



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

ESTUDIO DEL COMPORTAMIENTO DE LA CARGA COGNITIVA DE USUARIOS
QUE NAVEGAN EN UN SITIO WEB

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

CRISTIAN FELIPE RETAMAL CONTRERAS

PROFESOR GUÍA:
ÁNGEL JIMÉNEZ MOLINA

MIEMBROS DE LA COMISIÓN:
JAVIER RUIZ DEL SOLAR
HÉCTOR AGUSTO ALEGRÍA

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: CRISTIAN FELIPE RETAMAL CONTRERAS
FECHA: 2017
PROF. GUÍA: SR. ÁNGEL JIMÉNEZ MOLINA

ESTUDIO DEL COMPORTAMIENTO DE LA CARGA COGNITIVA DE USUARIOS QUE NAVEGAN EN UN SITIO WEB

En psico-fisiología la carga cognitiva se define como la cantidad total de información que el cerebro es capaz de procesar en un momento dado, cantidad que es finita y que se relaciona con la atención, percepción, memoria corto y largo plazo, control, motor, entre otros.

Cuando un usuario se enfrenta a tareas de interacción Humano Computador (HCI) necesariamente utiliza su capacidad mental, por ejemplo, al analizar una página web el usuario lee, discrimina en el contenido, concentra su atención, etc. Una sobreexigencia del recurso cognitivo genera frustración, desagrado y problemas para completar tareas definidas, lo que disminuye el nivel de experiencia del usuario que navega en un sitio web.

Bajo el contexto del proyecto Fondecyt “*A Cognitive Resource-Aware Mobile Service Framework to Support Human-Computer-Interactions in Ubiquitous Computing Environments*” liderado por el Profesor Ángel Jiménez, se desarrolla esta memoria cuya objetivo es comprobar la hipótesis de que es posible medir la carga cognitiva para actividades de navegación web frente a un computador mediante señales psico-fisiológicas y que, además, se produce una baja en la carga mental del usuario en momentos de transición entre el análisis de un elemento web y otro.

El diseño experimental considera la medición de 61 voluntarios utilizando 6 sensores psico-fisiológicos: Eye tracker, sensor de respuesta electrodermal (GSR), Sensor de Temperatura, Fotopletimógrafo (PPG), electrocardiógrafo (ECG) y electroencefalógrafo (EEG). Cada participante navega libremente por una página web que se presenta en 3 versiones manteniendo su diseño y variando su contenido.

Mediante un análisis de varianza con medidas repetidas (ANOVA-MR) y utilizando la media del diámetro pupilar como medida directamente relacionada con la carga cognitiva, se comprueba con un grado de significancia de $p - value = 0,00184$ en un intervalo de confianza de 95 % que es estadísticamente significativa la baja de la carga mental en los momentos de transición entre el análisis de un elemento web y otro.

Para determinar los niveles de carga cognitiva existentes y poder etiquetarlos se aplica el algoritmo de agrupamiento *k-means*. Los resultados de la clasificación con redes neuronales profundas demuestran que sí es posible medir la carga cognitiva durante la navegación web. Para la combinación de los sensores GSR, PPG y EEG se obtiene el mejor resultado con una exactitud de 95,73 % al clasificar. Además, se determina que el EEG es el sensor que más aporta a la clasificación (por sí solo alcanza una exactitud del 88,78 %), con lo que se concluye que es posible replicar el experimento y medir la carga mental utilizando solamente el EEG; o bien, el EEG acompañado del sensor de GSR y PPG que son fáciles de utilizar y de bajo costo en relación a otros sensores.

Agradecimientos

Gracias a mis padres, Carmen y Juan, que me han acompañado siempre y desde pequeño me han hecho un hombre de esfuerzo y que cumple sus metas.

Gracias a mi hermano Sebastian con quien compartimos tantas ñoñadas que me hacían reír.

Gracias a mi hoy novia y futura esposa Camila por su gran apoyo y paciencia hasta el final.

Y mil gracias al Dios verdadero, mi Dios Jehová que le dio propósito a mi vida cuando me enseñó “la vida que realmente lo es” (1 Timoteo 6:17-19).

Tabla de Contenido

Tabla de Contenidos	vi
Índice de Tablas	ix
Índice de Ilustraciones	xi
1. Introducción	1
1.1. Motivación	1
1.2. Definición del problema	2
1.3. Contexto institucional	3
1.4. Objetivos	4
1.5. Hipótesis de trabajo y metodología	5
1.6. Resultados esperados y alcances	6
2. Revisión Bibliográfica y Contextualización	7
2.1. Sistema Nervioso	7
2.1.1. Sistema Nervioso Autónomo	8
2.2. Medición de carga cognitiva	9
2.2.1. Medición por rendimiento	9
2.2.2. Medición subjetiva	9
2.2.3. Medición psico-fisiológica	10
2.3. Medidas psico-fisiológicas	11
2.3.1. Seguimiento ocular	11
2.3.2. Actividad pupilográfica	12
2.3.3. Electroencefalografía	14
2.3.4. Actividad electrodérmica	16
2.3.5. Actividad del sistema cardiovascular	18
2.3.6. Temperatura corporal	19
2.4. Selección de características	21
2.4.1. Criterio de correlación	21
2.4.2. Eliminación recursiva de características	22
2.5. Clusterización	24
2.5.1. <i>k-means</i>	24
2.6. Criterios de validación de clusters	26
2.6.1. Visualización de datos	27
2.6.2. WSS y Calinski & Harabasz	27
2.6.3. Bootstrap	28

2.7.	Clasificadores	30
2.7.1.	Naive Bayes	30
2.7.2.	Máquina de vector de soporte	31
2.7.3.	Redes Neuronales Artificiales	35
2.8.	Deep Learning	39
2.8.1.	Redes neuronales de alimentación profunda	39
2.8.2.	Regularización	40
2.9.	Estado del Arte	42
3.	Estudio Experimental e Implementación	47
3.1.	Instrumentación	47
3.1.1.	Unidad <i>GSR+</i> de Shimmer	48
3.1.2.	Sensor óptico de pulso de Shimmer	50
3.1.3.	Unidad <i>Bridge Amplifier+</i> de Shimmer	52
3.1.4.	ECG de tarjeta biomédica todo-en-uno de BITalino	54
3.1.5.	EEG Emotiv EPOC	56
3.1.6.	<i>Eye Tracker</i> Tobii T120	58
3.2.	Grupo experimental	60
3.3.	Diseño experimental	60
3.3.1.	Escenarios	60
3.3.2.	Preparación del experimento	62
3.4.	Procedimiento experimental	63
4.	Análisis de Datos	65
4.1.	Definición de observaciones	65
4.2.	Descripción y procesamiento de datos	65
4.2.1.	Ventanas de tiempo	65
4.2.2.	Diámetro pupilar	67
4.2.3.	Respuesta electrodérmica (GSR)	67
4.2.4.	Electrocardiograma (ECG)	69
4.2.5.	Fotopletismograma (PPG)	69
4.2.6.	Temperatura (T)	70
4.2.7.	Electroencefalograma (EEG)	71
4.3.	Extracción de características	71
4.3.1.	Diámetro pupilar	72
4.3.2.	Respuesta electrodérmica (GSR)	72
4.3.3.	Electrocardiograma (ECG)	73
4.3.4.	Ritmo cardíaco (HR)	73
4.3.5.	Temperatura (T)	73
4.3.6.	Electroencefalograma (EEG)	73
4.4.	Clusterización y número óptimo de clusters	74
4.4.1.	Clusterización por participante	74
4.5.	Selección de características	76
4.5.1.	Umbral en Matriz de correlación	76
4.5.2.	Eliminación Recursiva de Características (RF-RFE)	77
4.6.	Clasificación por modelos de aprendizaje	82
4.6.1.	Preparación de conjuntos de entrenamiento y prueba	82

4.6.2.	Naive Bayes y Matriz de correlación	83
4.6.3.	Naive Bayes y RFE	84
4.6.4.	Redes Neuronales Artificiales y Matriz de correlación	84
4.6.5.	Redes Neuronales Artificiales y RFE	86
4.6.6.	SVM y RFE	87
4.6.7.	Deep Learning	88
5.	Resultados	89
5.1.	Análisis estadístico	89
5.2.	Clasificación	90
5.2.1.	Por algoritmo de aprendizaje	90
5.2.2.	Por sensores	91
	Conclusión	92
	Bibliografía	96
A.	Grupo experimental.	101
B.	Resultados de la clasificación. Por participante.	103
B.1.	Naive Bayes y selección con matriz de correlación.	103
B.2.	Naive Bayes y selección con RFE.	105
B.3.	Redes Neuronales Artificiales y selección con matriz de correlación.	107
B.4.	Redes Neuronales Artificiales y selección con RFE.	109
B.5.	SVM con kernel lineal y selección con RFE.	111
B.6.	SVM con kernel polinomial de orden 2 y selección con RFE.	111
B.7.	SVM con kernel polinomial de orden 3 y selección con RFE.	112
B.8.	SVM con kernel polinomial de orden 4 y selección con RFE.	112
B.9.	SVM con kernel polinomial de orden 5 y selección con RFE.	113
B.10.	SVM con kernel gaussiano y selección con RFE.	113
B.11.	Deep Learning.	114

Índice de Tablas

2.1.	Comparación entre los efectos del SN Simpático y parasimpático.	8
2.2.	Tabla comparativa de distintas investigaciones de medición de carga cognitiva.	42
4.1.	Características extraídas por cada señal.	71
4.2.	Validación clusters usando Media y Varianza.	75
4.3.	Lista de características obtenidas con selección RFE para cada participante.	77
4.4.	Resultados de clasificación usando Naive Bayes y matriz de correlación.	83
4.5.	Promedio de clasificación con Naive Bayes y matriz de correlación por cantidad de clases.	83
4.6.	Resultados de clasificación usando Naive Bayes y RFE.	84
4.7.	Promedio de clasificación con Naive Bayes y RFE por cantidad de clases.	84
4.8.	Resultados de clasificación usando Redes Neuronales Artificiales y matriz de correlación.	85
4.9.	Promedio de clasificación con Redes Neuronales Artificiales y matriz de correlación por cantidad de clases.	85
4.10.	Resultados de clasificación usando Redes Neuronales Artificiales y RFE.	86
4.11.	Promedio de clasificación con Redes Neuronales Artificiales y RFE por cantidad de clases.	87
4.12.	Promedio de clasificación con SVM y RFE para 2 de clases.	87
4.13.	Resultados de clasificación usando Deep Learning.	88
4.14.	Promedio de clasificación con Deep Learning por cantidad de clases.	88
5.1.	Resultados de ANOVA con medidas repetidas para la data completa.	89
5.2.	Resultados de ANOVA con medidas repetidas. Media del diámetro pupilar para ventanas de transición y ventanas activas.	90
5.3.	Resumen de los resultados de clasificación por algoritmo de aprendizaje.	90
5.4.	Resumen de los resultados de clasificación por sensor.	92
A.1.	Detalle del grupo experimental.	101
B.1.	Clasificación con Naive Bayes y matriz de correlación.	103
B.2.	Clasificación con Naive Bayes y RFE.	105
B.3.	Clasificación con Redes Neuronales Artificiales y matriz de correlación.	107
B.4.	Clasificación con Redes Neuronales Artificiales y RFE.	109
B.5.	Clasificación con SVM con kernel lineal y RFE.	111
B.6.	Clasificación con SVM con kernel polinomial de orden 2 y RFE.	111
B.7.	Clasificación con SVM con kernel polinomial de orden 3 y RFE.	112

B.8. Clasificación con SVM con kernel polinomial de orden 4 y RFE.	112
B.9. Clasificación con SVM con kernel polinomial de orden 5 y RFE.	113
B.10. Clasificación con SVM con kernel gaussiano y RFE.	113
B.11. Clasificación con Deep Learning.	114

Índice de Ilustraciones

2.1. Esquema básico del sistema nervioso.	8
2.2. Medidas psico-fisiológicas según su sistema biológico.	10
2.3. Modelo de Eye tracker glasses.	12
2.4. Método del Eye tracker.	12
2.5. Estructura del globo ocular.	13
2.6. Ondas cerebrales.	14
2.7. Sistema internacional 10-20 para la colocación de los electrodos para EEG.	15
2.8. Ejemplo de registro de EEG.	15
2.9. Esquema de posición de los electrodos para medir la señal GSR.	17
2.10. Señal característica GSR y sus componentes fásica y tónica.	17
2.11. Configuración de electrodos ECG Lead II.	18
2.12. Ciclo cardíaco ideal.	19
2.13. Diferencia de temperatura en imagen térmica.	20
2.14. Ilustración del algoritmo <i>k-means</i>	25
2.15. Generación de hiperplano SVM caso separable.	32
2.16. Generación de hiperplano SVM caso no separable.	33
2.17. Mapeo de características para el caso no lineal.	34
2.18. Ejemplo de red neuronal artificial.	35
2.19. Esquema de una neurona artificial.	36
2.20. Funciones de activación más comunes.	37
2.21. Función de activación lineal rectificadas.	40
3.1. Diagrama de pasos en un experimento de <i>machine learning</i>	47
3.2. Unidad <i>GSR+</i> de Shimmer.	48
3.3. Ejemplo de medición de señal GSR.	49
3.4. Ejemplo de posición de electrodos para el dispositivo <i>GSR+</i> de Shimmer.	49
3.5. Conexión de foto-sensor a la unidad <i>GSR+</i> de Shimmer.	50
3.6. Ejemplo de señal PPG medida con el sensor óptico de pulso de Shimmer.	51
3.7. Posición sensor en lóbulo de la oreja.	52
3.8. Unidad <i>Bridge Amplifier+</i> de Shimmer.	53
3.9. Sonda del sensor de temperatura corporal de Shimmer.	53
3.10. Tarjeta biomédica todo-en-uno de BITalino.	54
3.11. Ejemplo de posicionamiento de los electrodos para la medición de ECG.	55
3.12. Emotiv EPOC EEG.	56
3.13. Verificación de la posición del sensor Emotiv EPOC EEG.	57
3.14. Posición del sensor Emotiv EPOC EEG.	57

3.15. Tobii T120. Eye Tracker usado en los experimentos.	58
3.16. Ejemplo de mediciones en Tobii Studio.	59
3.17. Posición del Eye tracker Tobii T120.	59
3.18. Ejemplo de página ficticia para el experimento.	61
3.19. Laboratorio donde se llevan a cabo los experimentos.	62
3.20. Diagrama de sensores que se conectan a los participantes.	63
3.21. Participante con los sensores mientras ejecuta el experimento.	64
4.1. Histogramas de la cantidad de ventanas según su largo.	66
4.2. Señal de dilatación pupilar procesada.	67
4.3. Señal GSR procesada.	68
4.4. Componente física obtenida de la señal GSR.	68
4.5. Señal ECG procesada.	69
4.6. Señal PPG procesada.	70
4.7. Señal de temperatura corporal procesada.	70
4.8. Ejemplo de clusterización.	75

Capítulo 1

Introducción

1.1. Motivación

"Hay más información disponible en nuestras yemas de los dedos durante un paseo en el bosque que en cualquier sistema informático, sin embargo, la gente encuentra un paseo entre los árboles relajante mientras que las computadoras frustrantes. Máquinas que se ajustan al entorno humano en lugar de obligar a los seres humanos a entrar en el suyo, harán que usar una computadora sea tan refrescante como dar un paseo en el bosque." [M. Weiser The Computer for the 21st Century, 1991] [2]

El trabajo a realizar se contextualiza en el marco de computación ubicua. La computación ubicua es un concepto acuñado hace varios años por Mark Weiser [2] [3] mediante el cual, y gracias a que existen diversos dispositivos sensoriales y servicios computacionales en red distribuidos invisiblemente por el entorno, una persona puede llegar a sentir que el entorno es inteligente y se adapta a sus necesidades de una forma natural.

El desarrollo de la computación ubicua nos trae complejas redes de dispositivos sensoriales y de procesamiento de la información que en parte estarán dedicados a una sola persona para facilitar su vida, y todo ello de una forma amigable en la que la interacción hombre-máquina se convierta en totalmente humana, buscando como máxima la cita del principio.

Para alcanzar este objetivo es necesario que el entorno inteligente tenga una información sensorial muy rica que permita aplicar técnicas de inteligencia artificial y servicios ubicuos sobre distintas actividades cotidianas, como en el uso la navegación de sitios Web, para ir haciendo cada vez más natural el uso de la tecnología para el usuario.

De este modo, se espera no interrumpir, irritar o sobrecargar cognitivamente al usuario con servicios innecesarios, inoportunos o irrelevantes; prestando soluciones computacionales ubicuas pertinentes como: al contar con la información de la localización, la hora, el estado fisiológico del usuario, por ejemplo, deshidratado y hambriento; el estado psicológico, por ejemplo, contento; el estado económico, por ejemplo, no ha cobrado su sueldo; un teléfono inteligente pueda entregar una notificación adecuada como mostrar el mapa de la localización,

con la iluminación de pantalla adecuada para la hora, con propuestas de bares o restaurantes frecuentados por el usuario y baratos. Otro ejemplo sería que, al navegar por un sitio web, se detecte el nivel de carga cognitiva y el estado emocional y exista un servicio capaz de proveer recomendaciones y contenido relacionado a los intereses del usuario en el momento adecuado para no generar frustración.

Uno de los desafíos que se presenta es evitar la sobresaturación del usuario. Este punto se discute en la siguiente sección.

1.2. Definición del problema

La capacidad cognitiva es la cantidad total de información que el cerebro es capaz de retener en un momento determinado. Esta cantidad es finita. Junto con esta definición, en la psicología cognitiva, la carga cognitiva se refiere a la cantidad total de esfuerzo mental que se utiliza para aprender o procesar nueva información como: atención, percepción, memoria de corto y largo plazo, control motor, entre otros. [1].

Algunos factores que aumentan la carga cognitiva son: el aumento del estrés y la ansiedad, el hambre, la fatiga y procesar una gran cantidad de información. Mientras que algunos factores que disminuyen la carga cognitiva son: automatización, planificación, actividades rutinarias y un entorno organizado.

Cuando una persona ha alcanzado su capacidad cognitiva en un momento dado, se está saturado cognitivamente. Cuando se satura, el cerebro no puede procesar fácilmente cualquier información extra. Se necesita tiempo para que la saturación disminuya.

Aunque la tecnología informática suele estar justificada en términos de aumentar la productividad humana, a menudo tiene el efecto opuesto, creando situaciones de trabajo más complejas que reducen la productividad y aumentan la sobrecarga cognitiva (Landauer, 1995) [4].

Estudios han proporcionado pruebas de que la co-adaptación del usuario con la tecnología proporcionan un mecanismo eficaz para hacer frente a la sobrecarga cognitiva (Wendy, 2000) [5]. Considerar diseños apropiados de adaptación de la tecnología adyacente a los usuarios les ayudará a satisfacer sus propias necesidades y reducir el nivel de sobrecarga cognitiva.

En el contexto de páginas web un factor importante para medir la eficacia de un sitio es la experiencia del usuario (UX) que refiere a cómo la persona reacciona y se siente acerca de un sitio web. Un alto nivel de UX hace que el sitio sea más fácil e intuitivo de usar, lo que hace que la carga cognitiva sea menor, es decir, se relacionan inversamente. Mientras mayor sea el nivel de UX, menor es la carga cognitiva, lo que permite al usuario disfrutar más del sitio.

Aunque se busca disminuir la carga cognitiva es imposible eliminarla del todo. Toda página web posee dos tipos de carga mental: la intrínseca y la extraña.

La carga cognitiva intrínseca se relaciona con el esfuerzo natural y propio de absorber nueva información dada por el sitio, el proceso de aprendizaje de navegación y acostumbramiento al diseño de la página. Esta carga mental es imposible de eliminar.

Por otro lado, la carga cognitiva extraña corresponde a la carga mental originada por procesamientos que ocupan recursos mentales, pero que no ayudan a entender el contenido del sitio, usualmente originada por la adición de detalles innecesarios que hacen que el aprendizaje de navegación sea más difícil. Por ejemplo, diferentes estilos de fuente que no transmiten ningún significado. La carga cognitiva extraña puede tener efectos negativos en la UX lo que limita al usuario, que puede hasta abandonar el sitio. Por ello, los diseñadores se esfuerzan por eliminarla o minimizarla [57] [58].

Por ejemplo, Google tiene una carga cognitiva extraña muy baja. Al ir a su motor de búsqueda se sabe exactamente lo que sirve y qué acción tomar. Por otro lado, el programa Adobe Photoshop tiene una alta carga cognitiva extraña, porque se necesita tiempo para aprender a usarlo.

Si un sitio web es demasiado confuso y la carga cognitiva extraña es alta incluso llegando a la saturación cognitiva, los visitantes se sienten frustrados y se van. Por lo tanto, disminuir el problema de la sobresaturación cognitiva es un desafío dentro de la computación ubicua.

1.3. Contexto institucional

Este trabajo de título se desarrolla en el marco del proyecto Fondecyt otorgado en el año 2013 titulado: “*A Cognitive Resource-Aware Mobile Service Framework to Support Human-Computer-Interactions in Ubiquitous Computing Environments*” a cargo del profesor del Departamento de Ingeniería Industrial Ángel Jiménez Molina. En este proyecto se pretende desarrollar un mecanismo de ingeniería cognitiva que seleccione, componga y desarrolle funcionalidades durante el tiempo en que se ejecute, tomando en consideración el contexto situacional y los recursos cognitivos empleados por el usuario de acuerdo a las tareas de interacción humano-computador (en inglés *Human-Computer Interactions, HCI*) realizadas.

El objetivo general del proyecto Fondecyt es reducir el agotamiento de recursos cognitivos mientras se proveen servicios. Para esto los objetivos específicos que lo sustentan son: 1) Modelar relaciones entre recursos mentales, tareas HCI y actividades; 2) Definir métricas para evaluar agotamiento de recursos cognitivos; 3) Crear algoritmos de Binding (Algoritmo capaz de seleccionar el servicio idóneo a entregar según el contexto); 4) Diseñar un modelo computacional para analizar servicios según demandas cognitivas e interferencias; 5) Desarrollar un framework para dispositivos móviles que englobe todo lo anterior.

Esto se hace considerando la ubicación, los intereses y las preferencias de los actores de los procesos, así como la actividad en la que se encuentran inmersos, los factores ambientales, sociales y temporales y la información interpretada de sensores, entre otras fuentes de información contextual, y aplicables a diferentes dominios como mobile marketing, sistemas de recomendación, e-health y gestión en tiempo real de la cadena de suministros, entre otros.

La literatura principal para desarrollar este proyecto es: “*Cognitive resource-aware unobtrusive service provisioning in ambient intelligence environments*” [Angel Jimenez-Molina, and In-Young Ko, Journal of Ambient Intelligence and Smart Environments 7 (2015)] [44].

Esta fuente de información considera el estado del arte de la metodología de estudio de cargas cognitivas en usuarios de servicios tecnológicos. Es el trabajo reciente del profesor guía Ángel Jiménez y es un punto de partida para el proyecto Fondecyt, y por lo tanto para la memoria en curso.

Además, existe otro alumno desarrollando un tema paralelo del mismo Fondecyt. Su memoria ya finalizada se titula: “*Medición de cargas cognitivas durante actividades de interacción humano computador en ambiente móvil usando sensores psico-fisiológicos*” [Hernán Lira López (2015)] [46].

1.4. Objetivos

Dado el problema de la saturación cognitiva, se estudia el comportamiento de la carga cognitiva en un ambiente web.

Existe amplia evidencia en la literatura que las cargas cognitivas se pueden medir con señales psico-fisiológicas por medio de biosensores [27] [24][39][51] [52] [60] [56]. Por lo tanto, **el objetivo general de esta investigación es estudiar el comportamiento de la carga cognitiva de usuarios que navegan en un sitio Web, por medio de señales psico-fisiológicas capturadas por biosensores, para identificar patrones de estados de baja carga cognitiva.**

Los objetivos específicos de este trabajo son:

- Investigar sobre la teoría cognitiva, sensores psico-fisiológicos y la relación entre medidas fisiológicas con la carga cognitiva.
- Realizar un diseño experimental que permita evaluar el comportamiento de la carga cognitiva mediante sensores psico-fisiológicos.
- Procesar la señales psico-fisiológicas con el fin de disminuir el ruido, eliminar artefactos y obtener métricas respaldadas en la literatura para el análisis de la carga cognitiva.
- Evaluar la capacidad estadística de las señales psico-fisiológicas para medir la carga cognitiva realizando un tratamiento y análisis adecuado a los datos obtenidos desde los sensores psico-fisiológicos.
- Aplicar modelos de análisis no supervisados de minería de datos a las señales psico-fisiológicas, para identificar los distintos niveles de carga cognitiva a los cuales se enfrenta un usuario específico cuando navega en un sitio web.
- Aplicar modelos de análisis supervisados de minería de datos sobre las señales psico-fisiológicas, para clasificar niveles de carga cognitiva en un usuario específico cuando navega en un sitio web.
- Discutir los resultados y entregar posibles vías de desarrollo del trabajo futuro.

1.5. Hipótesis de trabajo y metodología

En la literatura existe evidencia que las cargas cognitivas se pueden medir con señales psico-fisiológicas por medio de biosensores para actividades estáticas. [27] [24][39][51] [52] [60] [56].

La hipótesis de esta memoria es que es posible medir la carga cognitiva para actividades de navegación web frente a un computador mediante señales psico-fisiológicas y que, además, se produce una baja en la carga mental del usuario en los momentos que pasa del análisis de un elemento web a otro mientras navega libremente.

La metodología para probar esta hipótesis es la siguiente (basada en las recomendaciones de Brouwer et al., 2015) [19]:

1. Estudio del estado del arte. Generar un respaldo teórico para entender el problema de la evaluación de la carga cognitiva, para conocer los sensores, métodos de obtención y procesamiento de señales psico-fisiológicas y análisis de datos con métodos de aprendizaje supervisados y no supervisados.
2. Desarrollo de diseño experimental y medición de las señales fisiológicas del usuario web con sensores como: Eye-tracker (Tobii), ECG (de BITalino), EEG (Emotiv EPOC), GSR, PPG y sensor de temperatura (de Shimmer Sensing).
3. Procesamiento de las señales usando filtros y otros métodos para disminuir el ruido de las señales y eliminar artefactos, usando el software disponible para cada sensor y, si es necesario, otro procesamiento de datos con Matlab.
4. Aplicación de modelos estadísticos y métodos de aprendizaje supervisado y no supervisados a los datos obtenidos con el objetivo de detectar y clasificar la carga mental de los usuarios.
5. Mediante análisis estadísticos adecuados junto con tablas y figuras de mérito como: exactitud (accuracy), precisión (precision) y sensibilidad (recall); evaluar el desempeño de la clasificación.
6. Conclusiones basadas en estos indicadores presentados en tablas comparativas para proporcionar información sobre la causa del éxito de la clasificación y para examinar el éxito de diferentes combinaciones de sensores/características.
7. Conclusiones que expliquen cómo las medidas psico-fisiológicas agregan valor sobre el uso de otras medidas más fáciles y baratas por sí solas. Proponer aplicaciones que se beneficiarán del uso de medidas psico-fisiológicas para la estimación del estado mental.

1.6. Resultados esperados y alcances

Para este trabajo se esperan los siguientes resultados:

1. Estado del arte de la medición de carga cognitiva con sensores psico-fisiológicos.
2. Diseño experimental apropiado y replicable para la estimación de carga cognitiva mediante señales psico-fisiológicas en un ambiente web.
3. Procesamiento de las señales para mejorar su calidad para la medición de carga cognitiva.
4. Modelo estadístico que permita aceptar o rechazar la hipótesis de investigación: Se debe comprobar que la carga cognitiva, en los momentos que el usuario pasa de un elemento web a otro, es significativamente menor que la carga cognitiva mientras está analizando un elemento.
5. Tabla comparativa de desempeño de distintos clasificadores de carga mental.
6. Tabla comparativa de desempeño de clasificación tomando diferentes combinaciones de sensores.

Este trabajo no considera aplicar modelos que no hayan sido utilizados en trabajos previos ni entregar en detalle aplicaciones específicas del conocimiento desarrollado ni de los resultados obtenidos.

Capítulo 2

Revisión Bibliográfica y Contextualización

El presente capítulo tiene como objetivo ubicar al lector en el entorno en el cual se desarrolla este trabajo de título, entregando los antecedentes previos y necesarios para su contextualización.

Se comienza con la descripción del sistema nervioso en la sección 2.1 mencionando brevemente su estructura. Luego, en la sección 2.2 se describen las principales metodologías de medición de carga cognitiva. En la sección 2.3 se describe cada una de las medidas psicofisiológicas a utilizar, sus formas de medición y cómo se relaciona con la carga cognitiva.

En la sección 2.4 se describen dos métodos de selección de características. En las secciones 2.5 y 2.6 se describe un método de clusterización y los criterios de validación. En las secciones 2.7 y 2.8 se describen los métodos de aprendizaje supervisado, que se utilizan en este trabajo.

Por último en la sección 2.9 se comenta el estado del arte sobre la investigación de la carga cognitiva en ambientes web, cómo afecta al usuario y algunas aplicaciones en esta área.

2.1. Sistema Nervioso

El sistema nervioso está formado por el Sistema Nervioso Central (SNC) que está compuesto principalmente por el cerebro y la espina dorsal, y funciona como el “centro de comandos”; y el sistema nervioso periférico (SNP) que corresponde a todos los nervios y ganglios fuera del SNC, cuya función es de “red de mensajería” enviando al SNC las señales obtenidas de los sensores nerviosos y respuestas motoras de éste a los músculos [29].

A su vez, el SNP se divide en el Sistema Nervioso Somático (SNS) que controla la actividad muscular voluntaria; y el Sistema Nervioso Autónomo (SNA) que se encarga de las funciones viscerales como la de glándulas u órganos. Esta relación se explica en la Figura 2.1.

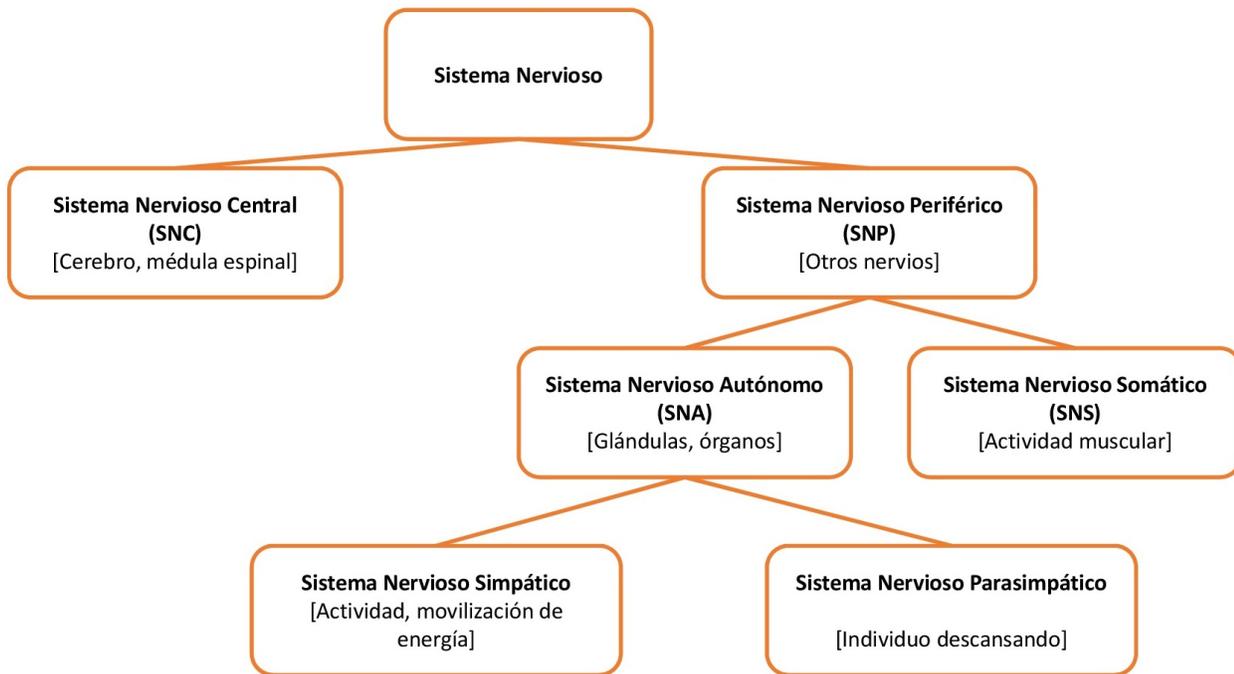


Figura 2.1: Esquema básico del sistema nervioso.

2.1.1. Sistema Nervioso Autónomo

En particular, el Sistema Nervioso Autónomo (SNA) es el encargado de controlar la acción involuntaria de los músculos (asociada a la musculatura lisa), del corazón y de varias glándulas. Posee dos subdivisiones: El sistema nervioso simpático y el parasimpático. Ambos suelen operar de forma antagónica. Por ejemplo, la activación del sistema simpático incrementa el ritmo cardíaco, dilata los bronquios (genera una respiración más rápida), estimula la glándula suprarrenal (liberando adrenalina) y prepara el cuerpo para la acción. En cambio, el parasimpático disminuye el ritmo cardíaco, baja la frecuencia respiratoria contrayendo los bronquios, relajando el cuerpo. En la Tabla 2.1 se compara la actividad de estos sistemas.

Tabla 2.1: Comparación entre los efectos del SN Simpático y parasimpático [20].

Función	Simpático	Parasimpático
Pupila	Dilatación	Contracción
Salivación	Inhibición	Estimulación
Pulso	Más rápido	Más lento
Respiración	Incremento	Disminución
Digestión	Inhibición	Estimulación

El sistema nervioso autónomo envuelve los reflejos y los comportamientos involuntarios, principalmente las respuestas de bajo nivel de conciencia. Esto es muy usado para interpretar niveles emocionales, de excitación o cognitivos, y existen múltiples señales psico-fisiológicas

que nos permite conocer el estado del SNA. Por ejemplo, podemos medir los cambios en la conductividad de la piel, que se relaciona con el grado de humedad dado por la actividad de glándulas sudoríparas que a su vez son controladas por el SNA.

2.2. Medición de carga cognitiva

La carga cognitiva se define como el nivel percibido de esfuerzo que se requiere para la realización de una tarea como una medida del trabajo mental producido durante su ejecución. El concepto de carga cognitiva está estrechamente relacionado con el concepto de los recursos mentales limitados.

La Teoría de Recursos Múltiples (Wickens, 2002) [6] dice que los seres humanos tienen un conjunto finito de recursos mentales y que son de varios tipos; que pueden utilizarse en forma paralela por múltiples tareas y que las tareas pueden usar varios recursos a la vez. Cuando la demanda de una tarea es alta, los recursos destinados a esa tarea no están disponibles a una segunda tarea si se requiere el mismo tipo de recursos mentales y en la misma etapa de procesamiento.

Para medir la carga cognitiva existen 3 enfoques: métodos por rendimiento, métodos subjetivos y medidas psico-fisiológicas [27]. Estos se describen a continuación.

2.2.1. Medición por rendimiento

Las mediciones de la carga cognitiva por rendimiento son métodos en donde el investigador debe medir el desempeño del sujeto durante la actividad, esto es midiendo el tiempo que demora en realizarla. Estos métodos tienen la restricción de que el sujeto debe estar enfocado en una actividad, con lo que realizar más de una tarea no es un escenario ideal [24].

2.2.2. Medición subjetiva

Los métodos subjetivos incluyen escalas de autopercepción, protocolos de pensamiento en voz alta, y las entrevistas post-tarea. Pueden arrojar luz sobre el estado cognitivo del usuario y son importantes para evaluar la percepción de los usuarios de la tarea. Mediciones subjetivas son típicamente tomadas después de completar una tarea y son, por tanto, estáticas. La naturaleza estática de estos métodos los hace inadecuados para evaluar los cambios dinámicos de la carga cognitiva [7].

A modo de ejemplo, uno de los instrumentos más conocidos y más utilizados es el índice de carga de tareas de la NASA (TLX, Task Load Index) (Hart & Staveland, 1988) [38]. En este método subjetivo de estimación de la carga cognitiva se evalúa el desempeño mental, físico, demandas según el tiempo, esfuerzo y frustración de la persona.

Consiste en un conjunto de ítems presentados en forma de afirmaciones o juicios, ante los cuales se pide la reacción de los participantes evaluando en una escala de 5 puntos. Se obtiene una puntuación por pregunta y al final una puntuación total dependiendo del peso relativo de cada variable. Sus ventajas son la rapidez y facilidad en su uso y su validez aceptada. Además, ofrece la ventaja frente a otros métodos, como los de valoración psico-fisiológica de no ser intrusivos [11].

2.2.3. Medición psico-fisiológica

Las técnicas de medición fisiológicas son objetivas e incluyen la electroencefalografía (EEG), el seguimiento ocular, entre otros. En particular, los cambios en diámetro de la pupila se han utilizado para la evaluación de la carga en los estudios en ciencias de la información y en HCI (por ejemplo, Iqbal, Zheng, y Bailey, 2004; Tungare y Pérez-Quñones, 2009) [7]. Las medidas psico-fisiológicas permiten detectar instantáneamente cambios en la carga cognitiva, por lo que es ideal si se quiere medir en tiempo real y para tareas duales o múltiples. Las dificultades que presenta son la incomodidad de utilizar los sensores y el complejo procesamiento de los datos [25].

En esta metodología se mide principalmente las reacciones del sistema nervioso (descrito en la sección 2.1). Cada parte del SN se relaciona con alguna(s) medida(s) psico-fisiológica(s) en particular. En la Figura 2.2 se presenta un mapa conceptual de las medidas psico-fisiológicas según el sistema biológico, que se pueden medir para estimar la carga cognitiva.

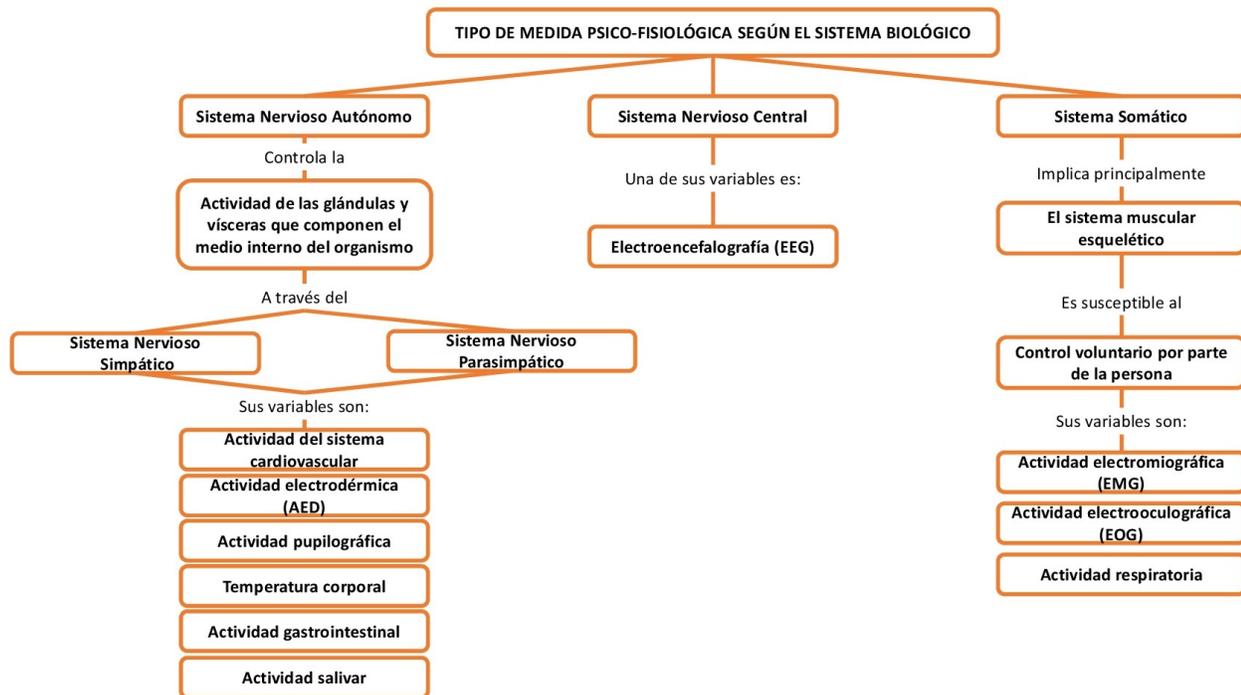


Figura 2.2: Mapa conceptual de medidas psico-fisiológicas según su sistema biológico.

2.3. Medidas psico-fisiológicas

Dentro de las medidas psico-fisiológicas que permiten estimar la carga cognitiva de una persona, se han escogido las siguientes para este trabajo:

- Seguimiento ocular
- Actividad pupilográfica
- Electroencefalografía
- Actividad electrodérmica
- Actividad del sistema cardiovascular
- Temperatura corporal

Cada una de ellas se describe teóricamente a continuación.

2.3.1. Seguimiento ocular [54] [15]

El seguimiento ocular (*eye tracking*, en inglés) es el proceso de determinar dónde una persona está mirando. Es posible medir las características de ese movimiento ocular y el ojo en sí mismo.

Algunas características del movimiento ocular son:

- Sacadas: Rápido movimiento o salto del ojo humano entre un punto de fijación y otro. (Los movimientos de sacadas siguen un patrón (o ruta) dependiendo de varios factores: qué se está mirando actualmente, el objetivo que se busca, la experiencia y las emociones).
- Fijaciones: Momento en el cual se mantiene la vista relativamente fija o enfocada. En este momento se extrae la información visual. Se produce debido a que la nitidez de la vista sólo se obtiene en un punto determinado del rango de visión humana llamado fóvea.

El seguimiento ocular permite obtener información sobre los momentos de fijación (captura la visión foveal). Esto es beneficioso porque, en la mayoría de los casos, los momentos de fijación coinciden con los de atención.

El dispositivo para realizar el seguimiento de la mirada es el *eye tracker* y existen distintos tipos, como por ejemplo, lentes o una cámara en la pantalla del computador. Éste funciona según el método “*corneal-reflection/pupil-centre*” (Goldberg & Wichansky, 2003) [32]. Este dispositivo también permite medir la variación del diámetro pupilar. Un ejemplo de un *eye tracker glasses* se muestra en la Figura 2.3.



Figura 2.3: Modelo de *Eye tracker glasses*.

En palabras sencillas, se mide la reflexión que se genera en la córnea y en la retina (que atraviesa la pupila) y se realiza una medición relativa entre éstas para determinar el punto donde está fijando la vista el usuario. Esto generalmente se observa con una luz infrarroja que emite el *eye tracker*. Un esquema de este método se presenta en la Figura 2.4.



Figura 2.4: Esquema de método de determinación de la mirada con el *eye tracker*.

2.3.2. Actividad pupilográfica [8]

Es importante conocer la estructura del sistema visual pues su reacción nos da información sobre distintos estados de un individuo. Entre las mediciones que se pueden realizar a este sistema está la respuesta pupilar.

Las estructuras principales del sistema visual son (ver Figura 2.5):

- Globo ocular: Estructura mayor que da forma al ojo, de forma esferoidal.
- Córnea: Estructura transparente que permite la refracción de la luz con el fin de que se pueda enfocar la imagen en la retina.
- Iris: Tiene la forma de un disco perforado en su centro por un orificio circular, la pupila.

- Pupila: Orificio de tamaño variable. Su diámetro normal se encuentra comprendido entre 2 y 4 mm, pudiendo alcanzar los 8 mm en dilatación máxima y 0,5 mm en contracción máxima.
- Retina: Capa interna del globo ocular. En esta se inicia el proceso de visión con múltiples terminaciones nerviosas que recogen, elaboran y transmiten sensaciones visuales.

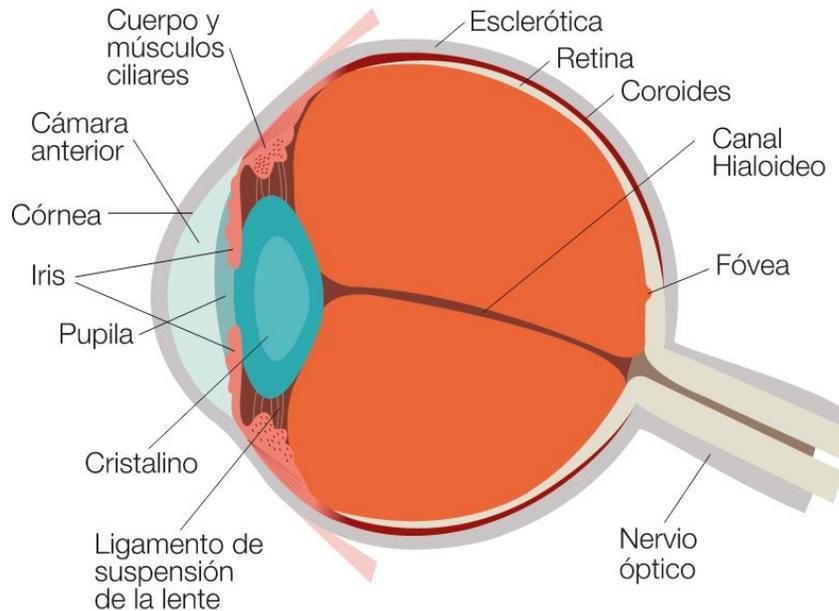


Figura 2.5: Estructura del globo ocular.

La pupila es la encargada de regular la luz entrante del exterior a nuestro globo ocular, de la que dependen sus cambios de tamaño, generalmente comprendido entre 2 mm (condiciones de alta luminosidad) y 8 mm (oscuridad). El iris, músculo y tejido pigmentado que rodea la pupila, es el responsable de los cambios en el tamaño pupilar a través de dos grupos de músculos (Beatty y Lucero-Wagoner, 2000) [13].

Según Beatty (1986) y Beatty y Lucero-Wagoner (2000) los cambios en el tamaño de la pupila se deben a tres tipos de reflejos: el reflejo de luz (se produce ante cambios de iluminación en el ambiente; los aumentos de niveles de luz producen una rápida contracción en la pupila), el reflejo de cercanía o acomodación (tiene como función mejorar el enfoque visual) y el reflejo psico-sensorial.

El reflejo psico-sensorial consiste en una dilatación pupilar asociada a una tarea o estímulo relevante presentado en cualquier modalidad sensorial. A diferencia de los dos reflejos anteriores, en este caso los cambios en el tamaño de la pupila son más sutiles, están mediados por el sistema nervioso simpático y no se deben a las propiedades físicas del estímulo sino a la implicación cognitiva y emocional del procesamiento.

En cualquier caso y de forma general, en la respuesta pupilar intervienen componentes de inhibición parasimpática (que explican la primera fase de dilatación) y de activación simpática (que explican la fase de contracción posterior) (Kuipers y Thierry, 2013; Steinhauer, Siegle, Condray y Pless, 2004).

2.3.3. Electroencefalografía (EEG) [27] [50]

La electroencefalografía (EEG) es el registro de la actividad eléctrica del cerebro medida en el cuero cabelludo. El dispositivo llamado electroencefalógrafo mide la tensión resultante de los cambios de flujo de corriente iónica dentro de las neuronas del cerebro producidos por la actividad sináptica cerebral.

Hay cinco grandes ondas cerebrales que se distinguen por sus diferentes bandas de frecuencia. Estas bandas de frecuencia se llaman, de baja a alta frecuencia: Delta (1-4 Hz); Theta (4-8 Hz); Alfa (8-12 Hz); Beta (12-25 Hz) y Gamma (sobre 25 Hz). En la Figura 2.6 se muestra un esquema de cada una de estas ondas cerebrales según su banda de frecuencia. [50].

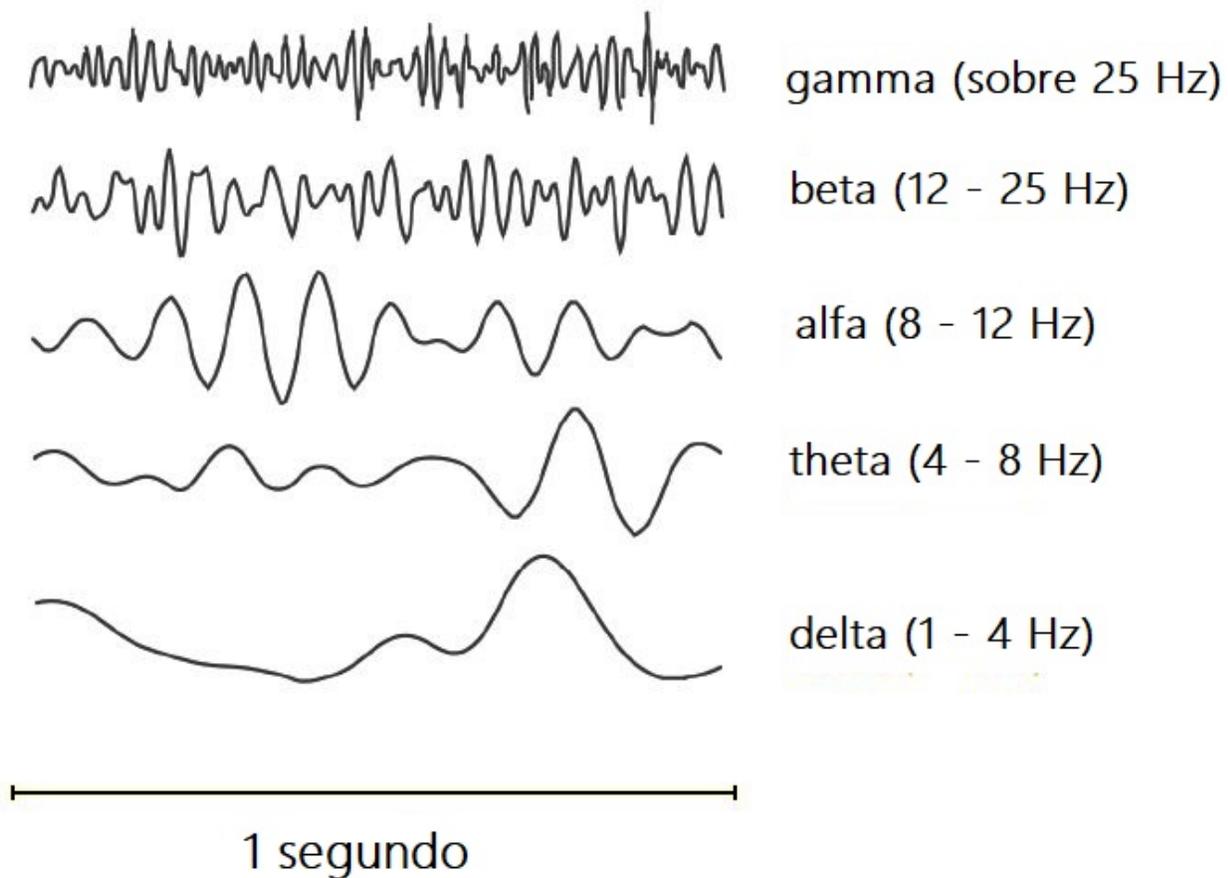


Figura 2.6: Esquema de ondas cerebrales clasificadas en bandas de frecuencia [40].

Existen sistemas internacionales estandarizados de colocación de electrodos para EEG. En la Figura 2.7 se muestra el sistema conocido como 10-20. Cada sitio tiene una letra para identificar el lóbulo cerebral y un número para identificar la ubicación dentro de los hemisferios. Cada uno de los electrodos registra una onda distinta como se observa en la Figura 2.8.

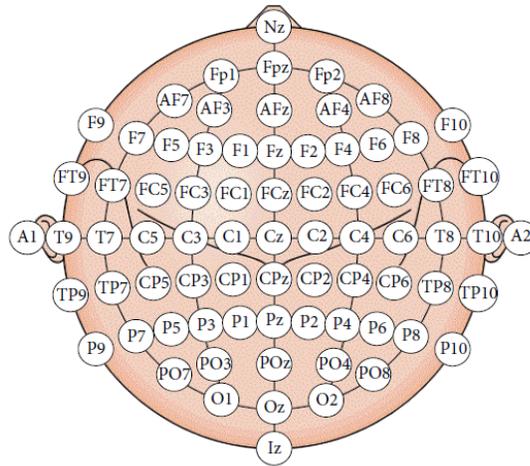


Figura 2.7: Sistema internacional 10-20 para la colocación de los electrodos para EEG.

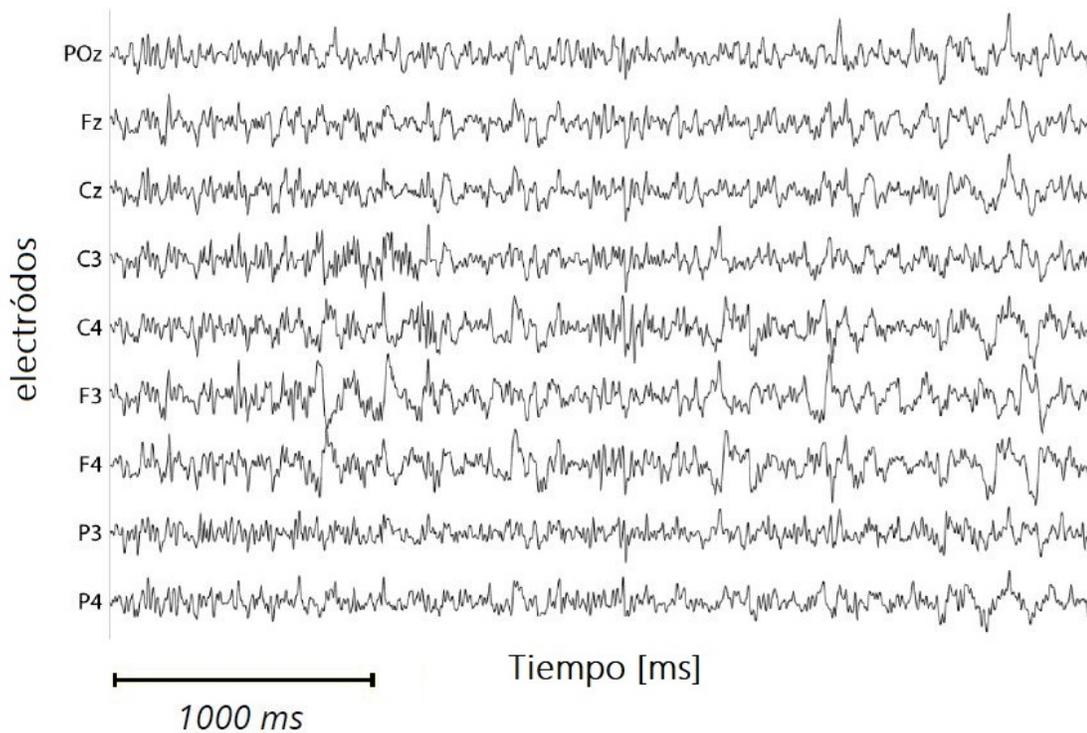


Figura 2.8: Ejemplo de registro de EEG. Se observa la actividad cerebral para distintos electrodos según el sistema internacional 10-20.

Para clasificar estados emocionales, Noppadon [50] somete a varios usuarios a una serie de estímulos de imágenes y música clásica de marcada valencia positiva y negativa, para obtener un estímulo feliz y uno triste. Para obtener las bandas de frecuencia descompone las ondas cerebrales usando wavelets. Luego, clasifica con máquina de vector de soporte (SVM) con kernel Gaussiano obtiene un acierto de 75,9%.

En su experimento determina que las bandas de alta frecuencia, como Beta y Gamma, son más efectivas para determinar el estado emocional de felicidad que las de baja frecuencia. Por lo tanto, concluye que para determinar estados emocionales se pueden omitir las bandas Delta y Theta.

Por otro lado, Fritz et al. [27], cuya hipótesis es que la señal EEG sirve para predecir si una tarea era fácil o difícil para el usuario, menciona que muchos estudios han encontrado una baja en la actividad de la banda Alfa junto con, la mayoría de las veces, un aumento en la actividad de la banda Theta va acompañado de un incremento en la demanda atencional y en la carga de trabajo de la memoria. Además menciona que, otros estudios han concluido que las bandas Theta y Delta son sensibles ante estímulos donde se manipula la dificultad.

Finalmente [27] utiliza la relación $Beta/(Alfa + Theta)$ basado en la evidencia de que aumenta Beta con la ejecución de la tarea, Theta se suprime y Alfa se bloquea. Ante el experimento de exponer al usuario a 8 tareas de comprensión de código variando la dificultad tuvo un 85 % de precisión en determinar la dificultad de una tarea usando *Machine Learning*.

Por lo tanto, las ondas cerebrales tienen un amplio espectro de trabajo para explorar en el campo de clasificación de estados emocionales y de carga mental. Se puede experimentar con variadas mediciones, relaciones y funciones sobre las bandas de frecuencia para la extracción de características y determinar los estados de carga cognitiva.

2.3.4. Actividad electrodérmica [56] [51]

La actividad electrodérmica se define como el cambio en las propiedades eléctricas de la piel. También conocida como respuesta galvánica de la piel (GSR, *Galvanic Skin Response*), es de bajo costo, capturada fácilmente y robusta. En este método, la conductancia eléctrica de la piel se mide a través de uno o dos sensores o electrodos usualmente unidos a una parte de la mano o el pie. La conductividad de la piel varía con los cambios en el nivel de humedad de la piel (sudor) y puede revelar cambios en el sistema nervioso simpático.

La resistencia eléctrica de la piel es compleja: depende por un lado de la frecuencia de la corriente eléctrica (por ejemplo, para la corriente continua presenta un valor entre 150 $k\Omega$ a 250 $k\Omega$) y también varía según la humedad de ésta que cambia según la reacción de las glándulas sudoríparas controladas por el sistema nervioso simpático. Es un indicador de la excitación psicológica y fisiológica, además sirve para reconocer estados emocionales. Por ejemplo, si aumenta la excitación, aumenta la actividad de las glándulas lo que disminuye, a priori, la resistencia eléctrica (aumenta la conductividad).

La ubicación de los electrodos requiere una zona con gran cantidad de glándulas sudoríparas, como por ejemplo la zona palmar de la mano y los dedos. Se proponen distintas opciones como la falange distal, falange medial o falange proximal. Lo importante es evitar al máximo los movimientos de los electrodos por lo que deben ser firmemente fijados. En la Figura 2.9 se presentan dos opciones: ubicar los electrodos en la falange medial en dedos distintos (electrodos A-B); o bien, en la palma de la mano (electrodos C-D).

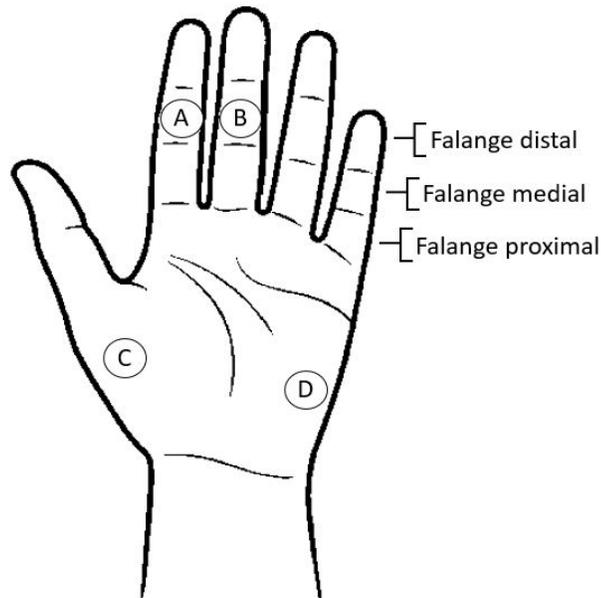


Figura 2.9: Esquema de posición de los electrodos para medir la señal GSR [42].

La señal GSR cuenta con 2 componentes: el componente fásico que tiene cambios más rápidos y está relacionado con los estímulos externos y la actividad poco específica; y la componente tónica, que es la señal base y varía más lentamente, fija el nivel de conductancia de la piel. Un ejemplo de la forma característica de la señal y de sus componentes se presenta en la Figura 2.10. El principal desafío es interpretar de esta señal las razones de los incrementos y disminuciones en los niveles de sudor. Alguno de los factores que afectan la señal son: Temperatura y sequedad de la piel. Por la baja tasa de variación de esta señal el tiempo de muestreo va entre 0 y 5 Hz.

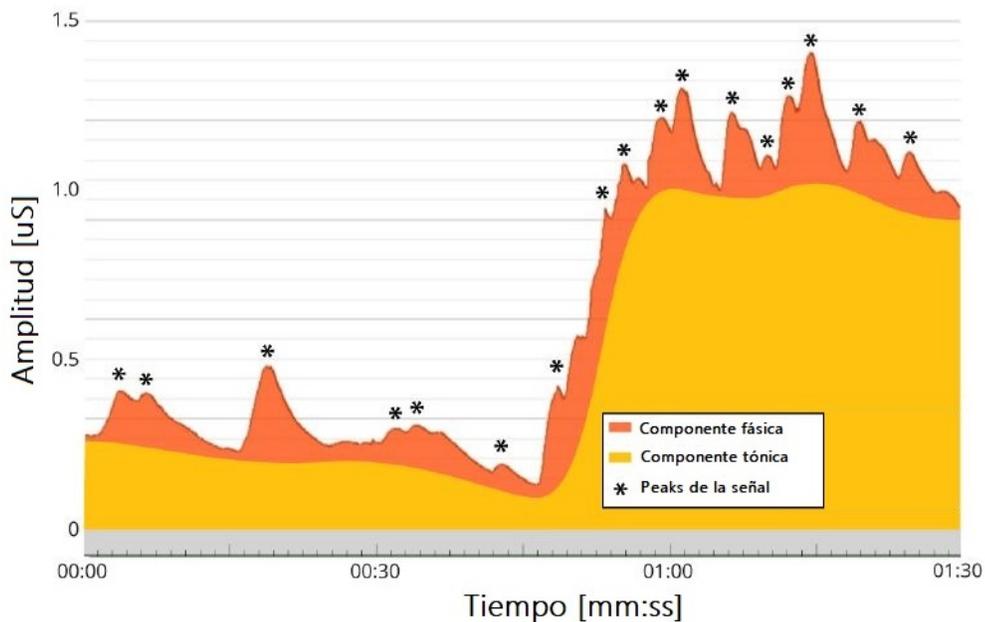


Figura 2.10: Señal característica GSR y sus componentes fásica y tónica [42].

2.3.5. Actividad del sistema cardiovascular

La electrocardiografía (ECG) consiste en el registro gráfico de la actividad eléctrica del corazón. El dispositivo que se utiliza para capturar esta señal es el electrocardiógrafo que consta de tres o más electrodos que se pegan en el torso del sujeto y que recopilan los datos necesarios para obtener las ondas que describen el ciclo cardíaco. Un ejemplo de la configuración de un electrocardiógrafo y la señal obtenida se observa en la Figura 2.11 donde se presenta el estándar Lead II de posición de electrodos.



Figura 2.11: Configuración de electrodos ECG Lead II y ejemplo de forma de señal obtenida con esta configuración.

El ciclo cardíaco corresponde a los eventos que ocurren en el corazón desde un latido hasta el siguiente. El ciclo está compuesto por dos etapas principales: diástole durante la cual el corazón se llena de sangre y sístole cuando el corazón bombea toda esa sangre al cuerpo [20].

El ciclo comienza cuando el nodo sinoauricular (SA) se despolariza en la aurícula derecha durante la última parte de la diástole. Esta primera onda de despolarización corresponde a la onda P en la señal eléctrica generada por el corazón. La onda P es seguida por una contracción auricular en donde aparece la onda QRS reflejando el comienzo de la sístole. Luego, se produce la contracción ventricular, en donde la presión ventricular es lo suficientemente alta para cerrar las válvulas auriculoventriculares entre las aurículas y los ventrículos. Luego, la presión ventricular disminuye a causa del aumento de presión auricular con lo que se abren las válvulas y los ventrículos se llenan de sangre. Cuando la presión ventricular es mayor que la presión de la arteria aorta se abre la válvula y la sangre se dirige a la circulación general. Ocurrido esto los ventrículos se repolarizan y se observa la onda T con lo que vuelve a comenzar la diástole [20].

El tiempo entre el comienzo de la onda P y el comienzo de la onda QRS es alrededor de 160 mseg. El intervalo Q-T es alrededor de 300 mseg. Un ciclo cardíaco basado en una tasa de 72 latidos por minuto dura 830 mseg por lo que hay aproximadamente 370 mseg entre el fin de la onda T y el comienzo de la siguiente contracción auricular [20]. Un ejemplo de un ciclo cardíaco ideal se presenta en la Figura 2.12.

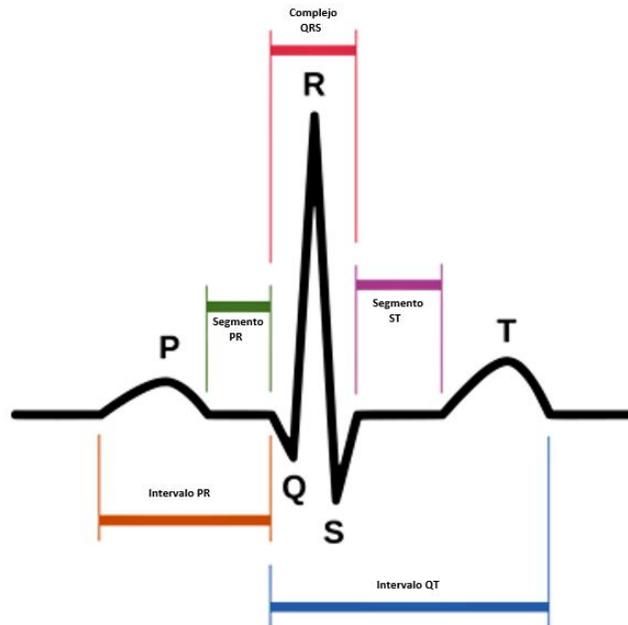


Figura 2.12: Ciclo cardíaco ideal.

En un ECG normal se producen ondas R cuyo rango de amplitud es aproximadamente 2 mV. El peak de 2 mV debe ser amplificado por 2500 para poder trabajar.

El período cardíaco se define como el tiempo en milisegundos entre dos latidos consecutivos. Típicamente se mide entre dos sucesivas ondas R en un ECG. Habitualmente el período cardíaco se convierte a pulso cardíaco en donde se mide la cantidad de latidos por minuto.

2.3.6. Temperatura corporal [31]

El cuerpo humano realiza constantes intercambios de temperatura con el ambiente, en forma de energía, como parte del proceso de autoregulación para mantener la homeostasis (equilibrio interno del cuerpo) que necesita (Ring, 2007). Esta intensidad de energía intercambiada crece y decrece en proporción a la temperatura corporal. La regulación del flujo vascular cutáneo y de la radiación térmica se considera una función del Sistema Nervioso Autónomo (Leal Brioschi, 2006).

Para el estudio de la relación entre la carga mental y los cambios fisiológicos del cuerpo humano (en particular empleando termografía, medición gráfica de la temperatura de un individuo, por ejemplo, una cámara infrarroja), se trata de relacionar cambios de la temperatura facial (respuesta fisiológica) como índice objetivo de la carga mental (indicador de lo que sucede en el cerebro). Entre las respuestas del Sistema Nervioso Autónomo que encontramos con esta medición están la fatiga y el estrés. Estudios realizados en esta área (Genno, et. al., 1997) consideran que la temperatura de la piel tiene un gran potencial como medida fisiológica sin necesidad de tener contacto físico con el individuo.

En sus experimentos Genno, et. al., (1997) considera que la nariz es una zona apropiada pues tiene la ventaja de no tener movimiento y se confirma que varía su temperatura en condiciones de estrés. También se encuentra que la frente es una zona de poca variación (Stoll, 1964) por lo que es un punto de comparación. Siempre se mide la variación de la temperatura de la piel debido a la respuesta del Sistema Nervioso Autónomo ante la respuesta de estrés. En estos estudios se observa que la temperatura de la nariz baja en dichas condiciones producto de la vasoconstricción de los vasos capilares periféricos de la nariz; y se detecta un ascenso de la temperatura en período de descanso. Se sugiere un control exhaustivo de la temperatura ambiental para evaluar los cambios en la temperatura corporal. Se propone que esta reacción fisiológica ante un aumento en la carga cognitiva es porque el individuo necesita mayor ventilación, aunque los resultados no son concluyentes.

Otros métodos de medición de la temperatura corporal es con los termistores, que miden la temperatura por conducción, a diferencia de la termografía que mide la temperatura a partir de las ondas electromagnéticas infrarrojas emitidas por el cuerpo.

En la Figura 2.13 se presenta un experimento realizado por Gomez (2015) [31] para comprobar lo estipulado por Genno, et. al. (1997). En la imagen térmica de un mismo individuo en dos escenarios, donde el primero es en estado de descanso o reposo y en el segundo sometido a un experimento de carga mental. Se observa lo relativamente constante que es la temperatura de la frente, que es la zona de referencia; y además, la caída de la temperatura de la nariz tras ser sometido a un estímulo que aumenta su carga mental (en la región de interés marcada se observa la diferencia en la punta de la nariz). También se marcan las regiones de interés para el análisis de imágenes.

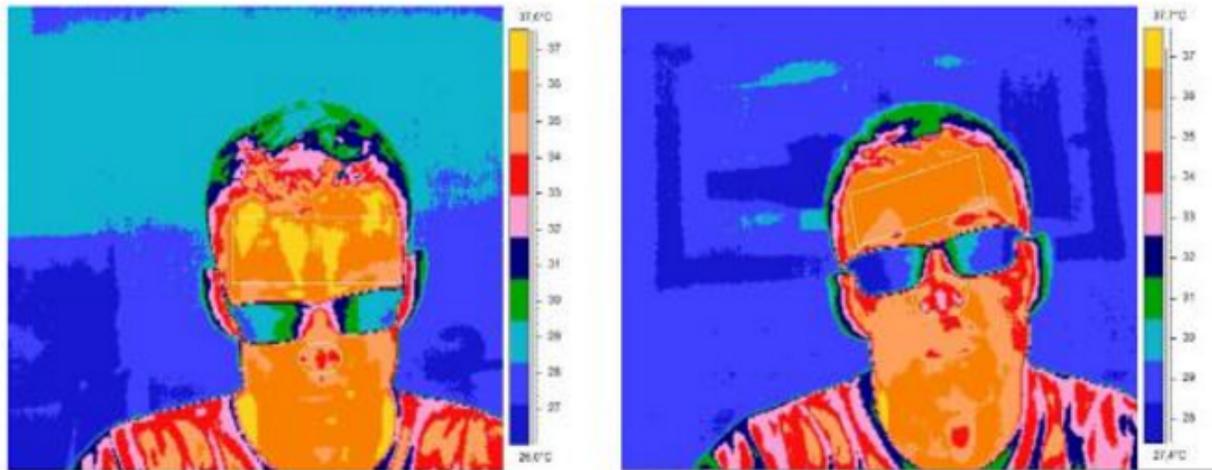


Figura 2.13: Ejemplo de imagen térmica de un sujeto en reposo (primera) y al finalizar un experimento de carga cognitiva (segunda). Se observa la diferencia en la temperatura de la nariz [31].

2.4. Selección de características [34][35]

El uso de métodos de selección de características tiene muchos beneficios potenciales, entre ellos: facilitar la visualización y comprensión de los datos, reducir los requisitos de medición y almacenamiento, reducir los tiempos de entrenamiento y utilización, desafiar la maldición de la dimensionalidad para mejorar el rendimiento de la predicción, mejorar el desempeño de predictores o clasificadores, entre otros.

Existen muchos métodos y algunos de ellos ponen más énfasis en un aspecto que en otro. A continuación se describirán brevemente dos métodos de selección de características.

2.4.1. Criterio de correlación

Un enfoque para la selección de características es observar su relevancia individual. Un ejemplo es el criterio de correlación. Este método es considerado rápido y eficaz, particularmente cuando el número de características es grande y el número de ejemplos de entrenamiento disponibles es comparativamente pequeño.

El **coeficiente de correlación de Pearson** es un índice de relevancia clásico utilizado para el ranking individual de características. El coeficiente de correlación de Pearson se define como (ver Ecuación 2.1):

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sum_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_i (x_i - \bar{x}_i)^2 \sum_j (y_j - \bar{y}_j)^2}} \quad (2.1)$$

Cuando las características son correlacionadas, la separación de clase no mejora significativamente mediante el uso de dos características en lugar de una. Sin embargo, el criterio de correlación sólo pueden detectar dependencias lineales entre la variables.

Para un conjunto de n características, del cual se busca reducir su dimensionalidad eliminando las características redundantes, se define una **matriz de correlación** que resume la correlación de cada par de características. Como se observa en la Ecuación 2.2 se mantiene una diagonal de 1 (correlación perfecta con sí misma) y en el resto de la matriz las distintas correlaciones para todas las combinaciones.

$$R = \begin{bmatrix} 1 & \rho(X, Y) \\ \rho(X, Y) & 1 \end{bmatrix} \quad (2.2)$$

El índice de correlación va desde 0 y 1 donde, por lo general, índices por sobre 0,6 se considera que las variables que se comparan están muy correlacionadas y se sugiere eliminar una. Con esto se cumple el objetivo de seleccionar las características que más aporten información y así reducir la dimensionalidad del conjunto.

2.4.2. Eliminación recursiva de características

La eliminación recursiva de características (RFE, *Recursive Feature Elimination*) es un procedimiento de eliminación hacia atrás, es decir, se comienza con todas las características y se eliminan progresivamente las menos útiles. Puede producir buenos resultados pero a expensas de conjuntos de características posiblemente más grandes.

El método RFE es un proceso iterativo que elimina características según un criterio. Consiste en 3 pasos: (1) entrenar un clasificador; (2) calcular el criterio de puntuación; (3) eliminar las características con menor puntaje [47].

La primera aproximación al método de eliminación recursiva de características fue propuesta por Guyon et al. (2002) [36] y está basada en la máquina de soporte vectorial (SVM). Se ha aplicado en varios estudios de expresión génica para seleccionar características relevantes. También se ha aplicado a señales EEG en estudios de BCI (Brain-Computer Interfaces) [37].

En este caso se usa SVM como clasificador (ver Sección 2.7.2). La idea es definir la importancia de una característica en términos de su contribución al SVM según una función de decisión. En cada paso del procedimiento RFE, el SVM es entrenado con el conjunto de datos dado, se calcula la función de decisión D (ver Ecuación 2.3), y la característica con menor contribución se descarta. SVM debe ser reentrenado en de cada iteración.

Por lo tanto, el método se reduce a la simple eliminación iterativa de la característica x_i con el menor peso en valor absoluto $|w_i|$. Los pesos w_i de la función de decisión se obtienen de un subconjunto de ejemplos de entrenamiento llamados vectores de soporte. La función de decisión se define como:

$$D(x_i) = w_i \cdot x_i + b \quad (2.3)$$

Como se observa en la Ecuación 2.3 el peso w_i es una combinación lineal de los datos de entrenamiento. (Comparar con Ecuación 2.14).

SVM-RFE usa la magnitud del peso como criterio de puntuación. A cada iteración, la característica con la puntuación más baja es eliminada. Guyon (2002) [36] propone como criterio $c_i = (w_i)^2$ y eliminar las características que contribuyen menos al margen de maximización [37].

RF-RFE [45]

Se ha desarrollado un método alternativo de RFE donde básicamente se reemplaza SVM por Random Forest (RF). Kai-Quan Shen (2007) [45] combina RF con RFE para mejorar la selección de características en la clasificación de fatiga mental usando señales EEG.

Random Forest es un conjunto de varios árboles de decisión. Cada uno de ellos se construye usando un conjunto bootstrap (es decir, datos obtenidos con remuestreo estadístico) que corresponden a dos tercios de los datos de entrenamiento originales; y el tercio restante se deja fuera como casos *out-of-bag* (OOB).

Durante el crecimiento del árbol, se usa una selección aleatoria de un subconjunto de características (con un tamaño fijo M menor a la dimensionalidad del vector de características) para la determinación de la división en cada nodo. El conjunto de estos árboles de decisión débiles (con baja correlación y bajo sesgo pero alta varianza) produce un clasificador fuerte, es decir, un RF con bajo sesgo y baja varianza.

RF resulta ser un clasificador sencillo de entrenar con pocos parámetros (M ya mencionado y N_{tree} el número de árboles de decisión dentro del RF) con lo que se pueden obtener buenos resultados en muchos conjuntos de datos [17].

Otra característica de RF es que puede dar una medición de importancia a cada característica según su contribución al desempeño de la clasificación.

Esa contribución es calculada mediante comparar el desempeño del predictor (árbol de decisión) sobre el conjunto original OOB (que da una puntuación N_i) y los casos donde los elementos OOB se permutan aleatoriamente (que da una puntuación $N_i^{m-permutado}$) para las m características.

Por lo tanto, se define la medida de importancia $IMP(m)$ para cada característica m según la Ecuación 2.4:

$$IMP(m) = \frac{\sum_i (N_i - N_i^{m-permutado})}{N_{tree}} \quad (2.4)$$

El método RF-RFE es similar a SVM-RFE propuesto por Guyon (2002) [36] pero con C_t (Ecuación 2.5) como criterio de puntuación, basado en RF.

$$C_t = \frac{IMP(m)}{\sigma_m / \sqrt{N_{tree}}} \quad (2.5)$$

Para tener un criterio de puntuación más robusto se calcula el C_t que es el Z-score de la medida $IMP(m)$, donde σ_m es la desviación estándar del error y N_{tree} es la cantidad de árboles de decisión del RF [45].

El primer paso es ordenar las características según su grado de importancia dado por el índice C_t , que se ordena de mayor a menor. Luego, se elimina el último elemento de la lista, es decir, el con menor valor de C_t . Como se generan RF con un conjunto distinto de características, se ordenan nuevamente las restantes según un nuevo cálculo de su valor C_t y se repite el proceso de eliminación de la característica con menor índice de importancia.

Tal como en SVM-RFE, se eliminan las características que contribuyan menos al margen de maximización.

2.5. Clusterización

El análisis de clusters o clusterización básicamente consiste en encontrar grupos de elementos de tal forma que: los elementos de un grupo sean similares (o que estén relacionados) entre sí y sean diferentes (o no estén relacionados) con los elementos de otros grupos.

A continuación se define la metodología *k-means*, un método de seteo de los parámetros del algoritmo de clusterización y un método de validación.

2.5.1. *k-means* [9]

K-means es un método iterativo de clustering que consiste en dividir el espacio en k grupos o clusters. El valor de k se define de antemano. Los clusters están definidos por su centroide o punto medio de los datos agrupados, por lo que cada punto se asigna al cluster que tenga el centroide más cercano. Los centroides iniciales se escogen de manera aleatoria.

Este algoritmo es fácil de implementar y converge rápidamente, aún así puede caer en mínimos locales. Los centroides obtenidos usualmente corresponden a la media de los puntos del cluster.

Dado un conjunto de muestras con puntos n -dimensionales dentro de un set de k clusters, el algoritmo *k-means* busca una partición tal que el error cuadrático entre la media empírica de un grupo y cada punto del mismo grupo sea mínima.

Por lo tanto, el objetivo de este algoritmo es minimizar la suma del error cuadrático para todos los clusters.

Los principales pasos que sigue el algoritmo *k-means* son:

1. Seleccionar una partición inicial con k clusters.
2. Generar una nueva partición asignando cada elemento a un cluster donde con centroide más cercano.
3. Calcular nuevos centroides hasta minimizar el error cuadrático.

En la Figura 2.14 se observan los pasos que realiza el algoritmo. En 2.14a se tiene datos de entrada bidimensionales con tres grupos; en 2.14b se tiene la asignación inicial de 3 puntos como centroides; en 2.14c y 2.14d ocurren las iteraciones intermedias en las que se van actualizando las etiquetas de cada punto; y en 2.14e se observa la convergencia del algoritmo en la minimización del error cuadrático y el resultado final de la clusterización.

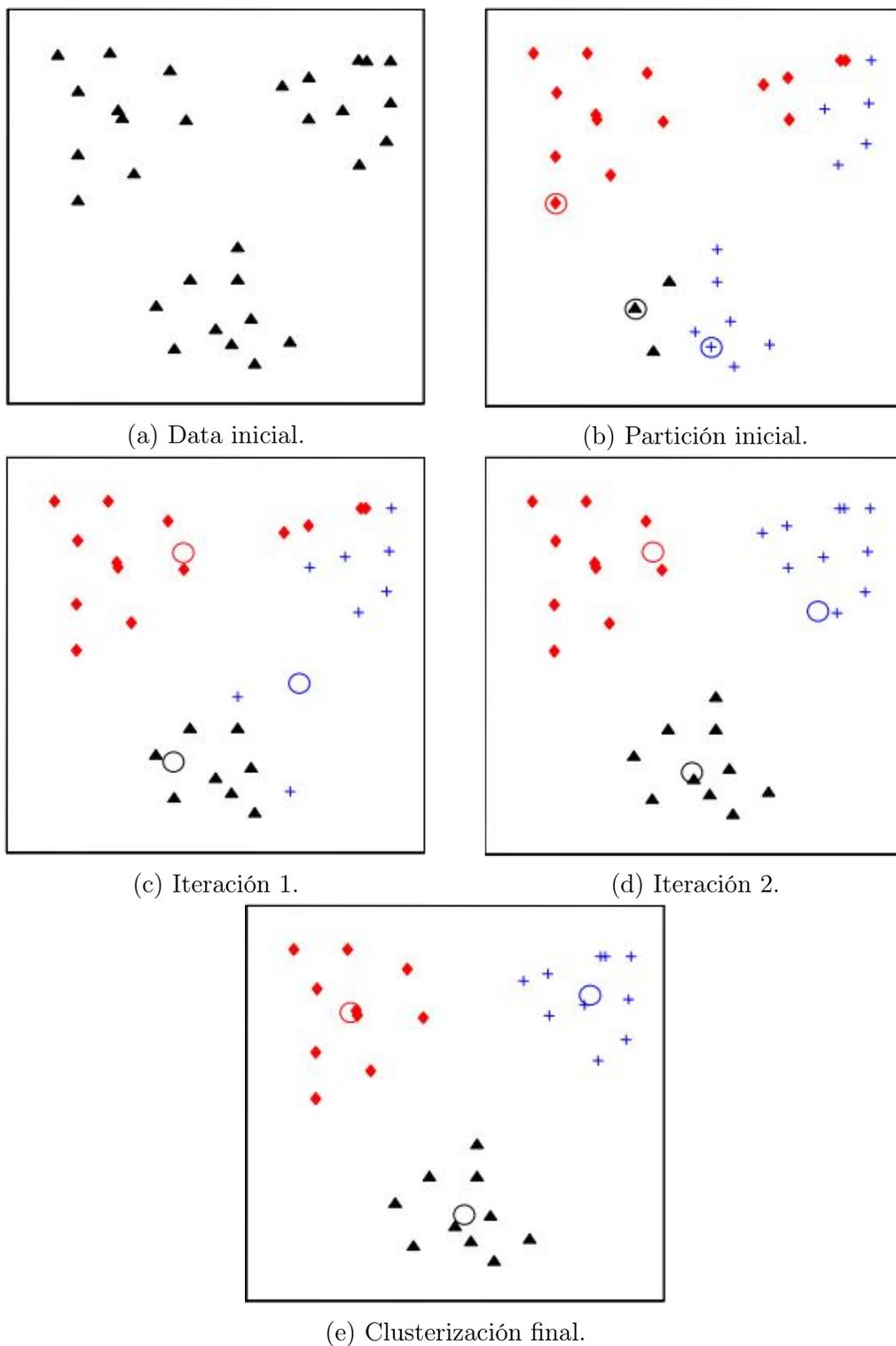


Figura 2.14: Ilustración del algoritmo k -means. (Imágenes de A.K. Jain. 2010. *Data clustering: 50 years beyond K-means* [9]).

Los parámetros del algoritmo k -means son: número de clusters k , puntos de inicialización y la métrica de distancia. Siendo la más crítica el número k de clusters.

La métrica típicamente usada es la distancia Euclidiana que funciona bien para definir clusters con forma esférica. Por otro lado, la distancia de Mahalanobis ha sido usada para detectar clusters de forma hiper elipsoidal, pero con alto costo computacional. Así mismo, se han explorado distintos tipos de métricas de distancia para distintas aplicaciones.

Diferentes inicializaciones pueden generar diferentes clusters finales porque el algoritmo puede caer en mínimos locales. Una forma de evitar esto es correr múltiples veces el algoritmo para un k dado con diferentes particiones iniciales y escoger la partición con el menor error cuadrático en el resultado.

Para escoger el valor de k no existe un criterio perfecto, más bien se utilizan métodos heurísticos, como por ejemplo, la observación de los datos. A continuación se presenta una opción más metódica para definir el número óptimo k de clusters.

2.6. Criterios de validación de clusters [55]

Existen muchos algoritmos de clusterización. Diferentes algoritmos o parámetros de entrada causan diferentes clusters. Es por ello que surge la necesidad de evaluar los clusters resultantes de manera objetiva y cuantitativa. El análisis se focaliza en la validez de la clusterización en base a la estructura del espacio de características, no necesariamente respecto a la estructura original de los datos.

Por ejemplo, si no hay estructura de clusters en un conjunto de datos, la salida de la clusterización no tendrá sentido, sólo será un artefacto creado por el algoritmo. Para estos casos es necesario efectuar algún tipo de validación para asegurar la existencia de una estructura de clusters antes de pasar a otros análisis.

Se pueden definir 3 tipos de criterios de prueba: criterios externos (que compara la estructura del clustering con una pre-especificada); criterios internos (evalúa la estructura de clustering solo a partir de las características, sin información externa) y los criterios relativos (compara con otras estructuras de clustering obtenidas con otros algoritmos o parámetros).

En particular, los criterios relativos no requieren pruebas estadísticas que aumentan el costo computacional, pues se enfoca en la comparación entre resultados de la clusterización. Un ejemplo es la determinación del número real k de clusters para *k-means*, parámetro que usualmente es fijado a priori por el usuario.

Aunque en algunos casos k puede ser estimado por el conocimiento previo del usuario, en último término depende siempre de los datos mismos.

Una sobre-estimación o sub-estimación de k afectará la calidad de la clusterización resultante. Demasiados clusters complica la estructura real y dificulta el análisis del resultado; y una partición con muy pocos clusters causa pérdida de información y conclusiones erróneas.

Afortunadamente, existen diversos índices, métodos y criterios para abordar este problema. A continuación se describen algunos de utilidad para este trabajo.

2.6.1. Visualización de datos

Es el método más directo para estimar el valor de k . Se proyectan los puntos en un espacio bi- o tridimensional. Existen varios de estos métodos.

La limitación está cuando se tienen conjuntos de datos reales tan complejos que su estructura no puede visualizarse adecuadamente en 2 o 3 dimensiones.

2.6.2. WSS y Calinski & Harabasz

Para estimar el valor adecuado de k se obtiene una secuencia de estructuras de clusterización aplicando repetidamente el algoritmo desde un mínimo k_{min} a un máximo k_{max} .

Estas estructuras se evalúan con ciertos índices. La solución se determina eligiendo la de mejor índice. Estos índices combinan la información de la capacidad intra-cluster y de aislamiento inter-cluster.

A continuación, se describen 2 índices, WSS y Calinski & Harabasz, que se usan como método para determinar el número k óptimo de clusters. De ellos se obtienen 2 curvas que al intersectarse dan el número k como resultado.

Suma total interna de cuadrados (WSS) [49]

Una heurística simple es calcular la suma total interna de cuadrados (WSS, *Total within sum of squares* en inglés) para diferentes valores de k y buscar un codo en la curva.

La suma interna de cuadrados para un solo cluster es la distancia cuadrada media de cada punto del cluster a su respectivo centroide. El total dentro de la suma de cuadrados es la suma de la suma de cuadrados de todos los grupos. Se muestra el cálculo en la Ecuación 2.6.

$$WSS = \sum_{i=1}^k \sum_{j=1}^{n_i} \|x_j - \bar{C}_i\|^2 \quad (2.6)$$

El WSS total disminuirá a medida que el número de clusters aumenta, ya que cada grupo será más pequeño y más apretado. La esperanza es que la velocidad a la que disminuye el WSS disminuirá para k más allá del número óptimo de grupos. En otras palabras, la gráfica de WSS frente a k debe aplanarse más allá del k óptimo, por lo que el k óptimo estará en el codo de la gráfica. Desafortunadamente, este codo puede ser difícil de ver. Por ello, se busca la intersección con la curva Calinski-Harabasz, índice descrito a continuación.

Calinski & Harabasz (CH) [21]

El índice de Calinski-Harabasz (CH) de una agrupación es la relación de la varianza *entre-cluster* y la variación total *intra-cluster*. El índice de CH se calcula según la Ecuación 2.7:

$$CH(k) = \frac{[traza(\mathbf{B}/k - 1)]}{[traza(\mathbf{W}/N - k)]} \quad (2.7)$$

Donde \mathbf{B} denota la matriz de la suma de los errores cuadráticos entre diferentes grupos (entre-cluster). Se calcula como se muestra en 2.8:

$$traza(\mathbf{B}) = \sum_{j=1}^k |C_j| \|\bar{C}_i - \bar{x}\|^2 \quad (2.8)$$

Y \mathbf{W} denota la diferencia al cuadrado de todos los objetos de un cluster con respecto a su centroide (intra-cluster). Se calcula como se muestra en 2.9:

$$traza(\mathbf{W}) = \sum_{i=1}^k \sum_{j=1}^k w_{i,j} \|x_j - \bar{C}_i\|^2 \quad (2.9)$$

El valor que maximiza $CH(k)$ sugiere un valor estimado para k . En teoría, la relación de B a W debe ser maximizada en el k óptimo.

Una característica importante del índice es el hecho de que, por un lado, la traza W comenzará en un valor comparativamente grande. Con un número creciente de clusters k , irá acercándose a la solución de agrupamiento óptima. Pasado el valor óptimo se producirá una disminución notablemente menor.

Por otro lado, la traza B debe comportarse en la dirección opuesta, aumentando a medida que aumenta el número de clusters k . Del mismo modo, ese aumento se amortigua al pasar el valor óptimo.

2.6.3. Bootstrap [49]

El algoritmo de bootstrap evalúa la estabilidad de un cluster dado. Utiliza el coeficiente de Jaccard como medida de similitud entre conjuntos.

El **índice de similitud de Jaccard (J)** (“coincidencia estricta”) entre dos conjuntos, A y B, es la relación entre el número de elementos en la intersección de A y B sobre el número de elementos en la unión de A y B. Se calcula como se presenta en la Ecuación 2.10:

$$J = a/(a + b + c); \quad (2.10)$$

Donde:

- a : x_i y x_j pertenecen al mismo cluster de \mathbf{C} y a la misma categoría de \mathbf{P} .
- b : x_i y x_j pertenecen al mismo cluster de \mathbf{C} pero a diferentes categoría de \mathbf{P} .
- c : x_i y x_j pertenecen a diferentes cluster de \mathbf{C} pero a la misma categoría de \mathbf{P} .

Donde \mathbf{P} es una partición pre-especificada del conjunto de datos \mathbf{X} , con N datos, y es independiente de la estructura de clustering \mathbf{C} resultado de aplicar algoritmo bootstrap cuya estrategia básica se describe a continuación:

1. Agrupar los datos como de costumbre.
2. Dibujar un nuevo conjunto de datos (del mismo tamaño que el original) remuestreando el conjunto original con reemplazo (es decir, que algunos de los puntos de datos pueden aparecer más de una vez y otros no). Agrupar el nuevo conjunto de datos.
3. Para cada cluster del conjunto original se busca el conjunto más similar del nuevo conjunto de clusters, del cual se obtiene el coeficiente de Jaccard. Si el coeficiente máximo de Jaccard es menor a 0,5 el cluster original se considera disuelto, es decir, no apareció en el nuevo agrupamiento. Un clusters que se disuelve con demasiada frecuencia probablemente no es un cluster real.
4. Repetir los pasos 2 y 3 varias veces.

El índice de estabilidad de cada cluster en el conjunto original es el promedio de sus coeficientes de Jaccard sobre todas las iteraciones de bootstrap.

Como regla general, los grupos con un índice inferior a 0,6 deben considerarse inestables. Valores entre 0,6 y 0,75 indican que el grupo está midiendo un patrón en los datos, pero no existe mucha certeza de qué puntos deben agruparse. Los valores de estabilidad sobre 0,85 se consideran altamente estables (muy probablemente un cluster real).

Por otro lado, diferentes algoritmos de clustering pueden dar diferentes valores de estabilidad, incluso si dan resultados similares, por lo que este algoritmo también está midiendo cuán estable es el algoritmo de agrupación. Además, bootstrap asume que el valor de k es conocido.

2.7. Clasificadores

2.7.1. Naive Bayes

El clasificador Naive Bayes es un algoritmo de aprendizaje estadístico que aplica una versión simplificada de la regla de Bayes para calcular la probabilidad de cada clase (ver Ecuación 2.11 donde C es la variable aleatoria que denota la clase; \mathbf{X} la variable aleatoria del vector de características; c una clase particular conocida y \mathbf{x} observación particular del vector de características). Este método está diseñado para tareas de aprendizaje supervisado, donde el objetivo es predecir la clase de las instancias de prueba y en que las instancias de entrenamiento incluyen información de las clases [30].

$$P(C = c|\mathbf{X} = \mathbf{x}) = \frac{P(C = c)P(\mathbf{X} = \mathbf{x}|C = c)}{P(\mathbf{X} = \mathbf{x})} \quad (2.11)$$

Naive Bayes se basa en modelos de probabilidad que incorporan fuertes supuestos de independencia condicional de las características dadas las clases. Gracias a esto, Naive Bayes permite simplificar el cálculo de las verosimilitudes aplicando las propiedades de probabilidades independientes como se observa en la Ecuación 2.12.

$$P(\mathbf{X} = \mathbf{x}|C = c) = P(X = x_1, \dots, x_n|C = c) = \prod_{i=1}^n P(X = x_i|C = c) \quad (2.12)$$

A pesar del supuesto poco realista de independencia, el clasificador ingenuo (*naive*) de Bayes es eficaz en la práctica y a menudo compite bien con clasificadores más sofisticados.

Dada una muestra de prueba $X = (x_1, x_2, \dots, x_n)$, el resultado de clasificación, C^* , está determinado por la probabilidad posterior $P(C = c|\mathbf{X} = \mathbf{x})$ (que se calcula usando la Ecuación 2.11 y 2.12), es decir, la probabilidad de que una muestra de prueba \mathbf{X} pertenezca a la clase C , según la Ecuación 2.13:

$$C^* = \underset{C}{\operatorname{arg\,m\acute{a}x}} P(C = c|\mathbf{X} = \mathbf{x}) \quad (2.13)$$

En resumen, Naive Bayes tiene 2 etapas:

1. En la etapa de entrenamiento, la probabilidad $P(X|C)$ (la probabilidad de la muestra de entrenamiento X dada su clase C) se estima con respecto a los datos de entrenamiento;
2. En la etapa de prueba, basada en la probabilidad posterior $P(C|X_{test})$, se toma una decisión de si la muestra de prueba X_{test} pertenece a una clase C , usando la Ec. 2.13.

El clasificador Naive Bayes es usado, por ejemplo, para detectar emociones desde expresiones faciales; y en la clasificación de carga cognitiva y monitoreo de niveles de estrés en tareas HCI mediante señales psico-fisiológicas [24][27][61][52][12].

2.7.2. Máquina de vector de soporte [59]

La máquina de vector de soporte (SVM, *Support Vector Machine*) fue propuesta primeramente por Boser, Guyon, & Vapnik (1992) [16] y se estableció rápidamente como un poderoso algoritmo para resolver problemas de clasificación dentro del contexto conocido como aprendizaje supervisado.

Hoy es usado, por ejemplo, para la clasificación de estados emocionales, la clasificación de carga cognitiva y estimación de niveles de estrés en tareas HCI usando señales psicofisiológicas [50][52][12]. A continuación se explica brevemente cómo funciona la máquina de vector de soporte.

Caso linealmente separable

Para entender su funcionamiento se analiza primero el caso ideal, donde los datos son **linealmente separables**. Para este caso se dice que existe un hiperplano definido por la Ecuación 2.14 que separa completamente los datos de entrenamiento.

$$f(\delta) \equiv w \cdot \delta + b \equiv 0 \quad (2.14)$$

Donde:

δ : Vector de característica de la muestra.

w : Vector normal al hiperplano.

b : Bías. Desplazamiento del hiperplano.

En general, existe más de un hiperplano capaz de realizar la separación, por eso se escoge el que:

- Minimice el error en la separación de las muestras, es decir, el error de clasificación.
- Maximice el margen de separación, es decir, mejore la generalización del clasificador.

Todos los puntos de entrenamiento con $y_i = 1$ (es decir cumplen $f(\delta) > 0$) se encuentran en un lado del hiperplano, mientras que todos los puntos de entrenamiento con $y_i = -1$ se encuentran en el otro lado (por lo tanto cumplen $f(\delta) < 0$). Para obtener el hiperplano óptimo se debe encontrar w y b . Usualmente se fija w y se mueve el valor de b .

Por lo tanto, $f(\delta)$ es la regla de decisión y los hiperplanos paralelos que separan los datos se pueden resumir en la Ecuación 2.15:

$$y_i(w \cdot \delta_i + b) \geq 1 \quad (2.15)$$

En la Figura 2.15 se observan los vectores de soporte, que son los que cumplen: $w \cdot \delta + b = 1$ y $w \cdot \delta + b = -1$. Las nuevas muestras que cumplan $w \cdot \delta + b < 0$ serán clasificadas como una clase y las que cumplan $w \cdot \delta + b > 0$ como la otra clase, por lo tanto, se discrimina sólo entre 2 clases.

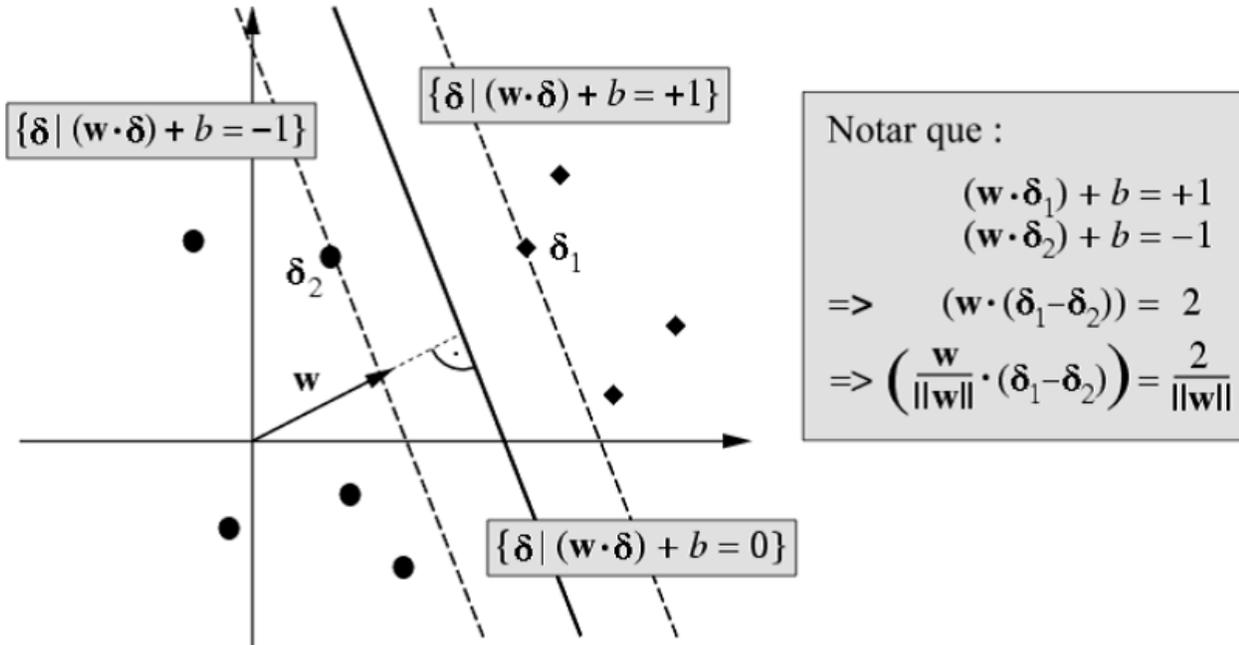


Figura 2.15: Generación de hiperplano SVM caso separable.

Para encontrar el margen máximo SVM, es decir, el mejor hiperplano de separación, se debe resolver el problema de minimización cuadrática que se muestra en la Ecuación 2.16:

$$\begin{aligned} & \text{mín } \frac{1}{2} w \cdot w \\ & \text{s.a : } y_i(w \cdot \delta_i + b) \geq 1 \end{aligned} \tag{2.16}$$

En resumen, la metodología general es:

1. En la etapa de entrenamiento, construir un hiperplano óptimo que separe los puntos en las clases respectivas.
2. En la etapa de prueba, clasificar un punto nuevo de acuerdo a su ubicación con respecto al hiperplano de separación.

Caso no separable

Generalizando un poco más, para el **caso no separable** es necesario escoger qué tan importantes son los outlier (puntos aislados) dentro de la clasificación. Para esto se introduce el término llamado variable de holgura ξ_i por cada punto δ_i de los datos, como se observa en la Ecuación 2.17.

$$y_i(w \cdot \delta_i + b) \geq 1 - \xi_i \quad (2.17)$$

Usando la misma filosofía del caso separable explicada anteriormente, el caso no separable considera en la minimización la importancia de los datos del conjunto de entrenamiento que están fuera del margen delimitado por el hiperplano clasificador. En la Figura 2.16 se observa este caso.

El problema de minimización ahora considera el aporte de los outliers que se regula mediante el ponderador λ que va entre $0 < \lambda < \infty$, donde cero significa que no se consideran. La Ecuación 2.16 queda de la forma que se presenta en la Ecuación 2.18:

$$\begin{aligned} \text{mín } & \frac{1}{2} w \cdot w + \lambda \sum_i \xi_i \\ \text{s.a. } & y_i(w \cdot \delta_i + b) \geq 1 - \xi_i \end{aligned} \quad (2.18)$$

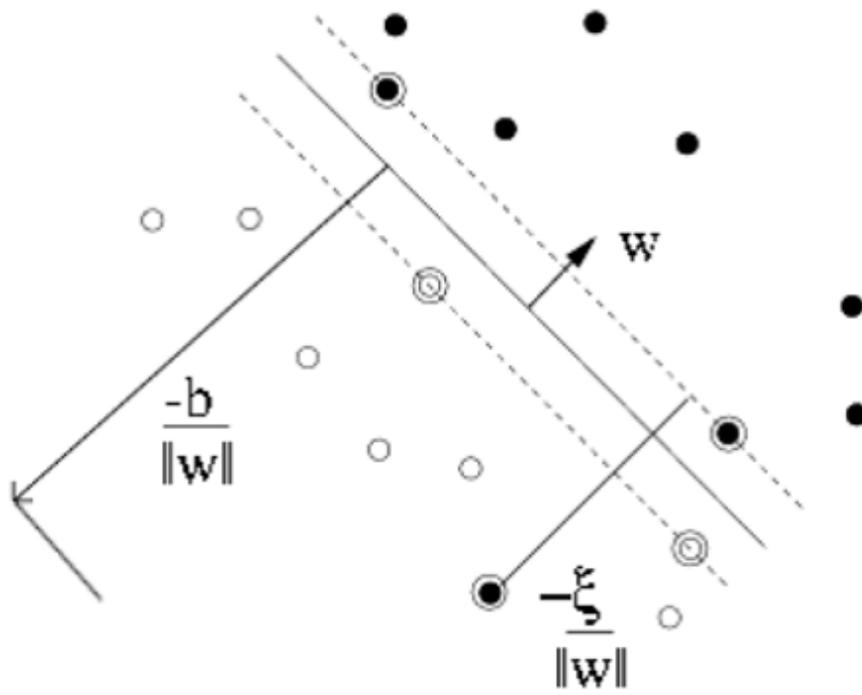


Figura 2.16: Generación de hiperplano SVM caso no separable.

Caso no lineal

Finalmente, la mayor generalización demuestra el gran poder de SVM. La idea básica, tal como se muestra en la Figura 2.17, es que para realizar la separación para el caso no lineal en un espacio m -dimensional se realiza un mapeo de los vectores de características con una función φ a un espacio mayor N -dimensional donde sea posible generar un hiperplano.

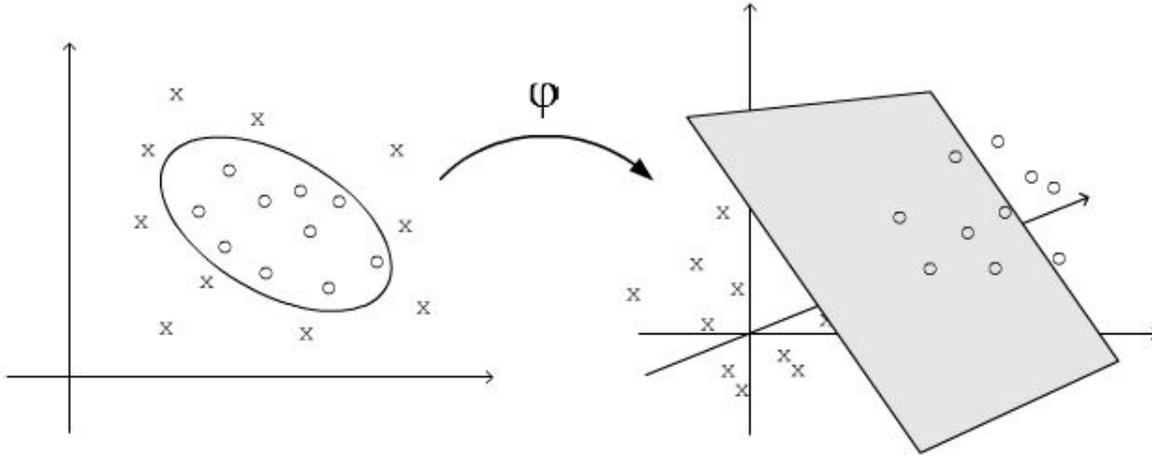


Figura 2.17: Mapeo de características para el caso no lineal. Usando la función φ se llevan los puntos de espacio m -dimensional a un espacio mayor N -dimensional donde son posibles de separar por un hiperplano [59].

Gracias al mapeo, la regla de decisión $f(x)$ que se construye como lineal en el espacio generado, se vuelve bastante general para incluir todas las formas lineales y cuadráticas (líneas, elipses, hipérbolas) en el espacio de características original, como se presenta en la Ecuación 2.19:

$$F(\delta) = F[\varphi(\delta)] = W \cdot \varphi(\delta) + B \quad (2.19)$$

Para encontrar W y B se resuelve el problema de minimización exactamente como en la Ecuación 2.18, pero sólo en el espacio de mayor dimensión, como se ve en la Ecuación 2.20:

$$\begin{aligned} \min & \frac{1}{2} W \cdot W + \lambda \sum_i \Xi_i \\ \text{s.a.} & y_i(W \cdot \varphi(\delta_i) + B) \geq 1 - \Xi_i \end{aligned} \quad (2.20)$$

2.7.3. Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (ANN, *Artificial Neural Networks*) están inspiradas en las redes neuronales biológicas del cerebro humano. Están constituidas por elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes. Estos elementos están organizados de una forma parecida a la que presenta el cerebro humano. Por ejemplo, las Redes Neuronales Artificiales aprenden de la experiencia, generalizan de ejemplos previos a ejemplos nuevos y abstraen las características principales de una serie de datos. Matemáticamente actúa como una caja gris, pues se conocen los datos con los que se entrena la red y podemos escoger hasta cierto grado las características y parámetros de ésta, pero no se logra recuperar la aproximación a la función que hace la misma red. Un ejemplo de red neuronal artificial se muestra en la Figura 2.18 cuya topología es de perceptrón multicapa.

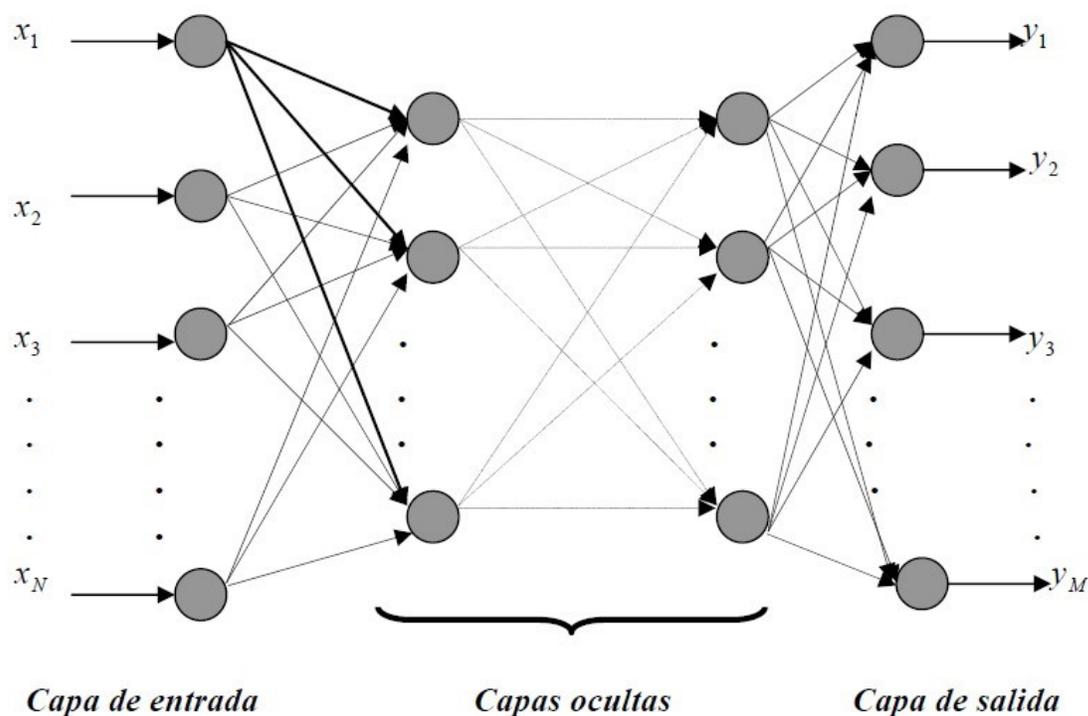


Figura 2.18: Ejemplo de red neuronal artificial.

La unidad básica es la Neurona Artificial, luego de definir qué es, se explican sus componentes principales.

Neurona Artificial

Es la componente básica de la red neuronal. La neurona artificial modela el comportamiento de la neurona biológica; como se muestra en la Figura 2.19, posee entradas (que representan las dendritas de la neurona biológica), salidas (axón) y pesos sinápticos (sinapsis de la neurona biológica). Una neurona por sí misma es muy pequeña y simple, pero, al igual

que el cerebro humano que posee billones de estas, al combinarse muchas de ellas pueden resolver problemas complejos.

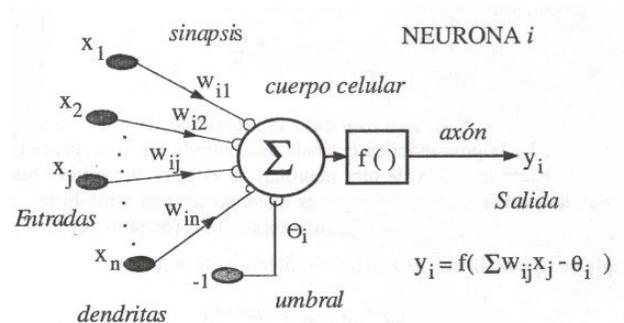


Figura 2.19: Esquema de una neurona artificial.

Capa de entrada

Es la capa que recibe directamente la información proveniente de las fuentes externas de la red. La cantidad de neuronas en la capa de entrada está determinada por el número de características medidas.

Capa Oculta

Son todas las capas internas de la red (entre la capa oculta y la capa de salida), no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales. Una red se dice totalmente conectada cuando todas las salidas desde un nivel llegan a todos los nodos o neuronas de la capa siguiente (se incluyen las capas de entrada y salida).

Capa de Salida

Transfiere información de la red hacia el exterior. La cantidad de neuronas en la salida queda determinada por el número de clases que existen.

Pesos sinápticos w_{ij} y bias θ_i

Los pesos sinápticos corresponden a la fuerza o grado de interconexión entre una neurona i (desde una capa) a la neurona j (de la capa siguiente). Los pesos sinápticos pueden tomar valores positivos, negativos o cero. En caso de una entrada positiva, un peso positivo actúa como excitador, mientras que un peso negativo actúa como inhibidor. En caso de que el peso

sea cero, no existe comunicación entre el par de neuronas. Mediante el ajuste de los pesos sinápticos la red neuronal realiza su proceso de aprendizaje. Además, cada neurona tiene asociado un número denominado bias o umbral, que puede verse como un número que indica a partir de que valor del potencial postsináptico la neurona produce una salida significativa, es un número que no cambia al entrenarse la red.

Función de Activación

La función de activación se utiliza para limitar el rango de valores de la respuesta de la neurona. Generalmente, los rangos de valores se limitan a $[0,1]$ o $[-1,1]$, sin embargo, otros rangos son posibles de acuerdo a la aplicación o problema a resolver. Existen diversas funciones de activación y la decisión entre una u otra dependerá nuevamente de la aplicación o problema a resolver. Existen funciones de activación comúnmente utilizadas y con las cuales se han obtenido resultados satisfactorios en diversas aplicaciones. En la Figura 2.20 se muestran las funciones de activación más comunes.

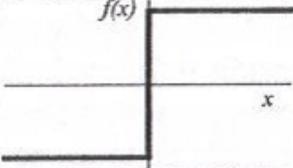
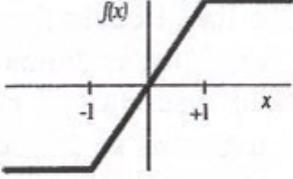
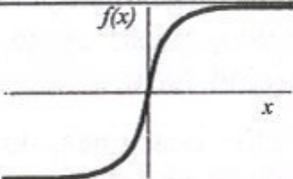
	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } +l \leq x \leq -l \\ +1, & \text{si } x > +l \end{cases}$	$[-1, +1]$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	

Figura 2.20: Funciones de activación más comunes.

Las funciones de activación más utilizadas son: la función sigmoidea, que son para obtener valores en el intervalo $[0,1]$ y la tangente hiperbólica $[-1,1]$.

Cantidad teórica de neuronas en la capa oculta

La mayoría de los problemas prácticos se resuelven con una sola capa oculta, por lo tanto solo se aborda este caso.

Surge la cuestión de cuál es el número óptimo de neuronas en esta capa. Si bien la cantidad óptima de neuronas en la capa oculta se obtiene de manera heurística y dependerá de cada caso, siempre se puede partir con un número “óptimo”.

Un número elevado de neuronas en la capa oculta puede generar sobreentrenamiento, es decir, no generaliza la respuesta del clasificador, y muy pocas producirán subentrenamiento, es decir, no se alcanza a resolver el problema.

Se han encontrado dos criterios para definir el número de neuronas en la capa oculta, los cuales son:

1. Regla de la Pirámide Geométrica: $h = \sqrt{n \cdot m}$ donde m es la cantidad de entradas de la red y n la cantidad de salidas.
2. En función del número de muestras al entrenar: $h = \log_2(T)$ donde T es la cantidad de muestras en el conjunto de entrenamiento.

Sobreentrenamiento

Dado un conjunto de datos, es posible que la red neuronal logre reproducir muy bien el comportamiento de dichos datos pero no el de datos nuevos, es decir, la red no logra generalizar la respuesta. Este problema se conoce como sobreentrenamiento, sobreaprendizaje o sobreajuste, además, acentúa cuando existe ruido o errores en el conjunto de datos.

Existen diferentes modos de evitar el sobreaprendizaje. El primero de ellos sería obtener más datos con más variabilidad para el entrenamiento, aunque esto no es siempre posible. Otra posibilidad es reducir el tamaño de la red (menor número de parámetros), de modo que la red neuronal es menos flexible y más robusta frente al ruido, aunque si se reduce demasiado, puede que no sea capaz de aprender o aproximar la función objetivo. Dotar a la red neuronal de suficientes parámetros para que sea capaz de aprender y evitar el sobreaprendizaje, es el principal aspecto a tener en cuenta en el dimensionamiento de una red neuronal.

En general, para evitar el sobreentrenamiento se utiliza un conjunto de validación luego de cada iteración en el proceso de entrenamiento para comprobar si existe sobreentrenamiento y corregirlo variando los parámetros. Durante el proceso iterativo de optimización de los parámetros de la red, se comparan los errores obtenidos con los datos de entrenamiento y con los datos de validación. En el caso de que durante sucesivas iteraciones, el error con los datos de entrenamientos disminuya, mientras que el error con los datos de validación aumente, se detiene el proceso de ajuste, como un criterio de parada adicional.

2.8. Deep Learning [33] [10]

Desde los años 80', el aprendizaje profundo (*deep learning*, en inglés) ha mejorado consistentemente en su capacidad de proporcionar reconocimiento o predicciones más precisas. Por otra parte, *deep learning* ha sido sistemáticamente aplicado con éxito a conjuntos de datos más y más amplios.

A continuación se hace una breve descripción de la arquitectura de redes neuronales profundas que se utilizan en esta memoria y de algunos métodos de regularización que existen en *deep learning* para evitar sobreentrenamiento.

2.8.1. Redes neuronales de alimentación profunda

Las redes neuronales de alimentación profunda (*Deep feedforward networks*), también llamadas perceptrón multicapas (*multilayer perceptrons, MLPs*), son por excelencia los modelos de aprendizaje profundo.

Se basa en las redes neuronales artificiales descritas en la Sección 2.7.3 con la diferencia de que están configuradas para trabajar con una gran cantidad de datos y dar estabilidad al entrenamiento y la generalización del resultado. Las redes neuronales de alimentación *profunda* se conocen así pues cuentan con una gran cantidad de capas ocultas interconectadas.

Esta configuración de redes neuronales de varias capas se puede utilizar para llevar a cabo tareas de aprendizaje profundo. Las arquitecturas de aprendizaje profundo son modelos de extracción jerárquica de características, típicamente involucrando múltiples niveles de no linealidad. Los modelos de aprendizaje profundo son capaces de aprender representaciones útiles de datos en bruto y han exhibido alto rendimiento en datos complejos como imágenes, voz y texto (Bengio, 2009).

Las neuronas de capa oculta requieren una **función de activación**, que es difícil determinar cuál es la mejor. La función de activación recomendada por defecto para la mayoría de las redes neuronales de alimentación profunda es la **función de activación lineal rectificada**.

La aplicación de la función de activación lineal rectificada produce una transformación no lineal. Sin embargo, la función permanece muy cerca de ser lineal, en el sentido de que es una función lineal por piezas, con dos piezas lineales como se observa en la Figura 2.21. Debido a esto, se conservan muchas de las propiedades que hacen que los modelos lineales sean fáciles de optimizar con los métodos basados en gradientes. También preservan muchas de las propiedades que generalizan bien los modelos lineales.

Estos elementos básicos descritos, adicionados a los mencionados en la Sección 2.7.3, permiten configurar la red neuronal profunda para el entrenamiento y posteriormente la clasificación. Sin embargo, es necesario utilizar un método para *regular* el sobreajuste o sobreentrenamiento de la red neuronal. A continuación, se describe el concepto de regularización que sirve para evitar este efecto.

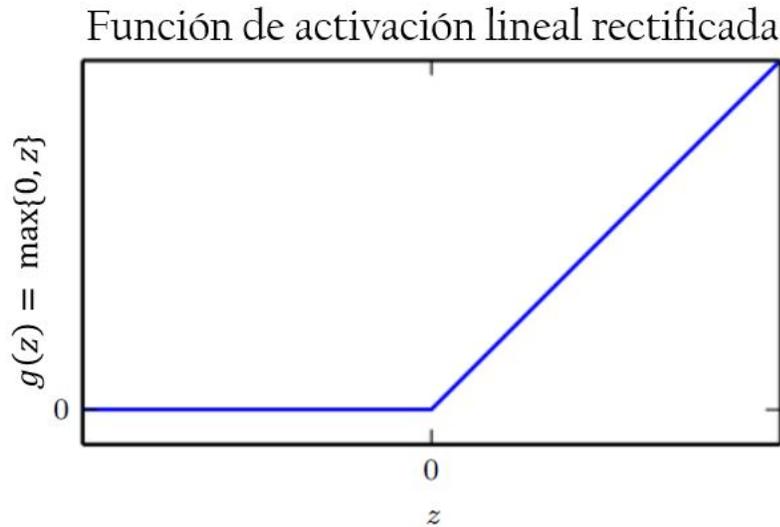


Figura 2.21: Función de activación lineal rectificada.

2.8.2. Regularización

En el aprendizaje automático (*machine learning*), donde se trata de crear programas capaces de generalizar comportamientos, puede ocurrir el problema de que el modelo de aprendizaje se subajuste, de modo que no llegue a generalizar lo suficiente el modelo, o bien, se sobreajuste, lo que hace que el modelo no funcione para futuros datos de entrada. Para evitar esto podemos regularizar el modelo de aprendizaje mediante una función de penalización llamada regularizador de la función objetivo.

La regularización, según Ian Goodfellow (2016) [33], es “cualquier modificación, que realizamos en un algoritmo de aprendizaje, que tenga la intención de reducir su error de generalización, pero no su error de entrenamiento”. Existen muchas estrategias de regularización. A continuación, se describen dos de ellas.

Parámetro de regularización L1

También conocida como *LASSO*. La regularización L1 representa la suma de todas las normas 1 de los pesos y sesgos de la red, es decir, la suma de los valores absolutos de los parámetros individuales. La forma de L1 se muestra en la Ecuación 2.21:

$$R_1(\mathbf{W}, \mathbf{B}|j) = \|\mathbf{W}\|_1 \quad (2.21)$$

Parámetro de regularización L2

La regularización L2 es también conocida como *Regresión Ridge* o *Regularización Tikhonov*. La regularización L2 representa la suma de los cuadrados de todos los pesos y sesgos de la red. Esta estrategia conduce los pesos más cerca del origen, por ello también se le conoce como regularización con decaimiento de los pesos. La forma de L2 se muestra en la Ecuación 2.22:

$$R_2(\mathbf{W}, \mathbf{B}|j) = \frac{1}{2} \|\mathbf{W}\|_2^2 \quad (2.22)$$

Función objetivo regularizada

Se define la función objetivo para la red neuronal artificial como: $L(\mathbf{W}, \mathbf{B}|j)$ donde \mathbf{W} representa la matriz de pesos y \mathbf{B} la columna de vectores de bias para cada ejemplo j de entrenamiento (comparar con Figura 2.19).

Muchos enfoques de regularización se basan en la limitación de la capacidad de los modelos mediante la adición de una función de penalización a la función objetivo $L(\mathbf{W}, \mathbf{B}|j)$. Denotamos la función objetivo regularizada por la Ecuación 2.23:

$$L'(\mathbf{W}, \mathbf{B}|j) = L(\mathbf{W}, \mathbf{B}|j) + \lambda_1 R_1(\mathbf{W}, \mathbf{B}|j) + \lambda_2 R_2(\mathbf{W}, \mathbf{B}|j) \quad (2.23)$$

Donde los valores de λ_1 y λ_2 son parámetros que ponderan la contribución relativa de los términos de penalización R_1 y R_2 en relación a la función objetivo $L(\mathbf{W}, \mathbf{B}|j)$. Ajustando los valores de λ a cero no se produce ninguna regularización. Por otro lado, con valores mayores de λ corresponde a una regularización mayor.

2.9. Trabajos realizados – Estado del Arte

En el desarrollo del área de interacción humano computadora (*Human-Computer Interaction*, HCI, en inglés) que tiene como foco la interfaz entre el usuario y la computadora o máquina que utilice, se busca constantemente mejorar el diseño de las tecnologías para hacer más natural y amigable esta interacción. En los últimos años se han centrado los estudios en el usuario, observándolo desde el punto de vista de su respuesta psico-fisiológica, muy ligado a sus emociones y actividad cognitiva.

Respecto a esta última, se han aprovechado el desarrollo de sensores psico-fisiológicos que, con mejor precisión, permiten estudiar a usuarios en distintas clases de tarea. Según la literatura expuesta las principales señales a considerar para el estudio de carga mental de un usuario son: EEG, ECG, GSR, Eye-tracking, temperatura corporal. En el Tabla 2.2 se presenta un resumen realizado por Lira [46] donde se muestran distintas referencias, los objetivos de su investigación, los sensores usados, los experimentos y resultados.

Tabla 2.2: Tabla comparativa de distintas investigaciones de medición de carga cognitiva (obtenidas de [46]).

Ref.	Autor	Objetivo	Sensores	Experimento	Resultados
[24]	Haapalainen et al.	Medición de cargas cognitivas en tareas cognitivas básicas.	Eyetracker, EEG, ECG, flujo de calor, pulso.	Problemas desktop-based de percepción visual y velocidad cognitiva.	ECG y flujo de calor juntos distinguen con un 80 % de precisión tareas de alta y baja demanda cognitiva.
[27]	Fritz et al.	Comprobar si a través de sensores biométricos se puede percibir la dificultad de una tarea mientras se está trabajando en ésta.	Eyetracker, EEG	8 tareas de comprensión de código variando la dificultad.	85 % precisión la dificultad de una tarea usando Machine Learning.

Ref.	Autor	Objetivo	Sensores	Experimento	Resultados
[56]	Shi et al.	Evaluar los niveles de estrés y excitación al incrementar niveles de dificultad.	GSR	Interfaz de control de tráfico con 3 escenarios. 1) Usando gestos y hablando, 2) Solo hablando y 3) Solo usando gestos. Además se varía la dificultad en 5 formas: Complejidad visual; Número de entidades; Número de distractores; Tiempo límite; Número de acciones para finalizar.	Resultados del GSR indican que existe un incremento significativo cuando aumento la dificultad de la tarea cognitiva.
[51]	Nourbakhsh et al.	Evaluar eficacia en la medición de GSR ante tareas de distinta dificultad.	GSR	8 tareas aritméticas con 4 niveles de dificultad (binario, 1 dígito, 2 y 3 debiendo juntar los números y responder el número de dígitos existente).	Se aprueba la hipótesis nula.
[52]	Nourbakhsh et al.	Medición de carga cognitiva.	GSR, eye blink	8 tareas aritméticas con 4 niveles de dificultad (binario, 1 dígito, 2 y 3 debiendo juntar los números y responder el número de dígitos existente).	Combinando los dos sensores se llega a una precisión de 75% para binario y 50% para 3 dígitos.

Ref.	Autor	Objetivo	Sensores	Experimento	Resultados
[60]	Xu et al.	Verificar si se puede medir carga cognitiva mediante dilatación pupilar si es que la luminosidad cambia.	Eyetracker	Tareas aritméticas: sumar 4 números que aparecen en pantalla. Varía en nivel de dificultad según el número de dígitos (4 grados de dificultad desde binario a 3 dígitos). Además, para objeto de la hipótesis se varía la luminosidad de la habitación.	Se aprueba la hipótesis nula.
[39]	Ikehara et al.	Evaluar carga cognitiva.	Eyetracker, presión del mouse, GSR, pulso oxímetro	Seleccionar en la pantalla las fracciones cuyo valor es mayor a $1/3$. Dos grados de dificultad. La dificultad baja son fracciones de rápida resolución. La dificultad alta se logra mediante un aumento en la dificultad de las tareas perceptivas (aumento del número de fracciones en pantalla y aumentando la velocidad en que son presentadas en pantalla) y aumento en la dificultad de tareas cognitivas (reducción de tiempo y aumento del rango de valores).	GSR y eyetracker tienen la mayor significancia para detectar la dificultad de la tarea.

Además de estudiar la efectividad de los sensores para evaluar la carga mental, se ha estudiado cómo varía según las distintas tareas que se realizan y cómo afecta al usuario. Por ejemplo, Bailey (2008) [18] estudió mediante la dilatación pupilar cómo varía la carga cognitiva al ejecutarse una tarea y bajo la premisa de que las interrupciones durante una tarea que produce bajas en la capacidad de ejecución de una tarea tiene un costo cognitivo.

Este costo de interrupción de una tarea envuelve considerables efectos negativos como: aumento del tiempo para completar la tarea (Monk, Boehm-Davis et al. 2002), mayor rango de error (Latorella 1998), necesidad de esfuerzos adicionales para la toma de decisiones (Speier, Valacich et al. 1999) y, en el estado afectivo, aumento de la frustración y ansiedad (Zijlstra, Roe et al. 1999; Adamczyk and Bailey 2004; Bailey and Konstan 2006). Por ejemplo, cuando aparece una interrupción en un momento al azar (por ejemplo, una tarea secundaria) mientras se realiza una tarea principal, el tiempo para completarla puede aumentar hasta en un 30 %, se pueden cometer hasta 2 veces más errores y aumenta hasta 2 veces el nivel de desagrado del usuario; a diferencia de si la interrupción apareciera en un momento previamente programado.

En base a esto Bailey (2008) [18] estudia empíricamente que las interrupciones podrían tener un menor costo si ocurre en un momento de baja carga mental, y que ese momento ocurre en los bordes entre sub-tareas al ejecutar la tarea general (Miyata & Norman 1986). Para probar esto, Bailey [18], mediante 3 tareas distintas con respectivas sub-tareas, evalúa la carga cognitiva mediante la dilatación pupilar. La primera tarea consiste en: evaluar 2 rutas distintas entre dos ciudades; el usuario debe medir la distancia y el costo de tramos de las rutas, tabular los datos y finalmente discriminar y escoger la más corta y económica. En la segunda: debe editar un documento y corregir la ortografía para 3 niveles de complejidad (editar una palabra, editar 2 palabras, editar una frase completa). Y la tercera consiste en: la clasificación de 9 correos electrónicos con asuntos explícitos (baja complejidad) y asuntos ambiguos (alta complejidad al clasificar) según 4 categorías. Cada uno de estos escenarios se aplicaron a 24 personas (7 mujeres) de entre 19 y 50 años de edad. Las principales conclusiones de Bailey [18] son: (i) La carga cognitiva varía durante la ejecución de una tarea; (ii) Efectivamente baja la carga cognitiva entre las sub-tareas en comparación con la tarea anterior y (iii) Diferentes sub-tareas demandan diferente cantidad en la carga cognitiva, es decir, se prueba la influencia de la complejidad entre sub-tareas.

Por otro lado, el estudio de la experiencia del usuario en el uso de páginas web ha sido un área llamativa para distintas aplicaciones, pues hoy en día el entorno web es cotidiano. Comprender la usabilidad del sitio web y la efectividad de sus distintas componentes con el objetivo de mejorar la experiencia del usuario, determinar la respuesta emocional, si es agradable o no el sitio web, y el comportamiento del usuario permite tomar decisiones respecto al diseño de la página web. En este respecto, gran parte de los estudios de la atención de usuarios en páginas web complejas (que contienen imágenes, texto y contenido variado) usan como principal metodología el eye tracking [48].

El estado del arte en la observación de experiencia de usuarios, que envuelve la medición de su comportamiento como reacción a distintos estímulos como notificaciones, permite predecir la respuesta del usuario según Navalpakkam & Churchill [48], quienes mediante el seguimiento del puntero del ratón lograron determinar una configuración del layout del sitio web más agradable al usuario y que mejora la efectividad de la notificación.

Navalpakkam & Churchill [48] lograron determinar los patrones de atención del movimiento del puntero del ratón en comparación a los movimientos de la mirada utilizando un eye tracker. El experimento consiste en presentar a 3 grupos de usuarios (cada grupo de 30 personas) distintos escenarios: Una página web compleja donde el primer grupo mantiene fija la posición de un anuncio publicitario (arriba-derecha), el segundo grupo tiene una variación de la posición del anuncio de un 50 % y el tercer grupo una variación del 100 % de la posición del anuncio, teniendo una posición aleatoria de las 6 opciones posibles (arriba-derecha, arriba-izquierda, al medio-derecha, al medio-izquierda, abajo-derecha, arriba-izquierda). En el experimento el usuario debe leer un artículo presentado. Finalmente, se concluye que la mirada y los patrones de movimientos del ratón contienen información importante para evaluar el estado del usuario, como por ejemplo, para determinar si está distraído de la tarea asignada o si está haciendo esfuerzos para cumplirla. Se confirma la correlación entre los movimientos oculares con el movimiento del puntero del ratón para predecir distintos estados psico-emocionales de un usuario web. También se concluye que el usuario es más propenso a observar las notificaciones cuando éstas varían de posición en el sitio web en vez de cuando están fijas. Se conoce, por lo tanto, dónde es el mejor lugar para ubicar el anuncio publicitario.

El aporte de este trabajo de título es determinar el momento propicio para el despliegue de servicios ubicuos. Se estudia la respuesta del sistema nervioso, aplicando conceptos de neurociencia, mediante distintas señales psico-fisiológicas para determinar cuándo es más apropiado presentar el servicio ubicuo según el estado de carga cognitiva de la persona con el objetivo de mejorar la experiencia del usuario.

En Dujovne & Velásquez (2009) [23] los web objects son definidos como “un grupo estructurado de palabras o recursos multimedia que se presenta en un sitio web que tiene metadatos que describen su contenido”. Cada sitio web consiste en un conjunto de *web objects* [43]. Como segundo aporte de este trabajo de título, basado en lo que dice Bailey [18], se desea comprobar que existe una baja en la carga cognitiva en el paso de un *web object* a otro en una página web. En otras palabras, se propone que el análisis visual de cada *web object* puede ser considerado como una sub-tarea dentro de la tarea principal que es navegar en un sitio web. Para ello se evalúa la carga cognitiva con sensores psico-fisiológicos escogiendo varios de los mencionados (Tabla 2.2) para abarcar más clases de señales y obtener un mayor conjunto de datos para un mejor análisis y mejores conclusiones.

Capítulo 3

Estudio Experimental e Implementación

Por lo general, los estudios experimentales de *machine learning* siguen los pasos del diagrama de la Figura 3.1. Se adquieren datos u observaciones del mundo real; dichos datos se procesan y de ellos se extraen características, que luego de una selección, sirven para aplicar algún algoritmo de aprendizaje. Este capítulo y en el siguiente se explica el experimento, su implementación y el análisis de datos llevado a cabo, siguiendo el mismo orden lógico.



Figura 3.1: Diagrama de pasos en un experimento de *machine learning*.

3.1. Instrumentación

Para los adquisición de datos se utilizaron los siguientes sensores, que luego se describen:

- Unidad *GSR+* de Shimmer.
- Sensor óptico de pulso de Shimmer.
- Unidad *Bridge Amplifier+* de Shimmer.
- ECG de tarjeta biomédica todo-en-uno de BITalino.
- EEG Emotiv Epoc.
- *Eye Tracker* Tobii T120.

3.1.1. Unidad *GSR+* de Shimmer

La unidad *GSR+* de Shimmer mide la respuesta galvánica de la piel, descrita en la Sección 2.3.4, entre dos electrodos ubicados en dos dedos de la misma mano. En la Figura 3.2 se observa la unidad *GSR+* de Shimmer donde las entradas en verde y rojo son para los electrodos. Además, este dispositivo cuenta con una entrada Jack (en negro) para una medición auxiliar como el sensor óptico de pulso presentado en la siguiente sección.



Figura 3.2: Unidad *GSR+* de Shimmer.

Mediciones

La resistencia típica de la piel varía entre $47\text{ k}\Omega$ a $1\text{ M}\Omega$ (conductividad entre $21\ \mu\text{S}$ a $1\ \mu\text{S}$) según Cacioppo, Tassinari & Berntson (2007). La unidad *GSR+* de Shimmer ha sido diseñada para medir la resistencia de la piel entre $10\text{ k}\Omega$ a $4,7\text{ M}\Omega$ ($100\ \mu\text{S}$ a $0,2\ \mu\text{S}$).

Dado que la señal GSR tiene una bajísima variación, la frecuencia de muestreo que propone el manual es entre: DC – 15,9 Hz. Literatura asociada (Geddes & Baker, 1989) propone valores entre 0,03 – 5 Hz para una correcta medición. Si se selecciona una frecuencia mayor al rango existe una opción de “auto-range”. Frecuencias mayores producen un efecto negativo en la señal en la transición que puede ser eliminado con un filtro pasa bajos.

La señal GSR cuenta con 2 componentes: la componente fásica, que tiene cambios más rápidos y está relacionado con los estímulos externos y la actividad poco específica; y la componente tónica, que es la señal base y varía más lentamente, fija el nivel de conductancia de la piel. El principal desafío es interpretar de esta señal las razones de los incrementos y disminuciones en los niveles de sudor.

Algunas perturbaciones o factores que afectan la medición de la señal GSR son: temperatura ambiental, sequedad de la piel y el movimiento de los electrodos. En la Figura 3.3 se muestra un ejemplo de la medición de GSR.

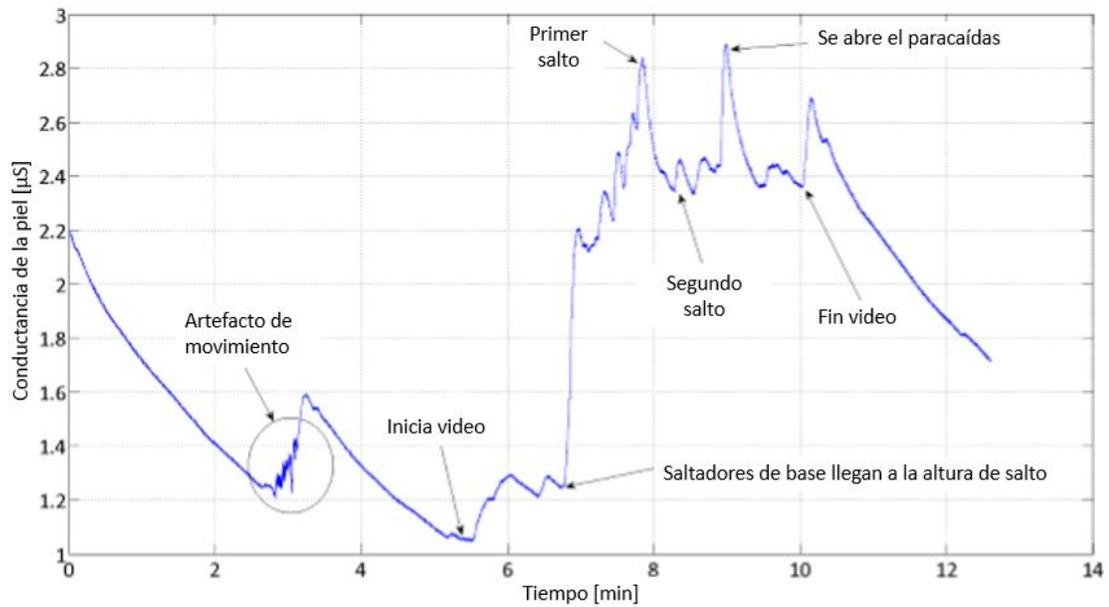


Figura 3.3: Ejemplo de medición de señal GSR. Se muestran distintos estados, como el ruido por el movimiento de los electrodos y distintas respuestas corporales de un sujeto que observa un video “que induce el nerviosismo”.

Posición de los electrodos

La posición de los electrodos requiere una zona con gran cantidad de glándulas sudoríparas, como por ejemplo la zona palmar de la mano y los dedos (ver Figura 2.9). Algo importante es evitar al máximo los movimientos de los electrodos, pues producen un ruido considerable, por lo que deben ser firmemente fijados. En la Figura 3.4 se muestra un ejemplo de conexión de una unidad *GSR+* de Shimmer.



Figura 3.4: Ejemplo de posición de electrodos para el dispositivo *GSR+* de Shimmer.

Recomendaciones

- No conectar los electrodos a la piel cuando el dispositivo esté conectado a un puerto de carga o USB.
- Para eliminar el ruido de alta frecuencia generado por el movimiento de los electrodos se puede utilizar un filtro pasa bajos con una frecuencia de corte de 5 Hz, sin afectar mayormente los datos de interés.
- Mantener un control de la temperatura ambiental para evitar perturbaciones.
- Al comparar entre sujetos tomar en cuenta la sequedad de la piel que diferencia a cada sujeto de prueba.

3.1.2. Sensor óptico de pulso de Shimmer

El sensor óptico de pulso de Shimmer mide el fotopletismograma (PPG) de un dedo, lóbulo de la oreja o cualquier otra parte del cuerpo con tejido capilar. Esta medición permite, mediante un algoritmo, estimar el ritmo cardíaco. La adquisición se realiza conectando el foto-sensor (Figura 3.5) a la entrada Jack 3,5 mm en la unidad *GSR+* (de la Figura 3.2, entrada negra).



Figura 3.5: Conexión de foto-sensor a la unidad *GSR+* de Shimmer.

Mediciones

Un PPG es un pletismograma obtenido ópticamente, que es una medición del cambio de volumen de un órgano. El PPG, en particular, mide los cambios en la absorción de la luz. El cambio de volumen provocado por el pulso de presión sanguínea se detecta mediante iluminar la piel con una luz LED y con un fotodiodo se mide la cantidad de luz transmitida o reflejada. Cada ciclo cardíaco aparece como un pico, como se ve en la Figura 2.12. Esta forma característica se puede detectar a partir de un PPG. Esta señal puede también ser usada para monitorear la respiración, hipovolemia y otras condiciones circulatorias. En la Figura 3.6 se presenta un ejemplo de una medición de la señal PPG con el sensor óptico de pulso de Shimmer. La forma de onda PPG difiere de sujeto a sujeto y la forma en que está conectado el sensor.

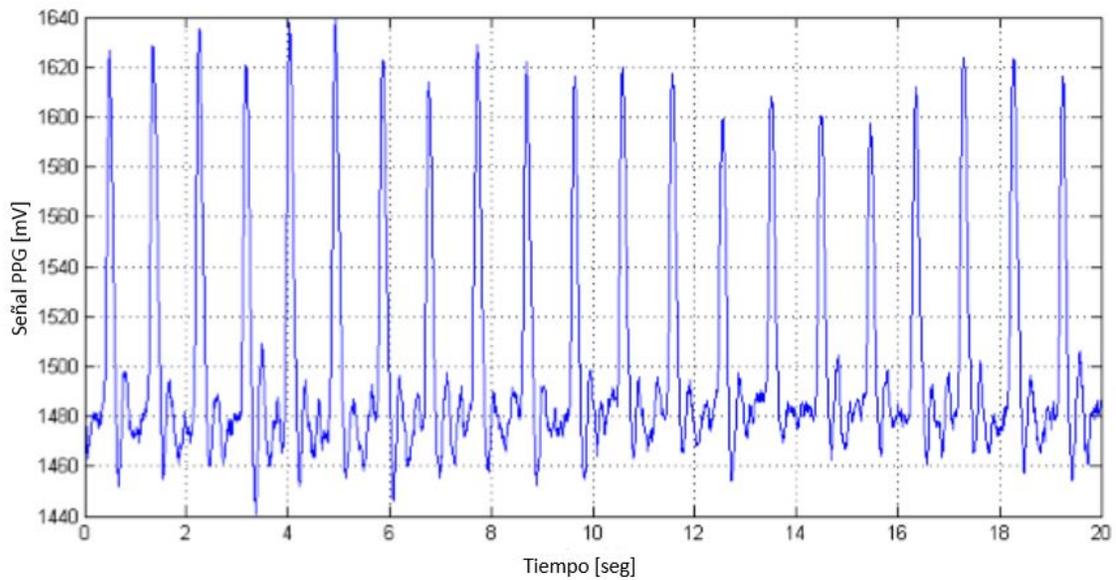


Figura 3.6: Ejemplo de señal PPG medida con el sensor óptico de pulso de Shimmer.

La frecuencia de muestreo debe ser lo suficientemente alta como para permitir que la frecuencia del pulso sea capturada correctamente. La frecuencia cardíaca normal en reposo oscila entre los 50 y 100 latidos por minuto. A través de la experimentación se ha encontrado que la frecuencia de muestro de 100 Hz o mayor proporciona un buen rendimiento. Sin embargo, siempre depende del uso o aplicación que se le dará a esta señal.

Conversión de PPG a ritmo cardíaco

Como se observa en la Figura 3.6 la señal PPG tiene una forma muy parecida a la forma característica del ciclo cardíaco (Figura 2.12). También se observa variación en la amplitud de los picos y es debido a la influencia que tiene en la señal PPG la respiración, la actividad muscular y el movimiento del sensor.

Es posible identificar el ritmo cardíaco a partir de la señal PPG. Existen muchos algoritmos para la conversión PPG-to-HR disponibles en la literatura publicada, como por ejemplo Lao, et al. (2013), Fu, Liu, & Tang, (2008) y Shin, Lee, & Lee (2009). Afortunadamente, el software de procesamiento que Shimmer provee incluye una aplicación para la conversión PPG-to-HR.

Posición del sensor

El sensor óptico puede ponerse tanto en el dedo como en el lóbulo de la oreja. Sin embargo, se considera que, para obtener una señal de mejor calidad, el lóbulo es la mejor opción, pues, da libertad a los movimientos de la mano, movimientos que aumentarían el ruido al capturar la señal.

Además del mínimo movimiento del lóbulo, lo que reduce el ruido, esta zona no tiene actividad muscular que cause interferencia en la medición PPG. El sensor debe colocarse en la parte inferior de lóbulo de la oreja, como se muestra en la Figura 3.7. Esto facilita la obtención de la información relacionada con el ritmo cardíaco.



Figura 3.7: Posición correcta en el lóbulo de la oreja del sensor óptico para mediciones de PPG.

Recomendaciones

- No conectar el sensor a la piel cuando el dispositivo esté conectado a un puerto de carga o USB.
- Asegurarse de que el sensor esté en la posición correcta y pedir el usuario que se mueva lo menos posible.
- Escoger una frecuencia de muestro de mayor a 100 Hz.

3.1.3. Unidad *Bridge Amplifier+* de Shimmer

La unidad *Bridge Amplifier+* de Shimmer por su diseño permite obtener una gran variedad de mediciones. Cuenta con una resistencia altamente sensible que permite conectar un sensor de temperatura. Se puede obtener la temperatura ambiental y la temperatura de la superficie de la piel. En la Figura 3.8 se observa la unidad donde se conecta el sensor de temperatura a la entrada amarilla.

Posición del sensor

La sonda del sensor de temperatura corporal de Shimmer se muestra en la Figura 3.9. Es importante asegurarse de que el sensor esté en la posición correcta. Se recomienda en el dedo pulgar aunque también puede medirse en la axila, ya que es una zona con movimiento

mínimo que no interviene en las tareas que realice con las manos el sujeto. Esto reduce el ruido por el movimiento del sensor.



Figura 3.8: Unidad *Bridge Amplifier+* de Shimmer.



Figura 3.9: Sonda del sensor de temperatura corporal de Shimmer.

Recomendaciones

- No conectar el sensor a la piel cuando el dispositivo esté conectado a un puerto de carga o USB.
- Se recomienda una baja frecuencia de muestreo. Por ejemplo, 10 Hz.

3.1.4. ECG de tarjeta biomédica todo-en-uno de BITalino

De la tarjeta biomédica de BITalino se utiliza el sensor para obtener electrocardiograma (ECG), dispositivo desarrollado para obtener la actividad eléctrica del corazón. Consta de una tarjeta de adquisición que transmite por un Bluetooth integrado. Al sensor se conecta un cable triple con tres electrodos desechables (Figura 3.10) que se pegan en el torso del sujeto y que recopilan los datos necesarios para obtener las ondas descritas en el ciclo cardíaco (ver Sección 2.3.5).

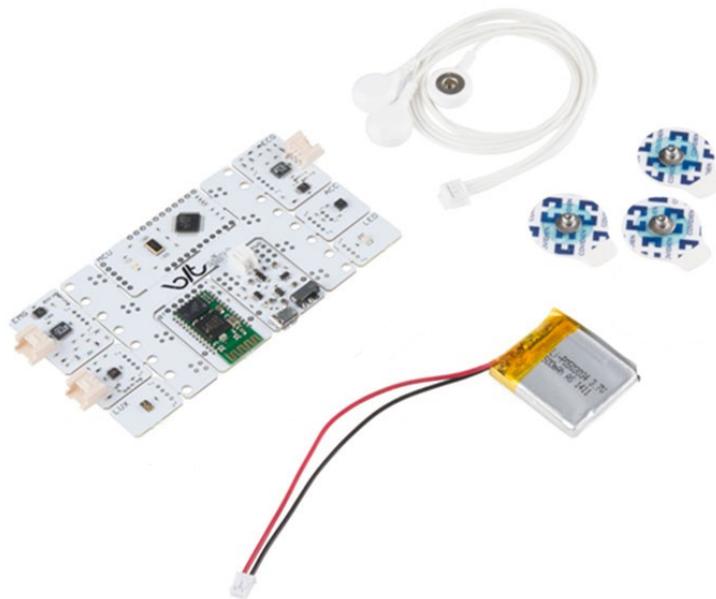


Figura 3.10: Tarjeta biomédica todo-en-uno de BITalino.

Mediciones

Para las mediciones de ECG se recomienda, aunque no es esencial, la preparación de la piel. Un buen contacto con la piel garantiza mejor calidad de la señal. Se debe reducir al mínimo la posibilidad de interferencia de la señal de los electrodos por el contacto esporádico causado por pelos, grasa corporal o partículas de suciedad. Los pelos en el punto de contacto pueden afeitarse y la piel ser limpiada con alcohol o desinfectante para eliminar la grasa y el sudor. También se puede utilizar gel conductor apropiado para ECG.

Respecto a la frecuencia de muestreo se recomienda que sea de 512 Hz para la adquisición de datos de ECG de grado clínico. Tasas menores pueden utilizarse en mediciones de ECG ambulatorio y según el requerimiento del usuario. El dispositivo BITalino puede medir a 100 Hz y 1000 Hz.

Posición de los electrodos

En la Figura 3.11 se muestra un ejemplo de cómo pueden ser conectados los electrodos al cuerpo. Esta configuración obedece el estándar Lead-II (compare con Figura 2.11).

Aunque la práctica común es colocar los electrodos en las extremidades correspondientes (brazos y piernas) de acuerdo con sus nombres, en realidad, todos los electrodos de las extremidades pueden ubicarse en el pecho. Lo importante es que cada electrodo quede lejos del corazón y en la dirección de la extremidad correspondiente. Por ejemplo, el electrodo positivo se puede colocar en cualquier lugar del pecho, siempre y cuando se coloque lejos del corazón en la dirección al hombro derecho, como se observa en la Figura 3.11 .

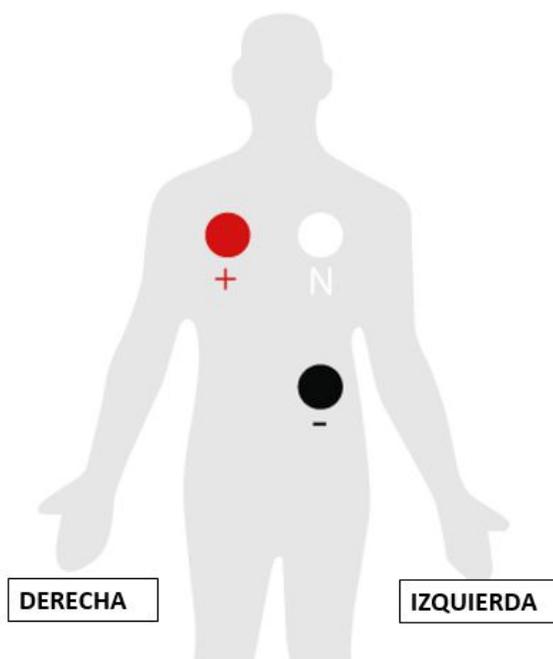


Figura 3.11: Ejemplo de posicionamiento de los electrodos para le medición de ECG.

Recomendaciones

- No conectar los electrodos a la piel cuando el dispositivo esté conectado a un puerto de carga o USB.
- Para minimizar las interferencias (de la red eléctrica o de señales inalámbricas) se recomienda escoger cables del menor largo posible.
- Si se requiere utilizar cables muy largos se recomienda utilizar filtros para eliminar el posible ruido generado, como filtro pasa banda para eliminar el ruido a la frecuencia de la red (50/60 Hz). También un filtro pasa alto a fin de eliminar las componentes de baja frecuencia donde se propone una frecuencia de corte de 0,05 Hz para ECG de diagnóstico, mientras que una frecuencia de corte de 0,5 Hz es adecuado para la monitorización de ECG a largo plazo.

3.1.5. EEG Emotiv EPOC

Es un EEG inalámbrico de 16 canales (AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, y O2). De ellos 14 son canales efectivos, 2 son de posicionamiento. Sigue al sistema internacional 10-20 (ver Figura 2.7). La frecuencia de muestreo del sensor de 128 Hz. La adquisición de los datos es en dos formatos: .txt o .edf (European Data Format). En la Figura 3.12 se muestra el dispositivo.



Figura 3.12: Emotiv EPOC EEG. Dispositivo inalámbrico para adquisición de electroencefalograma.

Mediciones

Para la medición del EEG se debe hidratar cada electrodo para mejorar la conductividad de la señal eléctrica y ubicar directamente el sensor en el cuero cabelludo. Esto ayuda a evitar el ruido generado por el movimiento y roce con el pelo. En la Figura 3.13 se muestran las mediciones en tiempo real tomadas con el software *Emotiv Xavier TestBench* para medir EEG. Se observan los 14 canales medidos y el panel de control de conexión de los electrodos.

Posición del sensor

El EEG se debe poner en la cabeza de arriba hacia abajo, ajustando el soporte por detrás de la oreja y ajustando los electrodos de la frente a tres dedos de la ceja, como se muestra en la Figura 3.14. Luego, el electrodo de referencia, señalado con la flecha, deben estar en la posición correcta para que los demás electrodos sean detectados por el software. Se debe conseguir que todos los electrodos estén en verde, como se observa en la Figura 3.13.

Recomendaciones

- En lo posible escoger usuarios con pelo corto para facilitar el contacto de los electrodos con el cuero cabelludo.
- Para cuidar los electrodos de la oxidación se deben secar y guardar en su caja al finalizar su uso.
- Evitar los movimientos de la cabeza para evitar el ruido y los artefactos por movimiento y electricidad muscular.



Figura 3.13: Software Xavier Testbench para la verificación de la posición del sensor Emotiv EPOC EEG.

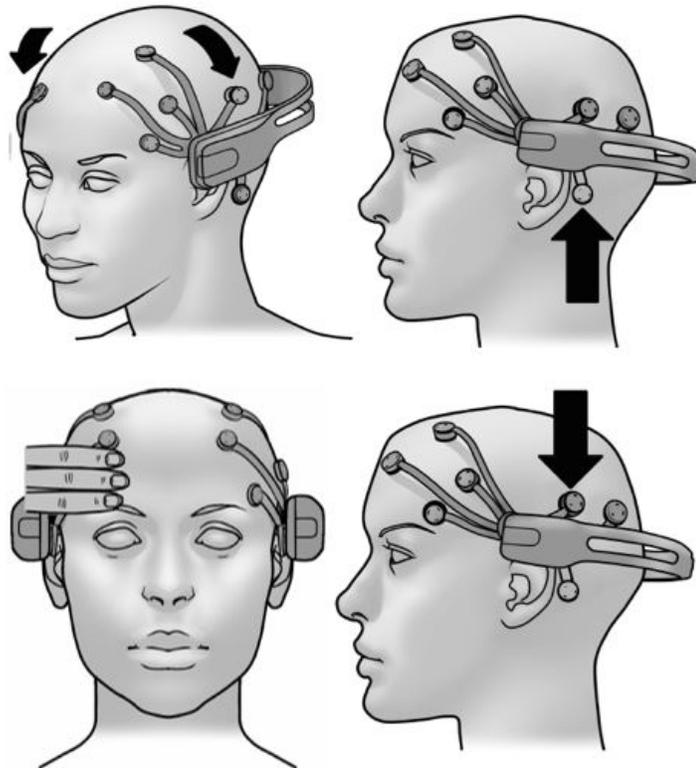


Figura 3.14: Posición del sensor Emotiv EPOC EEG. Instrucciones para poner correctamente el EEG.

3.1.6. *Eye Tracker Tobii T120*

Eye Tracker Tobii T120 es un instrumento de medición que permite seguir la mirada del sujeto de prueba con alta precisión. El Eye Tracker viene integrado a una pantalla de 17 pulgadas lo que permite un entorno más natural para tener mediciones más realistas. En la Figura 3.15 se muestra el dispositivo.



Figura 3.15: Tobii T120. Eye Tracker usado en los experimentos.

Mediciones

Durante el seguimiento, Tobii utiliza diodos infrarrojos para generar patrones de reflexión sobre las córneas de los ojos del usuario. Estos patrones de reflexión, junto con otra información visual acerca de la persona son recogidos por los sensores. Sofisticados algoritmos de procesamiento de imágenes identifican características pertinentes, incluidos los ojos y los patrones de reflexión de la córnea. Con cálculos complejos se calcula la posición tridimensional de cada globo del ojo, y, finalmente, el punto de mirada en la pantalla; es decir dónde el usuario está mirando.

Se pueden obtener distintas métricas como: distintas características de las fijaciones y las sacadas y el diámetro pupilar. También se pueden observar mapas de calor, clusters, entre otras, como se muestra en la Figura 3.16.



(a) Mapa de la mirada.



(b) Mapa de calor.

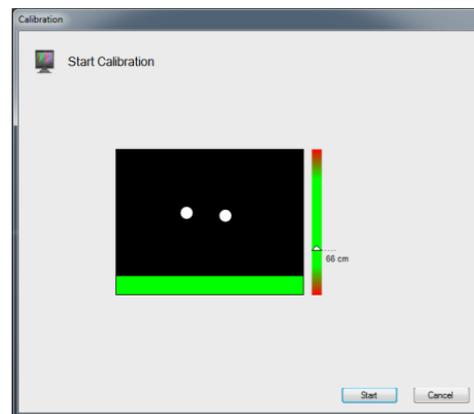
Figura 3.16: Ejemplo de mediciones en Tobii Studio.

Posición del dispositivo

Para asegurar una correcta medición el usuario debe sentarse en frente con la vista en medio de la pantalla. El asiento debe ser cómodo y con pocos grados de movimiento (sin ruedas y con respaldo fijo) y la distancia a la pantalla debe ser aproximadamente del largo de su brazo (ver Figura 3.17a). El software Tobii Studio permite en la calibración ajustar la posición del usuario (Figura 3.17b).



(a) Posición del usuario frente al Eye Tracker Tobii T120.



(b) Detección de ojos y calibración de posición de usuarios en Tobii Studio.

Figura 3.17: Posición del Eye tracker Tobii T120.

Recomendaciones

- Evitar que el usuario se mueva de la posición de calibración: usar una silla fija (es decir, sin ruedas), ajustar el alto de la pantalla para evitar que la persona se incline hacia adelante, dar instrucciones explícitas de no moverse.
- Controlar la iluminación variable para evitar que afecte la medición del diámetro pupilar.

3.2. Grupo experimental

Se realiza el experimento con 61 participantes (19 mujeres y 42 hombres) entre las edades de 19 a 35 años. Todos los participantes estaban relacionados al contexto de la educación universitaria, entre sus ocupaciones: un enfermero, una licenciada de literatura hispánica, algunos ingenieros y mayoritariamente estudiantes universitarios. De estos se utilizan 53 participantes válidos. Los 8 restantes se descartan por diversos problemas de medición. Detalles de género, edad y ocupación del grupo experimental se presentan en el Anexo A.

Además, se consideran las siguientes características según lo expuesto en [46]:

- Personas saludables: Que no presenten enfermedades que perjudiquen los resultados del experimento y que no estén bajo tratamiento farmacológico ni que presenten antecedentes de enfermedades neurológicas o psiquiátricas.
- Sin uso de medicamentos o drogas recreativas durante las últimas 24 horas.
- Con visión o visión corregida (uso de lentes) normal.
- Que hayan firmado el consentimiento informado declarando que cumplen los requisitos anteriores y que permiten el uso de los datos recolectados.

3.3. Diseño experimental

El objetivo del experimento es demostrar que el paso entre el análisis de un web object a otro por parte de un usuario que navega libremente por una página web, resulta en una baja de la carga cognitiva del participante.

La variable independiente o manipulada es la configuración del sitio web en tres escenarios expuestos aleatoriamente. Los elementos que se varían son: encabezados de noticias, foto representativa de la noticia y los anuncios publicitarios (ads). Las variables dependientes son: la carga cognitiva que se mide mediante señales psico-fisiológicas obtenidas de: ECG, EEG, GSR, PPG, temperatura corporal y dilatación pupilar.

3.3.1. Escenarios

Se escoge para el estudio un sitio web ficticio, creado específicamente para el experimento, que es corrido en forma local para evitar retardos del proveedor de Internet.

Los elementos dentro del sitio web son: 7 encabezados de noticias con su respectiva imagen representativa, 4 anuncios de publicidad, una barra de navegación típica con menú, logo de la página en la esquina superior izquierda y un barra al pie de página.

Estos elementos siguen una configuración básica. Un ejemplo se presenta en la Figura 3.18. Esta disposición de los elementos web se mantiene para todos los escenarios generados.

A cada participante se le presenta este escenario 3 veces y para ellos existe un total de 21 noticias y 12 anuncios disponibles. Para evitar sesgos, tanto los anuncios como los encabezados de noticias aparecen siempre de forma aleatoria y sin repetirse entre los 3 escenarios por usuario, esto es para disminuir la probabilidad de que aparezcan páginas web idénticas tanto para un mismo usuario como entre usuarios.

El contenido escogido, tanto para noticias como para anuncios, se relaciona con la vida estudiantil. Dentro de los temas se encuentran: viajes, cursos, cervezas, educación, música, tecnología y misceláneos.

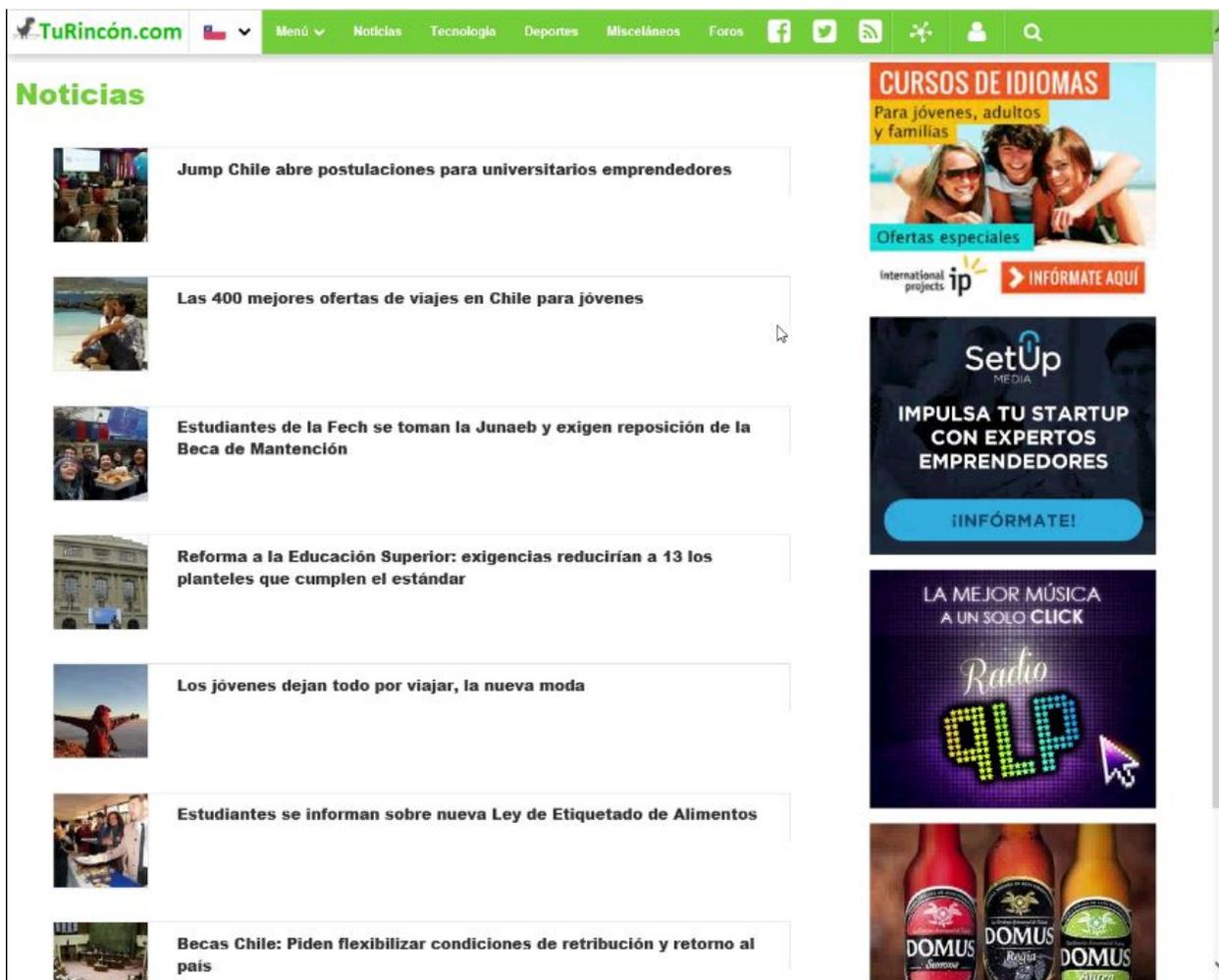


Figura 3.18: Ejemplo de página ficticia para el experimento.

3.3.2. Preparación del experimento

Se adquieren todos los insumos necesarios para llevar a cabo los muchos experimentos como: electrodos desechables, cinta Micropore, suero fisiológico para hidratar los electrodos del EEG.

Para tener un buen ambiente experimental es necesario contar con un espacio dedicado para el funcionamiento del estudio. Idealmente una habitación aislada que no sea utilizada por otros, para mantener la configuración experimental y el entorno físico tan constante como sea posible [41].

Para el seguimiento de los ojos, las condiciones de iluminación son esenciales. Se debe evitar la luz directa del sol ya que la luz solar contiene luz infrarroja que afecta la calidad de las mediciones de seguimiento ocular. También se debe evitar los cuartos altamente iluminados, es decir, no debe haber ninguna luz desde arriba [41].

Para evitar estos efectos de la luz, se aísla el laboratorio de la luz externa con cortinas *black-out* que se observan en la Figura 3.19. Por lo tanto, el experimento se lleva a cabo a oscuras. Además, se toman las medidas para evitar todo tipo de interrupciones externas durante el experimento.



Figura 3.19: Laboratorio donde se llevan a cabo los experimentos. Se observan las cortinas *black-out* a ambos lados de la sala para bloquear las ventanas.

3.4. Procedimiento experimental

Una vez llegada la persona se le explica el experimento, se pide que lea y firme el consentimiento informado. Se sienta a la persona frente a la pantalla. Se le conectan los sensores en el siguiente orden: (1) ECG, (2) temperatura axilar, (3) EEG, (4) GSR, (5) PPG y se calibra el eye tracker (6); como se muestra en el diagrama de la Figura 3.20.

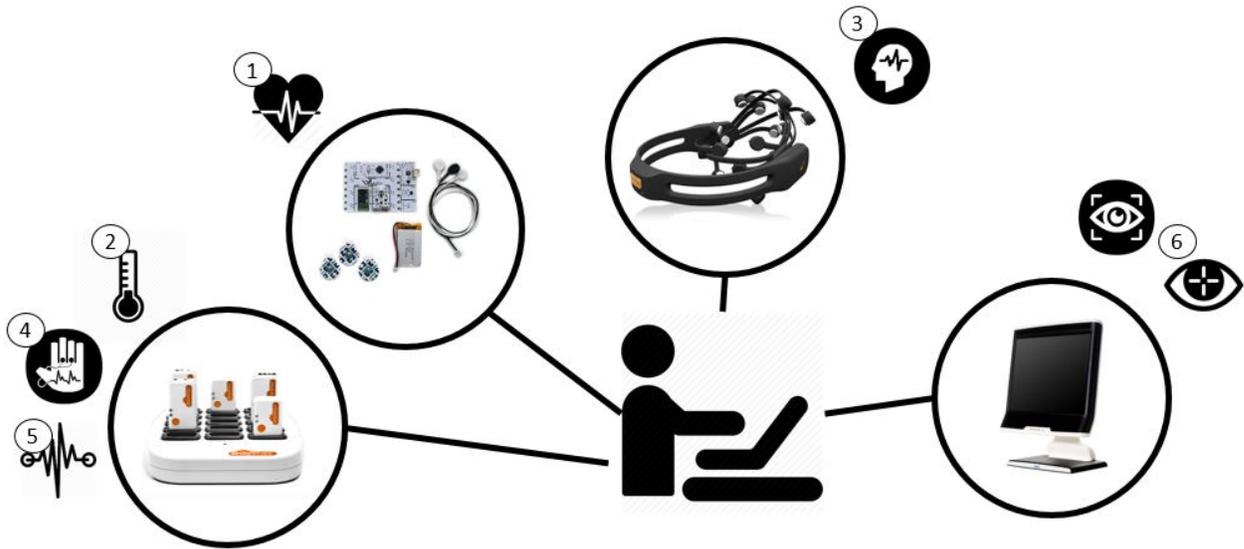


Figura 3.20: Diagrama de sensores que se conectan a los participantes.

Previo a las pruebas se somete a cada usuario a una etapa de relajación consistente en la visualización de tres videos de 4 minutos cada uno, de bellos paisajes con música instrumental de fondo. La segunda parte de esta etapa se pide al participante que respire profundo durante un minuto con los ojos cerrados y con música instrumental suave de fondo. Esta etapa tiene como objetivo eliminar el efecto Hawthorne (modificación en la conducta como consecuencia de saber que están siendo estudiados) y efectos fisiológicos similares al “bata blanca” en las señales medidas. Además, se pide al participante que mantenga una postura fija sentado frente al computador, sin mover la cabeza ni la mano derecha donde van conectados algunos sensores.

Los tres escenarios (descritos en la Sección 3.3.1) son presentados en forma aleatoria para cada participante. Esto se realiza con una opción disponible en el software del Eye Tracker. La indicación es que navegue libremente por la página web el tiempo que quiera y que indique cuando desee pasar al siguiente escenario. Una vez navegado por los 3 escenarios se da por finalizado el experimento.

Configuración de los sensores

Para medir el EEG se utiliza el sensor *Emotiv EPOC EEG* con una frecuencia de muestreo de 128 Hz. El sensor se pone sobre la cabeza posicionando primero los sensores de referencia. Para mejorar la conducción de las señales eléctricas del cerebro se hidrata cada electrodo

previamente. Para capturar los datos y verificar que el sensor quedó bien puesto se utiliza el software provisto por el fabricante llamado *Emotiv Xavier Testbench*.

Para medir el ECG se utiliza el dispositivo *BITalino BioMedical Development All-in-One Board* con una frecuencia de muestreo de 1000 Hz. La configuración de los tres electrodos sigue el estándar lead II. Antes de poner los electrodos se prepara la piel limpiándola de la grasa e impurezas con el fin de disminuir el ruido. Además, se utiliza un gel para ECG para la medición. Se dispone del software *OpenSignals (r)evolution* provisto por el fabricante.

Para medir las señales GSR y PPG se utiliza el sensor *Shimmer GSR+ Unit* con una frecuencia de muestreo de 120 Hz. La posición de los electrodos para la medición de la respuesta galvánica es en la zona palmar de la falange proximal del dedo índice y anular de la mano derecha. El sensor de PPG se ubica en el lóbulo de la oreja derecha.

Para medir la temperatura corporal se utiliza el sensor *Shimmer Bridge Amplifier+ Unit* con una frecuencia de muestreo de 50 Hz. El sensor se posiciona bajo la axila derecha. Ambos sensores *Shimmer* se sincronizan mediante una base que facilita el proveedor junto con el software *Consensys*.

Para medir la dilatación pupilar y el seguimiento de la mirada se utiliza el dispositivo *Eye Tracker Tobii T120* con una frecuencia de muestreo de 120 Hz. Para calibrar y realizar la toma de datos se utiliza el software provisto *Tobii Studio*.

Finalmente, se muestra en la Figura 3.21 un participante con todos los sensores puestos mientras ejecuta el experimento de analizar la página web descrita en la Sección 3.3.1.

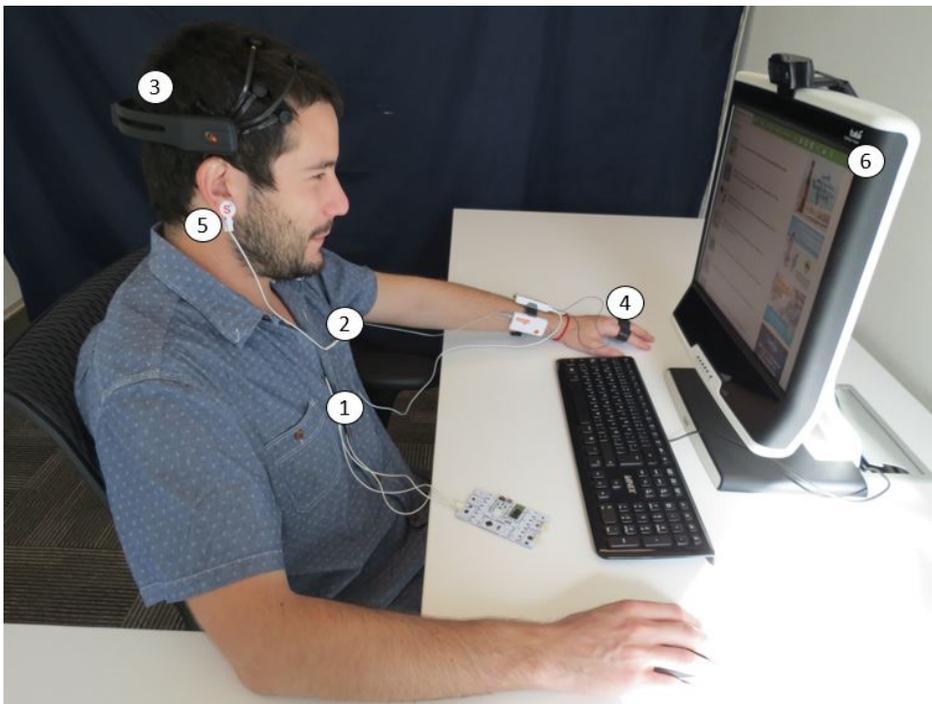


Figura 3.21: Participante con los sensores ejecuta el experimento. Los sensores son: (1) ECG, (2) temperatura axilar, (3) EEG, (4) GSR, (5) PPG y (6) eye tracker

Capítulo 4

Análisis de Datos

4.1. Definición de observaciones

En base a la investigación de Bailey (2008) [18] de que en la transición entre sub-tareas se genera una baja en la carga cognitiva, se ha considerado el análisis de cada *web object* como una sub-tarea específica y el paso entre ellos como el período de transición. Es por ello que en este estudio se evalúa la carga cognitiva en dos tipos de ventanas de tiempos:

- **Ventana activa sobre un elemento:** Es el tiempo en que el usuario se mantiene dentro de un área de interés específica que puede corresponder a un encabezado de noticias, un anuncio publicitario o la barra de menú del sitio web.
- **Ventana de transición:** El tiempo que transcurre mientras el usuario no se mantiene observando ningún área de interés, por lo que se toma como una transición entre dos elementos, o bien, entre el mismo elemento (una revisita).

4.2. Descripción y procesamiento de datos

4.2.1. Ventanas de tiempo

Las ventanas de tiempo definen el período en que el usuario inspecciona un elemento, o bien, está en una transición. Para discriminar entre tipos de ventanas, el archivo de los datos exportados desde el programa *Tobii Studio* genera una columna con el área de interés (AoI, *Area of Interest*) que el participante está inspeccionando para cada muestra. Discrimina entre 3 valores: cuando el usuario no está mirando la pantalla (inactivo), cuando el usuario mira cierta AoI (ventana activa) y cuando el usuario está inspeccionando fuera del AoI (lo que se considera como ventana de transición).

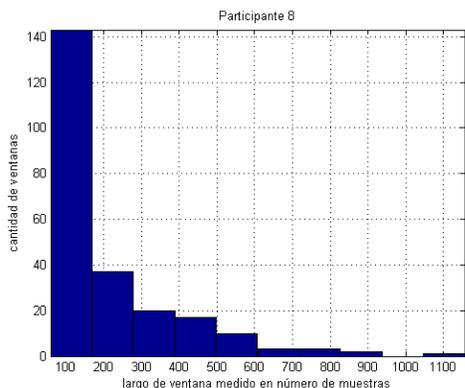
Se establece un largo mínimo de tiempo de 500 milisegundos para definir una ventana de tiempo válida. Esto se basa en la investigación de Loyola (2015) [53] que analiza con *eye*

tracking la identificación de *web objects* claves en un sitio web. Este tiempo se elige para evitar la posible contaminación de la señal de la pupila por el análisis de un objeto anterior.

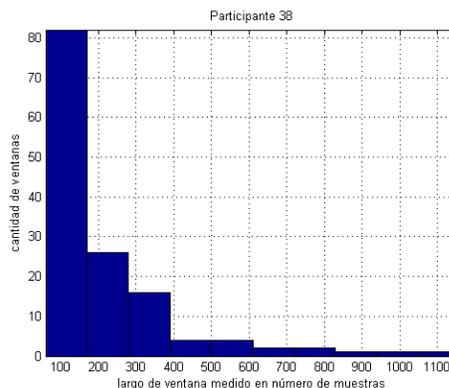
Las ventanas de tiempo de largo menor al umbral no se consideran para el análisis, y por lo tanto, se eliminan. Si coincide que antes y después de la ventana eliminada resulta en el análisis del mismo *web object* se juntan generando una ventana de largo mayor.

Al eliminarse las ventanas de tiempo menores a 500 ms (60 muestras): Se obtiene solo el 25% del total de ventanas, en promedio para todos los participantes. A su vez sólo el 15% (del 25 por ciento que quedan) son ventanas de transición.

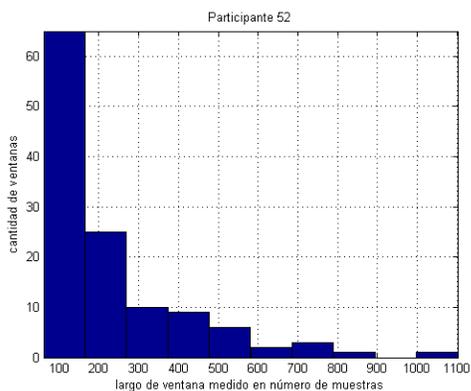
Un hallazgo que podría darnos información del comportamiento del usuario web es que se observa una tendencia en la distribución de la cantidad de ventanas versus su duración, tendencia que es independiente del tiempo de navegación del usuario como se observa en la Figura 4.1.



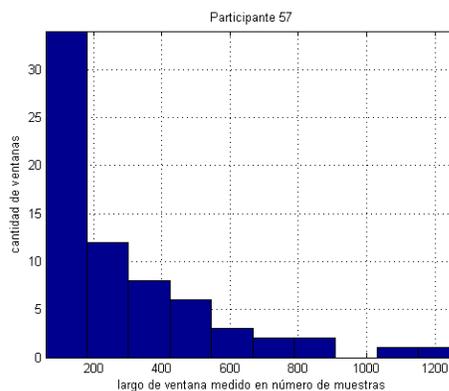
(a) Participante 8 con 237 ventanas.



(b) Participante 38 con 140 ventanas.



(c) Participante 52 con 123 ventanas.



(d) Participante 57 con 70 ventanas.

Figura 4.1: Histogramas de la cantidad de ventanas según su largo para distintos participantes con distinto tiempo de navegación. Dado que la tasa de muestreo del *eye tracker* es de 120 Hz, se considera entonces que 120 muestras corresponden a 1000 milisegundos.

4.2.2. Diámetro pupilar

La data exportada desde *Tobii Studio* contiene el diámetro de la pupila izquierda, el diámetro de la pupila derecha (en milímetros) y la validación de la confiabilidad de la captura de cada pupila valuada entre 0 (alta confiabilidad) y 4 (no se encontró el ojo). En promedio, para todos los participantes y considerando sólo las ventanas válidas, existe una confiabilidad para la captura de la pupila izquierda de 0.2469; y para la pupila derecha de 0.22036; valores suficientemente buenos para validar la captura de los datos del diámetro de la pupila. Como estos valores son un promedio de los participantes se escogió por cada caso los datos de la pupila con mayor grado de confiabilidad.

A continuación, se procede a eliminar los artefactos que distorsionan la señal como las sacadas y los pestaños. La data extraída cuenta con una columna que determina si la muestra es una fijación o una sacada, dicha información es usada para filtrar las sacadas. Por otro lado, se utiliza una interpolación lineal entre los valores de los pestaños detectados. Un ejemplo del resultado de este procesamiento se observa en la Figura 4.2. Además, se aplica como filtro pasa bajos una ventana de Blackman con una frecuencia de corte de 2 Hz.

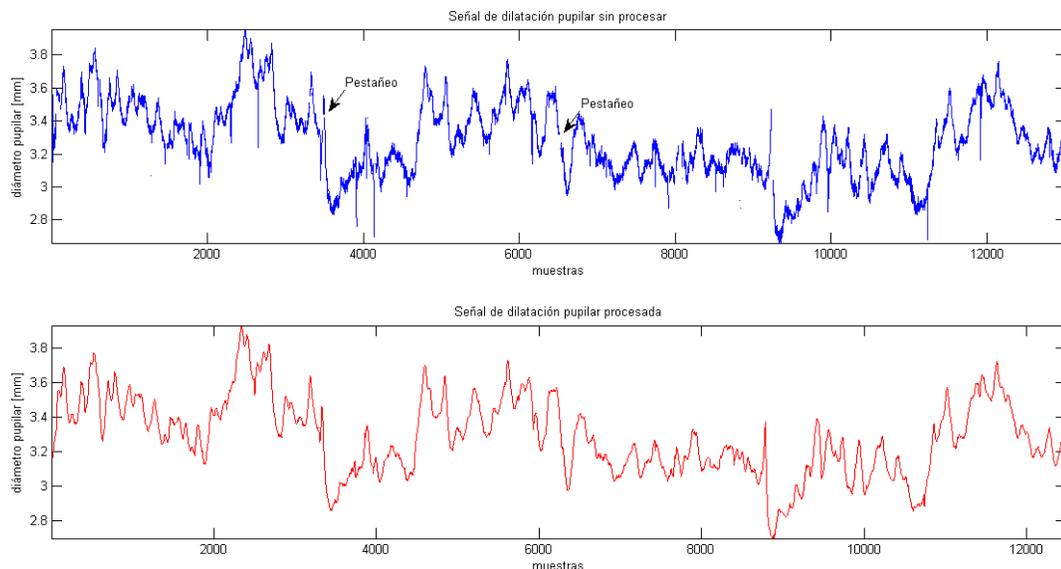


Figura 4.2: Señal de dilatación pupilar procesada. Se señalan artefactos como el pestaño y cómo la interpolación reconstruye la señal y, además, cómo el filtraje disminuye el ruido.

4.2.3. Respuesta electrodérmica (GSR)

La respuesta galvánica de la piel o actividad electro-dermal se obtiene desde *Consensus*, software provisto por *Shimmer Sensing* para la configuración de sus dispositivos y el manejo de datos. La data cruda entrega los valores de resistencia eléctrica de la piel en Kiloohmios [$k\Omega$]. Para disminuir el ruido y eliminar los artefactos de movimiento se realizan dos cosas: primero, se da la instrucción estricta a cada participante de no mover la mano ni los dedos donde van conectados los electrodos; y segundo, se filtra la señal con un pasa bajos de frecuencia de corte de 5 Hz como se recomienda en la sección 3.1.1. Además, por recomendaciones

de *GSR Pocket Guide*, de iMotions (2016) [42] se reduce la resolución sin riesgo de pérdida de datos. La señal GSR que se mide con frecuencia de muestreo de 120 Hz se reduce a 10 muestras por segundo. En la Figura 4.3 se muestra el resultado del procesamiento descrito.

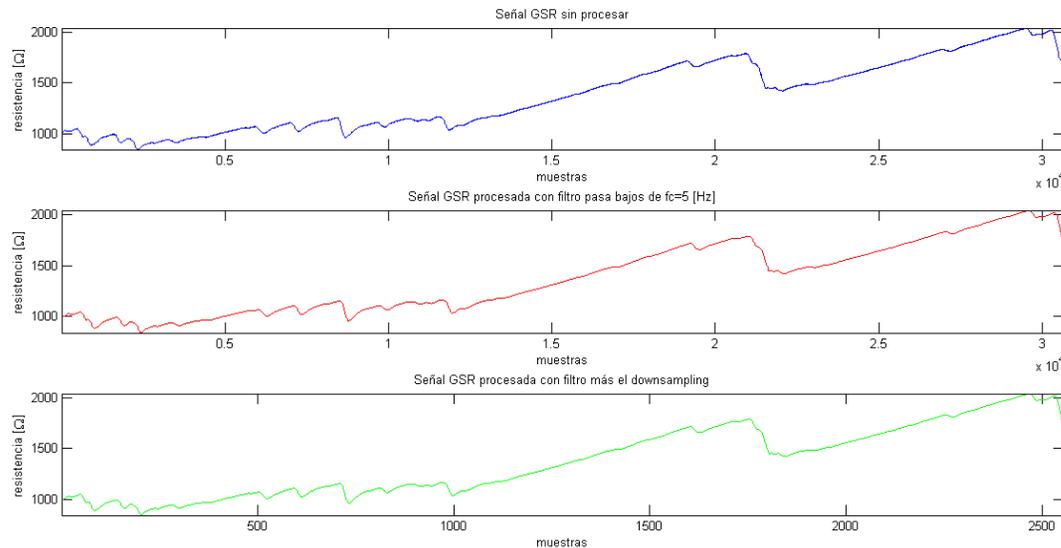


Figura 4.3: Señal GSR procesada. Se presenta la señal original (arriba), la señal filtrada con un pasa bajos de 5 Hz (en medio) y la señal con resolución disminuida a 10 Hz (abajo). Se demuestra el mejoramiento de la señal y que no produce pérdida de datos.

La señal GSR cuenta con dos componentes, como se ha explicado en la sección 2.3.4. Se extrae la componente fásica mediante recorrer la señal aplicando un filtro de mediana con ancho de ventana de $-4/+4$ y restando la media de la muestra actual [42]. El resultado se observa en la Figura 4.4. Esta componente permite detectar los peaks de la señal GSR. Ante transiciones lentas, la componente fásica no presenta mayores variaciones.

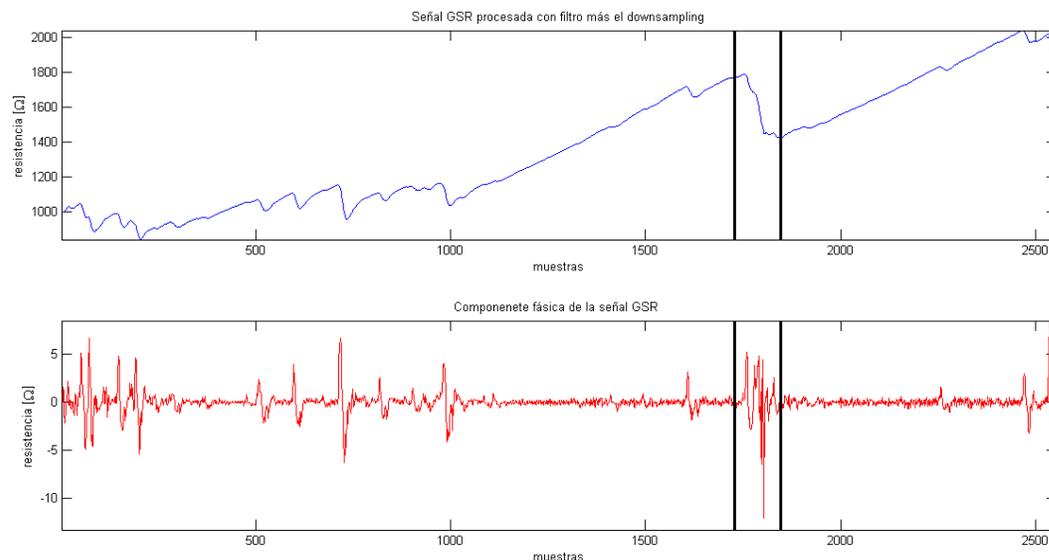


Figura 4.4: Componente fásica obtenida de la señal GSR. Se obtiene la componente fásica de la señal GSR procesada. Esta componente ayuda a encontrar los peaks de la señal GSR como se observa en la ventana de tiempo marcada.

4.2.4. Electrocardiograma (ECG)

La señal ECG se obtiene desde el software *OpenSignals* provisto por *BITalino* para la toma de datos en tiempo real. La data cruda arroja valores que deben ser escalados para obtenerlos en milivolts [mV]. El procesamiento de esta señal consiste en utilizar un filtro pasa bajos de frecuencia de corte de 100 Hz. El resultado del procesamiento de esta señal se muestra en la Figura 4.5.

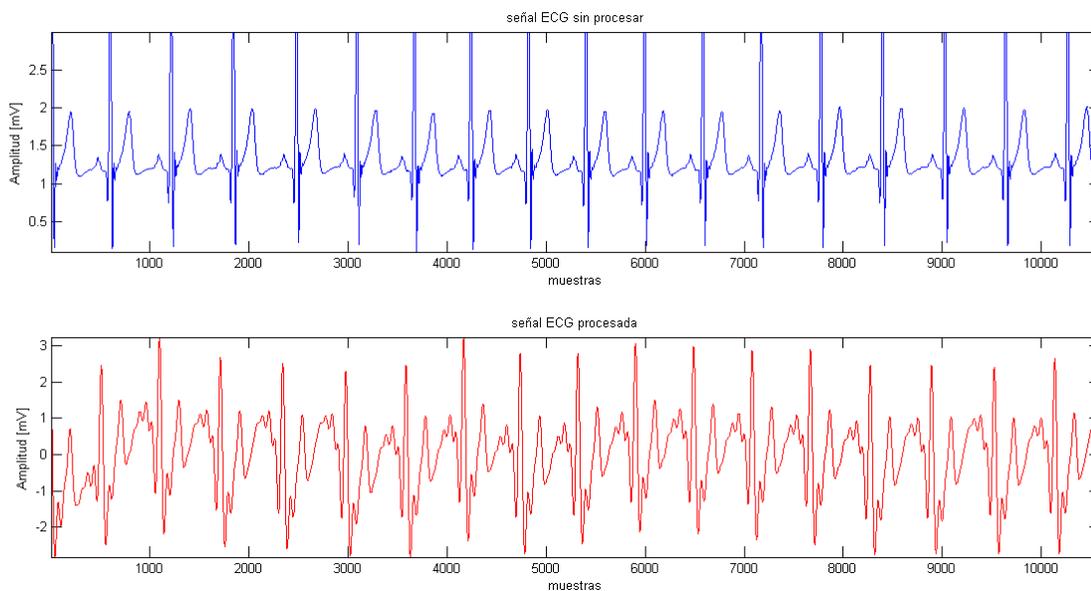


Figura 4.5: Señal ECG procesada. Se presenta la data cruda (arriba) y además el resultado de procesar la señal ECG.

4.2.5. Fotopletismograma (PPG)

La señal PPG se obtiene desde el software *Consensus* provisto por *Shimmer Sensing* para la toma y manejo de datos. La data cruda arroja valores de la señal en milivolts [mV]. De esta señal es posible obtener el ritmo cardíaco (HR). Previamente, se procesa la señal PPG con un filtro pasabajos con frecuencia de corte de 16 Hz con ventana de Blackman obteniéndose una señal más limpia. Luego, se obtiene el ritmo cardíaco siguiendo los siguientes pasos: primero, se debe encontrar los máximos; segundo, se resta el tiempo entre ellos; tercero, se convierte de centésimas a segundos y se convierte desde [segundos/pulso] a [pulsos/segundo], y por último, se multiplica por 60 para convertir a [pulsaciones/minuto].

El resultado de este procesamiento se observa en la Figura 4.6 donde se presenta la data cruda (arriba), el resultado de procesar la señal PPG con el filtro (al medio) y la señal HR (abajo) para el mismo conjunto de datos.

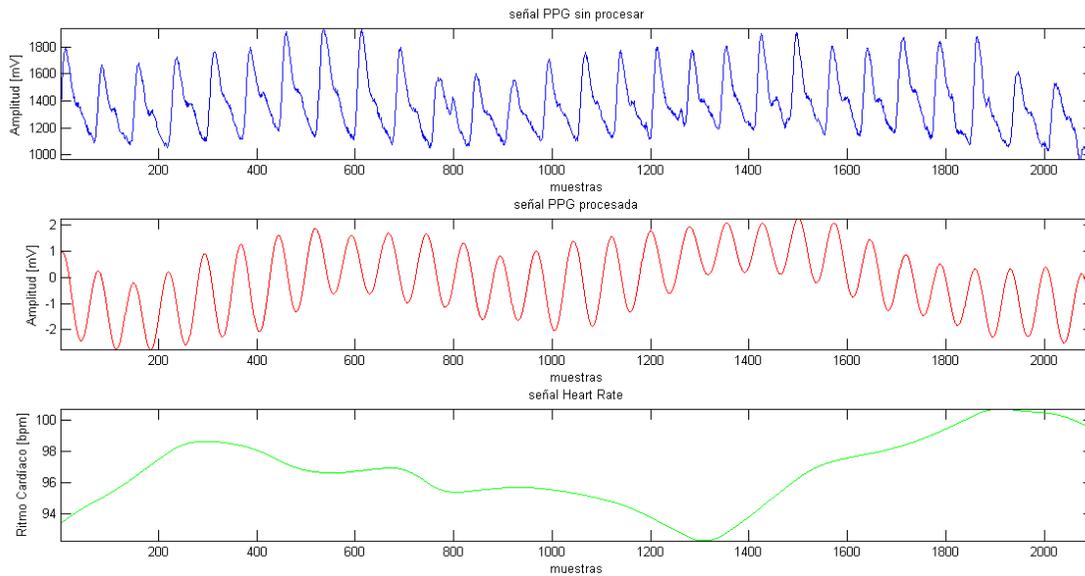


Figura 4.6: Señal PPG procesada. Se observa el resultado del filtraje y la obtención de la señal de ritmo cardíaco (HR) a partir de PPG.

4.2.6. Temperatura (T)

La señal de temperatura corporal medida en la axila se obtiene desde *Consensys* software provisto por *Shimmer Sensing* para el manejo de la data de este sensor. La data cruda arroja los valores de temperatura corporal en grados Celsius [$^{\circ}C$]. El procesamiento de esta señal consiste en utilizar un filtro pasa bajos de frecuencia de corte de 1 Hz, como se concluye de la toma de datos de Haapalainen et al. (2010) [24]. En la Figura 4.7 se observa el resultado de este procesamiento.

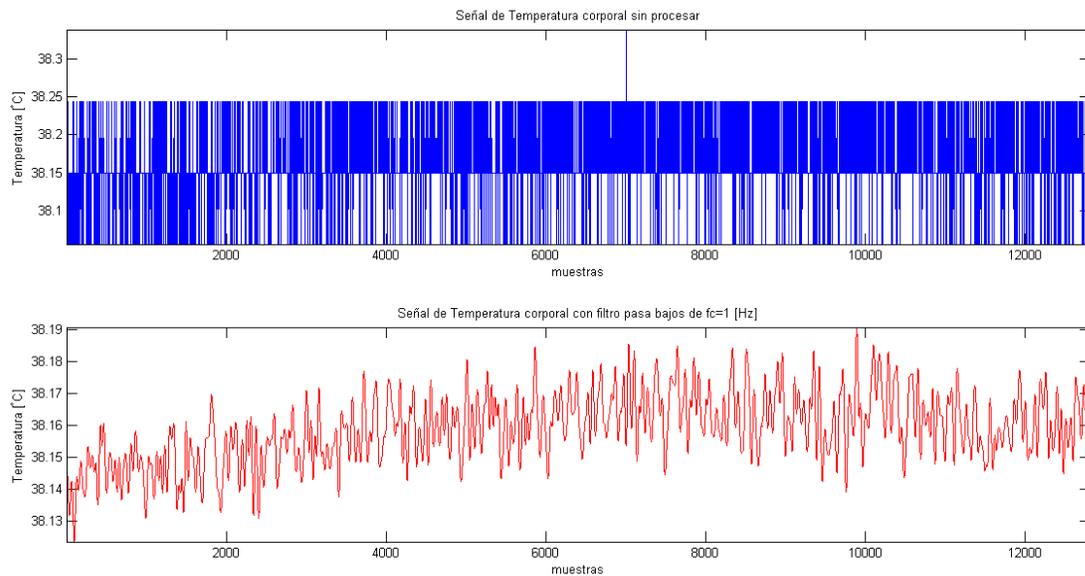


Figura 4.7: Señal de temperatura corporal procesada. Se presenta la data cruda (arriba), y además, el resultado de aplicar un filtro pasa bajos con frecuencia de corte de 1 Hz (abajo).

4.2.7. Electroencefalograma (EEG)

La señal EEG está sujeta a una gran variedad de artefactos y ruido que muchas veces es difícil de controlar y eliminar. Dentro de los elementos que generan artefactos están: los parpadeos, actividad oculomotora, los movimientos de la cabeza, las expresiones faciales que agregan ruido por la señal eléctrica muscular, movimiento de los electrodos, entre otros.

Para eliminar el efecto del balanceo de la cabeza se utiliza un filtro pasa alto de frecuencia de corte de 0.5 Hz. Además, se usa un filtro pasa bajos de frecuencia de corte de 40 Hz, para elimina ruido de la red eléctrica (50 - 60 Hz).

Para eliminar outliers y disminuir el efecto del artefacto parpadeo (como no se cuenta con la señal EOG para restarla mediante una regresión lineal al EEG y eliminar el parpadeo) se usa un filtro Hampel como opción [26].

4.3. Extracción de características

La extracción de características se realiza por ventanas de tiempo (ver Sección 4.2.1) considerando los 3 escenarios. Se obtienen las características de las señales procesadas.

Las señales tienen diferentes escalas. Para que sean objetos comparables es necesario estandarizar las señales antes de extraer características de ellas, como propone Guyon (2006) [35]. Para realizar la estandarización se utiliza la forma clásica $(\mathbf{x} - \mu)/\sigma$, donde \mathbf{x} es el vector correspondiente a la señal, μ y σ la media y la desviación estándar de la señal, respectivamente.

Se extrae un total de 44 características de las distintas señales: 2 de la dilatación pupilar, 6 del GSR, 2 de temperatura corporal, 3 de ECG, 3 de PPG-HR, y 2 por cada uno de los 14 canales del EEG. En la Tabla 4.1 se muestra un resumen de las características, y luego, se presenta más en detalle las características obtenidas.

Tabla 4.1: Características extraídas por cada señal.

Señal	Características extraídas
Pupila	media y varianza
GSR	data acumulada, promedio en función del tiempo y potencia espectral
Comp. Fásica	media, valor absoluto del máximo y cantidad de peaks
ECG	media, mediana y varianza de ECGMAD (desviación absoluta media)
PPG(HR)	media, desviación estándar y media cuadrática (RMS) del HR
T	media y mediana
EEG	potencia y fase de la señal analítica obtenida con la Transf. de Hilbert

4.3.1. Diámetro pupilar

La relación entre el tamaño de la pupila y la actividad mental ha sido ampliamente estudiada y se ha demostrado cómo la reacción pupilar refleja información en el proceso cognitivo. Producto de que se ha probado que la respuesta pupilar es un importante indicador del esfuerzo mental envuelto en resolver una tarea, se escoge como “*gold standard*” para clusterizar las ventanas y generar etiquetas de niveles cognitivos. Existe en la literatura casos de clusterización en el desarrollo de tareas web como en el trabajo de Loyola (2015) [53]. Las características seleccionadas son: la media y la varianza del diámetro pupilar del ojo con mayor confiabilidad en su medición.

4.3.2. Respuesta electrodérmica (GSR)

Basándose en las conclusiones de Nourbakhsh (2012) [51] y Shi et al. (2007) [56] se extraen las siguientes características a la señal procesada del GSR: Data acumulada normalizada, promedio en función del tiempo normalizado y potencia espectral sin componente continua normalizada.

Nourbakhsh [51] demuestra que la data del GSR normalizada tiene mejores resultados. Esta normalización de cada segmento de ventana se debe realizar con respecto a la señal completa, es decir, considerando todos los escenarios. En la Ecuación 4.1 se muestra cómo se calcula la señal GSR normalizada, se suma cada punto en el tiempo t , donde i corresponde al participante, k y m es la cantidad total de tareas, $m = 3$ en este caso.

$$GSR_{normalizado}(i, k, t) = \frac{GSR(i, k, t)}{\frac{1}{m} \sum_{j=1}^m \sum_{t=1}^{T_{ij}} GSR(i, j, t)} \quad (4.1)$$

Por lo tanto, se normalizan los datos de cada participante dividiendo la señal de la tarea por el valor medio de todas las tareas del sujeto. Luego, se calculan las características GSR acumulado como se muestra en la Ecuación 4.2 y GSR promedio según la Ecuación 4.3 donde T es el tiempo total de todas las tareas.

$$GSR_{acumulado}(i, k) = \sum_t GSR_{normalizado}(i, k, t) \quad (4.2)$$

$$GSR_{promedio}(i, k) = \frac{\sum_t GSR_{normalizado}(i, k, t)}{T} \quad (4.3)$$

De la componente física obtenida (ver Sección 4.2.3) se extraen las características: número de peaks, módulo del máximo y promedio de la componente física de la ventana [27].

4.3.3. Electrocardiograma (ECG)

Basándose en lo propuesto por Haapalainen et al. [24] y Lira [46], se escogen las siguientes características para esta señal: La mediana, la media y la varianza de la ECG MAD (ECG *median absolute deviation*, desviación absoluta media) que se calcula según la Ecuación 4.4 donde a cada muestra se le resta la mediana de la señal de la ventana analizada.

$$ECGMAD = |ECG_i - mediana(ECG)| \quad (4.4)$$

La desviación absoluta media es una medida de la dispersión estadística. Además, el MAD es una estadística robusta, siendo más resistente a los valores atípicos en un conjunto de datos que la desviación estándar. En la desviación estándar, las distancias de la media están al cuadrado, por lo que las desviaciones grandes se ponderan más fuertemente, y por lo tanto, los valores extremos pueden influir en ella. En el MAD, las desviaciones de un pequeño número de valores atípicos son irrelevantes.

4.3.4. Ritmo cardíaco (HR)

Las características del ritmo cardíaco, obtenido de la señal PPG, se escogen basándose en las características en el dominio del tiempo usadas en el trabajo de Betella (2014) [14]. Estas son: la media, la desviación estándar y la media cuadrática (RMS) del HR.

4.3.5. Temperatura (T)

Basándose en lo propuesto por Haapalainen et al. [24] y Lira [46], se escogen las siguientes características para esta señal: La mediana y la media de la temperatura.

4.3.6. Electroencefalograma (EEG)

Para el análisis de la señal del EEG existen principalmente dos enfoques: análisis por potencial de eventos relacionados (ERPs) y el análisis en base a tiempo-frecuencia de la señal. Se escoge este último porque está más relacionado con el proceso psico-fisiológico y estructural del cerebro, es particularmente usado para estudio de estados emocional-cognitivo y porque es más recomendable cuando se estudia un tiempo limitado o una cantidad de datos relativamente baja, como es el caso del estudio por ventanas de tiempo en esta memoria [40].

Dentro de las diversas formas de estudiar la señal EEG en tiempo-frecuencia existe: bandas de frecuencia con Transformada de Fourier (FFT), Morlet Wavelets y Transformada de Hilbert. Las 3 arrojan resultados similares según explica Cohen [22].

Se explora la opción de la Transformada de Hilbert que tiene como ventaja mayor control sobre el filtrado en frecuencia. Previo a aplicar esta transformada se debe utilizar un filtro. Se escoge un filtro pasa bandas de entre 2 - 15 Hz para centrar el estudio en las bandas de frecuencia Theta (4 - 8 Hz), relacionada a la actividad mental y Alfa (8 - 12 Hz), relacionada al estado de relajación, pues se suprime ante carga cognitiva [40]. Luego de aplicarla, se obtiene una señal compleja llamada “señal analítica”, de la cual se extraen dos características: la potencia y la fase. Esto se realiza para cada uno de los 14 canales de la señal EEG.

4.4. Clusterización y número óptimo de clusters

Existe la intuición de que existen dos niveles de carga mental: uno alto y uno bajo. Con el objetivo de comprobar esta intuición, se realiza la clusterización para buscar cuántos niveles de carga cognitiva presenta el usuario a partir de la medición del diámetro pupilar y también para etiquetar esos niveles.

4.4.1. Clusterización por participante

Se realiza la clusterización por participante para el análisis intra-sujetos considerando todos los escenarios. Se utilizan las características media y varianza del diámetro de la pupila por ventana de tiempo (ver Sección 4.3.1) pues son ampliamente usadas en la literatura obteniendo buenos resultados [27] [24].

Existe evidencia de casos de clusterización en el desarrollo de tareas web como en el trabajo de Loyola (2015)[53] donde se utiliza el método *k-means* descrito en la Sección 2.5.1. Como una sobre-estimación o sub-estimación del número de clusters (K) afecta la calidad del agrupamiento, se busca el valor óptimo de clusters. Dado que se debe analizar una gran cantidad de usuarios se escoge una procedimiento más metódico que la visualización de datos para obtener este valor.

Se prueban los valores de K desde dos en adelante para obtener dos curvas. Se escoge para ello el índice de Calinski & Harabasz (CH), un índice que ha dado buenos resultados y la medida interna de cohesión de suma de los cuadrados dentro del grupo (WSS) según la metodología descrita en la Sección 2.6.2. La regla de detención es el valor más cercano a la vecindad donde las curvas se intersectan. En la Figura 4.8b se muestra un ejemplo de esta metodología para el participante 59 donde la intersección se genera en $K=3$.

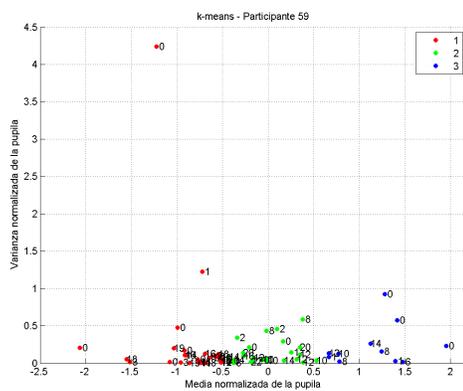
Se utiliza como criterio externo de validación de clusters el coeficiente de Jaccard obtenidos mediante el método de bootstrap, que evalúa cuán estable es el cluster, como se describe en la Sección 2.6.3. Los valores entre 0,6 y 0,75 indican que el grupo está midiendo un patrón en los datos, pero no hay certeza de cuáles puntos deben agruparse. Los grupos con valores de estabilidad por encima de aproximadamente 0,85 pueden considerarse altamente estables (clusters reales).

Utilizando esta metodología se obtienen buenos resultados en la clusterización con altos y aceptables índices de cohesión como se resume en la Tabla 4.2, donde se presenta el promedio, el máximo y el mínimo para todos los participantes de los coeficientes. Existen participantes que presentan clusters bien definidos con coeficientes de Jaccard muy cercanos a uno; y otros con valores muy al margen de ser aceptados. En promedio, los coeficientes están por sobre 0,75 por lo que se acepta como válida la clusterización.

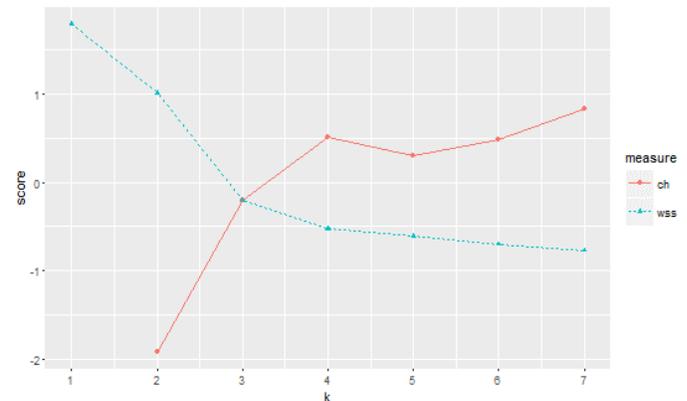
Tabla 4.2: Validación clusters usando Media y Varianza. Índices de Jaccard por clusters obtenidos con k-means utilizando el método de bootstrap.

	Cantidad clusters	coef. Jaccard					
Promedio	4,3	0,7773	0,7593	0,7541	0,7685	0,7350	0,7377
Máximo	6	0,9697	0,9763	0,9740	0,9612	0,8920	0,9249
Mínimo	3	0,6010	0,5753	0,5923	0,6169	0,6009	0,6297

Se concluye, fuera de la intuición inicial, que la cantidad de clusters va entre 3 y 6 niveles de carga cognitiva, y que en promedio existen alrededor de 4 niveles de carga cognitiva validado con aceptables índices de cohesión. En la Figura 4.8 se muestra un ejemplo de la clusterización para el participante 59 y el gráfico del método para obtener el número óptimo de clusters. En la Figura 4.8a se ve un alto grado de cohesión en los clusters cuyos índices de Jaccard son respectivamente: 0,6288 (cluster 1), 0,9024 (cluster 2) y 0,8517 (cluster 3).



(a) Agrupamiento de ventanas según su carga cognitiva para el participante 59.



(b) Número óptimo de clusters según el método del intersección de curvas CH y WSS.

Figura 4.8: Ejemplo de clusterización.

4.5. Selección de características

Paralelo al proceso de clusterización, que se realiza con el fin de obtener las etiquetas de las ventanas de tiempo (a partir de las características obtenidas de la dilatación pupilar) asignando un nivel de carga cognitiva a cada una de ellas para posteriormente entrenar los clasificadores con las características del resto de las señales, se trabaja con los métodos de selección de características que se explica en esta sección.

4.5.1. Umbral en Matriz de correlación

Para realizar la selección de características de entre las 42 obtenidas de las señales: GSR, EEG, ECG, PPG(HR) y Temperatura; se utiliza la matriz de correlación (descrita en la Sección 2.4.1) para eliminar los datos redundantes. Para ello se sugiere que el umbral para seleccionar características sea de 0,6 para comenzar.

Con el objetivo de mejorar la clasificación se itera con valores de umbral entre 0 y 0,6 en pasos de 0,01 hasta encontrar el mejor resultado al clasificar (para ver el valor del umbral óptimo por participante consulte los Anexos B.1 y B.3). Como el valor de correlación perfecta es 1, se elimina una de las dos características cuyo valor de correlación esté por sobre el umbral óptimo.

A modo de ejemplo, se presenta el procedimiento para las 5 primeras características del participante 1 que son: 'data acumulada del GSR', 'promedio de GSR', 'potencia espectral del GSR', 'promedio de la componente fásica' y 'máximo de la componente fásica'. En la Ecuación 4.5 se muestra la matriz original donde, marcado en negro, aparece el valor de correlación por sobre el umbral óptimo de 0,4415 para este participante según el Anexo B.1.

$$\begin{bmatrix}
 & 'accGSR' & 'promGSR' & 'potEspGSR' & 'promFasica' & 'maxFasica' \\
 'accGSR' & 1 & 0,1925 & 0,9992 & 0,0916 & 0,3878 \\
 'promGSR' & 0,1925 & 1 & 0,1934 & 0,0732 & -0,084 \\
 'potEspGSR' & 0,9992 & 0,1934 & 1 & 0,0936 & 0,388 \\
 'promFasica' & 0,0916 & 0,0732 & 0,0936 & 1 & \mathbf{0,5157} \\
 'maxFasica' & 0,3878 & -0,084 & 0,388 & 0,5157 & 1
 \end{bmatrix} \quad (4.5)$$

Para este ejemplo las características cuya correlación está sobre el umbral son: 'promedio de la componente fásica' y 'máximo de la componente fásica'. Por lo tanto, se debe eliminar una de ellas.

Luego de aplicar el método de selección, las características serían: 'data acumulada del GSR', 'promedio de GSR', 'potencia espectral del GSR' y 'promedio de la componente fásica'.

4.5.2. Eliminación Recursiva de Características (RF-RFE)

Existe evidencia en la literatura del uso del método RF-RFE (*Random Forest and Recursive Feature Elimination*) para eliminación de características con buenos resultados aplicado a la clasificación de fatiga mental con señales de EEG (Kai-Quan Shen et al. (2007) [45]). Para aplicar RF-RFE a este trabajo se sigue la metodología descrita en la Sección 2.4.2.

De las 42 características extraídas desde los sensores psico-fisiológicos se deben eliminar aquellas que están correlacionadas y que por ende no aportan información relevante al modelo y aumentan el error en la comparación por parejas.

Se obtiene el subconjunto de características mediante algoritmos de filtrado, esto permite trabajar con el mismo subconjunto de datos en todos los modelos a aplicar.

La metodología es la siguiente:

- Se realiza un análisis de correlación entre todas las características para cada participante.
- De los pares de características, si la correlación es mayor a 0,6 se elimina una de ellas.
- Al subconjunto preliminar se le aplica el método de selección de características RF-RFE.
- Se usa un algoritmo *Random Forest* (RF) para evaluar el modelo. Se configura de manera de evaluar todas las posibles combinaciones de características.
- Para cada participante se obtiene el subconjunto final de características que se presentan en la Tabla 4.3.

Tabla 4.3: Lista de características obtenidas con selección RFE para cada participante.

Part.	Características seleccionadas
1	maxPhasic, meanECGMAD, meanHR, powerEEG channel1, phaseEEG channel1, powerEEG channel3, phaseEEG channel12
2	medianTemp, avgGSR, powerEEG channel7, phaseEEG channel7, powerEEG channel1, powerEEG channel11
3	meanHR, meanPhasic, phaseEEG channel8, meanECGMAD, maxPhasic, rmsHR, accGSR, powerEEG channel5, phaseEEG channel4, phaseEEG channel6, phaseEEG channel12, avgGSR, stdHR, phaseEEG channel2, phaseEEG channel7, meanTemp, phaseEEG channel5, phaseEEG channel13, powerEEG channel13, phaseEEG channel1, medianECGMAD, powerEEG channel2, phaseEEG channel14, varECGMAD
4	meanHR, avgGSR, powerEEG channel1, phaseEEG channel9, powerEEG channel13, phaseEEG channel14, medianTemp, varECGMAD, phaseEEG channel2, rmsHR, meanPhasic, phaseEEG channel8, medianECGMAD, phaseEEG channel1, phaseEEG channel12, numpksPhasic, phaseEEG channel11, phaseEEG channel13, maxPhasic, powerEEG channel8
5	avgGSR, meanTemp
6	medianTemp, powerEEG channel6, powerEEG channel5, meanPhasic

Part.	Características seleccionadas
7	meanHR, powerEEG channel11, phaseEEG channel7, powerEEG channel1, meanTemp, avgGSR, powerEEG channel12, maxPhasic, meanPhasic, stdHR, varECGMAD, meanECGMAD, medianECGMAD, powerEEG channel4
8	avgGSR, powerEEG channel10, sumPowSpGSR, powerEEG channel11, phaseEEG channel4, varECGMAD, medianTemp, phaseEEG channel6, rmsHR, phaseEEG channel3, stdHR, phaseEEG channel8, phaseEEG channel1, phaseEEG channel2, powerEEG channel8, maxPhasic, phaseEEG channel13, phaseEEG channel14, phaseEEG channel7, meanECGMAD, phaseEEG channel11, meanHR, phaseEEG channel10, powerEEG channel5, powerEEG channel4
9	stdHR, maxPhasic, powerEEG channel9, numpksPhasic, phaseEEG channel13, phaseEEG channel3, phaseEEG channel9, powerEEG channel4, powerEEG channel14, avgGSR, phaseEEG channel11, phaseEEG channel8, rmsHR
10	avgGSR, powerEEG channel10, rmsHR, maxPhasic, powerEEG channel11, medianTemp, phaseEEG channel8, powerEEG channel13, meanHR, powerEEG channel3, powerEEG channel12, phaseEEG channel12, powerEEG channel9, powerEEG channel1, powerEEG channel5, meanECGMAD, phaseEEG channel5, powerEEG channel6, phaseEEG channel10, phaseEEG channel9, phaseEEG channel13, phaseEEG channel7, medianECGMAD, powerEEG channel7, powerEEG channel2, phaseEEG channel2, phaseEEG channel3, phaseEEG channel14, phaseEEG channel4, stdHR, numpksPhasic
12	rmsHR
13	avgGSR, accGSR, powerEEG channel5, powerEEG channel7, powerEEG channel9, medianTemp, phaseEEG channel3, phaseEEG channel10, phaseEEG channel7, phaseEEG channel8, phaseEEG channel13, phaseEEG channel1, phaseEEG channel5, phaseEEG channel11, phaseEEG channel6, phaseEEG channel14, meanECGMAD, powerEEG channel2, powerEEG channel14
14	phaseEEG channel9, powerEEG channel4, powerEEG channel2, numpksPhasic, powerEEG channel12, phaseEEG channel1, powerEEG channel3, phaseEEG channel6, powerEEG channel14
15	meanTemp, meanHR, phaseEEG channel7, rmsHR, phaseEEG channel5, powerEEG channel5, accGSR, phaseEEG channel3, stdHR, meanPhasic, phaseEEG channel9, powerEEG channel14, maxPhasic, medianECGMAD, phaseEEG channel8, powerEEG channel11
17	medianTemp, phaseEEG channel13, phaseEEG channel12, phaseEEG channel7
18	phaseEEG channel13, phaseEEG channel7, powerEEG channel11
19	powerEEG channel9, meanPhasic
20	meanTemp, phaseEEG channel11, rmsHR, powerEEG channel8, maxPhasic
21	powerEEG channel5, phaseEEG channel14, powerEEG channel7, powerEEG channel6, phaseEEG channel12, meanHR, powerEEG channel8, meanTemp, maxPhasic, powerEEG channel11, numpksPhasic, phaseEEG channel4, avgGSR, stdHR, medianECGMAD, phaseEEG channel8, phaseEEG channel9, powerEEG channel1, phaseEEG channel7, phaseEEG channel3, phaseEEG channel2, rmsHR, powerEEG channel4, phaseEEG channel13, meanPhasic, varECGMAD, powerEEG channel14, meanECGMAD, phaseEEG channel6, phaseEEG channel5, powerEEG channel9

Part.	Características seleccionadas
22	stdHR, phaseEEG channel11, powerEEG channel7, numpksPhasic, powerEEG channel9, powerEEG channel3, rmsHR, phaseEEG channel14, powerEEG channel6, powerEEG channel5, avgGSR, powerEEG channel12, phaseEEG channel1, powerEEG channel8, phaseEEG channel9, powerEEG channel4
23	meanHR, meanPhasic, powerEEG channel3, numpksPhasic, stdHR, meanECGMAD, powerEEG channel5, avgGSR, meanTemp, phaseEEG channel12, phaseEEG channel2, phaseEEG channel10, powerEEG channel7, maxPhasic, phaseEEG channel7, phaseEEG channel4, powerEEG channel8, phaseEEG channel6
24	powerEEG channel8, avgGSR, medianTemp, maxPhasic, phaseEEG channel12, phaseEEG channel8, powerEEG channel11, phaseEEG channel10, phaseEEG channel7, meanPhasic
25	phaseEEG channel1, powerEEG channel1, meanECGMAD, powerEEG channel6, powerEEG channel8, avgGSR, phaseEEG channel9, meanTemp, phaseEEG channel3, stdHR, phaseEEG channel8, powerEEG channel7, phaseEEG channel2, phaseEEG channel6, meanHR, powerEEG channel9, meanPhasic, phaseEEG channel5, maxPhasic, phaseEEG channel11, powerEEG channel2, rmsHR, phaseEEG channel12, phaseEEG channel10, medianECGMAD, accGSR, powerEEG channel5, varECGMAD, phaseEEG channel7, phaseEEG channel4, powerEEG channel13, phaseEEG channel14
26	phaseEEG channel4, powerEEG channel8, phaseEEG channel5, phaseEEG channel1, phaseEEG channel3, stdHR, meanPhasic, meanTemp, maxPhasic, medianECGMAD, phaseEEG channel11, rmsHR, meanHR, numpksPhasic, powerEEG channel3, meanECGMAD, phaseEEG channel13, phaseEEG channel12, powerEEG channel14, phaseEEG channel14, phaseEEG channel9, powerEEG channel11, sumPowSpGSR, avgGSR, phaseEEG channel2, phaseEEG channel10, powerEEG channel7
27	maxPhasic, meanTemp, phaseEEG channel14, powerEEG channel5, phaseEEG channel10
28	phaseEEG channel1, phaseEEG channel8, phaseEEG channel5, powerEEG channel11, powerEEG channel8, medianTemp, maxPhasic, phaseEEG channel4, phaseEEG channel12, phaseEEG channel11, stdHR, varECGMAD, phaseEEG channel9, powerEEG channel5, phaseEEG channel7, avgGSR, powerEEG channel13, phaseEEG channel6, phaseEEG channel3, meanHR, numpksPhasic, phaseEEG channel13, phaseEEG channel14, phaseEEG channel10, meanPhasic, medianECGMAD, rmsHR, phaseEEG channel2
29	sumPowSpGSR, avgGSR, powerEEG channel3, powerEEG channel11, phaseEEG channel1, phaseEEG channel14, powerEEG channel12, powerEEG channel9, phaseEEG channel10, phaseEEG channel5, phaseEEG channel7, meanHR, phaseEEG channel3, powerEEG channel6, powerEEG channel10, powerEEG channel8, phaseEEG channel6, phaseEEG channel11, meanPhasic, phaseEEG channel4, maxPhasic, powerEEG channel2
30	medianTemp, rmsHR, phaseEEG channel5, maxPhasic, phaseEEG channel14, powerEEG channel11

Part.	Características seleccionadas
31	meanTemp, varECGMAD, powerEEG channel12, powerEEG channel14, phaseEEG channel4, powerEEG channel6, phaseEEG channel6, meanECGMAD, phaseEEG channel5, medianECGMAD, numpksPhasic
32	phaseEEG channel1, meanHR, maxPhasic, accGSR, powerEEG channel3, powerEEG channel2, stdHR, phaseEEG channel13, numpksPhasic, powerEEG channel5
33	phaseEEG channel7, meanHR, rmsHR, phaseEEG channel9, powerEEG channel2, powerEEG channel1, meanECGMAD, stdHR, powerEEG channel9, phaseEEG channel11, avgGSR, varECGMAD, numpksPhasic, powerEEG channel11, meanTemp, powerEEG channel13, medianECGMAD, phaseEEG channel3, powerEEG channel10, powerEEG channel8, powerEEG channel3, phaseEEG channel10, phaseEEG channel1, phaseEEG channel13, phaseEEG channel5
34	powerEEG channel10, meanTemp, powerEEG channel7, meanHR, powerEEG channel8
35	powerEEG channel5, avgGSR, sumPowSpGSR, meanHR, powerEEG channel3, maxPhasic, powerEEG channel13, phaseEEG channel7, powerEEG channel1, rmsHR, meanECGMAD, phaseEEG channel14, powerEEG channel4, phaseEEG channel4, medianTemp, medianECGMAD, powerEEG channel7, powerEEG channel11, phaseEEG channel5, phaseEEG channel8, powerEEG channel6, phaseEEG channel2, phaseEEG channel11, phaseEEG channel1, powerEEG channel8, varECGMAD
37	phaseEEG channel5, phaseEEG channel11, medianTemp, phaseEEG channel12, phaseEEG channel4, avgGSR, powerEEG channel5
39	phaseEEG channel11, powerEEG channel8, phaseEEG channel14, meanHR, phaseEEG channel6, stdHR, avgGSR, phaseEEG channel5, powerEEG channel7, phaseEEG channel3, phaseEEG channel1, phaseEEG channel9, powerEEG channel11
40	meanHR, stdHR, medianTemp, avgGSR, powerEEG channel3, phaseEEG channel2, phaseEEG channel4
42	maxPhasic, phaseEEG channel13, phaseEEG channel3, avgGSR, powerEEG channel9, powerEEG channel5, powerEEG channel12, medianECGMAD, powerEEG channel10, phaseEEG channel5, meanHR, numpksPhasic, powerEEG channel8, rmsHR, phaseEEG channel2, varECGMAD, phaseEEG channel14, phaseEEG channel7, powerEEG channel6, phaseEEG channel10, powerEEG channel11, medianTemp, meanECGMAD, phaseEEG channel12, phaseEEG channel1, phaseEEG channel8, phaseEEG channel11, phaseEEG channel6, powerEEG channel4, phaseEEG channel4, phaseEEG channel9, powerEEG channel14
43	meanHR, powerEEG channel2, medianTemp, phaseEEG channel10, phaseEEG channel14, phaseEEG channel2, sumPowSpGSR, phaseEEG channel4, stdHR, varECGMAD, avgGSR, meanECGMAD, phaseEEG channel12, phaseEEG channel5, maxPhasic, phaseEEG channel11, phaseEEG channel8, phaseEEG channel3
44	meanTemp, phaseEEG channel6, phaseEEG channel11, varECGMAD, maxPhasic, meanECGMAD, phaseEEG channel3, powerEEG channel2, phaseEEG channel7, powerEEG channel13, avgGSR, meanHR, phaseEEG channel14, stdHR, phaseEEG channel12, meanPhasic

Part.	Características seleccionadas
45	medianTemp, powerEEG channel6, meanHR, phaseEEG channel2, phaseEEG channel8, stdHR, powerEEG channel9, phaseEEG channel4, medianECGMAD, phaseEEG channel6
46	phaseEEG channel12, meanHR, powerEEG channel13, avgGSR, phaseEEG channel2, phaseEEG channel14, powerEEG channel4, powerEEG channel8, phaseEEG channel6, rmsHR, meanTemp, maxPhasic, phaseEEG channel11, powerEEG channel14, stdHR, powerEEG channel3, meanPhasic, phaseEEG channel3, varECGMAD, medianECGMAD, powerEEG channel5, phaseEEG channel7, powerEEG channel11, phaseEEG channel1, phaseEEG channel8, phaseEEG channel5, powerEEG channel1, phaseEEG channel10, numpksPhasic, phaseEEG channel13, phaseEEG channel4
48	phaseEEG channel4, phaseEEG channel7, powerEEG channel8, powerEEG channel7, phaseEEG channel10, meanHR, rmsHR, powerEEG channel11, phaseEEG channel5, meanTemp, avgGSR, phaseEEG channel3, phaseEEG channel14, phaseEEG channel1, powerEEG channel1, phaseEEG channel2, varECGMAD
50	meanHR, stdHR, powerEEG channel5, powerEEG channel9, powerEEG channel1, powerEEG channel7, phaseEEG channel4, powerEEG channel13, phaseEEG channel8, powerEEG channel12, phaseEEG channel1, powerEEG channel3, sumPowSpGSR, meanECGMAD
51	powerEEG channel14, phaseEEG channel4, varECGMAD, avgGSR, powerEEG channel6, meanTemp, meanPhasic, phaseEEG channel10, phaseEEG channel3, phaseEEG channel1
52	powerEEG channel13, powerEEG channel14, phaseEEG channel7, phaseEEG channel2, powerEEG channel8, phaseEEG channel10, powerEEG channel12, phaseEEG channel13, meanTemp, accGSR, phaseEEG channel5, phaseEEG channel9, phaseEEG channel4, phaseEEG channel8, phaseEEG channel1, phaseEEG channel14
53	powerEEG channel1, phaseEEG channel4, avgGSR, phaseEEG channel7, phaseEEG channel11, phaseEEG channel6, phaseEEG channel5, meanPhasic, phaseEEG channel14, phaseEEG channel12, powerEEG channel10, phaseEEG channel2, stdHR, medianECGMAD, phaseEEG channel13, meanECGMAD, phaseEEG channel8, rmsHR, meanTemp, phaseEEG channel10, meanHR, phaseEEG channel1
55	phaseEEG channel11, rmsHR, meanHR, medianECGMAD
56	powerEEG channel7, rmsHR, sumPowSpGSR, phaseEEG channel5, maxPhasic
57	medianTemp, powerEEG channel6
58	meanTemp, powerEEG channel5, powerEEG channel13, phaseEEG channel6, powerEEG channel10, powerEEG channel8, powerEEG channel3, avgGSR, sumPowSpGSR, meanHR, varECGMAD, powerEEG channel1, powerEEG channel4, phaseEEG channel9, powerEEG channel14, powerEEG channel6, powerEEG channel11, rmsHR, powerEEG channel9, phaseEEG channel2, phaseEEG channel3, medianECGMAD, stdHR, phaseEEG channel7, phaseEEG channel4, phaseEEG channel10, phaseEEG channel11
59	numpksPhasic

Part.	Características seleccionadas
60	meanHR, rmsHR, powerEEG channel6, powerEEG channel4, meanECGMAD, powerEEG channel11, sumPowSpGSR, phaseEEG channel11, varECGMAD, powerEEG channel14, powerEEG channel12, phaseEEG channel2, avgGSR, phaseEEG channel12, powerEEG channel7, phaseEEG channel6, powerEEG channel9, phaseEEG channel13, powerEEG channel10, powerEEG channel13, meanPhasic, maxPhasic, meanTemp, phaseEEG channel10, phaseEEG channel9, phaseEEG channel14, phaseEEG channel8, phaseEEG channel1
61	powerEEG channel14, maxPhasic, powerEEG channel6, powerEEG channel3, meanPhasic, meanTemp

4.6. Clasificación por modelos de aprendizaje

En esta sección se presenta el procedimiento de cada algoritmo de aprendizaje supervisado utilizado en combinación con los métodos de selección de características de la Sección 4.5. Los modelos probados para evaluar su desempeño son:

- Clasificador Naive Bayes y método de la matriz de correlación.
- Clasificador Naive Bayes y método RFE.
- Clasificador de redes neuronales artificiales y método de la matriz de correlación.
- Clasificador redes neuronales artificiales y método RFE.
- Clasificador SVM y método RFE (para 2 clases).
- Deep learning.

4.6.1. Preparación de conjuntos de entrenamiento y prueba

Para realizar la clasificación se genera previamente un conjunto de entrenamiento con el 70% de las observaciones y un conjunto de prueba con el 30% restante. Para evitar sesgos, se distribuyen uniformemente las clases dentro de cada conjunto, y además, se seleccionan de manera aleatoria manteniendo la proporción.

Debido a que la clusterización genera algunas clases con muy pocos elementos, a veces una o dos ventanas que no tiene sentido clasificar pues quedarían sólo en el conjunto de entrenamiento y no en el de prueba, se eliminan las clases que tengan 5 ventanas (elementos) o menos. Esto se realiza con el objetivo de mejorar la clasificación.

Para evitar sesgos, cada resultado de clasificación se obtiene del promedio de correr 100 veces cada algoritmo.

Todo lo señalado se aplica a cada uno de los algoritmos de aprendizajes utilizados que se describen en esta sección.

4.6.2. Naive Bayes y Matriz de correlación

Para la clasificación multiclase se prueba el algoritmo Naive Bayes (descrito en la Sección 2.7.1) con las características seleccionadas según el método descrito en la sección 4.5.1 basado en la matriz de correlación. Los resultados de la clasificación para cada uno de los participantes se presentan en el Anexo B.1.

En la Tabla 4.4 se observa que en el peor caso la medida de exactitud (accuracy) de la clasificación es de 35% (participante 7) que tiene 6 clases; y el mejor caso es de 90,67% (participante 4) que, finalmente, tras la eliminación de los clusters pequeños, queda con 2 clases.

El promedio de clasificación para los 53 participantes es de 51,48% con una desviación estándar de 11,62%. No es un valor razonablemente aceptable pues es como lanzar una moneda, y además, denota una gran variabilidad en el accuracy de la clasificación para los participantes.

Tabla 4.4: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas usando la matriz de correlación como método de selección.

-	Accuracy (%)	Recall (%)	Precision (%)
Promedio	51,67	44,29	43,55
Desviación estándar	11,58	12,95	16,23
Máximo	90,67	91	91,73
Mínimo	35	27,73	19,1

Dado el amplio rango de los resultados de la clasificación por participante, se estudia el promedio de clasificación según la cantidad de clases finales. Como se observa en la Tabla 4.5, existe una clara tendencia para estos resultados de que a mayor cantidad de clases, menor es el desempeño de la clasificación.

Tabla 4.5: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas usando la matriz de correlación como método de selección. Promedio por cantidad de clases.

N clases	Accuracy (%)	Recall (%)	Precision (%)
2	75,35	74,17	79,67
3	55	47,23	47,44
4	45,98	36,56	33,58
5	41,72	34,83	32,42
6	39,96	31,9	25,63

4.6.3. Naive Bayes y RFE

Para comparar la clasificación multiclase con Naive Bayes (descrito en la Sección 2.7.1) respecto a las características seleccionadas se usa el método descrito en la sección 4.5.2 basado en la eliminación recursiva de características (RFE). Los resultados para todos los participantes se presentan en el Anexo B.2.

En la Tabla 4.6 se observa que en el peor caso la medida de exactitud (accuracy) es 39,29 % (participante 21) que tiene 6 clases; y el mejor caso es de 96 % (participante 13) que, finalmente, tras la eliminación de los clusters pequeños, queda con 2 clases. Tanto para el peor como para el mejor caso la clasificación mejoró con el método RFE respecto al método anterior de la matriz de correlación.

El promedio de clasificación también mejora para los 53 participantes llegando a un 64,94 % con una desviación estándar de 13,53 %. Es un valor razonablemente aceptable pero que denota una gran variabilidad en el accuracy de la clasificación para los participantes.

Tabla 4.6: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas usando RFE como método de selección.

-	Accuracy (%)	Recall (%)	Precision (%)
Promedio	64,94	60,76	68,91
Desviación estándar	13,53	15,54	16,38
Máximo	96	95	96,88
Mínimo	39,29	34,44	36,79

Nuevamente se estudia el promedio de clasificación según la cantidad de clases finales. Como se observa en la Tabla 4.7, se mantiene la tendencia que a mayor cantidad de clases, menor es el desempeño de la clasificación. Además, la selección con RFE mejora en mayor grado cuando existe un número de clases menor.

Tabla 4.7: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas usando RFE como método de selección. Promedio por cantidad de clases.

N clases	Accuracy (%)	Recall (%)	Precision (%)
2	88,16	87,55	91,61
3	70,57	67,31	76,17
4	55,87	50,98	61,36
5	55,01	48,84	55,07
6	46,15	34,44	38,89

4.6.4. Redes Neuronales Artificiales y Matriz de correlación

Se prueba otra forma de clasificación, las redes neuronales artificiales: Se implementa con una capa oculta con todas las neuronas artificiales totalmente interconectadas y se entrena

con el algoritmo *backpropagation*. Para calcular la cantidad de neuronas en la capa oculta se utiliza el método heurístico de la regla de la pirámide geométrica descrito en la sección 2.7.3.

El algoritmo se lleva a cabo con la ayuda del *Neural Network Toolbox* de Matlab con la aplicación *Neural Net Pattern Recognition* (*nprtool*) de dicho toolbox.

Este clasificador se combina con el método de la matriz de correlación para la selección de características, descrito en la sección 4.5.1. Los resultados para todos los participantes se presentan en el Anexo B.3.

En la Tabla 4.8 se observa que en el peor caso la medida de exactitud es de 37,17% (participante 48) que clasifica 5 clases; y el mejor caso es de 88% (participante 4) que clasifica con 2 clases.

El promedio de la clasificación, considerando los 53 participantes, es de 55,65% con una desviación estándar de 11,29% para la medida accuracy. El resultado mejora respecto a la clasificación con el algoritmo Naive Bayes y matriz de correlación. Aún así, el valor promedio de clasificación no es suficientemente bueno.

Tabla 4.8: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas usando la matriz de correlación como método de selección.

-	Accuracy (%)	Recall (%)	Precision (%)
Promedio	55,65	49,8	49,37
Desviación estándar	11,29	12,79	14,81
Máximo	88	86,88	90,2
Mínimo	37,17	30,65	25,83

Se analizan los resultados de la clasificación según la cantidad de clases finales. Como se presenta en la Tabla 4.9 se mantiene la tendencia observada de que a mayor número de clases el porcentaje de clasificación baja, siendo el mejor resultado un 79,08% para 2 clases, lo que se considera un buen valor, y 60,07% para 3 clases, valor razonable de clasificación.

Tabla 4.9: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas usando la matriz de correlación como método de selección. Promedio por cantidad de clases.

N clases	Accuracy (%)	Recall (%)	Precision (%)
2	79,08	78,47	81,22
3	60,07	53,93	55,1
4	49,88	42,49	40,78
5	42,79	37,22	33,39
6	40,19	32,08	30

4.6.5. Redes Neuronales Artificiales y RFE

En este caso al igual que el anterior, el clasificador con redes neuronales artificiales se implementa con una capa oculta con todas las neuronas artificiales totalmente interconectadas y se entrena con el algoritmo *backpropagation*. Para calcular la cantidad de neuronas en la capa oculta se utiliza el método heurístico de la regla de la pirámide geométrica descrito en la sección 2.7.3.

El algoritmo se lleva a cabo con la ayuda del *Neural Network Toolbox* de Matlab con la aplicación *Neural Net Pattern Recognition* (`nprtool`) de dicho toolbox.

Este clasificador se combina con el método RFE para la selección de características. Los resultados para todos los participantes se presentan en el Anexo B.4.

En la Tabla 4.10 se observa que la peor medida de exactitud obtenida es de 45,24 % (participante 48) que clasifica con 5 clases; y el mejor resultado es de 95,24 % (participante 23) que clasifica con 2 clases.

El promedio de clasificación, considerando los 53 participantes es de 68,94 % con una desviación estándar de 11,54 % para la medida de exactitud. El valor de clasificación alcanzado es razonablemente aceptable y mejor que el caso del clasificador Naive Bayes y RFE.

Tabla 4.10: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas usando RFE como método de selección.

-	Accuracy (%)	Recall (%)	Precision (%)
Promedio	68,94	66,62	76,92
Desviación estándar	11,54	13	11,62
Máximo	95,24	95,45	95,45
Mínimo	45,24	39,99	46,29

Se analizan también para este caso los resultados de la clasificación según la cantidad de clases finales. Como se presenta en la Tabla 4.11 se mantiene la tendencia observada de que a mayor número de clases el porcentaje de clasificación baja.

En este caso se obtienen muy buenos resultados para la clasificación de 2 clases (90,61 %) y aceptables resultados para 3, 4 y hasta 5 clases.

Tabla 4.11: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas usando RFE como método de selección. Promedio por cantidad de clases.

N clases	Accuracy (%)	Recall (%)	Precision (%)
2	90,61	89,99	92,15
3	73,34	71,69	83,56
4	63,29	59,57	70,75
5	57,94	54,92	64,35
6	53,85	53,61	66,31

4.6.6. SVM y RFE

Para clasificar entre dos clases se utiliza la máquina de vector de soporte (SVM) descrito en la Sección 2.7.2. Dados los resultados anteriores donde la selección con RFE presenta mejor desempeño que la matriz de correlación, se utiliza sólo este método para obtener las características para el clasificador.

Se prueba SVM con distintos kernel buscando el que tenga mejor desempeño. En los Anexos B.5 a B.10 se muestran los resultados para cada participante. Cabe señalar que solo se utilizaron los participantes que finalmente cuentan con 2 clases para clasificar.

En la Tabla 4.12 se compara al promedio de clasificación de los participantes con 2 clases para distintos kernels. Se observa que el mejor resultado se obtiene con un kernel polinomial de orden 3 cuya exactitud es de 89,58 % y el peor caso es con kernel gaussiano cuya exactitud llega al 73,58 %.

Tabla 4.12: Resultados de clasificación usando SVM y distintos kernels a partir de características de señales psico-fisiológicas usando RFE como método de selección. Promedio para todos los participantes que finalmente tienen 2 clases.

kernel	Accuracy (%)	Recall (%)	Precision (%)
lineal	88,14	88,24	91,13
pol. 2	86,79	87,69	89,73
pol. 3	89,58	89,4	90,59
pol. 4	88,04	87,95	89,61
pol. 5	87,95	87,86	89,73
gaussiano	73,58	68,22	74,88

Si comparamos con los otros clasificadores, en los casos que discriminan entre 2 clases, el mejor resultado de SVM (89,58 %) está en dentro de los mejores, aunque no supera la combinación Redes Neuronales Artificiales y RFE. Con el argumento que al aumentar las clases disminuye la exactitud de la clasificación, como es la tendencia en los resultados anteriores, se desestima intentar realizar un clasificador SVM multiclase con una cascada de SVMs para verificar si se obtienen mejores resultados para los participantes con más clases finales.

4.6.7. Deep Learning

Para aplicar la metodología *deep learning* se utiliza una red neuronal profunda (MLP), descrita en la Sección 2.8. Para la implementación se utiliza el paquete H2O para Deep Learning de R [10].

La red neuronal programada cuenta con 2 capas ocultas con 800 neuronas cada una según lo recomendado por Hinton (2012) [28]. Se utiliza la función de activación por defecto, la función lineal rectificadora descrita en la Sección 2.8.1. En este caso, no hay selección de características por lo que se aplican todas las características que aparecen en la Tabla 4.1.

Además, para evitar el sobreajuste del modelo se utiliza el método de regularización de la Ecuación 2.23 con los valores de $\lambda_1 = 10^{-5}$ y $\lambda_2 = 10^{-5}$ como recomienda el manual de H2O [10]. Para reducir el sobreajuste se incluye también un “dropout” (o “abandonar”) de 50% por cada capa, que previene las co-adaptaciones de las neuronas artificiales a los datos de entrenamiento. Esto significa que cada neurona de las capas ocultas se omite al azar de la red con una probabilidad de 0,5 [28].

En la Tabla 4.10 se observa los resultados. La peor medida de exactitud obtenida es de 72,16% (participante 17) que clasifica con 4 clases; y el mejor resultado es de 99,9% (participante 9) que clasifica con 3 clases. El promedio de clasificación, considerando los 53 participantes es de 88,46% con una desviación estándar de 7,94% para la medida de exactitud. El valor de clasificación alcanzado es muy bueno y el mejor entre todos los modelos probados. El detalle por participante se presenta en el Anexo B.11.

Tabla 4.13: Resultados de clasificación usando Deep Learning.

-	Accuracy (%)	Recall (%)	Precision (%)
Promedio	88,46	88,84	88,85
Desviación estándar	7,94	7,67	7,64
Máximo	99,9	100	100
Mínimo	72,16	71,45	74,35

Se realiza el análisis de la clasificación según la cantidad de clases finales. Como se presenta en la Tabla 4.14 se mantiene la tendencia observada de que a mayor número de clases el porcentaje de clasificación baja pero con un quiebre en la cantidad de 4 y 5 clases. Además, se observa que las variaciones entre clases es menor. Por último, en este caso se obtienen muy buenos resultados para la clasificación de todas clases.

Tabla 4.14: Resultados de clasificación usando Deep Learning. Promedio por cantidad de clases.

N clases	Accuracy (%)	Recall (%)	Precision (%)
2	90,37	88,11	90,79
3	89,81	90,2	89,48
4	86,13	87,16	87,86
5	88,42	89,46	88,75
6	78,53	76,32	78,85

Capítulo 5

Resultados

5.1. Análisis estadístico

Para verificar la hipótesis de que existe una baja en la carga cognitiva en las ventanas de transición se escoge como característica de estudio el promedio del diámetro pupilar dentro de cada ventana de un largo mínimo de 500 milisegundos. El objetivo es ver si la media del diámetro pupilar varía dependiendo de si está en una ventana activa (*web object*) o en una ventana de transición.

Para ello se realiza un análisis de varianza ANOVA (*Analysis of variance*, en inglés) de medidas repetidas (MR). Este tipo de análisis se realiza para estudiar el efecto de uno o más factores cuando al menos uno de ellos es un factor intra-sujetos, es decir, que todos los niveles del factor se aplican a todos los sujetos. En este experimento el grupo completo de participantes es sometido a los 3 niveles (escenarios) del factor a evaluar que es la página web. Por ello es apropiado realizar un análisis ANOVA-MR.

El diseño más simple de medidas repetidas consiste en medir dos variables en una misma muestra de sujetos. Para este análisis se utiliza la media del diámetro pupilar en una ventana activa y la media del diámetro pupilar en una ventana de transición, como ya se ha mencionado. La señal pupilar ha sido estandarizada y procesada para hacer la comparación como se explica en la sección 4.3.

Para el ANOVA-MR se considera el universo completo de ventanas de todos los participantes. La hipótesis nula es que la media del diámetro de la pupila en las ventanas activas no es significativamente diferente del diámetro en las ventanas de transición. En la Tabla 5.1 se presentan los resultados del ANOVA-MR dando un $P - value = 0,00184$ en un intervalo de confianza del 95 %, por lo que se rechaza la hipótesis nula.

Tabla 5.1: Resultados de ANOVA con medidas repetidas para la data completa.

Suma de cuadrados	Media cuadrática	F	Significancia (P-value)
1,8	1,7829	9,717	0,00184

Además, como se muestra en la Tabla 5.2, la media del diámetro pupilar es menor en las ventanas de transición que en las ventanas activas (*web object*). Por lo tanto, **se concluye para todos los datos que, la diferencia entre la media del diámetro pupilar en las ventanas activas y las ventanas de transición es estadísticamente significativa; siendo menor en las ventanas de transición.**

Tabla 5.2: Resultados de ANOVA con medidas repetidas. Media del diámetro pupilar para ventanas de transición y ventanas activas.

Factor	Media diámetro pupilar	Desviación estándar
Transición	0,01924	0,97
Activa	0,10737	0,84

Con estos resultados, más la relación entre el diámetro pupilar y la carga cognitiva [53], es posible concluir que para la data completa, **la carga mental en las ventanas de transición es significativamente menor a la carga mental en las ventanas activas, comprobándose la hipótesis de esta memoria.** Además, esto da sentido a los resultados de clasificación obtenidos que se discuten en la Sección 5.2.

5.2. Clasificación

En esta sección se presentan y se discuten los resultados de la clasificación para evaluar el desempeño de los distintos algoritmos de aprendizaje probados y luego para evaluar el aporte de cada sensor.

5.2.1. Por algoritmo de aprendizaje

Se probaron 6 distintos modelos de aprendizaje (descritos en la Sección 4.6), en la Tabla 5.3 se presentan como resultados el promedio de clasificación para los 53 participantes válidos para cada uno de los algoritmos usados.

Tabla 5.3: Resumen de los resultados de clasificación por algoritmo de aprendizaje.

Metodología	Accuracy (%)	Recall (%)	Precision (%)
Naive Bayes + Mat. Corr	51,67	44,29	43,55
Naive Bayes + RFE	64,94	60,76	68,91
NN+ Mat. Corr	55,65	49,8	49,37
NN+ RFE	69,24	66,99	77,15
SVM (pol.3) + RFE**	89,58	89,4	90,59
Deep Learning (MLP)	88,46	88,84	88,85

** Sólo 2 clases.

Dentro de los peores resultados se encuentra el algoritmo que utiliza el clasificador Naive Bayes (con un 51,67 % y un 64,94 %). Los resultados no tan buenos se deben a que este tipo de clasificador hace el supuesto de independencia de las características seleccionadas. Por lo tanto, no se puede concluir que es un mal clasificador, más bien, se concluye que las características no son necesariamente independientes entre sí. Cabe señalar también la importancia de verificar previamente la independencia de las variables si se desea probar este clasificador.

Respecto a los métodos de selección de características, el que da mejores resultados es el RFE. Para los datos y las características utilizadas RF-RFE es superior al método con la matriz de correlación, porque utiliza distintas combinaciones de características, a diferencia de la eliminación simple del método por correlación, y además, utiliza un sistema de puntuación basado en mejorar el desempeño de un clasificador, RF en este caso.

Los resultados de SVM para la clasificación de los casos de 2 clases arroja su mejor resultado utilizando un kernel polinomial de orden 3 con un 89,58 %. Si comparamos con los resultados de NN+RFE para 2 clases con un 90,61 % (Tabla 4.11) y de *deep learning* para 2 clases con un 90,37 % (Tabla 4.14) está dentro de los mejores resultados obtenidos. Se plantea la opción de considerar el clasificador SVM cuando desee, por ejemplo, separar el nivel más bajo de carga cognitiva del resto de los niveles, que puede ser una aplicación.

Los mejores resultados se obtienen con la red neuronal profunda (MLP) al aplicar *deep learning* llegando a un promedio de 88,46 % en la exactitud de la clasificación. Se consigue una clasificación muy buena utilizando esta metodología, además, existe un alto grado de confianza de que no existe sobreentrenamiento, pues se aplica el método de regulación. En particular, para los métodos de MLP con gran cantidad de neuronas en las capas ocultas, como este caso, es absolutamente necesario realizar la regulación.

5.2.2. Por sensores

Con el objetivo de conocer el aporte particular de cada sensor, se evalúa el desempeño de clasificación de las características extraídas de cada uno de ellos. Se escoge como modelo de aprendizaje supervisado la red neuronal profunda (MLP) que cuenta con el mayor desempeño de los métodos antes probados, con un 88,46 % como se observa en la Tabla 5.3. Se sigue la misma metodología descrita en la Sección 4.6.7.

En la Tabla 5.4 se presentan los resultados de evaluar el desempeño de cada sensor por separado. El sensor con mejor desempeño es el EEG con un 88,78 % de exactitud en la clasificación, levemente mejor a la clasificación considerando todos los sensores. Por otro lado, el resto de los sensores por separado tiene un nivel de clasificación muy pobre.

Luego, se prueban las combinaciones de los 3 sensores con mayor desempeño: EEG, GSR y PPG(HR). Como se observa en la Tabla 5.4 las combinaciones donde existe el EEG son las que arrojan los mejores resultados. La de mayor desempeño es la combinación de GSR, PPG(HR) y EEG con un 95,73 %.

Una diferencia importante entre el sensor EEG y los demás es que permite extraer mayor cantidad de características debido a que los 14 electrodos aportan 2 cada uno, dando un total de 28 características del EEG. Este factor puede ser condicionante para que el desempeño de este sensor sea mejor que el resto. Por lo tanto, se concluye que es posible obtener buenos resultados de clasificación para este diseño experimental con menos de los 5 sensores, incluso solamente con el EEG. También se pueden descartar el sensor de temperatura y el ECG.

Tabla 5.4: Resumen de los resultados de clasificación por sensor.

Sensor(es)	Accuracy (%)	Recall (%)	Precision (%)
GSR	51,41	50,48	51,73
ECG	45,25	45,25	45,09
PPG(HR)	45,28	45,01	45,54
T	41,54	41,65	41,43
EEG	88,78	88	88,49
PPG(HR)+EEG	91,51	90,30	91,06
GSR+EEG	93,38	92,91	92,85
GSR+PPG(HR)	61,19	61,86	62,23
GSR+PPG(HR)+EEG	95,73	94,25	95,60

Conclusión

El estudio del comportamiento humano y su fisiología cuando realiza actividades HCI tiene un carácter complejo debido a los múltiples factores que afectan a cada persona, su desempeño y su comportamiento frente a esta clase de tareas. En este trabajo se ha estudiado el comportamiento de un usuario frente a la tarea simple de navegar libremente por un página web ficticia, hecha específicamente para este estudio.

Se cumple el objetivo específico de realizar un diseño experimental que permite evaluar el comportamiento de la carga cognitiva mediante sensores psico-fisiológicos. En un ambiente controlado, se expone a los participantes a 3 páginas web con la misma configuración, pero con elementos distintos mientras porta una serie de sensores (Eye Tracker, GSR, PPG, ECG, EEG y temperatura). Este diseño permite evaluar la carga cognitiva, mediante los sensores psico-fisiológicos, en ventanas de tiempo de las cuales se definen dos: las ventanas activas (donde el usuario examina algún web object) y las ventanas de transición (cuando el usuario no examina ningún elemento de la página).

De los 61 usuarios medidos sólo se utilizaron para el análisis 53, quienes fueron válidamente medidos. Para estos participantes se obtienen los resultados que validan la hipótesis inicial de que existe una variación en la carga cognitiva en el paso entre una ventana de transición y ventana activa mientras navegan por la página web.

Se cumple, según el test ANOVA de medidas repetidas, que para la data completa, es decir, considerando el universo completo de ventanas de todos los participantes, el diámetro pupilar (como medida de la carga cognitiva) es significativamente menor en las ventanas de transición que en las ventanas activas con una significancia de $P - value = 0,00184$ para un intervalo de confianza de 95 % (ver Tablas 5.1 y 5.2). Por lo tanto, se cumple el objetivo general de identificar patrones de estados de baja carga mental y, además, se verifica la hipótesis de esta memoria de que efectivamente es posible medir la carga cognitiva en actividades de navegación web y que la carga mental del usuario baja en la transición del análisis de un elemento web a otro mientras navega libremente.

Se realiza un proceso de extracción de características fundado en la literatura, resumido en la Tabla 4.1. Para esto existen amplias opciones que se relacionan con la carga cognitiva. Por ejemplo, en el caso de la señal GSR se obtiene la componente fásica cuyos peaks explican los cambios abruptos de la señal GSR que se relaciona con esfuerzo mental. De la señal PPG es posible obtener el ritmo cardíaco (HR) cuyas variaciones se relacionan con la respuesta cognitiva en distintas tareas. La señal EEG se filtra para trabajar en las bandas de frecuencia Theta (que aumenta al subir la carga mental) y Alfa (que disminuye) que se relacionan al

comportamiento cognitivo. Se escogen características como media, mediana y varianza de ECG y Temperatura. Cada una de estas características se utilizan para evaluar la carga cognitiva.

Se aplica el modelo de análisis no supervisado de minería de datos *k-means* con el cual se encuentra que existen varios niveles de carga cognitiva en la tarea de navegar por Internet. En promedio se obtienen 4 niveles de carga según la clusterización con la señal pupilar. Con esto se concluye que, fuera de la intuición inicial de que solo existen 2 niveles de carga mental (uno alto y uno bajo), existen varios estados de carga mental que pueden obtenerse y clasificarse. La clusterización permite etiquetar el nivel de carga cognitiva de cada ventana de tiempo.

De los 2 métodos de selección de características probados, la selección en base a la matriz de correlación y eliminación recursiva de características (RF-RFE), se concluye que la mejor opción es RF-RFE, pues presenta un mayor desempeño. Sin embargo, tiene como contra que se ajusta el modelo al usuario, pues cada uno de los participantes queda finalmente con distintas características utilizando este método.

Se aplican varios modelos de aprendizaje supervisados cuyos resultados se resumen en la Tabla 5.3. El que presenta mayor desempeño es *deep learning* usando una red neuronal profunda (MLP), con un resultado muy bueno de 88,46 % de exactitud utilizando todas las características (es decir, no se aplica la selección). Además, se tiene el cuidado de aplicar métodos de regularización para evitar el sobreajuste del modelo. Por lo tanto, se prueba que es posible clasificar de buena manera los niveles de carga cognitiva con las señales psicofisiológicas.

Además, se evalúa el aporte de cada sensor por separado utilizando el modelo de *deep learning* antes mencionado, cuyos resultados se resumen en la Tabla 5.4. El sensor con mejor desempeño es el EEG con una exactitud de 88,78 %. De este sensor se extraen 28 características (2 por electodo). Esta cantidad puede ser el motivo por el cual este sensor aporta más a la clasificación que el resto. Cuando se combina EEG con los sensores que le siguen en el aporte que entregan a la clasificación, GSR (con un desempeño de 51,41 %) y PPG(HR) (con un desempeño de 45,28 %), ambos con desempeño bajo; se genera un gran aumento en la exactitud de clasificación, llegando a 95,73 %.

Como conclusión general, el método de redes neuronales profundas con *deep learning*, sin selección de características para no ajustar el modelo al usuario, utilizando como sensores: el *Eye tracker*, el EEG (el que de mayor desempeño), el GSR y el PPG(HR) (que son fáciles de utilizar y más baratos, que mejoran la clasificación al combinarse con EEG) se considera el mejor modelo desarrollado en esta investigación, modelo que se puede replicar.

A continuación se mencionan algunos puntos que pueden mejorar el diseño experimental y los resultados.

- **Diseño experimental y toma de datos:** El hecho de darle varios grados de libertad al usuario al navegar por el sitio y no tareas específicas impide medir la dificultad de la tarea y la carga cognitiva in situ, como por ejemplo con un Test Nasa TLX, que permita comparar con la medición psicofisiológica realizada. El objetivo de esta decisión es darle más realismo a la navegación por Internet. Se escoge como medida

maestra (*gold standard*) el diámetro pupilar medido con el *eye tracker* Tobii T120 lo que obliga a tomar medidas para evitar los efectos externos sobre esta señal. Junto con ésta, todas las demás señales se ven afectadas por múltiples factores. Uno de los principales es el movimiento del usuario y los sensores. Una forma de mejorar la captura de cada señal es evitando lo más posible el artefacto de movimiento que genera ruidos y outliers. Además, para mejorar el procesamiento y el análisis, se podría tomar una medida base de cada señal del sujeto en reposo, previo a la realización del experimento para tener un punto de comparación del estado fisiológico de la persona.

- **Clusterización:** En este trabajo se optó por usar el método *k-means* y como medida de validación el índice de Jaccard, obtenido por el algoritmo de bootstrap. Una de las limitaciones de *k-means* es que su modelo se basa en conjuntos esféricos que son separables y que los valores convergen al centro del grupo, y además, asume que son del mismo tamaño. Como opción existen otros tipo de agrupamiento como la clusterización jerárquica o usando redes SOM. Probar con otro tipo de método de agrupamiento podría dar mejores resultados, así como con otros tipos de características de la señal de la pupila.
- **Procesamiento de señal EEG:** En este trabajo se utilizó el enfoque basado en la Transformada de Hilbert, la razón, como se explicó en la sección 4.3.6 es que tiene mayor flexibilidad al utilizar el filtro y determinar las bandas de frecuencia con las que se desea trabajar. Sin embargo, el enfoque clásico del análisis de esta señal para medir la carga cognitiva es de bandas de frecuencia obtenidas con Transformada de Fourier (FFT) o Wavelets. A pesar del buen desempeño de la Transformada de Hilbert, aplicar estos enfoques podría mejorar el aporte de esta señal a la clasificación, o bien permitiría comparar su desempeño con el enfoque actual. Además, sería interesante comprobar qué electrodos son los que más aportan, pues se podría realizar un análisis más específico del cerebro y también disminuir el costo computacional al reducir las características al escoger ciertos electrodos, o bien, darle mayor ponderación a los que más aporten.

Como trabajo futuro se propone mejorar los resultados obtenidos en este estudio aplicando los puntos señalados y, además, aprovechar la gran cantidad de datos para estudiar el comportamiento emocional en junto con el cognitivo, del usuario web. Además, se propone centrar la investigación en el sensor EEG que mostró mayor desempeño; y en lo posible, con un sensor EEG con mayor grado de confiabilidad de medición que el utilizado en esta memoria.

Como aplicación, dado que se probó que la carga mental es menor en las ventanas de transición, se podría generar sistemas de recomendación durante la navegación web que apoye al usuario en la búsqueda de contenido afín a sus intereses en el momento apropiado, es decir, cuando tenga baja carga cognitiva, que por lo demostrado ocurre en los tiempos de transición, porque son los momentos donde es más probable capturar la atención del usuario y la interrupción tiene menor costo cognitivo.

Bibliografía

- [1] John Sweller. 1988. *Cognitive Load During Problem Solving: Effects on Learning*. Cognitive Science 12(2):257-285 (1988).
- [2] M. Weiser. 1991. *The computer for the 21st century*.
- [3] M. Weiser. 1993. *Ubiquitous Computing*.
- [4] Thomas K. Landauer. 1996. *Review of: The Trouble with Computers Usefulness, Usability, and Productivity*. Journal Computational Linguistics archive 22 (2):267-270. MIT Press Cambridge, MA, USA.
- [5] Wendy E. Mackay. 2000. *Responding to cognitive overload: Co-adaptation between users and technology*. Intellectica, 2000/1, 30, pp. 177-193.
- [6] C. Wickens. 2002. *Multiple resources and performance prediction*. Theoretical issues in ergonomics science, 3(2):159-177.
- [7] Gwizdka J. 2010. *Distribution of Cognitive Load in Web Search*. Journal of the American Society for Information Science and Technology (JASIST) 61(11):2167-2187.
- [8] A. Duque y C. Vázquez. 2013. . *Implicaciones clínicas del uso del tamaño pupilar como indicador de actividad psicológica: una breve revisión*. Clínica y Salud 24(2):95-101.
- [9] A. K. Jain. 2010. *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters 31 (2010):651-666.
- [10] Arno Candel, Jessica Lanford, Erin LeDell, Viraj Parmar & Anisha Arora. 2015. *Deep Learning with H2O*. Published by H2O.ai, Inc. Third Edition.
- [11] Arquer I. & Nogareda C. *Estimación de la carga mental de trabajo: el método NASA TLX. Nota Técnica de Prevención 544*. Barcelona: INSHT.
- [12] Barreto, A., Zhai J., Adjouadi, M. 2007. *Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction*. HCI 2007. vol. 4796, pp. 29-38. Springer Heidelberg.
- [13] Beatty, Jackson; Lucero-Wagoner. 2000. *The Pupillary System*. Handbook of psychophysiology, 2nd ed., (pp. 142-162). New York, NY, US: Cambridge University Press.

- [14] Betella et al. 2014. *Inference of human affective states from psychophysiological measurements extracted under ecologically valid conditions*. *Frontiers in Neuroscience* 8(286).
- [15] Bojko Aga. 2013. *Eye tracking the user experience. A Practical Guide to Research*. Rosenfeld Media.
- [16] Boser, B.E., Guyon, I., & Vapnik, V.N. 1992. *A training algorithm for optimal margin classifiers*. In *Proceedings of the 5th Annual Workshop of Computational Learning Theory*, 5, 144-152, Pittsburgh, ACM.
- [17] Breiman, L. 2001. *Random forests*. *Machine Learning*, (45):5–32, 2001.
- [18] Brian P. Bailey & Shamsi T. Iqbal. 2008. *Understanding Changes in Mental Workload During Execution of Goal-directed Tasks and Its Application for Interruption Management*. *Journal ACM Transactions on Computer-Human Interaction (TOCHI)* 14 (4) Article No. 21-.
- [19] Brouwer A-M, Zander TO, van Erp JBF, Korteling JE and Bronkhorst AW. 2015. *Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls*. *Frontiers in Neuroscience* (9) 136.
- [20] Cacioppo, Tassinary, Berntson. 2007. *The Handbook of Psychophysiology*. Cambridge University Press. Third Edition.
- [21] Calinski RB, Harabasz J. 1974. *A dendrite method for cluster analysis*. *Communications in Statistics* 3:1–27.
- [22] Mike X. Cohen. *Analyzing neutral time series data. Theory and Practice*. Massachusetts Institute of Technology., 2014.
- [23] Dujovne, L. E., Velásquez J. D. 2009. *Design and Implementation of a Methodology for Identifying Website Keyobjects*. *Knowledge-Based and Intelligent Information and Engineering Systems Volume 5711 of the series Lecture Notes in Computer Science* pp 301-308.
- [24] E. Haapalainen, S. Kim, J. F. Forlizzi, and A. K. Dey., 2010. *Psychophysiological measures for assessing cognitive load*. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pp. 301–310.
- [25] Chen et al. 2011. *A comparison of four methods for cognitive load measurement*. *Proc. of the 23rd Australian Computer-Human Interaction Conf.*, ACM, 2011, pp. 76-79.
- [26] Antonio Quintero Rincón et al. 2012. *Preprocesamiento de EEG con Filtros Hampel*. *IEEE Latin American Transactions* 2012(89).
- [27] Fritz, Begel, Müller, Yigit-Elliot, Züger. 2014. *Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development*. *International Conference on Software Engineering (ICSE)*, Hyderabad, pp 402-413.

- [28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov. 2012. *Improving neural networks by preventing co-adaptation of feature detectors*. Neural and Evolutionary Computing (cs.NE).
- [29] Gazzaniga Michael S. 2014. *Cognitive Neuroscience: The Biology of the mind*. University of California, fourth Edition.
- [30] George H. John, Pat Langley. 1995. *Estimating Continuous Distributions in Bayesian Classifiers*. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, San Mateo.
- [31] Gómez Milán E. et. Al. 2015. *Neurotermografía y termografía psicósomática*. Ediciones Fundación Internacional Artecittà. Granada. Abril 2015.
- [32] Goldberg, J. H., & Wichansky, A. M. 2003. *Eye tracking in usability evaluation: A practitioner's guide*. In Hyona, J., Radach, R., & Deubel, H. (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 493-516). Amsterdam:Elsevier.
- [33] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.
- [34] Guyon, I., Elisseeff, A. 2003. *An Introduction to Variable and Feature Selection*. Journal of Machine Learning Research 3 (2003):1157-1182.
- [35] Guyon, I., Elisseeff, A. 2006. *An Introduction to Feature Extraction*. Feature Extraction. Foundations and Applications 207:1-25.
- [36] Guyon, I. et al. 2002. *Gene selection for cancer classification using support vector machines*. Machine Learning, 46:389-422.
- [37] Gysels, E., Renevey, P., Celka, P. 2005. *SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces*. Signal Processing 85 (2005):2178-2189.
- [38] Hart, S.G. and Staveland. 1988. *Development of NASATLX (Task Load Index): results of empirical and theoretical research*. In Hancock, P.A. and Meshkati, N. (Eds.), *Human Mental Workload*. Amsterdam, North-Holland, 139-183.
- [39] Ikehara et al. 2005. *Assessing Cognitive Load with Physiological Sensors*. Proceedings of the 38th Annual HICSS 2005:295a.
- [40] iMotions Biometric Research Platform. 2016. *EEG Pocket Guide*.
- [41] iMotions Biometric Research Platform. 2016. *Eye Tracking Pocket Guide*.
- [42] iMotions Biometric Research Platform. 2016. *GSR Pocket Guide*.
- [43] J.D. Velásquez. 2013. *Combining eye-tracking technologies with web usage mining for identifying Website Keyobjects*. Engineering Applications of Artificial Intelligence 26(56):1469-1478.

- [44] Angel Jimenez-Molina and In-Young Ko. 2015. *Cognitive resource-aware unobtrusive service provisioning in ambient intelligence environments*. Journal of Ambient Intelligence and Smart Environments 7:37–57.
- [45] Kai-Quan Shen et al. 2007. *A Feature Selection Method for Multilevel Mental Fatigue EEG Classification*. IEEE Transactions on Biomedical Engineering 54(7):1231-1237.
- [46] Lira L., Hernán. 2015. *Medición de cargas cognitivas durante actividades de interacción humano computador en ambiente móvil usando sensores psico-fisiológicos*. Memoria de Ingeniero Civil Industrial. Santiago, Universidad de Chile.
- [47] Mao, Y. et al. 2006. *Accelerated Recursive Feature Elimination Based on Support Vector Machine for Key Variable Identification*. Chinese J. Chem. Eng., 14(1):65-72.
- [48] Navalpakkam V. & Churchill E. 2012. *Mouse Tracking: Measuring and Predicting Users' Experience of Web-based Content*. In Proceedings of the ACM SIGCHI conference on human factors in computing systems (pp. 2963-2972). ACM Press.
- [49] Nina Zumel and John Mount. 2014. *Practical Data Science with R*. Manning Publications Co. 1st Edition.
- [50] Noppadon J. et al. 2013. *Real-Time EEG-Based Happiness Detection System*. The Scientific World Journal 2013(4).
- [51] Nourbakhsh et al. 2012. *Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks*. Conference: Proceedings of the 24th Australian Computer-Human Interaction Conference.
- [52] Nourbakhsh, Wang, Chen. 2013. *GSR and Blink Features for cognitive load classification*. Human-Computer Interaction – INTERACT 2013 Volume 8117 of the series Lecture Notes in Computer Science pp 159-166.
- [53] P.Loyola et al. 2015. *Combining eye-tracking and pupillary dilation analysis to identify Website Key Objects*. Neurocomputing 168:179–189.
- [54] Poole, Alex and Linden J. Ball. 2005. *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects*. Encyclopedia of human computer interaction, Idea Group.
- [55] Rui Xu, Donald C. Wunsch II. 2009. *Clustering*. IEEE Press, 2009, John Wiley & Sons, Inc., Hoboken, New Jersey, USA. (pp. 263-278).
- [56] Shi et al. 2007. *Galvanic Skin Response (GSR) as an Index of Cognitive Load*. CHI Extended Abstracts on Human Factors in Computing Systems.
- [57] Molly Talbert. *Reducing Cognitive Load: The Best UX Design for Your Community Site [en línea]*. Higher Logic, Febrero, 2016. <<http://blog.higherlogic.com/reducing-cognitive-load-the-best-ux-design-for-your-community-site>>[consulta: 1 Noviembre, 2016].

- [58] Kathryn Whitenton. *Article: Minimize Cognitive Load to Maximize Usability [en línea]*. Nielsen Norman Group, Diciembre, 2013. <<https://www.nngroup.com/articles/minimize-cognitive-load/>>[consulta: 1 Noviembre, 2016].
- [59] William H.; Teukolsky, Saul A.; Vetterling, William T.; Flannery, B. P. 2007. *Section 16.5. Support Vector Machines*. Numerical Recipes: The Art of Scientific Computing (3rd ed.). New York: Cambridge University Press.
- [60] Xu, J. et al. 2011. *Pupillary Response Based Cognitive Workload Measurement under Luminance Changes*. Interact 2011, Part II. LNCS, vol. 6947:178-185. Springer, Heidelberg.
- [61] Ziheng Wang, Ryan M. Hope, Zuoguan Wang, Qiang Ji, Wayne D. Gray. 2012. *Cross-subject workload classification with a hierarchical Bayes model*. NeuroImage 59 (2012) 64–69.

Anexo A

Grupo experimental.

Tabla A.1: Detalle del grupo experimental.

N	Género	Edad	Ocupación
1	Hombre	27	Estudiante
2	Hombre	24	Estudiante
3	Mujer	19	Estudiante
4	Hombre	25	Estudiante
5	Mujer	25	Licenciada en literatura hispánica
6	Hombre	20	Estudiante
7	Mujer	28	Estudiante
8	Hombre	24	Estudiante
9	Mujer	28	Estudiante
10	Hombre	26	Estudiante
11	Hombre	23	Estudiante
12	Mujer	24	Estudiante
13	Mujer	23	Estudiante
14	Mujer	26	Ingeniero
15	Hombre	21	Estudiante
16	Hombre	24	Enfermero
17	Hombre	19	Estudiante
18	Hombre	25	Estudiante
19	Hombre	25	Estudiante
20	Hombre	23	Estudiante
21	Mujer	24	Ayudante de investigación
22	Hombre	25	Estudiante
23	Hombre	20	Estudiante
24	Hombre	26	Estudiante
25	Hombre	21	estudiante
26	Hombre	19	estudiante
27	Hombre	25	Estudiante
28	Hombre	25	Ingeniero

N	Género	Edad	Ocupación
29	Hombre	24	Estudiante
30	Hombre	20	Estudiante
31	Hombre	24	Estudiante
32	Mujer	21	Estudiante
33	Mujer	28	Estudiante
34	Mujer	25	Estudiante
35	Mujer	24	Estudiante
36	Hombre	20	Estudiante
37	Mujer	21	Estudiante
38	Hombre	24	Tesista
39	Hombre	23	Estudiante
40	Hombre	26	Estudiante
41	Hombre	24	Estudiante
42	Mujer	25	Estudiante
43	Mujer	23	Estudiante
44	Mujer	29	Estudiante
45	Hombre	35	Ingeniero
46	Hombre	23	Estudiante
47	Hombre	24	Estudiante
48	Hombre	33	Estudiante
49	Mujer	23	Estudiante
50	Hombre	24	Estudiante
51	Hombre	26	Ingeniero
52	Hombre	23	Estudiante
53	Hombre	23	Estudiante
54	Hombre	27	Estudiante
55	Hombre	21	Estudiante
56	Hombre	19	Estudiante
57	Mujer	22	Estudiante
58	Hombre	20	Estudiante
59	Hombre	25	Estudiante
60	Mujer	19	Estudiante
61	Hombre	19	Estudiante

Anexo B

Resultados de la clasificación. Por participante.

B.1. Naive Bayes y selección con matriz de correlación.

Tabla B.1: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas seleccionadas con el método de matriz de correlación.

Part.	Umbral de selección	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
1	0,4415	5	3	59	53,78	54
2	0,3627	5	5	36,89	30,57	29,24
3	0,3368	3	3	56,28	41,04	35,9
4	0,4096	5	2	90,67	91	91,73
5	0,3543	3	3	60,1	57,55	52,85
6	0,3746	3	3	52,31	45,52	49,3
7	0,3909	5	4	35	35	27,68
8	0,3344	5	4	40,34	31,3	30,53
9	0,3825	3	3	60,1	47,41	49,34
10	0,3863	6	5	35,67	31,3	26,78
12	0,3716	4	4	39,23	36,54	34,98
13	0,3983	3	2	80,08	75,52	86,19
14	0,4154	3	3	60,55	47,22	43,36
15	0,4136	3	3	52,19	43,92	43,46
17	0,3946	4	4	45,04	30,97	27,49
18	0,3721	3	3	46,2	46,36	42,67
19	0,3649	5	3	61,25	55,73	45,09
20	0,3883	4	3	53,36	47,11	40,48
21	0,3797	4	4	40,18	33,53	33,77
22	0,3479	6	5	47,18	40,55	39,46
23	0,36	3	2	69,33	68,52	74,08

Part.	Umbral de selección	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
24	0,3798	4	4	56,14	44,18	43,87
25	0,3604	5	5	37,97	31,26	37
26	0,4028	5	5	36,24	27,73	19,1
27	0,371	5	4	40,55	37,9	28,45
28	0,3859	3	2	66,23	67,05	70,03
29	0,3695	4	3	61,04	51,77	48,36
30	0,3822	5	4	46,77	39,9	42,01
31	0,3854	4	4	65,19	54,4	54,99
32	0,3651	6	5	47	39,99	40,85
33	0,4604	3	3	49,48	37,74	36,19
34	0,3648	5	5	55,83	42,11	38,06
35	0,4645	6	4	44,1	31,38	19,61
37	0,3536	3	3	50,2	50,22	57,49
39	0,3992	6	5	37,11	37,95	30,29
40	0,3612	6	5	37,07	29,68	26,88
42	0,4683	3	3	52,48	38,7	46,38
43	0,3653	3	3	54,69	52,76	59,03
44	0,3668	5	4	41,72	33,85	30,62
45	0,3909	5	3	46,08	40,09	42,33
46	0,364	6	6	39,96	31,9	25,63
48	0,334	6	5	39,1	31,79	30,75
50	0,5495	3	3	52,87	34,53	20,41
51	0,4356	6	4	48,61	30,18	31,52
52	0,3907	4	4	54,92	36,14	31,05
53	0,3812	4	3	60,58	49,65	40,33
55	0,383	4	3	60,56	57,41	61,79
56	0,394	3	3	54,09	42,86	45,85
57	0,4614	6	5	48,86	40,16	38,25
58	0,3744	3	3	50,59	44	57,46
59	0,3744	3	3	61,37	56,15	62,27
60	0,3735	3	3	49,53	44,79	56,71
61	0,4386	3	2	70,46	68,76	76,3
Prom.	0,3901	4,2	3,6	51,67	44,29	43,55
Std.	0,0392	1,2	1	11,58	12,95	16,23
Máx.	0,5495	6	6	90,67	91	91,73
Mín.	0,334	3	2	35	27,73	19,1

B.2. Naive Bayes y selección con RFE.

Tabla B.2: Resultados de clasificación usando Naive Bayes a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
1	5	3	78,57	77,78	85,71
2	5	5	55,56	49,81	45,5
3	3	3	68,75	53,59	54,17
4	5	2	94,5	92,25	96,43
5	3	3	71,43	71,63	82,42
6	3	3	84,62	71,67	89,18
7	5	4	56,25	56,25	76,11
8	5	4	43,66	36,61	61,08
9	3	3	76,19	81,48	82,42
10	6	5	52,38	50	56,25
12	4	4	65,38	66,43	75,29
13	3	2	96	95	96,88
14	3	3	68,18	55,83	52,94
15	3	3	80,95	82,01	88,1
17	4	4	70,83	50	36,79
18	3	3	72	71,76	83,33
19	5	3	75	77,38	81,82
20	4	3	92,86	94,44	93,33
21	4	4	39,29	42,36	56,25
22	6	5	54,55	50	56,33
23	3	2	76,19	76,82	84,38
24	4	4	53,57	44,97	61,46
25	5	5	51,72	45,86	68,57
26	5	5	52	40	47,62
27	5	4	48,28	50	73,96
28	3	2	81,82	80,83	87,5
29	4	3	69,23	60,61	64,76
30	5	4	61,54	54,46	57,17
31	4	4	68,75	57,5	68,89
32	6	5	50	49	43,33
33	3	3	62,07	49,88	50,67
34	5	5	61,11	50	53,33
35	6	4	60	56,25	60
37	3	3	70	72,5	84,21
39	6	5	72,22	73,33	80,83
40	6	5	55,56	49,29	57,56
42	3	3	47,62	57,62	64,44
43	3	3	73,08	70,91	84,21

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
44	5	4	53,85	49,59	57,43
45	5	3	72,22	71,12	83,95
46	6	6	46,15	34,44	38,89
48	6	5	47,62	39,22	49,07
50	3	3	65,22	51,59	53,33
51	6	4	52,17	55,18	60
52	4	4	52,78	43,13	53,27
53	4	3	68,42	63,33	70,91
55	4	3	61,11	62,38	77,78
56	3	3	63,64	65	78,75
57	6	5	52,38	40,76	47,37
58	3	3	61,54	60,71	76,74
59	3	3	68,42	61,11	85,71
60	3	3	72,09	63,82	83,13
61	3	2	92,31	92,86	92,86
Prom.	4,2	3,6	64,94	60,76	68,91
Std	1,2	1	13,53	15,54	16,38
Máx.	6	6	96	95	96,88
Mín.	3	2	39,29	34,44	36,79

B.3. Redes Neuronales Artificiales y selección con matriz de correlación.

Tabla B.3: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas seleccionadas con el método de matriz de correlación.

Part.	Umbral de selección	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
1	0,3897	5	3	64,57	61,67	59,92
2	0,3565	5	5	37,74	34,62	31,06
3	0,3848	3	3	59,97	47,42	50,46
4	0,3903	5	2	88	86,88	90,2
5	0,3824	3	3	62,24	61,43	64,59
6	0,3782	3	3	63,73	55,47	60,14
7	0,3742	5	4	47	47	44,21
8	0,3893	5	4	42,14	34,6	31,23
9	0,3322	3	3	60,52	49,74	48,52
10	0,406	6	5	42,19	41,45	38,83
12	0,3685	4	4	45,42	44,42	43,07
13	0,3524	3	2	83	81,45	84,04
14	0,2657	3	3	60,73	52,04	51,83
15	0,3255	3	3	63,38	62,71	64,45
17	0,3404	4	4	53,83	40,49	36,87
18	0,3344	3	3	52,56	52,1	51,22
19	0,3691	5	3	61,69	57,3	59,3
20	0,379	4	3	65,29	62,31	61,93
21	0,3723	4	4	47,89	41,36	40,07
22	0,337	6	5	42,36	37,2	32,97
23	0,3629	3	2	77,1	76,89	78,85
24	0,3594	4	4	52,29	43,4	42,36
25	0,3223	5	5	44,69	33,57	31,34
26	0,3405	5	5	40,88	31,93	27,2
27	0,3037	5	4	42,24	41,52	38,88
28	0,3235	3	2	68,86	68,48	71,52
29	0,3965	4	3	59,92	52,7	51,66
30	0,3795	5	4	50,81	46,23	46,63
31	0,3888	4	4	57,94	51,37	51,95
32	0,4007	6	5	51,56	49,47	45,25
33	0,3951	3	3	57,83	48,46	48,84
34	0,3333	5	5	45,67	36,5	32,24
35	0,3928	6	4	54,9	48	43,22
37	0,3552	3	3	54,87	52,63	54,51
39	0,3337	6	5	41	40,55	35,69
42	0,3609	3	3	65,19	56,04	55,82
43	0,3419	3	3	56,19	50,75	55,59

Part.	Umbral de selección	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
44	0,3497	5	4	49,31	37,59	36,31
45	0,3623	5	3	57,28	56,14	60,11
46	0,3499	6	6	40,19	32,08	30
48	0,2934	6	5	37,17	30,65	25,83
50	0,3632	3	3	61,52	47,39	49,59
51	0,3706	6	4	53,57	40,31	37,66
52	0,3116	4	4	51,11	36,1	37,73
53	0,403	4	3	65,68	57,3	61,1
55	0,365	4	3	54,06	52,76	52,74
56	0,357	3	3	59,86	48,97	45,32
57	0,3835	6	5	44,62	36,29	33,48
58	0,3661	3	3	55,15	48,83	50,36
59	0,3049	3	3	63,26	55,64	57,71
60	0,3373	3	3	56,23	50,59	51,56
61	0,2706	3	2	78,46	78,65	81,48
Prom.	0,3559	4,2	3,6	55,65	49,8	49,37
Std	0,0327	1,1	1	11,29	12,79	14,81
Máx.	0,406	6	6	88	86,88	90,2
Mín.	0,2657	3	2	37,17	30,65	25,83

B.4. Redes Neuronales Artificiales y selección con RFE.

Tabla B.4: Resultados de clasificación usando Redes Neuronales Artificiales a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
1	5	3	78,57	77,78	84,85
2	5	5	51,85	52,24	66,07
3	3	3	71,88	64,05	86,21
4	5	2	91,67	87,5	95
5	3	3	76,19	75,79	85,71
6	3	3	80,77	80	88,89
7	5	4	68,75	68,75	70,83
8	5	4	49,3	41,96	74,51
9	3	3	76,19	74,07	85,42
10	6	5	57,14	57,67	68,67
12	4	4	61,54	64,29	77,29
13	3	2	92	93,33	94,12
14	3	3	72,73	72,5	86,27
15	3	3	80,95	82,01	84,24
17	4	4	66,67	52,5	57,6
18	3	3	68	68,98	80,95
19	5	3	62,5	66,67	81,82
20	4	3	85,71	86,11	88,57
21	4	4	60,71	54,86	69,64
22	6	5	59,09	51,67	66,33
23	3	2	95,24	95,45	95,45
24	4	4	71,43	62,91	77,98
25	5	5	55,17	49,24	52,75
26	5	5	56	47,5	53
27	5	4	51,72	54,76	74,09
28	3	2	81,82	80,83	83,33
29	4	3	76,92	71,52	85,96
30	5	4	69,23	66,96	75,95
31	4	4	75	78,33	82,86
32	6	5	75	77	85,83
33	3	3	68,97	67,46	84,72
34	5	5	61,11	55	68
35	6	4	70	68,75	77,08
37	3	3	73,33	75,83	78,97
39	6	5	55,56	56,67	62,22
40	6	5	59,26	59,05	66,1
42	3	3	76,19	68,57	85,19
43	3	3	69,23	67,58	77,78

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
44	5	4	56,41	56,98	69,7
45	5	3	72,22	74,68	78,5
46	6	6	53,85	53,61	66,31
48	6	5	45,24	39,99	46,29
50	3	3	69,57	62,3	86,67
51	6	4	60,87	55,89	59,24
52	4	4	61,11	47,53	53
53	4	3	84,21	81,67	87,5
55	4	3	66,67	67,94	80
56	3	3	68,18	70	84,21
57	6	5	61,9	58,1	72,56
58	3	3	66,67	64,58	79,01
59	3	3	73,68	63,89	83,33
60	3	3	67,44	64,8	77,12
61	3	2	92,31	92,86	92,86
Prom.	4,2	3,6	68,94	66,62	76,92
Std	1,2	1	11,54	13	11,62
Máx.	6	6	95,24	95,45	95,45
Mín.	3	2	45,24	39,99	46,29

B.5. SVM con kernel lineal y selección con RFE.

Tabla B.5: Resultados de clasificación usando SVM con kernel lineal a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	83,33	87,5	90
13	3	2	88	85	91,67
23	3	2	95,24	95	95,83
28	3	2	81,82	80,83	85,29
61	3	2	92,31	92,86	92,86
Prom.	3,4	2	88,14	88,24	91,13
Std	0,9	0	6,03	5,95	4,36
Máx.	5	2	95,24	95	95,83
Mín.	3	2	81,82	80,83	85,29

B.6. SVM con kernel polinomial de orden 2 y selección con RFE.

Tabla B.6: Resultados de clasificación usando SVM con kernel polinomial de orden 2 a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	83,33	87,5	90
13	3	2	84	85	87,5
23	3	2	85,71	85,45	89,29
28	3	2	86,36	85,83	86,75
61	3	2	94,54	94,69	95,09
Prom.	3,4	2	86,79	87,69	89,73
Std	0,9	0	1,42	1,09	1,51
Máx.	5	2	86,36	87,5	90
Mín.	3	2	83,33	85	86,75

B.7. SVM con kernel polinomial de orden 3 y selección con RFE.

Tabla B.7: Resultados de clasificación usando SVM con kernel polinomial de orden 3 a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	87,5	86,25	89,73
13	3	2	88	88,33	88,19
23	3	2	90,48	90,91	92,31
28	3	2	86,36	85,83	86,75
61	3	2	95,54	95,68	95,99
Prom.	3,4	2	89,58	89,4	90,59
Std	0,9	0	1,74	2,33	2,38
Máx.	5	2	90,48	90,91	92,31
Mín.	3	2	86,36	85,83	86,75

B.8. SVM con kernel polinomial de orden 4 y selección con RFE.

Tabla B.8: Resultados de clasificación usando SVM con kernel polinomial de orden 4 a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	85,83	85	88,17
13	3	2	88	88,33	89,47
23	3	2	90,48	90,45	90,45
28	3	2	81,82	81,67	85,29
61	3	2	94,08	94,32	94,65
Prom.	3,4	2	88,04	87,95	89,61
Std	0,9	0	3,67	3,85	2,24
Máx.	5	2	90,48	90,45	90,45
Mín.	3	2	81,82	81,67	85,29

B.9. SVM con kernel polinomial de orden 5 y selección con RFE.

Tabla B.9: Resultados de clasificación usando SVM con kernel polinomial de orden 5 a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	85,83	85	88,14
13	3	2	88	86,67	88,19
23	3	2	90,48	90,45	92,31
28	3	2	81,82	83,33	85,71
61	3	2	93,62	93,83	94,29
Prom.	3,4	2	87,95	87,86	89,73
Std	0,9	0	3,67	3,05	2,74
Máx.	5	2	90,48	90,45	92,31
Mín.	3	2	81,82	83,33	85,71

B.10. SVM con kernel gaussiano y selección con RFE.

Tabla B.10: Resultados de clasificación usando SVM con kernel gaussiano a partir de características de señales psico-fisiológicas seleccionadas con el método de RFE.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
4	5	2	66,67	50	33,33
13	3	2	68	60	82,61
23	3	2	71,43	72,73	82,35
28	3	2	68,18	65	81,58
61	3	2	93,62	93,36	94,54
Prom.	3,4	2	73,58	68,22	74,88
Std	0,9	0	2,02	9,52	24,43
Máx.	5	2	71,43	72,73	82,61
Mín.	3	2	66,67	50	33,33

B.11. Deep Learning.

Tabla B.11: Resultados de clasificación usando Deep Learning.

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
1	5	3	88,35	92,14	86,79
2	5	5	85,24	96,14	94,26
3	3	3	99,23	98,52	100
4	5	2	96,29	93,35	95,9
5	3	3	90,42	90,21	85,66
6	3	3	95,93	100	92,95
7	5	4	84,27	96,29	92,94
8	5	4	76,35	71,45	83,52
9	3	3	99,9	96,83	100
10	6	5	82,4	86,48	81,46
12	4	4	86,98	87,5	89,01
13	3	2	90,22	85,64	93,01
14	3	3	94,69	95,13	98,47
15	3	3	93,52	94,77	95,73
17	4	4	72,16	77,52	76,44
18	3	3	83,95	83,92	82,77
19	5	3	98,69	99,79	98,61
20	4	3	81,29	82,88	82,01
21	4	4	80,58	82,51	82,23
22	6	5	91,27	91,02	92,45
23	3	2	81,28	79,99	81,56
24	4	4	97,71	99,37	99,61
25	5	5	91,66	91,91	91,63
26	5	5	86,89	86,97	85,43
27	5	4	73,36	74,82	74,35
28	3	2	86,39	86,1	86,38
29	4	3	99,86	96,45	100
30	5	4	90,14	88,35	88,59
31	4	4	91,96	89,75	90,21
32	6	5	98,23	95,71	96,67
33	3	3	76,77	78,3	75,56
34	5	5	93,74	93,36	91,09
35	6	4	90,51	90,71	90,43
37	3	3	85,97	82,56	87,04
39	6	5	95,63	94,4	94,83
40	6	5	80,06	79,45	79,81
42	3	3	80,37	86,05	86,67
43	3	3	99,33	100	97,19
44	5	4	95,06	94,36	93,54

Participante	N inicial de clases	N final de clases	Accuracy (%)	Recall (%)	Precision (%)
45	5	3	79,62	81,51	80,87
46	6	6	78,53	76,32	78,85
48	6	5	79,32	82,39	78,24
50	3	3	80,59	77,67	79,62
51	6	4	99,5	99,61	99,68
52	4	4	81,06	80,88	81,62
53	4	3	88,73	88,21	84,42
55	4	3	98,34	98,14	96,7
56	3	3	79,02	79,35	75,98
57	6	5	88,19	86,27	90,35
58	3	3	79,67	81,65	80,38
59	3	3	98,47	95,84	99,25
60	3	3	93,01	94,66	91,39
61	3	2	97,69	95,46	97,12
Prom.	4,2	3,6	88,46	88,84	88,85
Std	1,2	1	7,94	7,67	7,64
Máx.	6	6	99,9	100	100
Mín.	3	2	72,16	71,45	74,35