de cada variante a comparar. Es decir, si luego de comparar la IM se concluye que una variante o parámetro es óptimo, pero al comparar los desempeños del sistema se concluye que otra variante o parámetro es mejor, se escoge este último para el sistema final.

4.3.2. Pruebas sobre Variantes

4.3.2.1. Pruebas sobre Distancias

Esta prueba fue realizada para comparar las distintas distancias a partir de las cuales se pueden construir los RP's. Se consideraron sistemas donde solo se calcula un RP para una única distancia, como también sistemas donde se calculan 2 RP's distintos, cada uno para una distancia diferente, combinándose las medidas deducidas de cada RP en un solo vector de características. Los sistemas fueron comparados según la IM entre las características principales deducidas de los RP's. Además, se compararon los FP obtenidos por los distintos métodos en el CP, considerando, en cada caso, un umbral para la salida de la red neuronal tal que se obtuvieran los mismos TP que la salida del módulo 2 (sobre el mismo CP)

del sistema de [6] y [7]. Los sistemas eran entrenados usando CE y CV. Éste último conjunto se utilizó para evitar sobre entrenamiento.

Debido a que para cada distancia usada para construir los RPs se requiere tener un conjunto diferente de parámetros, se realizó primero, para cada una de ellas, una prueba donde éstos se ajustaban en forma individual. Dicha prueba consistía en lo siguiente: primero, dada una distancia, se van variando sus parámetros y se inspeccionan visualmente los RP obtenidos hasta que éstos presenten estructuras de líneas diagonales frente a HS. Con esto se obtuvieron valores preliminares de los parámetros para dicha distancia. Luego, cada uno de estos valores se varió individualmente, en torno a su valor preliminar, dejando los demás parámetros constantes. Para cada parámetro se consideraron alrededor de 5 valores distintos. Con las características deducidas a partir de estos parámetros, se calculó la información mutua entre los HS y dichas características, para el CE, escogiéndose el valor que maximizara la IM. Así, se obtuvo el set óptimo de parámetros para cada distancia. En la Tabla 6 se resumen estos valores óptimos.

Tabla 6: Valores Óptimos Obtenidos para cada Distancia.

Dichos parámetros se ajustaron visualmente, observando los RP obtenidos a partir de ellos, y seleccionado aquellos que marcaban de mejor manera los HS. Para los significados de estos parámetros, ver las ecuaciones de la sección 2.2.4.

Distancia	m	τ	ε	σ_d	σ_t	p
d_{en}	4	20	0,3	-	-	-
$d_{corren-n}$	4	20	0,2	0,7	-	-
d_{correl}	4	20	0,4	-	-	-
$d_{corren-kp}$	4	20	0,5	2	0,0095	0,08
$d_{correl-kp}$	4	20	40	-	0,0085	0,08

Luego de obtener estos valores óptimos, se procedió a comparar la IM, para el CV, entre las características deducidas de cada distancia y las marcas de HS del experto. Se consideraron las siguientes características: L_{max} , L_{medio} y L_{desv} (que corresponden a las características deducidas de RP que obtenían la IM de mayor valor). En la Tabla 7 se muestran estos resultados.

Tabla 7: Comparación de Información Mutua entre L_{max} , L_{σ} y L_{mean} y Marcas de HS de Experto.

Los parámetros usados para cada distancia eran los óptimos determinados anteriormente.

Distancia	Info. Mutua L_{max}	Info. Mutua L_{medio}	Info. Mutua L_{desv}
d_{en}	0,165	0,085	0,116
$d_{corren-n}$	0,174	0,075	0,111
d_{correl}	0,179	0,093	0,124
$d_{corren-kp}$	0,177	0,107	0,137
$d_{correl-kp}$	0,174	0,110	0,127

Finalmente, se probó sobre el CP el desempeño de diferentes sistemas que usaban una o dos distancias para determinar sus características, y se comparó su desempeño midiendo los FP para el u_h a la salida de la NN en que los TP eran iguales o mayores a los TP obtenidos con la salida del módulo 2 (M2) de [6] (aplicado sobre el mismo CP). En la Tabla 8 se detallan los resultados de esta prueba. Nótese que un sistema que utiliza dos distancias calcula dos RPs distintos (uno para cada distancia) y obtiene características de cada uno de ellos. Luego, las características deducidas de ambos RPs son consideradas en vector de características que ingresa a la NN.

Tabla 8: TP y FP Obtenidos sobre CP para Distintas Distancias Usadas para Construcción de RP's

En las columnas de más a la izquierda aparecen los TP de cada iteración, y posteriormente los FP de cada iteración con su promedio. Cada una de las tres iteraciones (it.1, it. 2. e it. 3) corresponden a un entrenamiento diferente de la NN. El sistema para el cual se obtienen los mejores resultados aparece destacado.

Sistema	TP It. 1	TP It. 2	TP It. 3	FP It. 1	FP It. 2	FP It. 3	Promedio FP
M2	11	11	11	48	48	48	48,0
<i>d_{en}</i>	11	11	11	23	28	24	25,0
<i>d_{corren-n}</i>	11	11	11	24	22	18	21,3
<i>d_{correl}</i>	11	11	11	26	22	27	25,0
<i>d_{corren-kp}</i>	11	11	12	27	23	27	25,7
<i>d_{correl-kp}</i>	11	11	11	24	26	23	24,3
<i>d_{correl-kp}</i> y <i>d_{corren-kp}</i>	11	11	11	25	20	25	23,3
<i>d_{correl-kp}</i> y <i>d_{corren-n}</i>	11	11	11	15	16	22	17,7
<i>d_{correl-kp}</i> y <i>d_{corren-kp}</i>	11	12	11	25	27	26	26,0
<i>d_{corren-kp}</i> y <i>d_{en}</i>	11	11	11	37	33	20	30,0

Como se puede apreciar en la Tabla 8, los mejores resultados se obtienen considerando como medida de distancia la $d_{correl-kp}$ y $d_{corren-n}$. Las características deducidas de estas dos distancias están dentro de las que mayor información mutua tienen con las marcas del experto, como se puede ver en la Tabla 7. Notar, sin embargo, que la única característica del RP en donde $d_{correl-kp}$ es la que tiene mayor información mutua con las marcas del experto es L_{medio} . Para las otras dos características (L_{max} y L_{desv}), la IM es máxima para otra distancia. Asimismo, la IM entre las características deducidas de $d_{corren-n}$ no es máxima para ninguna de las tres características principales. Por las razones que ya se explicaron en la sección 4.3.1, se le dio más prioridad a los resultados de la Tabla 8, escogiéndose un sistema en donde se construyen dos RPs, uno a partir de $d_{correl-kp}$ y otro a partir de $d_{corren-n}$.

4.3.2.2. Pruebas sobre Mediciones de RP's

La siguiente prueba fue realizada para ver el aporte de las distintas medidas deducidas del RP. Consistió en medir la IM de cada característica con las marcas de HS realizadas por experto para el CV. Luego, se seleccionaron las 3 características de mayor IM y se contrastó la clasificación en CP de un sistema que usaba solo estas tres características (además de las características que no

provenían del RP) versus un sistema que usara todas las características provenientes del RP. El desempeño de la clasificación se determinó midiendo los FP obtenidos por los distintos métodos en el CP, considerando, en cada caso, un umbral para la salida de la red neuronal tal que se obtuvieran los mismos TP que la salida del módulo 2 (M2), sobre el mismo CP. Los sistemas eran entrenados usando CE y CV. Éste último conjunto se utilizó para evitar sobre entrenamiento. Para construir los RPs, se utilizaron las distancias y los parámetros determinados en la sección anterior. En la Tabla 9 se muestran los resultados para la IM, y en la Tabla 10 se muestran los resultados para la clasificación.

Tabla 9: Comparación de Información Mutua entre Cada Medidas Deducidas del RP y Marcas de HS de Experto.

Las medidas que tenían mayor información mutua con las marcas del experto, y que fueron consideradas como las principales para esta prueba (L_{max} , L_{medio} y L_{desv}), aparecen destacadas.

Medida	Distancia	Información Mutua
L_{max}	$d_{corren-n}$	0,174
ENT	$d_{corren-n}$	0,010
RR	$d_{corren-n}$	0,012
L_{medio}	$d_{corren-n}$	0,075
L_{desv}	$d_{corren-n}$	0,111
L_{skew}	$d_{corren-n}$	0,003
L_{kurt}	$d_{corren-n}$	0,047
DET	$d_{corren-n}$	0,000
RA	$d_{corren-n}$	0,005
L_{max}	$d_{correl-kp}$	0,174
ENT	$d_{correl-kp}$	0,014
RR	$d_{correl-kp}$	0,000
L_{medio}	$d_{correl-kp}$	0,110
L_{desv}	$d_{correl-kp}$	0,127
L_{skew}	$d_{correl-kp}$	0,006
L_{kurt}	$d_{correl-kp}$	0,025
DET	$d_{correl-kp}$	0,000
RA	$d_{correl-kp}$	0,077

Tabla 10: Comparación de Clasificaciones sobre CP entre un Sistema que Usa las 9 Características de RP y un Sistema que Usa solo las 3 con Mayor IM.

En las columnas de más a la izquierda aparecen los TP de cada iteración, y posteriormente los FP de cada iteración con su promedio. Cada una de las tres iteraciones (it.1, it 2. e it. 3) corresponden a un entrenamiento diferente de la red neuronal. El sistema para el cual se obtienen los mejores resultados aparece destacado.

Sistema	TP It. 1	TP It. 2	TP It. 3	FP It. 1	FP It. 2	FP It. 3	Promedio FP
M2	11	11	11	48	48	48	48,0
Sistema Usando 3 Características de RP	11	11	11	18	19	23	20,0
Sistema Usando 9 Características de RP	11	12	11	17	29	19	21,7

Se puede apreciar en la Tabla 10 que utilizar sólo las 3 medidas de los RP que obtienen una mayor IM con los HS entrega mejores resultados. Vemos que eliminar las 6 características de menor IM no solo no daña la clasificación, sino que la mejora. Dado que las diferencias entre ambos sistemas no son significativas, se opta por el más sencillo.

4.3.2.3. Pruebas sobre Potencias de Bandas de Frecuencia de EEG

Esta prueba fue realizada para ver el aporte de las potencias en tres bandas de frecuencia. Consistió en medir la IM entre cada una de las 3 características: P_{σ}^n , P_{δ}^n y P_{HF}^n (ver sección 7.1.2.3), con las marcas de HS realizadas por experto para el CV. Luego, se contrastó la clasificación en el CP de un sistema que, de estas 3 características, solo consideraba P_{σ}^n , versus un sistema que usaba las tres características. Las demás características consideradas en este trabajo, tales como las deducidas del RP o de la función *env*, se mantenían en ambos sistemas. El desempeño de la clasificación se determinó midiendo los FP obtenidos por los distintos métodos en el CP, considerando, en cada caso, un umbral para la salida de la red neuronal tal que se obtuvieran los mismos TP que la salida del módulo 2 (M2), sobre el mismo CP, del sistema de [6] y [7]. Los sistemas eran entrenados usando CE y CV. Éste último conjunto se utilizó para evitar sobre entrenamiento. Se consideró la distancia y los parámetros determinados en las secciones anteriores. En la Tabla 11 se muestran los resultados para la información mutua, y en la Tabla 12 se muestran los resultados para la clasificación.

Tabla 11: Comparación de Información Mutua entre las Potencias de Bandas de Frecuencia del EEG y Marcas de HS de Experto.

Medida	Información Mutua
P_{σ}^n	0,01930
P_{δ}^n	0,00004
P_{HF}^n	0,00000

Tabla 12: Comparación de Clasificaciones sobre CV entre Sistema que Usa las 3 Características Basadas en Potencias de Bandas de Frecuencia del EEG y un Sistema que Usa solo las con Mayor IM.

Cada una de las tres iteraciones (it.1, it 2. e it. 3) corresponden a un entrenamiento diferente de la NN. El sistema para el cual se obtienen los mejores resultados aparece destacado.

Sistema	TP It. 1	TP It. 2	TP It. 3	FP It. 1	FP It. 2	FP It. 3	Promedio FP
M2	11	11	11	48	48	48	48,0
Sistema Usando 1 Característica	11	11	11	23	18	17	19,3
Sistema Usando 3 Características	11	11	11	23	23	25	23,7

Se puede apreciar en la Tabla 12 que eliminar P_{δ}^n y P_{HF}^n no solo no daña la clasificación, sino que la mejora. Debido a que un sistema sin estas características es además más sencillo, el sistema final no utiliza P_{δ}^n ni P_{HF}^n .

4.3.2.4. Pruebas sobre Método de Elección de u_{peaks}

Esta prueba fue realizada para ver qué método convenía más para calcular u_{peaks} (Ver sección 7.1.2.3). Debido a que este umbral determina dos características del sistema (*anch* y *fb*), se midió la IM de ambas características con las marcas del HS realizadas por el experto en el CV. Luego, se contrastó la clasificación en CP de tres sistemas. Todos estos sistemas tenían los mismos valores para todas las características, salvo *anch* y *fb*. En cada sistema, estas características eran calculadas con una de las tres variantes para calcular u_{peaks} . El desempeño de la clasificación se determinó midiendo los FP obtenidos por los distintos métodos en el CP, considerando, en cada caso, un umbral para la salida de la red neuronal tal que se obtuvieran los mismos TP que la salida del módulo 2 (M2), sobre el mismo CP. Los sistemas eran entrenados usando CE y CV. Éste último conjunto se utilizó para evitar sobre entrenamiento. Se utilizó la distancia y los parámetros determinados en las secciones anteriores. En la Tabla 13 se

muestran los resultados para la IM, y en la Tabla 14 se muestran los resultados para la clasificación.

Tabla 13: Comparación de IM entre *anch* y Marcas de HS, y entre *fb* y Marcas de HS, para distintas formas de calcular u_{peaks}

La medida con mayor información mutua para *anch* aparece destacada.

Método	IM <i>anch</i>	IM <i>fb</i>
$u_{peaks-1} = \bar{pe} - 2 \cdot std_{pe}$	0,213	0,006
$u_{peaks-2} = \bar{pe} - std_{pe}$	0,146	0,012
$u_{peaks-3} = \min_i pe_i$	0,146	0,012

Tabla 14: Comparación de Clasificaciones sobre CV entre 3 Sistemas Usando Distintos Métodos de Cálculo de u_{peaks}

Cada una de las tres iteraciones (it.1, it. 2. e it. 3) corresponden a un entrenamiento diferente de la red neuronal. El sistema para el cual se obtienen los mejores resultados aparece destacado.

Método	TP It. 1	TP It. 2	TP It. 3	FP It. 1	FP It. 2	FP It. 3	Promedio FP
M2	11	11	11	48	48	48	48,0
$u_{peaks-1} = \bar{pe} - 2 \cdot std_{pe}$	11	11	11	21	22	23	22,0
$u_{peaks-2} = \bar{pe} - std_{pe}$	11	11	11	20	30	20	23,3
$u_{peaks-3} = \min_i pe_i$	11	11	11	23	27	18	22,7

Se puede apreciar en la Tabla 13 que el mayor valor para la IM entre *anch* y las marcas de HS del experto se da para $u_{peaks-1}$. En general, la IM medida entre *anch* y los HS es mucho mayor que la IM de *fb*, independiente de cómo se calcule u_{peaks} . Por ello, se da más prioridad a la IM con *anch* para decidir qué método de cálculo del umbral es mejor. Nótese que $u_{peaks-1}$ resulta ser el mejor algoritmo también cuando se considera el desempeño del clasificador, como se observa en la Tabla 14. Aunque en este caso la diferencia entre el desempeño de los sistemas calculados con $u_{peaks-1}$ y $u_{peaks-3}$ es baja, se opta por $u_{peaks-1}$ debido a que es también el mejor para el criterio de IM.

4.3.3. Pruebas sobre Parámetros

Idealmente, los parámetros deberían optimizarse en conjunto, ajustándose todos a la vez para buscar la mejor combinación de sus valores. Sin embargo, debido al gran número de parámetros libres en el sistema, una prueba que realizara dicha optimización habría sido prohibitivamente larga. Debido a esto, los

parámetros menos importantes del método fueron ajustados en forma preliminar y sobre conjuntos de entrenamiento y de validación pequeños. En estas pruebas preliminares, los parámetros se ajustaron en forma individual, dejando fijos los valores para los demás parámetros. Dichas pruebas se describen a continuación.

4.3.3.1. Pruebas sobre Parámetros de Distancia

Los parámetros del kernel temporal, p y σ_t , para $d_{correl-kp}$, y el parámetro σ_d de $d_{corren-n}$, fueron ajustados en forma independiente en el conjunto de validación mediante IM. Se comenzó con los valores encontrados en la parte 4.3.2.1 (es decir, $p = 0.08$ y $\sigma_t = 0.0085$) y se designaron otros valores en torno a ellos. Con esto, se obtuvo un conjunto de valores para cada parámetro. Para cada uno de estos valores se calcularon las tres características basadas en RPs (L_{max} , L_{medio} y L_{desv}) y se calculó la IM entre dichas características y las marcas del experto, para CV. Nótese que primero se varió σ_t , dejando p constante, y para el valor óptimo de σ_t , se varió p . En la Tabla 15 se muestran los resultados al variar σ_t , en la Tabla 16 se muestran los resultados al variar p . Se hizo un proceso análogo determinar el valor de σ_d para la distancia $d_{corren-n}$. Los resultados de este proceso se muestran en la Tabla 17.

Tabla 15: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

Los RPs se construyeron utilizando $d_{correl-kp}$, para distintos valores de σ_t . Se usó un valor de p constante e igual a 0.08. El parámetro para el cual se obtienen los mejores resultados aparece destacado.

σ_t	IM L_{max}	IM L_{medio}	IM L_{desv}
0,0025	0,147	0,091	0,109
0,0045	0,176	0,109	0,135
0,0065	0,180	0,106	0,140
0,0085	0,174	0,110	0,127
0,0105	0,000	0,078	0,054
0,0125	0,000	0,071	0,056
0,0145	0,000	0,079	0,063

Tabla 16: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

Los RPs se construyeron utilizando $d_{correl-kp}$, para distintos valores de p . Se usó un valor de σ_t constante e igual a 0,0065. El parámetro para el cual se obtienen los mejores resultados aparece destacado.

p	IM L_{max}	IM L_{medio}	IM L_{desv}
0,02	0,163	0,086	0,118
0,04	0,175	0,079	0,115
0,08	0,169	0,093	0,125
0,10	0,163	0,086	0,118
0,12	0,146	0,055	0,082

Tabla 17: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

RP se construyeron utilizando $d_{corren-n}$, para distintos valores de σ_d . El parámetro para el cual se obtienen los mejores resultados aparece destacado.

σ_d	IM L_{max}	IM L_{medio}	IM L_{desv}
0,35	0,0950	0,0470	0,0673
0,5	0,1511	0,0794	0,1018
0,7	0,1742	0,0746	0,1106
0,9	0,0115	0,0230	0,0051
1,4	0,0000	0,0000	0,0000

Se puede observar que los mejores resultados se obtienen para $p = 0,08$, $\sigma_t = 0,0065$ y $\sigma_d = 0,7$. Estos fueron los valores de estos parámetros en el sistema final. Nótese que $p = 0,08$ [s] corresponde a una frecuencia de 12,5 [Hz].

4.3.3.2. Pruebas sobre Parámetros de Método de Retraso de Tiempo

Los parámetros del método de retraso de tiempo, τ y m , fueron determinados mediante pruebas distintas. Se había determinado visualmente que $\tau = 20$ y $m = 4$ era un buen punto de partida para estos parámetros. Para ello, se fueron variaron estos parámetros, observándose para qué valores se obtenían RPs que marcaran más claramente los HS. Luego, se estudió la influencia de τ en la IM entre las características deducidas de los RPs (L_{max} , L_{desv} y L_{medio}) y las marcas de HS del experto, para el CV. Los RP se calcularon utilizando las dos distancias del sistema final, es decir, $d_{correl-kp}$ y $d_{corren-n}$. Una vez calculado el valor óptimo de τ , m fue determinado mediante el método de los falsos vecinos (ver [24]).

Tabla 18: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

RPs se construyeron utilizando $d_{corren-n}$ y $d_{correl-kp}$, considerando distintos valores para τ . Se utilizó un valor de m constante e igual a 4. El parámetro para el cual se obtienen los mejores resultados aparece destacado.

τ	IM L_{max} $d_{correl-kp}$	IM L_{medio} $d_{correl-kp}$	IM L_{desv} $d_{correl-kp}$	IM L_{max} $d_{corren-n}$	IM L_{medio} $d_{corren-n}$	IM L_{desv} $d_{corren-n}$
5	0,1504	0,0755	0,1065	0,1429	0,0750	0,0957
20	0,1686	0,0932	0,1249	0,1742	0,0746	0,1106
30	0,0916	0,0538	0,0720	0,1440	0,0627	0,0958
40	0,1001	0,0635	0,0825	0,1118	0,0282	0,0656

Se puede apreciar en la Tabla 18 que el valor óptimo para τ corresponde a 20. Una vez fijado este valor, se aplicó el método de los falsos vecinos para calcular m . Este método se debe aplicar sobre los datos usados para calcular cada RP, de tal forma de deducir el m óptimo para ese RP. Como el sistema calcula varios RPs, uno para cada ventana de los datos originales, lógicamente al aplicar el método sobre estos datos se obtienen varios valores para m . Se decidió encontrar la distribución de estos valores en el CP. Dicha distribución se muestra en la Figura 28.

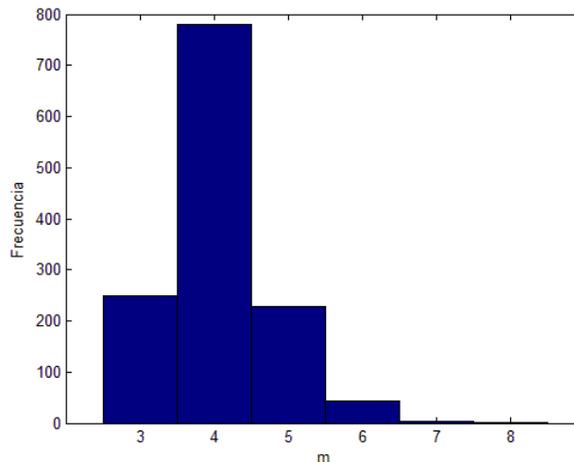


Figura 28: Distribución de m Según Método de Falsos Vecinos.
El valor de τ usado era 20. Los valores de m se calculan para cada ventana del CP.

Como se puede apreciar en la Figura 28, la media y la moda de m corresponden a 4. Ya que éste es el mejor valor en la mayoría de los casos, fue el utilizado en el sistema final.

4.3.3.3. Pruebas sobre Parámetros de Histograma de Largo de Líneas de RP

Los parámetros del histograma de largo de líneas diagonales del RP, L_{min} y tw , fueron ajustados en forma independiente en el conjunto de validación, mediante IM. Se comenzó con valores encontrados en forma visual, $tw = 7$ y $L_{min} = 4$, (corresponden a los valores para los que se obtuvieron los HS más claros en el RP) y se designaron otros valores en torno a ellos. Con esto, se obtuvo un conjunto de valores para cada parámetro. Para cada uno de estos valores se calcularon las tres características basadas en RPs (L_{max} , L_{medio} y L_{desv}) y se calculó la IM entre dichas características y las marcas del experto, para CV. Nótese que primero se varió tw , dejando L_{min} constante, y para el valor óptimo de tw , se varió L_{min} . En la Tabla 19 se muestran los resultados al variar tw , y en la Tabla 20 se muestran los resultados al variar L_{min} .

Tabla 19: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

RP's se construyeron utilizando $d_{corren-n}$ y $d_{correl-kp}$, considerando distintos valores para tw . Se utilizó un valor de L_{min} constante e igual a 4.

tw	IM L_{max} $d_{correl-kp}$	IM L_{medio} $d_{correl-kp}$	IM L_{desv} $d_{correl-kp}$	IM L_{max} $d_{corren-n}$	IM L_{medio} $d_{corren-n}$	IM L_{desv} $d_{corren-n}$
3	0,1686	0,0932	0,1249	0,1683	0,0717	0,1020
5	0,1686	0,0932	0,1249	0,1736	0,0748	0,1057
7	0,1686	0,0932	0,1249	0,1742	0,0746	0,1106
12	0,1686	0,0934	0,1242	0,1742	0,0731	0,1062
17	0,1599	0,0759	0,1041	0,1586	0,0683	0,0999

Tabla 20: IM entre Características Deducidas de RPs (L_{max} , L_{medio} y L_{desv}) y Marcas de HS de Experto.

RP's se construyeron utilizando $d_{corren-n}$ y $d_{correl-kp}$, considerando distintos valores para L_{min} . Se utilizó un valor de tw constante e igual a 7.

L_{min}	IM L_{max} $d_{correl-kp}$	IM L_{medio} $d_{correl-kp}$	IM L_{desv} $d_{correl-kp}$	IM L_{max} $d_{corren-n}$	IM L_{medio} $d_{corren-n}$	IM L_{desv} $d_{corren-n}$
2	0,1686	0,0748	0,1124	0,1742	0,0651	0,1011
4	0,1686	0,0932	0,1249	0,1742	0,0746	0,1106
8	0,1686	0,1456	0,1464	0,1742	0,1304	0,1259
12	0,1686	0,1541	0,1355	0,1742	0,1299	0,1283

Se puede observar que estos parámetros influyen muy poco en los resultados. Para el caso de tw , se mantuvo el valor original, es decir: $tw = 7$, debido a que la IM es casi igual para todos los valores. Lo mismo sucede para L_{min} , donde también se mantiene el valor original.

4.3.3.4. Pruebas sobre Parámetros de Entrenamiento

Los parámetros del entrenamiento de la red neuronal corresponden a: el número de veces que se repiten los positivos en el conjunto de entrenamiento, n_p , el número de neuronas en la capa oculta, N_o , la tasa de aprendizaje de la red neuronal, μ , y la tasa de momentum de la red neuronal, α . Dichos parámetros fueron ajustados heurísticamente, es decir, fueron modificándose hasta obtener una red que convergiera en un número aceptable de pasos para el CE. Los valores finales, obtenidos luego de realizar esta prueba, fueron: $n_p = 8$, $N_o = 20$, $\alpha = 0,2$ y $\mu = 0,05$. Nótese que es posible estos valores no correspondan a los parámetros óptimos, pero se decidió que, debido a que este no era el enfoque del trabajo, bastaba con que la red convergiera para obtener resultados satisfactorios. Esto, debido a que estos parámetros están enfocados hacia la convergencia de red neuronal, la cual ocurría, por lo que se concluyó que estaban bien seleccionados.

4.3.4. Análisis de Resultados

Los resultados obtenidos en esta prueba permitieron definir el algoritmo final que realiza el sistema implementado en este trabajo. A continuación se resumen y analizan los resultados de cada una de las partes de esta prueba preliminar.

Para la selección de la distancia, se puede apreciar que los mejores resultados se obtienen al considerar un sistema con características deducidas de dos distancias distintas, $d_{correl-kp}$ y $d_{corren-n}$. Se observa que en este caso el promedio de los FP a la salida del sistema es muy inferior comparado con las otras distancias. Pareciera ser que ambas medidas se complementan, ya que a pesar de que para ambas por separado se obtienen buenos resultados, cuando éstas se combinan en un solo sistema mejora enormemente el desempeño. Resulta una buena idea utilizar un kernel periódico (de periodo dentro de la banda sigma), ya que una de las distancias óptimas lo incluye. Se puede observar, además, que la IM entre los HS y tanto L_{desv} como L_{medio} , aumenta al considerar este kernel. Esto corrobora el hecho que al agregar un kernel periódico se resaltan los HS, haciendo que el sistema sea mejor para detectar ondas cuyas frecuencias se encuentren en la banda sigma. Parece ser también una buena idea tener una distancia que utilice un kernel periódico y otra que no, como sucedía para el par óptimo en este caso. Esto, de alguna manera permite consolidar dos enfoques distintos en un solo sistema, uno en donde se resaltan las frecuencias en la banda sigma, y otro donde no. Se debe notar que la única característica del RP cuya IM con respecto a las marcas del experto se maximiza para $d_{correl-kp}$ es L_{medio} . Para las otras dos características (L_{max} y L_{desv}), la IM es máxima para otra distancia. Asimismo, para ninguna de las tres características se cumple que su IM con respecto a las marcas del experto es máxima para $d_{corren-n}$. Sin embargo, es más importante el desempeño del sistema que la IM entre sus características y las

marcas de HS, ya que este desempeño es lo que se busca maximizar, y la IM no considera que dos características puedan aportar información redundante. Por ello, se da más prioridad a la salida del sistema que a la IM. Es importante notar que sería posible considerar sistemas que combinaran un mayor número de distancias distintas, en vez de solo una o dos, como en este caso. Sin embargo, se optó por limitar el número máximo a dos distancias, para mantener el sistema más simple y con menos características, y para no alargar demasiado esta prueba.

Para la selección de las características deducidas a partir del RP, resulta levemente mejor utilizar solamente las 3 mejores características derivadas de medidas sobre dicho gráfico. Las otras 6 medidas, las cuales en un principio se pensó que podían aportar a la detección, no entregan información adicional sobre la presencia de HS. Esto se observa tanto en la prueba en el cálculo de la información mutua entre dichas características y las marcas del experto, como en la prueba de clasificación, donde el sistema que utiliza las 9 características tiene un peor desempeño que aquel que solamente utiliza las 3 mejores. Notar que, a pesar que la información mutua de las 6 peores características es bastante baja, podría haber sido posible que estas logran aportar alguna información complementaria que las otras 3 características no entregaban. Sin embargo, esto no es el caso, pues al realizar la prueba sobre las detecciones del sistema éste no solo no mejora cuando se agregan dichas características, sino que empeora. La razón de por qué L_{max} , L_{medio} y L_{desv} resultan mejores que las otras 6 características se debe probablemente al enfoque de cada una de ellas. Se sabe que un RP, calculado a partir de una señal periódica, tiende a tener líneas diagonales claras, regulares y largas. Asimismo, debido a que un HS corresponde a un patrón bastante periódico, dentro de su duración, tiende a producir líneas con estas mismas características. Esto causa que los valores de L_{max} y L_{medio} sean mayores frente a la presencia de un HS. Mientras L_{max} mide la máxima línea presente, L_{medio} considera qué tan largas son en general todas las líneas. Así, mientras L_{max} determina qué tan grande es la recurrencia que abarca más estados del sistema, L_{medio} mide el tamaño promedio de las recurrencias del este sistema. Similarmente, al haber varias líneas diagonales en los RP es común que estas sean de diferentes largos. Esto se debe, en parte, a que el RP considera una ventana de tiempo finito, por lo que las periodicidades de mayor periodo solo se dan para pocos estados. Se puede dar que muchos estados del RP tengan esa periodicidad con otros estados, pero estos otros estados caen fuera del RP. Visto de otra forma, las líneas diagonales más alejadas de la LOI no pueden ser tan largas como aquellas cercanas a la LOI, pues topan antes con los bordes del RP. Así, para un sistema periódico la varianza de los largos de las líneas diagonales será mayor, aumentando el valor de L_{desv} , en contraste con el caso en que no existen señales periódicas, en donde todas las líneas serán relativamente cortas, y por ende su varianza será baja. Las otras medidas no están relacionadas tan directamente con el largo y la regularidad de las líneas, por lo que no resultan adecuadas para detectar HS. Por ejemplo, el RR solo mide la densidad del RP, sin preocuparse de las estructuras que en él se forman. En general, se ve que estas 6

características aportan más ruido que información a la detección, y por ende son completamente prescindibles. Por ello se privilegia un sistema más simple, que no incluye dichas características y que tiene un mejor desempeño.

Para la prueba de selección de características basadas en la potencia de tres diferentes bandas de frecuencia (P_{δ}^n , P_{σ}^n y P_{HF}^n), se aprecia que únicamente la potencia en la banda sigma es relevante en la clasificación. En primer lugar, se observan valores bajísimos para la IM entre los HS y tanto la potencia en la banda delta como la potencia de las altas frecuencias. Asimismo, el sistema que utiliza estas dos características obtiene resultados bastante peores sobre el CP que aquel que no las usa. Por ende, no conviene agregar estas dos características al sistema, pues ellas aportan más ruido que información a la clasificación. Esto implica una contradicción con lo mostrado en [6], en donde dichas potencias son usadas satisfactoriamente en el módulo 1 de la detección de HS. Una explicación podría ser que, debido a que los datos utilizados son bastante reducidos, éstos no son lo suficientemente representativos del problema en general, y que si se consideran más datos, dichas características pasen a ser relevantes. Sin embargo, haber proseguido por esta línea de investigación habría desviado demasiado el tema de este proyecto. Debido a que estas características no son el enfoque principal de esta memoria, se optó por confiar en los resultados obtenidos en esta parte, y eliminar estas dos características. Sin embargo, es importante destacar que ambas características son importantes en el módulo 1 del sistema mostrado en [6] para descartar trozos que no corresponden a HS. Quizás utilizarlas como características en un sistema que clasifique HS como no fue lo más adecuado, ya que podrían ser más útiles si se utilizaran para pre-procesar la señal y descartar trozos de registro antes de utilizar el clasificador, tal como lo hace el módulo 1.

En la prueba realizada para definir el método para determinar u_{peaks} , se observan mejores resultados para el caso en que $u_{peaks} = \bar{pe} - 2 \cdot std_{pe}$. Debido a que en este caso no existía ningún método previo para calcular este umbral, se debió determinar empíricamente. En general, la idea de este umbral de amplitud era que fuera superado por casi todos los HS marcados por el experto. Por ello, se consideró fijarlo como el mínimo de los pe_i . Sin embargo, el hecho que la IM para el caso en que el umbral se fija como $u_{peaks} = \bar{pe} - 2 \cdot std_{pe}$ sea mucho mayor que en los otros dos casos, demuestra que conviene elevar un poco este umbral, lo que implica no dejar pasar algunos HS, con tal de eliminar alzas en la amplitud de la señal que no asociadas con HS. De todos modos, el desempeño de los 3 sistemas es bastante similar, por lo que, cuál de estos tres métodos se elige no es fundamental para los resultados. Además, es posible que existan otras formas de calcular dicho valor que podrían resultar mejores y no que no fueron consideradas en este estudio. De todos modos, las tres formas utilizadas en esta parte son bastante adecuadas, y considerar otras formas de calcular u_{peaks} se escapa de los alcances de este proyecto.

Las pruebas para ajustar los parámetros menos importantes del sistema permitieron definir en forma sencilla y rápida estos parámetros. Es importante que dichos parámetros tengan valores adecuados, pero no es crucial para el desempeño del sistema que éstos sean exactamente el óptimo global. Así, los parámetros encontrados en esta parte corresponden a una solución suficientemente buena, sin ser necesariamente la mejor solución posible, la cual resultaría demasiado costosa de encontrar. Una razón de esto es que cuando se debían ajustar varios valores al mismo tiempo, resultaba menos costoso computacionalmente ir variando solamente uno de ellos, en forma sucesiva, y dejando los demás parámetros constantes. La alternativa a esto, que consiste en optimizar en el espacio multidimensional de todos los parámetros, es exponencialmente más costosa, y no resulta práctica para parámetros de menor importancia. Asimismo, tampoco es necesario ajustar estos parámetros en una base de datos demasiado grande. Esto también puede hacer que las pruebas demoren demasiado, y los resultados, aunque quizás sean mejores, no deberían tener un impacto significativo en el desempeño del sistema. Con la base de datos usada, aunque quizás no se encuentre el mejor valor para los parámetros, se encuentran valores suficientemente buenos. Un resultado interesante de estas pruebas es que el valor del parámetro p del kernel periódico para el cual se obtienen los mejores resultados corresponde a la frecuencia de 12.5 [Hz]. Dicha frecuencia, se encuentra cercana a la mitad de la banda sigma. Así, es posible concluir que el kernel periódico funciona como se espera, resaltando aquellos elementos de la señal cuyas frecuencias son cercanas a 12.5 [Hz] (es decir, frecuencias dentro de la banda sigma). Lógicamente, esto es bueno para la clasificación, ya que implica que los HS son resaltados.

En las pruebas anteriores se emplearon dos criterios para seleccionar las mejores variantes del sistema, y sus parámetros. Ambos criterios, aquel basado en IM y aquel basado en el desempeño del sistema sobre el CV, tenían ventajas y desventajas, las cuales justificaron el uso de ambos métodos. El método basado en IM permitía una evaluación rápida y confiable del aporte de las variantes, características o valores de los parámetros. Sin embargo, no consideraba la interacción entre las variables. Así, por ejemplo, es posible que dos características tengan cada poca IM con las marcas del experto pero que en conjunto sean un gran aporte a la detección de HS. Por otro lado, el criterio basado en el desempeño sí considera esta interacción y está más directamente relacionado con el problema que se busca resolver. Sin embargo, al estar basado en una NN, sus resultados son aleatorios, ya que la utilización de redes neuronales considera la inicialización aleatoria de sus pesos. Así, los resultados varían de una iteración a la otra, haciéndolos menos confiables. Debido a esto, se realizan 3 entrenamientos distintos para cada variante, y se promedian los resultados obtenidos en ellos. Además, cada uno de estos entrenamientos tiene un costo computacional considerable. Por lo tanto, resulta adecuado utilizar ambos métodos en forma complementaria. Así, en los casos en que ambos coincidan se obtienen resultados bastante confiables, y cuando no coinciden se da prioridad a la medida basada en el desempeño, ya que considera la interacción entre las características y está más ligada al problema a resolver.

Se debe tener presente que todos los resultados obtenidos para esta parte dependen del conjunto de entrenamiento y validación utilizados. Dicho conjuntos eran relativamente pequeños, ya que utilizar conjuntos más grandes habría demorado demasiado tiempo. Debido a que los ajustes realizados en esta parte no eran fundamentales para el desempeño del sistema, no se consideró necesario utilizar conjuntos mayores. De todos modos, esta parte establece un procedimiento para ajustar el sistema a datos diferentes. Podría ser válido en trabajos futuros considerar parámetros o variables descartadas en este trabajo, si es que se ajusta el sistema en otro conjunto, en donde dichas variables resulten superiores.

4.4. Prueba de Ajuste de Parámetros

Como se explicó con anterioridad, hay 6 parámetros del sistema que fueron considerados principales, y como tales fueron ajustados en una misma prueba que consideraba varios valores para ellos, combinados de distinta forma. Estos parámetros son: el traslape de las ventanas usadas para construir los RPs, tr ; el ancho de dichas ventanas, a ; los umbrales usados para la binarización de los RPs, ε_l y ε_n ; y las frecuencias de corte del filtro usado en el pre-procesamiento de la señal, f_{ch} y f_{cl} . A continuación se explica en detalle la forma en que fue llevada a cabo esta prueba, y posteriormente se muestran los resultados que se obtuvieron tras realizarla.

4.4.1. Metodología

Para cada parámetro principal, se determinó un conjunto de valores para aquel parámetro. En cada uno de estos conjuntos se incluyó el valor que había sido usado en las pruebas preliminares para cada uno de ellos. Estos valores son: $tr = 70\%$, $a = 200$, $\varepsilon_l = 40$, $\varepsilon_n = 0,2$, $f_{ch} = 20$ y $f_{cl} = 8$. Tras definir estos conjuntos, se realizó el producto cruz entre todos ellos, de tal forma de obtener todas las combinaciones posibles para los diferentes parámetros. Para cada una de estas combinaciones, se entrenó el sistema y se midió su desempeño, de tal forma de determinar la mejor combinación. En la Tabla 21 se muestran todas las combinaciones usadas en esta prueba.

Tabla 21: Valores de Parámetros Principales del Sistema para Cada Combinación de Parámetros Usada.

Cada una de las 24 combinaciones ha sido designada con un número identificador.

Identificador	ε_l	ε_n	a	tr	f_{cl}	f_{ch}
1	40	0.20	200	70%	8	20
2	50	0.25	200	70%	8	20
3	40	0.20	300	70%	8	20
4	50	0.25	300	70%	8	20
5	40	0.20	400	70%	8	20
6	50	0.25	400	70%	8	20
7	40	0.20	200	80%	8	20
8	50	0.25	200	80%	8	20
9	40	0.20	300	80%	8	20
10	50	0.25	300	80%	8	20
11	40	0.20	400	80%	8	20
12	50	0.25	400	80%	8	20
13	40	0.20	200	70%	6	25
14	50	0.25	200	70%	6	25
15	40	0.20	300	70%	6	25
16	50	0.25	300	70%	6	25
17	40	0.20	400	70%	6	25
18	50	0.25	400	70%	6	25
19	40	0.20	200	80%	6	25
20	50	0.25	200	80%	6	25
21	40	0.20	300	80%	6	25
22	50	0.25	300	80%	6	25
23	40	0.20	400	80%	6	25
24	50	0.25	400	80%	6	25

En esta prueba se utilizaron otros conjuntos de entrenamiento, prueba y validación, diferentes a los usados en las pruebas preliminares. Se tomó un registro mayor al utilizado en dichas pruebas, y se dividió en 5 subconjuntos de aproximadamente el mismo tamaño, llamados: C_1, C_2, C_3, C_4 y C_5 . Luego, se realizaban pruebas, en donde uno de estos subconjuntos era usado como conjunto de validación, y los otros 4 era usado como conjunto de entrenamiento. Esta técnica es conocida como validación cruzada. Debido a que existían 5 subconjuntos, se podían realizar 5 pruebas distintas. En la Tabla 22 se muestra cómo se escogieron estos 5 conjuntos.

Tabla 22: Inicio y Fin de Trozo de Tiempo para los 5 Subconjunto Usados en Validación Cruzada.
SQ2 corresponde a la fase 2 de sueño quieto, SQ a sueño quieto en general, SP a sueño paradójico y VI a vigilia. Cada subconjunto aparece numerado del 1 al 5.

Paciente	Registro	Inicio C_1 [s]	Fin C_1 [s]	Inicio C_2 [s]	Fin C_2 [s]	Inicio C_3 [s]	Fin C_3 [s]	Inicio C_4 [s]	Fin C_4 [s]	Inicio C_5 [s]	Fin C_5 [s]
1	SQ2	0	54	55	108	109	162	163	216	217	270
1	SQ3	0	90	91	180	181	270	271	360	361	450
1	VI	0	144	145	288	289	432	433	576	577	720
1	SP	0	54	55	108	109	162	163	216	217	270
2	SQ2	0	450	451	900	901	1350	1351	1800	1801	2250
2	SQ3	0	330	331	660	661	990	991	1320	1321	1650
2	VI	0	220	221	440	441	660	661	880	881	1100
2	SP	0	135	136	270	271	405	406	540	541	675

Para cada una de las 5 pruebas, realizadas para cada combinación de parámetros, se midió el desempeño en el CV contabilizando los FP. Para ello, el umbral a la salida de la red neuronal, u_h , fue elegido de tal forma de que los TP del sistema fueran iguales o superiores a los TP de los módulos 1 y 2 (m2) de [6] y [7], obtenidos para el mismo CV.

4.4.2. Resultados

En la Tabla 23 se muestran los resultados de esta prueba. Notar que dicha tabla fue construida sumando los resultados de las 5 iteraciones realizada en esta prueba. Es decir, los resultados para cada una de las 5 formas de escoger el CV y el CE fueron sumados, mostrándose en dicha tabla los resultados de esa suma.

Tabla 23: TP, FP, TPR y FDR Obtenidos para Prueba de Ajuste de Parámetros.
Los resultados mostrados en esta tabla corresponde a los eventos sumados para las 5 formas de seleccionar el conjunto de validación y el de prueba. Además de los 24 sistemas obtenidos para las distintas combinaciones de parámetros, se muestran los resultados para la salida del módulo 2 de [6], bajo el nombre de m2.

Sistema	FP	TP	TPR	FDR
1	656	483	76%	58%
2	596	480	75%	55%
3	714	478	75%	60%
4	706	478	75%	60%
5	599	478	75%	56%
6	556	478	75%	54%
7	701	481	76%	59%
8	567	481	76%	54%
9	716	480	75%	60%
10	589	483	76%	55%
11	758	479	75%	61%
12	645	481	76%	57%
13	674	484	76%	58%
14	635	481	76%	57%
15	786	481	76%	62%
16	794	480	75%	62%
17	651	483	76%	57%
18	547	484	76%	53%
19	764	482	76%	61%
20	563	481	76%	54%
21	729	483	76%	60%
22	566	483	76%	54%
23	791	481	76%	62%
24	647	482	76%	57%
m2	1082	478	75%	69%

Se observa de los resultados obtenidos que la mejor combinación de parámetros, si se considera la tasa de FP como el criterio de selección, corresponde a la asociada al número 18, la cual está determinada por los siguiente parámetros: $tr = 70\%$, $a = 400$, $\varepsilon_l = 50$, $\varepsilon_n = 0.25$ $f_{ch} = 25$ y $f_{cl} = 6$. Esta fue la combinación seleccionada para la prueba final.

4.4.3. Análisis de Resultados

Tras realizar esta prueba se encuentran y definen satisfactoriamente los mejores valores para los parámetros principales del sistema. Se observa que en general los resultados obtenidos para el sistema con diferentes parámetros resultan mejores que los obtenidos en la salida del módulo 2, para este subconjunto de datos. Sin embargo, se debe tener presente que en esta prueba, debido a la validación cruzada, el conjunto de entrenamiento, de validación y de prueba no quedan completamente separados. Así, es preferible dejar la comparación de ambos métodos para la prueba siguiente.

Debido a que los parámetros considerados en esta prueba eran considerados importantes para el sistema, sus valores fueron ajustados en forma global, es decir, definiendo un conjunto de valores posibles para cada parámetro, y realizando el producto cruz de dichos conjuntos. Esto causaba que la cantidad de combinaciones distintas para los parámetros creciera exponencialmente con el número de ellos. Por esto, fue necesario mantener la cardinalidad de cada conjunto de posibles valores para los parámetros relativamente pequeña. Existen formas alternativas de realizar esto, considerando métodos de computación evolutiva como optimización por enjambre de partículas o algoritmos genéticos, que corresponden a formas, de resolver problemas como éste, más eficientes que considerando todas las combinaciones posibles. Sin embargo, esto habría aumentado enormemente la complejidad de esta prueba, y habría salido completamente de los alcances de este trabajo. Así, los parámetros encontrados en esta parte corresponden a los mejores valores dentro de los considerados, y la única forma de encontrar mejores valores sería aumentar significativamente la complejidad del método de búsqueda o la duración de esta prueba.

Es importante notar que los resultados obtenidos para esta parte poseen un cierto grado de aleatoriedad, ya que requieren del entrenamiento de NN, cuyos pesos se inicializan al azar. Por lo tanto, es posible que si se volviera a realizar esta prueba se obtendrían otros resultados. De todas maneras, dicha aleatoriedad queda reducida con la validación cruzada, ya que se deben entrenar 5 redes neuronales (una para cada CE definido por validación cruzada) para cada combinación de parámetros. Así, es muy probable de que los parámetros seleccionados en esta prueba están dentro de las mejores combinaciones.

Se observa en esta prueba que los mejores resultados se obtienen para una ventana de mayor ancho. Esto se debe probablemente a que un RP de mayor ancho es capaz de capturar de mejor manera la información necesaria discriminar entre la presencia y ausencia de HS. Nótese que una ventana de ancho 400 corresponde a dos segundos (debido a la tasa de muestreo de 200 [Hz]). Dicha duración es la duración máxima de un HS. Por lo tanto, una ventana de ancho 400 o mayor asegura que todos los HS se vean representados, en su totalidad, en al menos un RP. Esto justifica el hecho de que los mejores resultados se obtengan para este ancho. Otro hecho destacable de los resultados de esta prueba es que

conviene fijar las frecuencias de corte, del filtro pasa-banda aplicado sobre el EEG, en los valores tales que dicha banda sea lo más grande posible. Esto reafirma la hipótesis de que no conviene filtrar el EEG exactamente en la banda sigma, si no que la clasificación mejora al expandir ligeramente dicha banda. Esto se debe probablemente al hecho de que si se filtra únicamente en la banda sigma, otros patrones tenderán a parecer HS aunque no lo sean. Así, al ampliar los límites del filtro pasa-bandas, se agrega una mayor capacidad de discriminación al sistema.

4.5. Prueba Final

Esta prueba fue realizada para comparar el desempeño del sistema desarrollado en este trabajo con el desempeño de la salida del módulo 2 de [6] sobre un conjunto menor de datos que el usado en [7]. Corresponde, por lo tanto, a la prueba final de este trabajo, y a la prueba más importante realizada, ya que es la que más se relaciona con los objetivos de esta memoria.

4.5.1. Metodología

Para esta prueba se escogieron 3 subconjuntos de todos los datos disponibles, y se dividieron en CE, CV y CP. Parte de CE son todos los datos usados en las pruebas anteriores, es decir, todos los que se usaron como CE, CV y CP en dichas pruebas. Debido a esto, en CV y CP solo aparecen datos nuevos, que nunca fueron usados en ninguna de las pruebas anteriores. En la Tabla 24 se muestra la forma en que se dividieron los registros de los seis pacientes para formar estos tres conjuntos.

Tabla 24: Inicio y Fin de Trozo de Tiempo de cada Registro Usado para los Conjuntos de Entrenamiento, Validación y Prueba. El guión indica que dicho registro no fue usado para conformar el conjunto respectivo. Los límites de inicio y fin son inclusivos

Paciente	Registro	Inicio CE [s]	Fin CE [s]	Inicio CV [s]	Fin CV [s]	Inicio CP [s]	Fin CP [s]
1	SQ2	0	270	271	292	-	-
1	SQ3	0	450	451	487	-	-
1	VI	0	270	271	292	-	-
1	SP	0	720	721	779	-	-
2	SQ2	0	2250	2251	2347	-	-
2	SQ3	0	1650	1651	1702	-	-
2	VI	0	1100	1101	1124	-	-
2	SP	0	675	676	700	-	-
3	SP	0	320	321	398	399	638
3	SQ1	0	89	90	110	111	179
3	SQ2	0	320	321	398	399	638
3	SQ3	0	179	180	221	222	359
3	SQ4	0	320	321	398	399	638
3	VI	0	320	321	398	399	638
4	SP	0	320	321	398	399	638
4	SQ1	0	89	90	110	111	179
4	SQ2	0	320	321	398	399	638
4	SQ3	0	179	180	221	222	359
4	SQ4	0	320	321	398	399	638
4	VI	0	320	321	398	399	638
5	SP	-	-	-	-	0	359
5	SQ2	-	-	-	-	0	638
5	SQ3	-	-	-	-	0	638
5	SQ4	-	-	-	-	0	638
5	VI	-	-	-	-	0	638
6	SP	-	-	-	-	0	638
6	SQ1	-	-	-	-	0	359
6	SQ3	-	-	-	-	0	359
6	SQ4	-	-	-	-	0	179

Esta prueba consistió en entrenar una red neuronal a partir del CE, utilizando el CV durante el entrenamiento para impedir el sobreajuste de la NN al CE. Los valores de los parámetros del sistema usados para esta prueba eran los

valores óptimos encontrados en la prueba anterior, es decir: $tr = 70\%$, $a = 400$, $\varepsilon_l = 50$, $\varepsilon_n = 0.25$ $f_{ch} = 25$ y $f_{cl} = 6$.

Una vez entrenado el sistema, se determinó que los TP de la salida del módulo 2 de [6] y [7] en el CE eran iguales a 86. Se fue variando el valor de u_h en la salida de la NN hasta que los TP del sistema sobre el CE fueran iguales o superiores a este valor. Con esto, se determinó el valor de u_h para utilizar en el CP. Posteriormente, el sistema ya entrenado fue utilizado en el CP, dejando u_h fijo y midiéndose los TP y los FP de su salida. Las detecciones de ambos sistemas fueron graficadas para poder ser comparadas. Luego, se construyó una curva ROC para la clasificación del sistema sobre CP variando el umbral u_h . Esto se realizó únicamente para construir la curva ROC, y no para ajustar u_h , pues este umbral ya había sido fijado en el CE.

4.5.2. Resultados

4.5.2.1. Resultados sobre CP

Para los resultados mostrados a continuación, el umbral u_h fue ajustado en el CE, de tal manera que los TP en ese conjunto fueran iguales o superiores a los TP obtenidos para la salida del módulo 2 de [6], para los mismos segmentos. En este caso, se tenía que $u_h = -0.225$. Luego, los desempeños de ambos sistemas fueron calculados para el CP. En la Tabla 25 se muestra la comparación de ambos desempeños, basada en las principales medidas de desempeño consideradas en este informe, es decir, FDR y el TPR.

Tabla 25: Comparación de TP, FP, FDR y TPR entre M2 y el Clasificador Basado en NN con $u_h = -0.225$.

En esta tabla, M2 corresponde a la salida del 2 del sistema implementado en [6] y el clasificador con NN corresponde al método propuesto en este trabajo. El CP tenía 105 HS positivos marcados por el experto.

Sistema	TP	FP	FDR	TPR
M2	86	362	85,7%	79,5%
Clasificador NN	90	350	81,9%	80,8%

En la Figura 29, Figura 30, Figura 31 y Figura 32, se muestran gráficamente ejemplos de marcas de HS del experto y de ambos sistemas, para el CP.

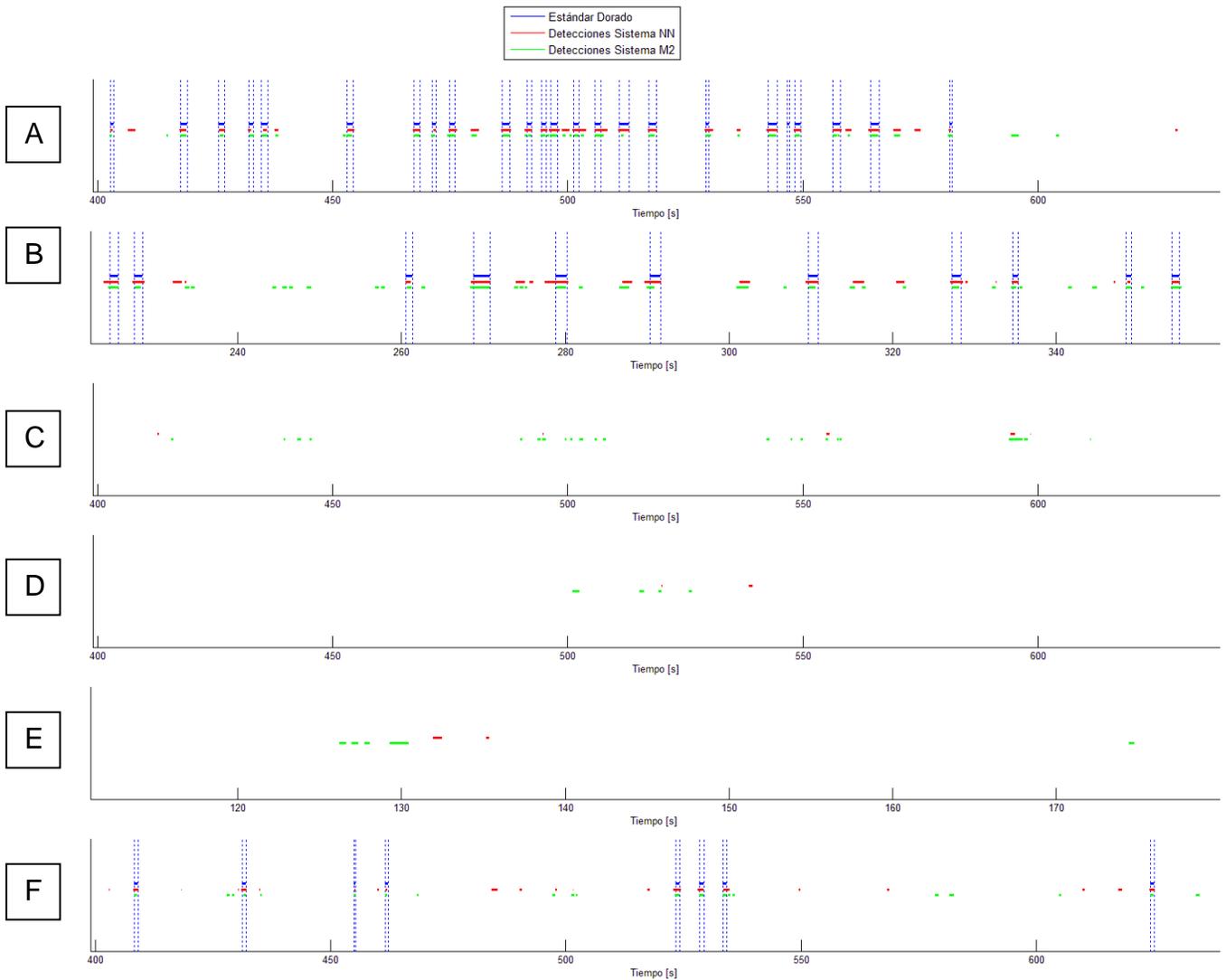


Figura 29: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $u_h = -0.225$ y Salida de módulo 2, para Paciente 4.
A: Registro perteneciente a SQ2. B: Registro perteneciente a SQ3. C: Registro perteneciente a SP. D: Registro perteneciente a VI. E: Registro perteneciente a SQ1.
F: Registro perteneciente a SQ4. Las marcas del clasificador basado en una red neuronal aparecen en rojo, las marcas a la salida del módulo 2 aparecen en verde, y el estándar dorado aparece en azul.

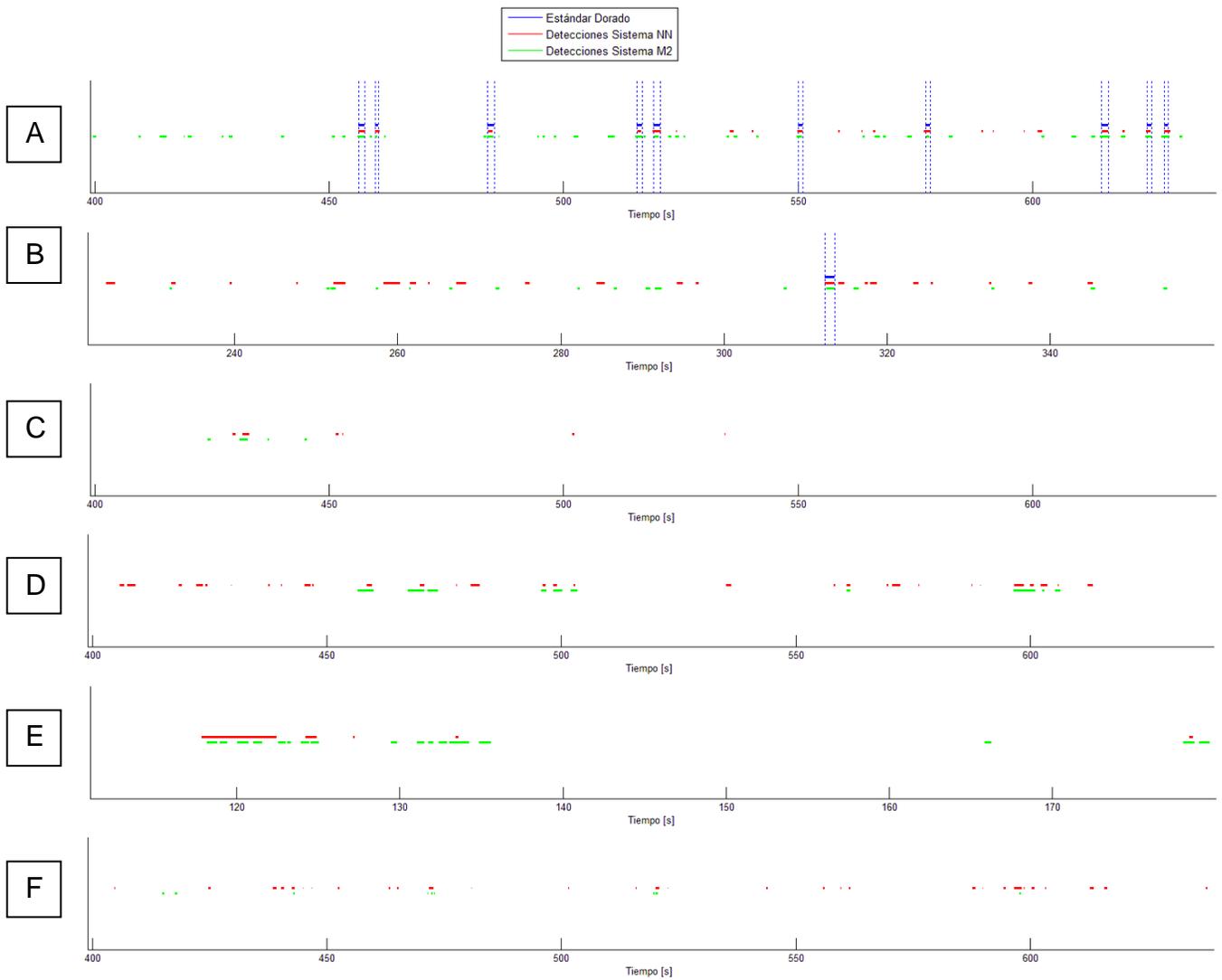


Figura 30: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $u_h = -0.225$ y Salida de módulo 2, para Paciente 5.

A: Registro perteneciente a SQ2. B: Registro perteneciente a SQ3. C: Registro perteneciente a SP. D: Registro perteneciente a VI. E: Registro perteneciente a SQ1.

F: Registro perteneciente a SQ4. Las marcas del clasificador basado en una red neuronal aparecen en rojo, las marcas a la salida del módulo 2 aparecen en verde, y el estándar dorado aparece en azul.

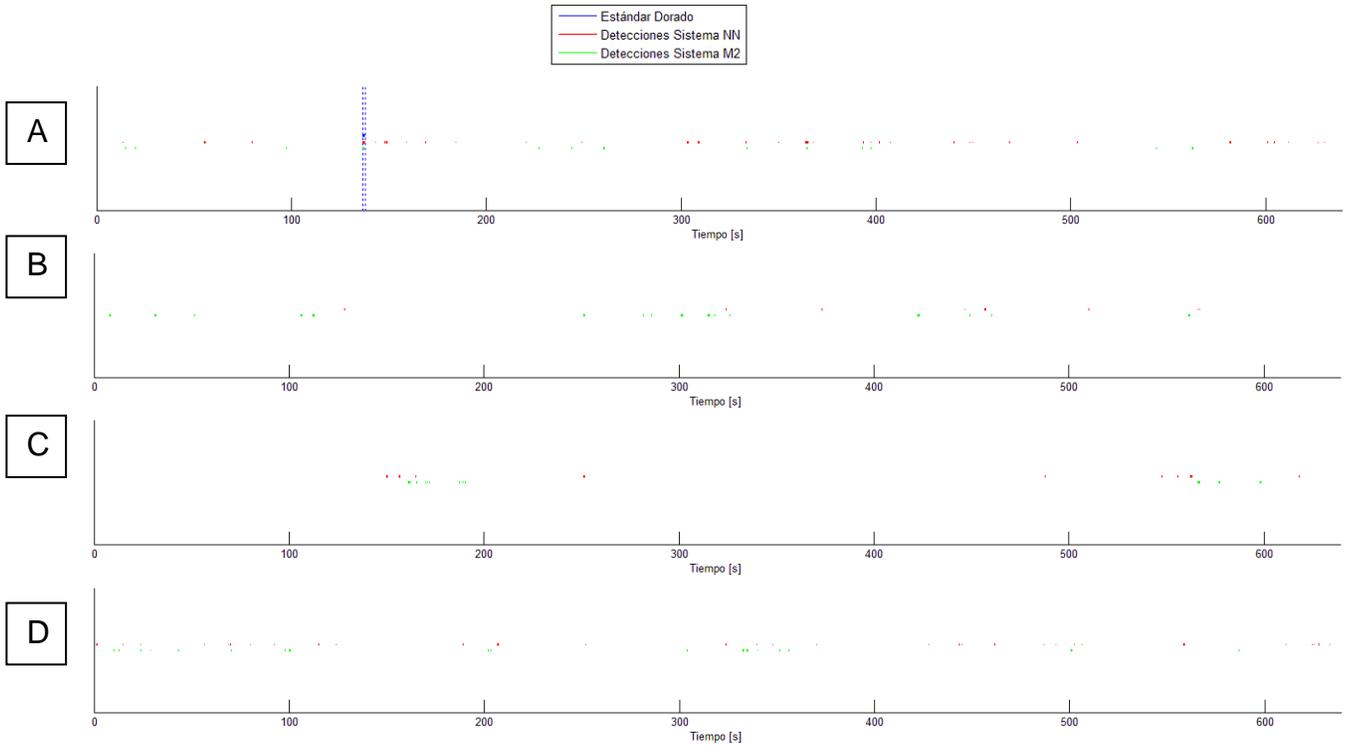


Figura 31: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $u_h = -0.225$ y Salida de módulo 2, para Paciente 6. A: Registro perteneciente a SQ3. B: Registro perteneciente a SP. C: Registro perteneciente a VI. D: Registro perteneciente a SQ4. Las marcas del clasificador basado en una red neuronal aparecen en rojo, las marcas a la salida del módulo 2 aparecen en verde, y el estándar dorado aparece en azul.

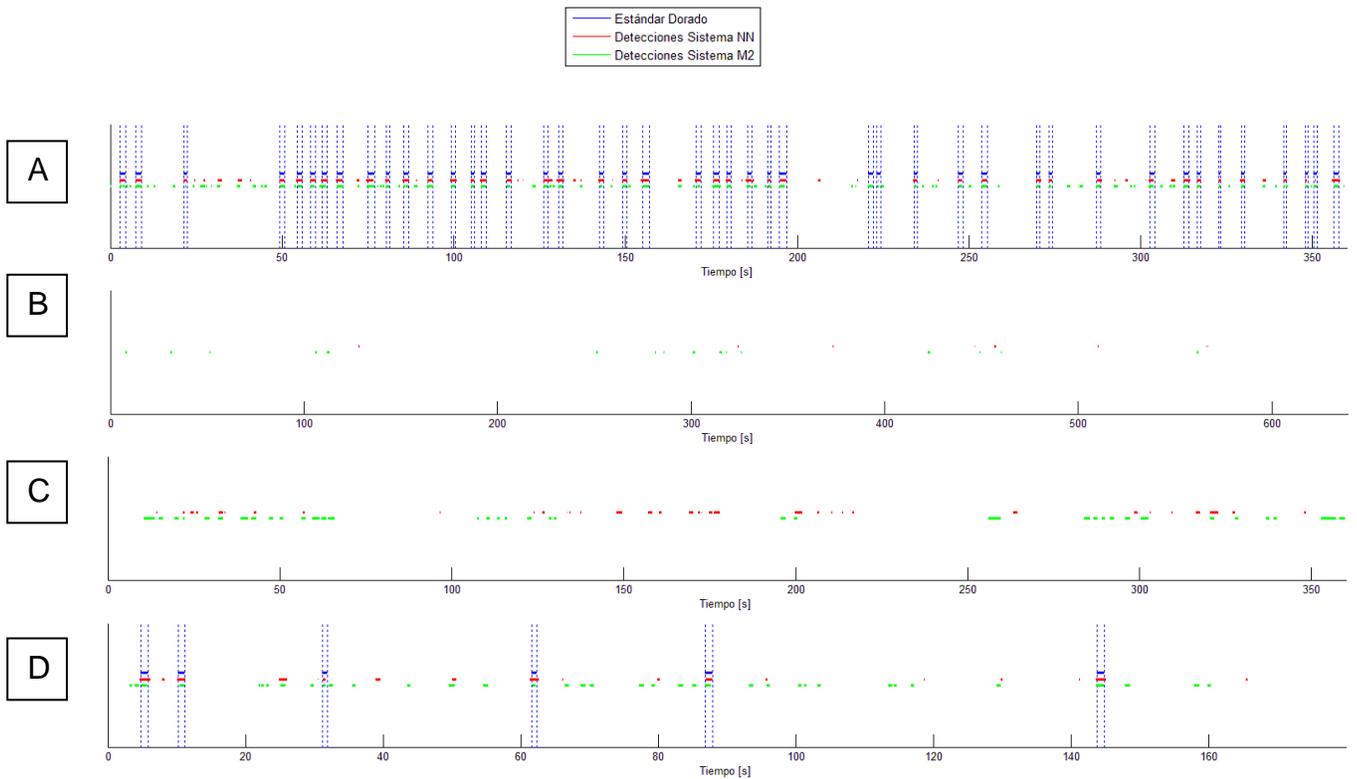


Figura 32: Comparación de Zonas Marcadas como HS por Clasificador basado en una Red Neuronal con $u_h = -0.225$ y Salida de módulo 2, para Paciente 7.

A: Registro perteneciente a SQ3. B: Registro perteneciente a SP. C: Registro perteneciente a SQ1. D: Registro perteneciente a SQ4. Las marcas del clasificador basado en una red neuronal aparecen en rojo, las marcas a la salida del módulo 2 aparecen en verde, y el estándar dorado aparece en azul.

4.5.2.2. Curva ROC

En esta sección se muestra la curva ROC asociada al clasificador NN, obtenida variando el umbral u_h y dejando el resto de los parámetros del sistema constante. Ya que no era posible medir la FPR, se muestra una curva ROC modificada en donde aparecen los VP en función de los FP. Para la salida del módulo 2 del sistema desarrollado en [6] no fue posible graficar una curva ROC, ya que no poseía el código de este método, sino que simplemente sus salidas, por lo que era imposible variar algún parámetro para obtener una curva ROC. Por ello, este sistema aparece simplemente como un punto en el plano VP-FP. En la Figura 33 se muestra esta curva ROC.

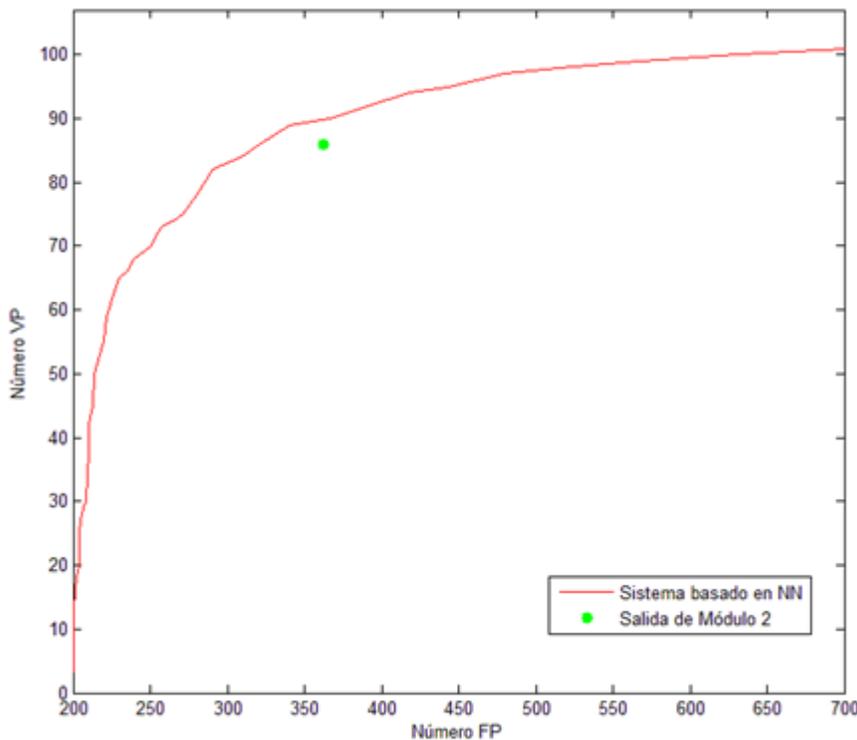


Figura 33: Curva ROC para Sistema Basado en Redes Neuronales sobre CP. La curva para el sistema aparece en color rojo, mientras que el punto verde representa el desempeño de los módulos 1 y 2 del sistema mostrado en [6] sobre el CP. El CP tenía 105 HS marcados por el experto.

4.5.3. Análisis de Resultados

Los resultados de esta prueba corresponden a la parte más importante de este trabajo, ya que dicha prueba está relacionada los objetivos más importantes de esta memoria. A continuación se analizan en forma detallada los resultados obtenidos para la prueba final de este informe.

La primera prueba realizada en esta parte implicaba clasificar los elementos del CP utilizando parámetros ajustados en el CE (incluido u_n). La razón por la que esta prueba se hizo así fue para simular el caso real de la aplicación del método, en donde es necesario ajustar todos los parámetros utilizando los datos originales que se tienen (los cuales corresponderían a CV y CE), ya que los datos donde se aplicará el método son desconocidos y no tienen asociados un estándar dorado. Se observa que el clasificador implementado en este trabajo obtiene una menor FDR con una TPR ligeramente mayor. Así, los resultados obtenidos para el sistema en el CP son mejores que los obtenidos a la salida de M2. Sin embargo, los desempeños son bastante similares, y no permiten asegurar que el sistema implementado en este trabajo sea significativamente mejor que los módulos 1 y 2 del sistema original. Es más, nada asegura que esta relación se mantenga para otra base de datos, pudiendo perfectamente la salida de M2 superar a la salida del clasificador implementado en otro CP. Lo que sí se puede asegurar es que ambos métodos son comparables.

Se debe tener presente que en [6] la TPR para la salida del módulo 2 es de aproximadamente 98%, mientras que la FDR es de aproximadamente 90%. Esto difiere en cierta medida con los resultados obtenidos para esta prueba, en donde para el M2 se tiene que una TPR de valor 80,8% y una FDR de valor 81,9%. Esto se debe a que los conjuntos utilizados en ambos casos difieren enormemente, pues para las pruebas realizadas en [6] se utilizaron 56 pacientes, mientras que en esta prueba se utilizan solo 4 trozos de registro. Esto hace que el entrenamiento de este método se realice con menos pacientes. Sería interesante, ver si los resultados del método implementado seguirían siendo comparables a los resultados obtenidos por M2 si es que se realizara una prueba que incluya los 56 pacientes.

Al observar los gráficos de las marcas de los HS obtenidos (Figura 29, Figura 30, Figura 31 y Figura 32) se observa que las detecciones de ambos sistemas difieren ligeramente. El sistema basado en NN tiende a sobrestimar los HS, lo que no sucede con la salida de M2. Esto puede deberse a que, debido a que la señal de EEG se divide en ventanas, a partir de las cuales se determinan vectores de características asociados a dicha ventana, la resolución del método implementado no es tan fina como la salida de M2. Esta resolución se podría mejorar si se consideran otro tamaño de las ventanas, otros traslapes, o ventanas con multiresolución. Sin embargo, se consideró que el estudio de dicha resolución se alejaba demasiado de los objetivos de esta memoria. Así, se puede ver que aunque para las tasas de TP son similares, la calidad de la detección de los HS es ligeramente mejor para la salida de M2, ya que es más precisa en el inicio y fin de estos eventos. Se observa visualmente que ambos sistemas presentan un número considerable de FP en casi todos los segmentos. Se debe tener presente que M2 no busca reducir los FP, si no que tener un TP lo más grande posible. En los módulos 3 y 4 del método implementado en [7] la mayoría de los FPs son reducidos. En este sentido, es más importante la comparación de la TPR de ambos sistemas, que resulta muy similar, a la comparación de la FDR. De todos

modos, es un aporte que el sistema mostrado sea capaz de reducir la FDR. Nótese que los FP de ambos sistemas coinciden en pocas oportunidades, mientras que los TP coinciden en la mayoría de las oportunidades. Además, ambos sistemas detectan la mayoría de los HS, pero no siempre son capaces de predecir correctamente el largo de dicho de eventos, lo que explica que la TPR no sea perfecta. Sería interesante probar con un nuevo método que combine ambos sistemas. Probablemente en este caso se obtengan mejores resultados. Si bien el método implementado por sí solo no presenta resultados significativamente mejores a la salida del módulo 2, si se combina con dicho módulo, escogiéndose como candidatos todos aquellos trozos marcados como HS por ambos sistemas, se pueden reducir sustancialmente los FP. Esto ayudaría a aliviar la carga de los módulos 3 y 4 del sistema original, encargados de reducir estos eventos. La razón por la cual los FP no coinciden se debe posiblemente a que los sistemas consideran características distintas a la hora de realizar la detección.

Se debe tener presente que parte de los registros de los pacientes 3 y 4 fueron utilizados en el CE, mientras que para los pacientes 5 y 6 todos los registros fueron utilizados en el CP. En este sentido, es de esperar que los resultados para los pacientes 3 y 4 sean mejores que para los pacientes 5 y 6. Esto se puede apreciar en la Figura 31 y Figura 32, donde aparece un número mayor de FP que para los otros dos pacientes. Esto se debe a que siempre existen diferencias entre las señales de distintos pacientes Sin embargo, se observa un buen desempeño para los pacientes 5 y 6, lo que permite concluir, de todos modos, que el sistema es capaz de generalizar adecuadamente a nuevas muestras.

En la curva ROC se puede apreciar que el sistema implementado en este trabajo es ligeramente mejor, existiendo una zona de dicha curva en donde el desempeño de dicho sistema es estrictamente mejor al desempeño de M2 (es decir, en esta zona los sistemas tienen una mayor TPR y una TDR). De todos modos, esta diferencia es pequeña, y la curva ROC encontrada es bastante cercana al punto que representa el desempeño de M2. Es importante destacar que esto no implica que el método basado en NN sea mejor que M2, ya que la diferencia es demasiado pequeña para llegar a esa conclusión. Sí se puede afirmar que representa un buen método complementario para M2, siendo ambos bastante similares, al menos para el CP considerado. Sería interesante tener una curva ROC para M2, de tal forma de comparar ambas curvas, pero esto no fue posible.

Finalmente, es importante destacar que el desempeño de ambos métodos no es adecuado para un clasificador de HS por sí solo, ya que ambos presentan demasiados FP para que resulte práctico aplicarlos. Por ello, ambos métodos sirven como primeras etapas de un detector mayor, por ejemplo, se les pueden agregar los módulos 3 y 4 del sistema descrito en [6], para eliminar los FP.

5. Conclusiones

5.1. Método

Al desarrollar el método presentado en este trabajo existieron una serie de decisiones de implementación. Cuando se estas decisiones fueron tomadas no se sabía bien qué efecto tendrían en el sistema. Debido a ello, fue necesario se realizar algunas pruebas para comparar las distintas variantes del método.

Una de las variantes más importantes de este método es la distancia utilizada, o las distancias utilizadas, para calcular el o los RPs, ya que ella determina cómo serán estos gráficos recurrentes, y por lo tanto las características deducidas de ellos. De entre las opciones consideradas se escogieron las mejores distancias. De todos modos, siempre existirán otras distancias, y formas de combinarlas, que pueden ser consideradas en trabajos futuros. Asimismo, se podría diseñar alguna distancia especial que se ajuste a este problema en particular. En este sentido, introducir un kernel periódico a las distancias permitió que éstas estuvieran más focalizadas en la detección de HS. Esto causó que las características deducidas a partir de estas distancias aumentaran su IM con las detecciones del experto, en comparación con las mismas distancias pero sin kernel periódico. Se concluye, entonces, que utilizar un kernel periódico fue beneficioso para el algoritmo. De las dos distancias seleccionadas en el método final, solo una tenía un kernel periódico. Esto permite concluir que de alguna forma ambos métodos de calcular distancias (con y sin kernel periódico) se complementan, entregando una información que la otra no entrega, y viceversa. Esto se podría comprobar comparando 2 sistemas, tal que cada uno utilice solo una de estas distancias, y comparar el desempeño de ambos sistemas.

El ajuste de parámetros llevado a cabo en el desarrollo de este trabajo era necesario para obtener buenos resultados. Sin embargo, resultó difícil realizar una prueba que optimizara todos los parámetros en conjunto, debido a que era muy costoso, en términos de tiempo, entrenar cada uno de los sistemas derivados de las diferentes combinaciones de parámetros. Otro problema que surge del ajuste de parámetros, y que también aparece al comparar variantes del método, es que resulta muy difícil realizar una comparación objetiva entre dos sistemas distintos, es decir, ser capaz de concluir a partir de una prueba cuáles valores de los parámetros o variantes del método son mejores. Esto, debido a que los resultados dependen de los datos usados para CP y CE, y además son aleatorios a causa del entrenamiento de la red NN. Por ello, solo se puede concluir que los parámetros del sistema, y las variantes del método, han sido seleccionados en forma satisfactoria, proporcionando una solución suficientemente correcta para el problema, capaz de entregar los resultados que se buscaban, pero no necesariamente la mejor. De cualquier modo, existe la convicción de que mejoras

significativas del método no serán producto de un mejor ajuste de parámetros, sino que más bien de cambios en las características utilizadas como entrada a la NN. Es importante destacar, sin embargo, que siempre será necesario ajustar estos parámetros frente a nuevos datos.

Se concluye que utilizar redes neuronales en el problema de clasificación es un buen método. No solamente resultó ser un algoritmo muy fácil de aplicar y con pocos parámetros, sino que además se observó empíricamente que la NN aprendía bien de los ejemplos. Además, permitió que un sistema relativamente sencillo, con unas pocas características experimentales, realizara una clasificación similar a un sistema mucho más complejo, al menos para los datos considerados. Esto fue posible gracias a la integración de información que realiza la red neuronal. Otra ventaja de la red neuronal es que permitió construir una curva ROC fácilmente ajustando el valor de u_h . Esta curva ROC cual fue útil para comparar el método implementado en este trabajo con M2. Además, como produce salidas continuas, fue posible interpolar estas salidas para mejorar la resolución del método y extender las clasificaciones puntos donde no se calculó un vector de características. La gran desventaja de NN es que sus resultados son, en cierta medida, aleatorios, es decir, varían de una aplicación a la siguiente. Esto dificultó las comparaciones. Sin embargo, este problema se eliminó, en parte, al realizar varios entrenamientos de la red y promediar sus resultados.

Se concluye que la utilización de RP para la detección de HS aparece como un método interesante y válido. Permiten desarrollar un algoritmo bastante sencillo para detectar HS, comparable a otro método mucho más complejo, al menos para los datos utilizados. Las otras características desarrolladas en este trabajo, en especial aquellas relacionadas con la amplitud y la duración de los HS (*anc*, *env* y *fb*), también parecieran ser un aporte para de resolver el problema planteado. Habría sido interesante realizar pruebas donde se compare un sistema sin estas características versus uno que sí las utiliza. Sin embargo, por consideraciones de tiempo, no fue posible realizar dichas pruebas. Es necesario desarrollar y refinar estas características, tanto las deducidas de los RP, como las otras introducidas en este trabajo. Es muy probable que si se perfeccionan se obtenga un sistema mejor.

En conclusión, el método implementado resuelve en forma apropiada el problema planteado, utilizando técnicas modernas para resolverlo en forma bastante adecuada. Es comparable con otras técnicas recientes, y representa un aporte en esta línea de investigación.

5.2. Resultados

A pesar de que los resultados fueron analizados en profundidad en la sección 4 de este informe, se presenta a continuación las principales conclusiones deducidas de los resultados de las pruebas que se realizaron.

Para las pruebas realizadas y las bases de datos utilizadas, el sistema planteado resulta ser ligeramente mejor que M2. Sin embargo, al no ser significativamente mejor, no se puede asegurar que siempre sea preferible utilizar este sistema en vez de los módulos 1 y 2 del sistema original. Además, fue imposible medir el verdadero aporte del método para todo el sistema original, ya que no se dispone del código de sus módulos 3 y 4. Por lo tanto, es difícil cuantificar qué tanto mejor, o peor, sería un método que utilizara el sistema implementado en este trabajo en vez de los módulos 1 y 2. Al menos, en términos de TPR, ambos métodos son similares, capturando correctamente la mayoría de los HS. Es importante destacar que los módulos 3 y 4 descartan la mayoría de los FP, por lo que la TPR resulta más importante en este caso que la FDR. De todos modos, puede resultar útil eliminar FPs antes de estos módulos. En ese sentido, un método que mantenga la TPR de M2 y que reduzca significativamente los FP sería interesante. El método implementado no es capaz de hacer esto, pero podría ser el primer paso en el desarrollo de un método que lo logre. Por otro lado, pareciera ser que los errores que comete el método implementado no coinciden con los errores que comete M2. Por lo mismo, un sistema híbrido que realice las detecciones en paralelo, y luego las combine, podría ser significativamente mejor que M2 por sí solo. En este caso, el método desarrollado sí podría ser un aporte significativo para reducir los FP, pero sería necesario realizar pruebas concretas para comprobar esta hipótesis. Es importante destacar, de todos modos, que ni el sistema desarrollado en esta memoria ni los módulos 1 y 2 del sistema original, son adecuados para detectar, por sí solos, husos sigma. Esto debido a que poseen demasiados FP. En general, debido a la alta TPR que tienen, están diseñados para ser las primeras etapas dentro de un sistema más complejo.

El sistema implementado pareciera no requerir un hipnograma, pues funciona correctamente en todas las etapas en las que se probó (VI, SQ2, SQ1, SQ3, SQ4 y SP). A pesar de que ambos sistemas se comportan en forma similar, M2 es más preciso en el inicio y fin de los HS detectados. En este sentido, el hecho de que el problema haya sido planteado como uno de clasificación, y que se hayan utilizado RP, implica una peor resolución, ya que se deben tomar ventanas de cierto tamaño (400 muestras o 2 [s]) para obtener los vectores de característica y posteriormente clasificar estos vectores como un todo, lo que hace que se pierda cierta información local de la señal.

Se puede apreciar que, al igual que el sistema original, el sistema implementado en esta memoria es capaz de generalizar a nuevos pacientes. Esto es de suma importancia, pues no tiene sentido desarrollar un método de detección de HS que solo funcione bien solo para unos pocos individuos. Resulta

fundamental, si se quiere aplicar un sistema de este estilo en la práctica, que el sistema sea capaz de generalizar a nuevos pacientes una vez que ya ha sido entrenado.

5.3. Objetivos

La mayoría de los objetivos planteados para este trabajo fueron cumplidos en forma adecuada. A continuación, se explica, para cada uno de ellos, de qué manera fueron satisfechos.

El objetivo principal de este trabajo fue desarrollar un sistema mejor a M2 pero que utilizara técnicas de aprendizaje de máquinas combinadas con RP. Este objetivo no se cumplió completamente, pues no se puede asegurar que el método propuesto sea significativamente mejor a M2. Sin embargo se desarrolló un sistema de desempeño muy similar a M2, que utiliza RPs y otras mediciones, combinadas mediante una NN, para detectar HS. Dicho sistema resultó ser bastante simple y básico, lo que implica que todavía puede ser mejorado sustancialmente. Si bien el método propuesto no sirve para reemplazar a M2, es posible que pueda ayudar a mejorarlo si se combina con el sistema implementado en este trabajo, pues como se vio, los errores de ambos tienden a no coincidir. Así, es posible que un sistema híbrido represente una mejora significativa a M2.

Los RP demostraron ser un método bastante adecuado para este problema, por lo que el objetivo de estudiar el aporte de estos gráficos a la detección de HS fue cumplido. Los RP resultan ser una alternativa bastante simple y comparable en calidad con otros métodos existentes, particularmente M2. Entre sus ventajas se incluye el hecho de que son capaces de resaltar visualmente los HS. Como se observa en la Figura 36, un HS provoca un patrón bastante claro de líneas regulares en el RP. Al combinar esto con otras técnicas como NN se obtiene un sistema con buen desempeño. Una de sus desventajas es que requiere dividir la señal en ventanas para poder funcionar, lo que hace que la resolución del método no sea buena, ya que la información de cada ventana de 2 [s] se resume en un único vector de características. Además, no detectan muy bien la duración ni la amplitud de los HS, si no que están más enfocados en la frecuencia y morfología de los eventos. Debido a esto, deben ser complementados con otras características, como las que se plantearon en este trabajo. Se observó que utilizar distancias basadas en la correntropía y en un kernel periódico dio buenos resultados. En general, la correntropía representa una buena medida de similitud entre dos vectores, al igual que al correlación. A diferencias de otras distancias, están basadas en el ángulo entre dichos vectores, lo que podría explicar su éxito en este trabajo. Lógicamente, utilizar un kernel periódico también ayuda bastante a la detección, pues tiende a resaltar las frecuencias en la banda sigma.

Se demostró que utilizar redes neuronales para detectar HS es una alternativa válida, por lo que el objetivo relacionado con esta premisa también se completó. A

diferencia de otros trabajos, en éste se intentó crear una entrada para la red neuronal que representara holísticamente las características del HS, tanto en amplitud, como en frecuencia, duración y morfología. Otros métodos que utilizan redes neuronales solo se centran en algunas de ellas. Por ejemplo, en [54] se utiliza la señal directamente del EEG, y en [11] y [15] se utiliza la SFT, es decir características relacionadas con la frecuencia de los HS. Se puede concluir, entonces, que siguen existiendo nuevas formas originales de aplicar NN en el problema de detección de patrones de señales de sueño.

El objetivo planteado de desarrollar nuevas características para los HS, que midan la amplitud, frecuencia, duración y morfología de los HS, fue satisfecho. Las nuevas características introducidas son las que presentan mayor IM con respecto a los HS. En particular, las características más importantes para el desempeño del sistema son: L_{max} , L_{medio} , L_{desv} , $anch$ y env .

Todos los objetivos de menor importancia planteados en la introducción de este trabajo se cumplieron, ya que eran básicos para el correcto desarrollo de esta memoria.

5.4. Trabajo Futuro

Existen muchas formas de expandir el trabajo presentado en esta memoria, las cuales han sido planteadas durante el desarrollo de este informe. A continuación se resumen algunas líneas de investigación que surgen a partir de este trabajo.

En primer lugar, este trabajo presenta varias características nuevas para detectar HS. Sin embargo, éstas podrían ser desarrolladas en mayor detalle. Además, sería interesante desarrollar nuevas características con el mismo objetivo que se tenía en este trabajo, es decir, que se enfoquen en alguno de los aspectos que definen a un HS, como su amplitud, morfología, duración o frecuencia.

Por otra parte, por consideraciones de tiempo, no fue posible realizar muchas pruebas que habría sido interesante realizar. Una de ellas sería probar el aporte del RP, por separado, a la detección. Similarmente, sería interesante estudiar la información que entregan las nuevas características desarrolladas a la detección. Por ejemplo, se podrían comparar sistemas sin algunas de las características con el sistema que las incluya todas, para determinar realmente cuál es el aporte de dichas características. Además, se podría ver cómo afecta al desempeño a la salida del módulo 4, del sistema mostrado en [7], el hecho de reemplazar los módulos 1 y 2 por el método implementado en este trabajo. Finalmente, es necesario realizar una prueba en que se considere un mayor número de pacientes, para ver si la relación entre los desempeños de ambos métodos se mantiene en este caso.

Otras ideas que fueron consideradas durante la implementación de este trabajo, pero que no se llevaron a cabo por consideraciones de tiempo, se describen a continuación. El sistema planteado en esta memoria utiliza NN para clasificar vectores de características, pero estas se podrían usar directamente sobre los RP, considerando cada punto del RP como una característica binaria. También, se podría implementar un RP que mida la simetría de los HS. Por otro lado, muchas características utilizadas fueron sub-muestreadas a la tasa en que se obtenían las características de los RP. A futuro, se podría realizar un método que utilice dos clasificadores, uno para las características de los RP, y otro para el resto de las características muestreadas a su tasa original, y luego combinar de alguna forma la salida de dichos clasificadores. Finalmente, la normalización realizada para la correntropía no fue la más adecuada. En otros trabajos podría investigarse cómo mejorar dicha clasificación.

Una línea de investigación interesante, a futuro, sería estudiar el desempeño de un sistema híbrido, que considere una combinación de los dos métodos comparados en este trabajo. También se podrían estudiar y caracterizar los casos en que el sistema implementado se equivoca, de tal forma de plantear mejoras enfocadas en corregir estos errores.

6. Bibliografía

- [1] N. Asghari, C. Broeg, L. Carone, R. Casas-Miranda, J.C. Castro Palacio, I. Csillik, R. Dvorak, F. Freistetter, G. Hadjivantsides, H. Hussmann, A. Khramova, M. Khristoforova, I. Khromova, I. Kitiashivilli, S. Kozlowski, T. Laakso, T. Laczkowski, D. Lytvinenko, O. Miloni, R. Morishima, A. Moro-Martin, V. Paksyutov, A. Pal, V. Patidar, B. Pecnik, O. Peles, J. Pyo, T. Quinn, A. Rodriguez, C. Romano, E. Saikia, J. Stadel, M. Thiel, N. Todorovic, D. Veras, E. Vieira Neto, J. Vilagi, W. von Bloh, R. Zechner, E. Zhuchkova, "Stability of terrestrial planets in the habitable zone of gl 777 a, hd 72659, gl 614, 47 uma and hd 4208", *Astronomy and Astrophysics*, vol. 426, pp. 353-365, 2004.
- [2] C. Bishop, "Neural Networks for Pattern Recognition", *Clarendon Press*, Oxford, 1995.
- [3] R. Bodizs, J. Kormendi, P. Rigo and A. Sandor Lazar, "The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm", *Journal of Neuroscience Methods*, vol. 178, num. 1, pp. 205-213, 2009.
- [4] S. Butterworth, "On the Theory of Filter Amplifiers", *Experimental Wireless and the Wireless Engineer*, vol. 71, pp. 536-541, 1930.
- [5] K. Campbell, A. Kumar and W. Hofmann, "Human and automatic validation of a phaselocked loop spindle detection system", *Electroencephalography and Clinical Neurophysiology*, vol. 48, num. 5, pp. 602-605, 1980.
- [6] L. Causa, C. Held, J. Causa, P. Estévez, C. Pérez, R. Chamorro, M. Garrido, C. Algarín y P. Peirano, "Automated Sleep-Spindle Detection in Healthy Children Polysomnograms" *IEEE Transactions on Biomedical Engineering*, vol. 57, num. 9, pp. 2135-2146, 2010.
- [7] L. Causa, "Detección Automática y Caracterización de Patrones de Husos Sigma en Registros Polisomnográficos de Niños", *Tesis para Optar al Título de Magister e Ingeniero Civil Electricista*, Universidad de Chile, 2011.
- [8] P. Corkum, R. Tannock and H. Moldofsky, "Sleep disturbances in children with attention deficit/hyperactivity disorder", *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 37, num. 6, pp. 637-646, 1998.
- [9] A.S. Elwakil, A.M. Soliman, "Mathematical models of the Twin-T, Wien-bridge and family of minimum component electronic chaos generators with demonstrative recurrence plots", *Chaos Solitons Fractals*, vol. 10 issue 8, pp. 1399-1411, 1999.
- [10] P. Estévez, C. Held, C. Holzmann, C. Pérez, J. Pérez, J. Heiss, M. Garrido y P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants", *Medical & Biological Engineering & Computing*, vol. 40, pp. 105-113, 2002.

- [11] P. A. Estévez, R. Zilleruelo-Ramos, R. Hernandez, L. Causa and C. M. Held, "Sleep spindle detection by using merge neural gas", *Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)*, Bielefeld, Germany, 2007.
- [12] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers", *HP Laboratories*, Kluwer Academic Publishers, Palo Alto, 2004.
- [13] I. Feinberg, M. Braun and E. Shulman, "EEG sleep patterns in mental retardation", *Electroencephalography and Clinical Neurophysiology*, vol. 27, num. 2, pp. 128-141, 1969.
- [14] A. Giuliani, C. Manetti, "Hidden peculiarities in the potential energy time series of a tripeptide highlighted by a recurrence plot analysis: a molecular dynamics simulation", *Physical Review E*, vol. 53 issue 6, pp. 6336-6340, 1996.
- [15] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen and K. Leblebicioglu, "Sleep spindles detection using short time Fourier transform and neural networks", *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN'02)*, Honolulu, Hawaii, USA, vol. 2, pp. 1631-1636, May 12-17, 2002.
- [16] S. Guiasu, "Information Theory with Applications", *McGraw-Hill*, New York, 1977.
- [17] A. Gunduz, A. Hegde, J. Principe, "Correntropy as a Novel Measure for Nonlinearity Tests", *Signal Processing*, vol. 89, num. 1, pp. 14-23, 2009.
- [18] S. Haykin, "Neural Networks: A Comprehensive Foundation", *MacMillan College Publishing*, New York, 1994.
- [19] C. M. Held, L. Causa, P. Estevez, C. Perez, M. Garrido, C. Algarin and P. Peirano, "Dual approach for automated sleep spindles detection within EEG background activity in infant polysomnograms", *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS'04)*, San Francisco, CA, USA, vol. 1, pp. 566-569, September 1-5, 2004.
- [20] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung y H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings Royal Society London A*, vol. 454, pp. 903-995, 1998.
- [21] J. R. Hughes, "Sleep spindles revisited", *Journal of Clinical Neurophysiology*, vol. 2, num. 1, pp. 37-44, 1985.
- [22] E. Huupponen, G. Gomez-Herrero, A. Saastamoinen, A. Varri, J. Hasan and S. Himanen, "Development and comparison of four sleep spindle detection methods", *Artificial Intelligence in Medicine*, vol. 40, num. 3, pp. 157-170. 2007.
- [23] E. Huupponen, A. Varri, S.-L. Himanen, J. Hasan, M. Lehtokangas and J. Saarinen, "Optimization of sigma amplitude threshold in sleep spindle detection", *Journal of Sleep Research*, vol. 9, pp. 327-334, 2000.

- [24] H. Kantz and T. Scheriber, "Nonlinear Time Series Analysis", University Press, Cambridge, 1997.
- [25] N. Kodama, Y. Aoki, H. Hiraga, T. Wada and J. Suzuki, "Electroencephalographic findings in children with moyamoya disease", *Archives of Neurology*, vol. 36, num. 1, pp. 16-19, 1979.
- [26] D. Kunz, H. Danker-Hopfe, G. Gruber, J. L. Lorenzo, S. L. Himanen, B. Kemp, T. Penzel, J. Roschke and G. Dorffner, "Interrater reliability between eight european sleep-labs in healthy subjects of all age-groups", *Journal of Sleep Research*, vol. 9, num. 1, pp.106, 2000.
- [27] J. Kurths, U. Schwarz, C.P. Sonett, U. Parlitz, "Testing nonlinearity in radiocarbon data", *Nonlinear Processes in Geophysics*, vol. 1 issue 1, pp. 72-75, 1994.
- [28] E. Limoges, L. Mottron, C. Bolduc, C. Berthiaume and R. Godbout, "Atypical sleep architecture and the autism phenotype", *Brain*, vol. 128, num. 5, pp. 1049-1061, 2005.
- [29] C. Manetti, A. Giuliani, M.A. Ceruso, C.L.Webber Jr., J.P. Zbilut, "Recurrence analysis of hydration effects on nonlinear protein dynamics: multiplicative scaling and additive processes", *Physics Letters A*, vol. 281 issue 5-6, pp. 317-323, 2001.
- [30] T.K. March, S. C. Chapman, R.O. Dendy, "Recurrence plot statistics and the effect of embedding", *Physica D: Nonlinear Phenomena*, vol. 200 issue 1-2, pp. 171-184, 2005.
- [31] N. Marwan, M. Romano, M. Thiel y J. Kurths, "Recurrence plots for the analysis of complex systems", *Physics Reports*, vol. 438, pp. 237-329, 2007.
- [32] N. Marwan, M. Thiel, N.R. Nowaczyk, "Cross recurrence plot based synchronization of time series", *Nonlinear Processes in Geophysics*, vol. 9 issue 3-4, pp. 325-331, 2002.
- [33] N. Marwan, M.H. Trauth, M. Vuille, J. Kurths, "Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods", *Climate Dynamics*, vol. 21 issue 3-4, pp. 317-326, 2003.
- [34] R. McCarley, "Neurobiology of REM and NREM sleep", *Sleep Medicine*, vol. 8, num. 4, pp. 302-330, 2007.
- [35] W. McCulloch y W. Pitts "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, vol. 7, pp. 115-133, 1943.
- [36] J.M. Nichols, S.T. Trickey, M. Seaver, "Damage detection using multivariate recurrence quantification analysis", *Mechanical Systems and Signal Processing*, vol. 20 issue 2, pp. 421-437, 2006.

- [37] P. Peirano, C. Algarin, R. Chamorro, S. Reyes, S. Duran, M. Garrido and B. Lozoff, "Sleep alterations and iron deficiency anemia in infancy", *Sleep Medicine*, vol. 11, num. 7, pp. 637-642, 2010.
- [38] P. Peirano and C. Algarin, "Sleep in brain development", *Biological Research*, vol. 40, num. 4, pp. 471-478, 2007.
- [39] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects", *UCLA, Brain Research Institute/ Brain Information Service*, Los Angeles, CA, 1968.
- [40] G. Rilling, P. Flandrin y P. Gonçalves, "On empirical Mode Decomposition and Its Algorithm", *IEEE-EURASIP Workshop on NonLinear Signal and Image Processing NSIP-03*, Grado, Italy, 2003.
- [41] M. Rosanova and D. Ulrich, "Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train" *Journal of Neuroscience*, vol. 25, num. 41, pp. 9398-9405, 2005.
- [42] T. Sauer, J. Yorke and M. Casdagli, "Embedology", *Journal of Statistical Physics*, vol. 65 num 3-4, pp.579-616, 1991.
- [43] B. Schölkopf, "SVMs - a practical consequence of learning theory", *IEEE Intelligent Systems*, num. 4, pp.18-21, 1998.
- [44] P. Schimicek, J. Zeitlhofer, P. Anderer and B. Saletu, "Automatic sleep-spindle detection procedure: aspects of reliability and validity" *Clinical Electroencephalography*, vol. 25, num. 1, pp. 26-29, 1994
- [45] M. Schabus, K. Hodlmoser, G. Gruber, C. Sauter, P. Anderer, G. Klosch, S. Parapatics, B. Saletu, W. Klimesch and J. Zeitlhofer, "Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities", *European Journal of Neuroscience*, vol. 23, num. 7, pp. 1738-1746, 2006.
- [46] S. V. Schonwald, E. L. de Santa-Helena, R. Rossatto, M. L. Chaves and G. J. Gerhardt, "Benchmarking matching pursuit to find sleep spindles" *Journal of Neuroscience Methods*, vol. 156, num. 1-2, pp. 314-321, 2006.
- [47] F. J. Schulte, H. J. Kaiser, S. Engelbart, E. F. Bell, R. Castell and H. G. Lenard, "Sleep patterns in hyperphenylalaninemia: a lesson on serotonin to be learned from phenylketonuria", *Pediatric Research*, vol. 7, num. 6, pp. 588, 1973.
- [48] M. Steriade, "Grouping of brain rhythms in corticothalamic systems.", *Neuroscience*, vol. 137, num. 4, pp. 1087-1106, 2006.
- [49] M. Steriade, "Coherent oscillations and short-term plasticity in corticothalamic networks", *Trends in Neurosciences*, vol. 22, num. 8, pp. 337-345, 1999.

- [50] M. Steriade and F. Amzica, "Coalescence of sleep rhythms and their chronology in corticothalamic networks", *Sleep Research Online*, vol. 1, num. 1, pp.1-10, 1998.
- [51] F. Takens, "Detecting strange attractors in turbulence", *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, Springer, Berlin, 1981, pp. 366-381.
- [52] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data", *Physical Reviews A*, vol. 34, num. 3, pp. 2427–2432, 1986.
- [53] V. Vapkin, "Statistical Learning Theory", *John Wiley*, New York, 1998.
- [54] E. M. Ventouras, E. A. Monoyiou, P. Y. Ktonas, T. Paparrigopoulos, D. G. Dikeos, N. K. Uzunoglu and C. R. Soldatos, "Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: a feasibility study" *Computer Methods and Programs in Biomedicine*, vol. 78, num. 3, pp. 191-207, 2005.
- [55] M. P. Walker and R. Stickgold, "Sleep, memory, and plasticity" *Annual Review of Psychology*, vol. 57, num. 1, pp. 139-166, 2006.
- [56] M. P. Walker, "The role of sleep in cognition and emotion", *Annals of the New York Academy of Sciences*, vol. 1156, num. 1, pp. 168-197, 2009.
- [57] E. Wan, "Finite Impulse Response Neural Networks with Applications in Time Series Prediction", *Thesis for the degree of Doctor of Philosophy*, Stanford University, Palo Alto, US., 1993.
- [58] S. Yoo, P. T. Hu, N. Gujar, F. A. Jolesz and M. P. Walker, "A deficit in the ability to form new human memories without sleep", *Nature Neuroscience*, vol. 10, num. 3, pp. 385-392, 2007.
- [59] N.V. Zolotova, D.I. Ponyavin, "Phase asynchrony of the north-south sunspot activity", *Astronomy and Astrophysics*, vol. 449 issue 1, pp. L1-L4, 2006.
- [60] J. Zygierewicz, K. J. Blinowska, P. J. Durka, W. Szelenberger, S. Niemcewicz and W. Androsiuk, "High resolution study of sleep spindles", *Clinical Neurophysiology*, vol. 110, num. 12, pp. 2136-2147, 1999.

7. Anexos

7.1. Variantes del Método

Durante la investigación y el desarrollo del sistema explicado en este trabajo, se probaron distintos métodos y variantes. Algunos de ellos resultaron más exitosos, y fueron utilizados en el sistema final, mientras que otros, por diversos motivos, fueron descartados. A continuación se explican algunas de las variantes que fueron consideradas y probadas en el transcurso de esta investigación, pero que no se usaron en el sistema final.

7.1.1. Variantes de Pre-Procesamiento de la Señal

7.1.1.1. EMD

Es necesario filtrar la señal del EEG de alguna manera, pues, debido a su origen fisiológico, existe ruido y diferencias entre señales de distintos individuos que hace muy difícil trabajar con las señales originales. Inicialmente se consideró utilizar EMD para pre-procesar la señal, al igual como se hace en [6], pues este método ya había sido probado con buenos resultados. Sin embargo, durante las pruebas preliminares surgieron varios inconvenientes que causaron que el pre-procesamiento mediante EMD fuera descartado.

Como es sabido, EMD funciona separando la señal original en un cierto número de componentes. Para realizar dicha descomposición el método EMD se basa en las variaciones de la señal misma, y no en límites de frecuencia fijos. Esto causa que los HS puedan aparecer a veces en una componente y otras veces en otra componente. Originalmente, se intentó construir los RP's a partir de las primeras 3 componentes de la descomposición, de forma similar a como se hacía en [6]. Sin embargo, como se explicó anteriormente, era posible que los HS se concentraran en cualquiera de estas tres componentes, o incluso en otras componentes de mayor orden que no estaban siendo consideradas. Esto tiene el inconveniente que si el sistema se entrenaba con datos en donde los HS se concentraban en una sola componente, el sistema construido tiende a enfocarse más en las características deducidas de dicha componente. Si posteriormente el clasificador se alimenta con otros datos, en donde la componente en la cual se concentraban los HS es otra, el sistema probablemente obtendría malos resultados. En la Figura 34 se puede apreciar un ejemplo en donde los HS de la señal original tienden a concentrarse en la segunda componente de EMD.

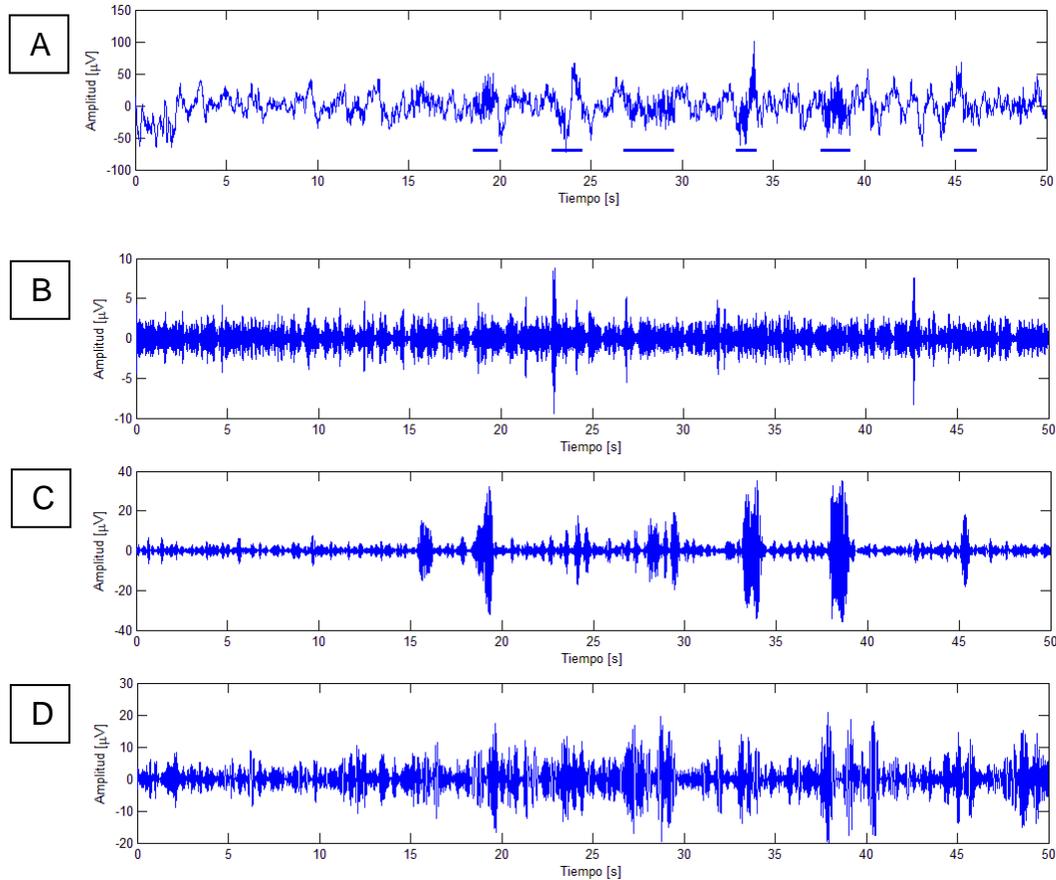


Figura 34: Ejemplo de Descomposición EMD.

A: Señal original para registro SQ2, con HS marcados por experto. B: Primera componente de descomposición EMD de señal A. C: Segunda componente de descomposición EMD de señal A. D: Tercera componente de descomposición EMD de señal A.

Considerando la naturaleza de los HS, se concluyó que no era conveniente usar EMD, ya que esta técnica está pensada para casos en los cuales una descomposición armónica no es la indicada. Sin embargo, se sabe que los HS poseen frecuencias en un rango bien definido. Debido a ello, se optó por usar un filtrado pasa banda de la señal. De esta manera, existe la certeza de que los HS están en la señal pre-procesada.

7.1.1.2. Filtrado

Considerando lo explicado en la parte 7.1.1, se decidió filtrar la señal en la banda sigma, que para los efectos de este trabajo corresponde a las frecuencias dentro de intervalo definido por 10 y 16 [Hz]. Sin embargo, en pruebas posteriores se descubrió que se obtenían mejores resultados si dicho intervalo se ampliaba

ligeramente. Por ello, se consideraron las frecuencias de corte del filtro, f_{cl} y f_{ch} , como parámetros del sistema, y se fueron modificando para determinar su valor óptimo.

También, se consideró que obtener características directamente de la señal original podría ayudar a complementar las características obtenidas de la señal pre-procesada. Sin embargo, las características deducidas de la envolvente de la señal sin filtrar no son útiles para la clasificación, ya que la señal original es demasiado ruidosa y posee grandes componentes de baja frecuencia que enmascaran los HS. Por otro lado, los RP obtenidos de la señal sin filtrar aparecen distorsionados y no aportan información nueva a la clasificación, por encima de los RP obtenidos de la señal filtrada. Debido a esto, la única característica del sistema final en la cual se utiliza la señal original es la potencia de la banda sigma, todas las demás están basadas en la señal pre-procesada.

7.1.2. Variantes en Construcción de RP's y Extracción de Características

7.1.2.1. Distancias para Construcción de RP's

Las distancias usadas para calcular los RP corresponden a $d_{correl-kp}$ y $d_{corren-n}$, las cuales se muestra en las ecuaciones (64) y (67), respectivamente. Se escogieron estas distancias tras pruebas preliminares sobre el conjunto de entrenamiento, donde resultaron ser superior a otras distancias. En la sección 4.3.2.1 se discuten dichas pruebas en mayor detalle. Los parámetros de estas distancias, vale decir, el periodo p y el sigma temporal σ_t para $d_{correl-kp}$, y el sigma espacial σ_d , para $d_{corren-n}$, también fueron escogidos mediante pruebas preliminares.

Las otras distancias que se consideraron en las pruebas preliminares son: d_{correl} , d_{en} y $d_{correl-kpn}$, las cuales se muestran en las ecuaciones (60), (58) y (68), respectivamente. A partir de estas distancias se construyeron sistemas mediante dos métodos:

1. Sistemas donde se calculaba un solo RP para una distancia, y se deducían características de este RP.
2. Sistemas donde se calculaban dos RPs, cada uno para una distancia distinta, y se agregaban características deducidas de ambos al vector de características.

Los sistemas construidos mediante ambos métodos fueron comparados en pruebas sobre un sub-conjunto del conjunto de entrenamiento. Dicha comparación

se llevó a cabo considerando el desempeño de la clasificación de cada sistema. Se observó que los mejores resultados ocurrían cuando se utilizaba un sistema que calculara dos RPs, uno para $d_{correl-kp}$ y otro para $d_{corren-n}$. Se observó que usando un kernel periódico cuya frecuencia estuviera cercana al centro de la banda sigma, los RPs tendían a contener información más acertada de la morfología de los HS, y por lo tanto las características deducidas de ellos tendían a ser más apropiadas. Las pruebas donde se compararon estas características se detallan en la sección 4.3.2.1.

7.1.2.2. Medidas de RP's

En pruebas iniciales se calcularon la mayoría de las medidas provenientes de los RPs. Dichas medidas son: la tasa de recurrencia o RR (2.2.6.1), largo máximo de líneas diagonales o L_{max} (2.2.6.3), media del histograma de líneas diagonales o L_{medio} (2.2.6.5), desviación estándar del histograma de líneas diagonales o L_{desv} (2.2.6.2), curtosis del histograma de líneas diagonales o L_{kurt} (2.2.6.2), coeficiente de asimetría de Fisher del histograma de líneas diagonales o L_{skew} (2.2.6.2), determinismo o DET (2.2.6.4), entropía de Shannon del histograma de líneas diagonales o E (2.2.6.6) y razón o RA (2.2.6.8). En la Figura 35 se muestra un ejemplo de una comparación entre el valor de estas medidas en función del tiempo y las marcas de HS del experto. Se puede observar que existe una alta correlación entre algunas de estas medidas y la presencia de HS.

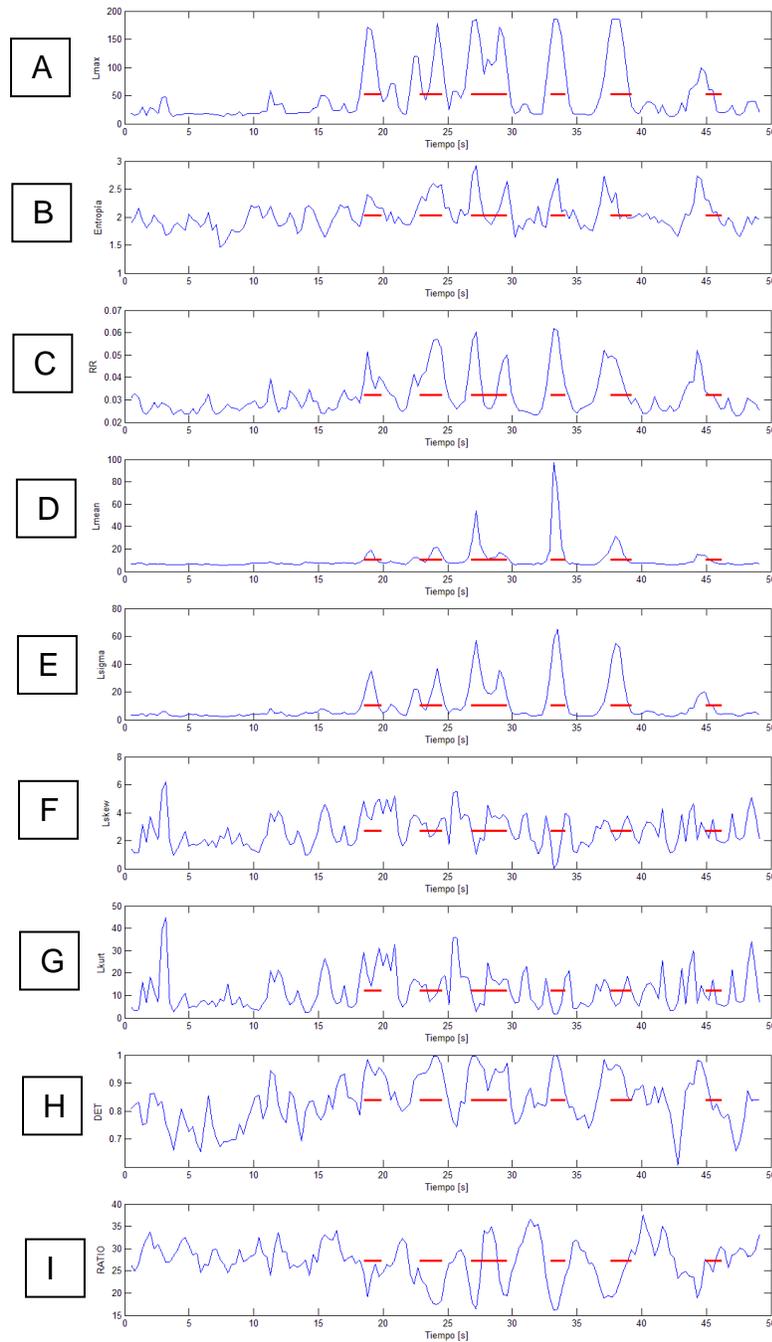


Figura 35: Contrastación de Medidas de RP con Marcas del Experto. En todos los gráficos la marca del experto está representada como líneas horizontales rojas. A: Gráfico de L_{max} versus tiempo. B: Gráfico de E versus tiempo. C: Gráfico de RR versus tiempo. D: Gráfico de L_{medio} versus tiempo. E: Gráfico de L_{desv} versus tiempo. F: Gráfico de L_{skew} versus tiempo. G: Gráfico de L_{kurt} versus tiempo. H: Gráfico de DET versus tiempo. I: Gráfico de RA versus tiempo.

Pruebas preliminares demostraron de algunos de las medidas de RP calculadas tienden a ser redundantes o aportan poca información para la clasificación, al tiempo que introducían ruido a la clasificación. En general, las medidas que están más directamente relacionadas con el histograma de líneas diagonales del sistema son las que más información entregan respecto a los HS. Esto se debe a que los RP's de la señal filtrada tienden a mostrar patrones muy regulares en instantes de tiempo donde existen HS, similares al patrón que se obtendría de una señal sinusoidal de periodo constante. En la Figura 36 se observa un RP con estas características. Como se puede apreciar, aparecen líneas diagonales muy largas y separadas por una distancia constante. Debido a este hecho, la presencia de HS está ligada a las estadísticas del largo de las líneas en los HS, en particular, el largo máximo, L_{max} , el largo promedio L_{medio} y la desviación de los largos L_{desv} . Otras medidas, por ejemplo RR , aportaban poca información a la clasificación. Esto se comprobó mediante la realización de pruebas para medir la IM entre cada característica deducida del RP y las marcas de HS. Además de esto, se realizaron pruebas en donde se comparaba el desempeño en la clasificación de distintos sistemas. La diferencia entre estos sistemas era que algunos incluían en su vector de característica todas las características deducidas del RP antes mencionadas, y otros solo ocupaban un subconjunto de ellas. Estas pruebas demostraron que lo más óptimo era considerar solo las tres características que fueron incluidas en el sistema final: L_{max} , L_{medio} y L_{desv} . Dichas pruebas se explican con mayor detalle en la sección 4.3.2.2.

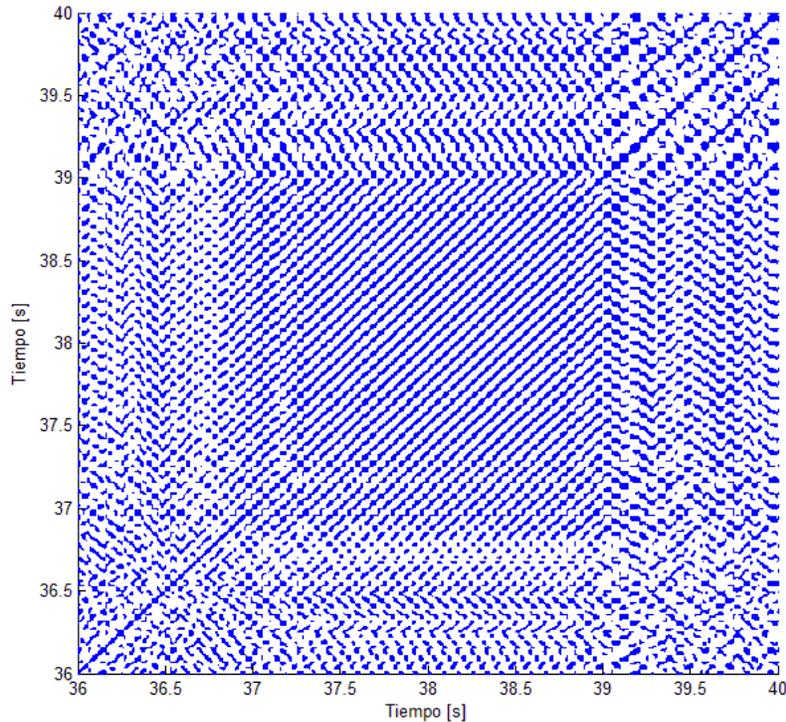


Figura 36: Ejemplo de RP Calculado para un Instante de Tiempo en Donde Existe un HS con Poco Ruido.

Este RP fue calculado usando la distancia d_{correl} a partir del EEG del paciente 1 para SQ2.

7.1.2.3. Variantes de Características no Deducidas de RP's

Inicialmente, además de medir la potencia en la banda sigma, para la señal del EEG sin filtrar, y dividirla por la frecuencia en la banda fisiológica, se medían potencias en otras dos bandas: la banda entre 0.5 y 3 [Hz], conocida como banda delta y la banda entre 30 y 60 [Hz], llamada banda de altas frecuencias. Al igual que para la banda sigma, para calcular dichas potencias, la señal original del EEG es dividida en las mismas ventanas utilizadas para calcular los RP. Luego, en cada una de estas ventanas, se calcula la FFT. Posteriormente, se suma el cuadrado de los coeficientes de Fourier asociados a frecuencias dentro de los intervalos. Con esto, se obtienen 2 potencias más: una para la banda delta, P_{δ} y una para las altas frecuencias, P_{HF} . La potencia en la banda fisiológica, que llamaremos P_{BF} , también se usa para normalizar estas dos potencias adicionales. Así, originalmente se obtenían 3 potencias normalizadas, dados por las siguientes fórmulas:

$$P_{\delta}^n = \frac{P_{\delta}}{P_{BF}}, \quad (126)$$

$$P_{\sigma}^n = \frac{P_{\sigma}}{P_{BF}}, \quad (127)$$

$$P_{HF}^n = \frac{P_{HF}}{P_{BF}}. \quad (128)$$

En pruebas preliminares se observó que solo la potencia P_{σ}^n aporta información a la clasificación. Se realizaron pruebas midiendo la IM entre las 3 potencias y la presencia de HS, observándose que la IM entre P_{δ} y las marcas de HS, y entre P_{HF} y las marcas de HS, eran bastante bajas. Además, se realizaron pruebas en el conjunto de entrenamiento, comparando el desempeño de un sistema (calidad de la clasificación) cuyo vector de características contenía las 3 potencias con un sistema cuyo vector de características no contenía las potencias P_{δ}^n y P_{HF}^n . Se observaron mejores resultados en el último caso. Los detalles de esta prueba se encuentran en la sección 4.3.2.3. A causa de esto, el sistema final no tenía las potencias P_{δ}^n y P_{HF}^n en su vector de características.

Para estimar el ancho de los *peaks* se define un umbral deducido a partir de mediciones llamadas pe_i (ver sección 3.2.2.1). Se probaron diferentes métodos de definir este umbral, a partir de los valores de los pe_i . En primer lugar, se calcula el promedio y la desviación estándar de los pe_i , dados por las siguientes ecuaciones:

$$\bar{pe} = \frac{1}{H} \sum_i^H pe_i, \quad (129)$$

$$std_{pe} = \sqrt{\frac{1}{H-1} \sum_i^H (pe_i - \bar{pe})^2}. \quad (130)$$

A partir de estos valores, se estudiaron tres opciones para definir el umbral, dadas por las siguientes ecuaciones:

$$u_{peaks-1} = \bar{pe} - 2 \cdot std_{pe}, \quad (131)$$

$$u_{peaks-2} = \bar{pe} - std_{pe}, \quad (132)$$

$$u_{peaks-3} = \min_i pe_i. \quad (133)$$

La idea de u_{peaks} era que fuera inferior al pe_i de cualquier HS marcado por el experto, tanto en el conjunto de prueba como en el conjunto de entrenamiento (aún cuando este parámetro se calcula únicamente a partir de los pe_i del conjunto de entrenamiento). Sin embargo, no podía ser demasiado bajo pues debía permitir discriminar correctamente entre HS y no HS. Como no existía certeza de cuál de estos valores resultaba mejor para ambos propósitos, se realizaron pruebas en el conjunto de entrenamiento para comparar las características deducidas a partir de cada umbral. Notar que para valor de $u_{peaks-n}$ (con $n = 1, 2$ ó 3) se deducen dos características: fb^n y $anch^n$. Se midió en el conjunto de entrenamiento la IM entre estas características y las marcas de HS, observándose que ésta era mayor para el caso en que el umbral era $u_{peaks-1}$. Además, se realizaron pruebas comparando entre sí el desempeño (calidad de la clasificación) de tres sistemas distintos. En cada uno de estos sistemas, se utilizaba en el vector de características uno de los tres pares de características del tipo fb^n y $anch^n$. Se obtuvieron los mejores resultados para el caso $n = 1$. Así, se usó $u_{peaks-1}$ como umbral en el sistema final, es decir, se definió $u_{peaks} = u_{peaks-1}$.

7.1.2.4. Variantes de Clasificación

En el sistema final la clasificación se realiza utilizando una red neuronal. Sin embargo, originalmente se pensó usar una SVM, debido a su simpleza y a que requería menos parámetros de entrenamiento. Una SVM con kernel gaussiano solo requiere dos parámetros: el γ del kernel y el C del entrenamiento. En pruebas preliminares sobre el conjunto de entrenamiento se observó que estos parámetros influían enormemente en los resultados. Al variarlos, era posible mover el sistema en distintos puntos del plano TPR-FPR. Sin embargo, los puntos obtenidos no representaban una curva que fuera monótonamente creciente. Esto hacía muy complicada la comparación de diferentes sistemas, necesaria inicialmente para tomar decisiones respecto a los valores de los otros parámetros y las diferentes variantes en la construcción del vector de características. En este sentido, resulta más conveniente usar una NN, pues su salida se puede pasar por un umbral, y modificar dicho umbral permite construir una curva ROC que necesariamente es creciente. Con dichas curvas, determinar qué sistema es mejor es relativamente sencillo, basta ver para cual de ellos la curva se acerca más al extremo superior izquierdo del gráfico. Debido a que clasificador en sí no debería afectar sustancialmente los resultados, y tampoco era el enfoque de este trabajo, se optó finalmente por una NN. Otra ventaja de usar una NN, por sobre la SVM, es que la salida de la NN es un valor real entre -1 y 1, mientras que la salida de la SVM es una clase, es decir, un valor discreto. Esto permite interpolar la salida de la red para expandir los resultados de la clasificación a los otros instantes de tiempo en donde no se había calculado un vector de características. Esto habría sido poco efectivo con una SVM, pues la interpolación habría sido muy básica.

La salida de la red neuronal se interpola de tal forma de expandir sus resultados a puntos vecinos en donde el vector de características no ha sido calculado, y luego se aplica un umbral para determinar el inicio y el fin de los HS. Esto, sin embargo, no siempre fue el caso, pues en las primeras versiones del sistema la salida de la red neuronal era transformada directamente a una clase mediante el umbral, y luego dicha clase era asignada a toda la ventana a partir de la cual se había calculado el vector de características. En el caso en que dos ventanas contiguas pertenecieran a clases distintas, se consideraba que el instante tiempo en que terminaba una clase y empezaba la otra era el promedio entre los tiempos medios de cada ventana. Una de las desventajas principales de este método era que los tiempos en donde podían comenzar o terminar los HS eran pocos y estaban definidos de antemano. Esto hacía que la resolución de las marcas fuera bastante pobre. Debido a ello, en el sistema final se optó por introducir una interpolación de las salidas de la NN.