

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

ESCUELA DE POSTGRADO

MAESTRÍA EN CIENCIAS MENCIÓN INFORMÁTICA



TESIS

**“DETERMINAR LA EFICIENCIA DE ALGORITMOS DE CLASIFICACIÓN
PARA ANALIZAR SEÑALES DE ELECTROENCEFALOGRAMAS DE
MOVIMIENTO IMAGINARIO Y PROPONER UN PROTOTIPO DE SOFTWARE
INTERACTIVO CEREBRO COMPUTADOR”**

Para optar el grado académico de:

MAESTRO EN CIENCIAS

MENCIÓN INFORMÁTICA

Presentada por el:

BR. CARLOS RAMÓN QUISPE ONOFRE

Asesor:

MGT. WALDO ELIO IBARRA ZAMBRANO

Cusco – Perú

2019

INDICE DE CONTENIDO

ÍNDICE DE FIGURAS	iv
ÍNDICE TABLAS	v
SIGLAS	vi
RESUMEN	vii
ABSTRACT.....	ix
INTRODUCCIÓN	xi
PLANTEAMIENTO DEL PROBLEMA	1
1.1. Situación problemática	1
1.2. Formulación del problema.....	2
1.2.1. Problema general	2
1.2.2. Problemas específicos.....	3
1.3. Justificación de la investigación	3
1.4. Objetivos de la investigación.....	3
1.4.1. Objetivo general	3
1.4.2. Objetivos Específicos	4
1.5. Delimitación	4
MARCO TEÓRICO	6
2.1. Bases teóricas conceptuales.....	6
2.1.1. Señales biomédicas.....	6

2.2.	Marco conceptual	10
2.2.1.	Arboles de decisión	10
2.2.2.	Naive Bayes.....	17
2.2.3.	Secuencia Mínima optima (SMO).....	22
2.2.4.	Support Vector Machine (SVM)	23
2.2.5.	Análisis Discriminante Lineal (LDA)	28
2.2.6.	Brain Computer Interface – Interfaz Cerebro Computador BCI.....	30
2.2.7.	Electroencefalograma	32
2.2.8.	Transformadas de Wavelet	36
2.3.	Antecedentes empíricos de la investigación	46
	METODOLOGÍA.....	50
4.1.	Ámbito de estudio: localización política y geográfica	50
4.2.	Tipo y nivel de la Investigación	50
4.3.	Unidad de Análisis.....	50
4.4.	Técnicas de recolección de información.....	50
4.5.	Técnicas de análisis e interpretación de la información	51
4.6.	Técnicas para medir la eficacia de los algoritmos de clasificación	51
	DESARROLLO DEL PROYECTO	54
4.1.	Desarrollo del proyecto.....	54
4.1.1.	Adquisición de datos.....	55
4.1.2.	Pre procesamiento de datos	56

4.1.3. Extracción de características.....	59
4.1.4. Selección de características	61
4.1.5. Clasificación	62
RESULTADOS Y DISCUSION	63
5.1. Presentación de resultados.....	63
5.2. Discusión de resultados	66
CONSTRUCCION DEL PROTOTIPPO	68
6.1. Construcción del prototipo de software interactivo cerebro computador....	68
6.1.1. Planificación.	68
6.1.2. Diseño:.....	69
6.1.3. Codificación.....	70
6.1.4. Pruebas.....	72
6.1.5. Lanzamiento.	72
Conclusiones.....	73
Recomendaciones	74
Bibliografía	75

ÍNDICE DE FIGURAS

Figura 1 Clasificador Naive Bayes – Independencia asumida.	18
Figura 2 Naive bayes - distribución de probabilidades.....	19
Figura 3 Clasificación binaria con SVM.	24
Figura 4 Interfaz Cerebro Computador. Fuente:.....	31
Figura 5 Sistema 10 – 20 para la toma de EEG.	35
Figura 6 Ejemplo de Wavelet madre.	39
Figura 7 Función coseno en diferentes escalas.....	41
Figura 8 Traslación y escalado.	42
Figura 9 Transformada de Wavelet continua.....	44
Figura 10 Tipos de curvas ROC.	53
Figura 11 Esquema de desarrollo del proyecto de investigación.....	54
Figura 12 Esquema de captura de datos EEG.....	56
Figura 13 Datos de los canales C3 y C4 de a sesión 1.....	57
Figura 14 Determinando la frecuencia y tiempo de la señal.....	57
Figura 15 Señales Canal C3.....	58
Figura 16 Señales del Canal C4.....	58
Figura 17 Canal C3 en el intervalo de 4 a 7 segundos.....	59
Figura 18 Canal C4 en el intervalo de 4 a 7 segundos.....	59
Figura 19 Descomposición de la señal EEG mediante la transformada discreta de Wavelet.	60
Figura 20 Prototipo de Simulador de Interfaz cerebro computador	70
Figura 21 Software de clasificación de LDA.....	70
Figura 22 Código para clasificar el algoritmo LDA	71

ÍNDICE TABLAS

Tabla 1: Matriz de confusión	51
Tabla 2: Resultados del algoritmo J48.....	63
Tabla 3: Resultados del algoritmo Naive Bayes	64
Tabla 4: Resultados del algoritmo SMO.....	64
Tabla 5: Resultados del algoritmo LDA.....	65
Tabla 6: Resumen de resultados	66
Tabla 7 Historia	68
Tabla 8 Tarea 1	69
Tabla 9 Tarjeta CRC	69
Tabla 10 Caso de prueba.....	72

SIGLAS

BCI	: Brain computer Interfaz – Interfaz cerebro computador
EEG	: Electroencefalograma
MI	: Movimiento Imaginario
CWT	: Transformada de Wavelet continua
SWT	: Transformada de Wavelet semidiscreta
DWT	: Transformada de Wavelet Discreta
J48	: Árbol de decisión de la familia J48
SMO	: Secuencia mínima óptima.
SVM	: Máquina de soporte de vectores
LDA	: Análisis Discriminante lineal.

RESUMEN

El presente trabajo busca determinar la eficiencia de algoritmos de clasificación para analizar señales de electroencefalogramas de movimiento imaginario. Para este objetivo se compararon los resultados de exactitud de los algoritmos de clasificación J48, Naive Bayes, Secuencia Mínima óptima (SMO) y el análisis lineal discriminante (LDA) por ser estos los más utilizados en las diferentes investigaciones consultadas.

Los métodos para clasificar señales de electroencefalogramas son: Adquisición de datos, pre procesamiento de datos, extracción de características, selección de características y la clasificación propiamente dicha. El conjunto de datos se obtuvo del III BCI Competition, que es un conjunto de datos públicos sobre electroencefalogramas de movimiento imaginario hacia la derecha y hacia la izquierda. Para el pre procesamiento se determinó que el tiempo de intensión de movimiento debe de estar entre el segundo 4 a 7 y que solo se utilizarían los electrodos C3 y C4, ya que los resultados con Cz no son significativos. Para la extracción de características se utilizó la transformada discreta de wavelet; que permitió descomponer la señal de los electroencefalogramas dentro de los rangos de las ondas alfa y beta, que son las encargadas de la intensión motora en el cerebro.

En la selección de características se consideró tres estadísticos como son: potencia promedio, valor medio absoluto, varianza, junto con sus combinaciones; Cada una de estas fue clasificada mediante los algoritmos J48, Naive Bayes, SMO y LDA.

De los algoritmos seleccionados el que tuvo mejor eficiencia fue el LDA con un 88.57%, seguido del SMO con un 87.86%. También se obtuvo un 0.7143 de concordancia en el estadístico Kappa.

Palabras claves

- Brain computer Interface (BCI), Electroencefalograma, transformada Discreta de Wavelet, J48, Naive Bayes, SMO, LDA

ABSTRACT

The present work seeks to determine the efficiency of classification algorithms to analyze electroencephalogram signals of imaginary movement. For this purpose, the accuracy results of the classification algorithms J48, Naive Bayes, Optimal Minimum Sequence (SMO) and discriminant linear analysis (LDA) were compared because these are the most used in the different investigations consulted.

The methods for classifying electroencephalogram signals are: Data acquisition, data preprocessing, feature extraction, feature selection and classification itself. The data set was obtained from the III BCI Competition, which is a set of public data on electroencephalograms of imaginary movement to the right and to the left. For the preprocessing it was determined that the movement intension time must be between the second 4 to 7 and that only electrodes C3 and C4 would be used, since the results with Cz are not significant. For discrete feature extraction, the discrete wavelet transform was used; which allowed to break down the signal of the electroencephalograms within the ranges of alpha and beta waves, which are responsible for motor intension in the brain.

In the selection of characteristics, three statistics were considered, such as: average power, absolute average value, variance, together with their combinations; Each of these was classified using the J48, Naive Bayes, SMO and LDA algorithms.

Of the selected algorithms, the one with the best efficiency was the LDA with 88.57%, followed by the SMO with 87.86%. A 0.7143 agreement was also obtained in the Kappa statistic.

Key words

- Brain computer Interface (BCI), Electroencephalogram, Wavelet Discrete Transform, J48, Naive Bayes, SMO, LDA

INTRODUCCIÓN

Los algoritmos de clasificación dentro del aprendizaje automático buscan que las computadoras aprendan a partir de comportamientos o características de los datos almacenados.

La clasificación de las señales de electroencefalogramas busca con sus resultados crear dispositivos o interfaces para personas con parálisis total o parcial de su cuerpo. A estos dispositivos los conocemos como BCI (brain computer interface o interfaces cerebro computador).

El presente trabajo busca comparar y analizar las señales de electroencefalogramas de movimiento imaginario mediante los algoritmos de clasificación J48, Naive Bayes, SMO y LDA, para determinar cuál de estos es más eficiente al predecir un movimiento imaginario hacia la derecha o hacia la izquierda.

Para realizar la clasificación de señales de electroencefalogramas no existe un único método de desarrollo y la cantidad de ondas emitidas por el cerebro hacen que este proceso sea muy complejo. Para esta investigación se considera como un filtro de ondas la transformada discreta de Wavelet y como vectores característicos se utilizan la potencia promedio, el valor promedio absoluto con sus respectivas combinaciones para luego ser clasificados mediante los algoritmos J48, Naive Bayes, SMO y LDA. Resultando el más eficiente el LDA.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Situación problemática

Los algoritmos de clasificación dentro del aprendizaje automático buscan que las computadoras aprendan a partir de comportamientos o características de los datos almacenados. Un sistema de clasificación genera un modelo predictivo.

Las técnicas de aprendizaje automático son muy eficientes para poder predecir el comportamiento de las señales de los electroencefalogramas, analizarlos en tiempo real y predecir su comportamiento; los algoritmos de clasificación para aprendizaje supervisado destacan por su alto nivel eficiencia.

En los estudios de clasificación de señales de electroencefalogramas de intención de movimiento existen diferentes algoritmos para clasificarlos entre estos tenemos: J48, Naive Bayes, la secuencia mínima óptima (SMO) implementación de John Platt que sirve para entrenar máquina de soporte vectorial (SVM), el análisis discriminante lineal (LDA), entre otros.

El electroencefalograma es una técnica no invasiva que sirve para medir la actividad cerebral, existen en el mercado diversos modelos que varían en costo y número de electrodos.

En un electroencefalograma se colocan un conjunto de electrodos en la cabeza de la persona las que captan las diferentes emisiones de voltajes que son representados en función al tiempo. Para la colocación de los electrodos se utiliza el estándar 10/20. La ubicación de los electrodos, los movimientos involuntarios de los ojos, los movimientos musculares de la mandíbula y otras partes de la cabeza y cara producen

señales denominadas como ruido. El registro del electroencefalograma se clasifica por frecuencia teniendo cuatro bandas representativas. Beta (13 y 30 Hz), alfa (8 y 12 Hz), theta (4 a 7 Hz) y delta (ondas menores de 4 Hz). En el presente estudio veremos el análisis de señales alfa y beta ya que estas se encuentran dentro del espectro sensorio motriz. Estos datos son almacenados y representados mediante ondas en la pantalla del computador. La interpretación de estas ondas solo las puede realizar un especialista.

Con los resultados de la clasificación de electroencefalogramas se construyen interfaces que ayudan a las personas que tienen discapacidades motrices totales o parciales para elevar su calidad de vida mediante el uso de aparatos prostéticos que respondan a las instrucciones dadas por su cerebro.

De los algoritmos previamente mencionados; el LDA es el algoritmo de clasificación que obtiene mejores resultados, esto según la literatura revisada, que serán corroborados en este estudio según los resultados obtenidos.

La presente investigación busca comparar los algoritmos de clasificación J48, Naive Bayes, optimización secuencial mínima (SMO) y análisis discriminante lineal (LDA) para determinar cuál es el más eficiente para analizar las señales de electroencefalogramas de movimiento imaginario. Con los resultados que se obtengan se propone construir un software interactivo.

1.2. Formulación del problema

1.2.1. Problema general

¿Se podrá determinar la eficiencia de los algoritmos de clasificación en el análisis de señales de electroencefalogramas de movimiento imaginario y construir un prototipo de software interactivo a partir de los resultados?

1.2.2. Problemas específicos

- ¿Será más eficiente el algoritmo de clasificación LDA en el análisis de señales de electroencefalogramas de movimiento imaginario?
- ¿Se podrá aplicar los resultados en la construcción del prototipo de software interactivo cerebro computador?

1.3. Justificación de la investigación

La presente investigación busca comparar los resultados de la clasificación de las señales de electroencefalogramas mediante diferentes algoritmos como J48, Naive Bayes, SMO y LDA para poder determinar cuál es el más eficiente mediante el uso. Al existir diferentes métodos de procesar las señales de electroencefalogramas se pretende buscar una manera eficiente de clasificarlas.

Con los resultados obtenidos mediante la comparación de los algoritmos de clasificación se podrán construir diferentes interfaces que ayudarán a las personas que sufren discapacidad motora completa o parcial y de esta manera mejorar su calidad de vida.

Al tener un prototipo interactivo de software se demostrará si los resultados obtenidos con el software Weka son concordantes o similares a los resultados obtenidos mediante su implementación en un lenguaje de programación de alto nivel.

1.4. Objetivos de la investigación

1.4.1. Objetivo general

Determinar el algoritmo de clasificación más eficiente para analizar señales de electroencefalogramas de movimiento imaginario y construir un prototipo de software interactivo a partir de los resultados.

1.4.2. Objetivos Específicos

- Determinar cuál de los algoritmos de clasificación: J48, Naive Bayes, optimización mínima secuencial o análisis discriminante lineal es el más eficiente en el análisis de señales de electroencefalogramas de movimiento imaginario.
- Construir un prototipo de software interactivo cerebro computador.

1.5. Delimitación

El presente trabajo de investigación busca determinar la eficiencia de los algoritmos de clasificación en el análisis de electroencefalogramas para movimiento imaginario para lo cual se tomó el conjunto de datos público de la Universidad de Graz proporcionada para el concurso de Berlin Brain computer II (BCI Competition II, 2003).

Los algoritmos de clasificación seleccionados para este estudio son: J48, Naive Bayes, LDA (análisis lineal discriminante), y SMO (secuencia mínima optima). Se utilizan estos algoritmos ya que pertenecen a diferentes grupos de funciones dentro del programa de aprendizaje automático, Weka, y son los que dan los mejores resultados en investigaciones previas de acuerdo a la revisión bibliográfica previa.

El J48 es un árbol de decisión que a partir de un conjunto de datos fabrica diagramas de construcciones lógicas, muy parecidas a las reglas de decisión.

Naive Bayes es un clasificador caracterizado por ser probabilístico fundamentado en el teorema de Bayes.

LDA (análisis discriminante lineal) es un algoritmo utilizado en estadística, reconocimiento de patrones y aprendizaje automático para encontrar una combinación lineal de rasgos que caracterizan o separan dos grupos de clases.

SMO (secuencia mínima óptima) es un algoritmo de clasificación está basado en las máquinas de soporte vectorial (SVM) que transforma los datos a un plano de dimensión superior y los clasifica en un hiperplano en esa dimensión. El entrenamiento es muy costoso, pero da muy buenos resultados.

CAPÍTULO II

MARCO TEÓRICO

2.1. Bases teóricas conceptuales

2.1.1. Señales biomédicas

El cuerpo humano emite diferentes tipos de señales, entre ellas tenemos las señales de origen químico, acústico o eléctrico, las cuales transmiten información. “Las señales reflejan propiedades de los sistemas biológicos asociadas a ellas”. (Carrión Pérez, 2007).

El procesamiento de señales biomédicas es un trabajo interdisciplinar. “El conocimiento de la fisiología del cuerpo resulta crucial para evitar el riesgo de diseñar métodos de análisis que distorsionen o, eliminen información significativa” (Carrión Pérez, 2007). Es también necesario tener conocimiento de otros temas como álgebra lineal, cálculo estadístico, minería de datos entre otros.

El procesamiento de señales biomédicas a través de modelos es un esfuerzo sistemático de representarlo mediante algoritmos. La complejidad de del modelo de señal depende de la complejidad a resolver.

Las señales emitidas por el corazón, cerebro y músculos tienen un origen bioeléctrico. Algunos se generan de manera automática continua y otros necesitan un estímulo externo. Dentro de estas señales tenemos:

El Electroencefalograma (EEG). El electroencefalograma es un método no invasivo que se utiliza para la captura de las señales eléctricas emitidas en el cerebro. Detecta la comunicación existente entre las neuronas mediante la emisión de señales eléctricas. (Alam Iztac, 2014).

Los Potenciales evocados (EPs). Un potencial evocado es una respuesta del sistema nervioso central o periférico a un estímulo visual, en el caso de los potenciales visuales, auditivo en el caso de los potenciales de tallo o nervio periférico en el caso de los potenciales somato-sensoriales. Su importancia en el diagnóstico de lesiones ocultas, difíciles de localizar mediante otros métodos paraclínicos, lo hacen de gran utilidad en pacientes con esclerosis múltiple, tumores de fosa posterior, trauma, accidentes cerebrovasculares y crisis conversivas. Su importancia en neurocirugía y especialmente en cirugía medular y de fosa posterior es extraordinaria. Existen tres tipos de potenciales evocados: visuales (PEV), de tallo (PET) y somato-sensoriales (PSS). (Toro, 1984) .

El Electrocardiograma (ECG). “El ECG es un registro de la actividad eléctrica del corazón. No limitado a la zona de conducción, sino a todo el corazón de forma global”. (Robledo Carmona, Jiménez Navarro, & Robledo Carmona, 2014). Es un estudio de rutina que se realiza para observar la actividad eléctrica del corazón.

El electrocardiograma puede suministrar mucha información sobre el corazón y su funcionamiento. Con este estudio es posible averiguar más sobre el ritmo cardíaco, el tamaño y funcionamiento de las cavidades del corazón y el músculo cardíaco (Zela Ari & Mamani Chambi, 2015).

El electrocardiograma es llevado a cabo a partir de unos electrodos (conductores eléctricos) aplicados en la superficie de la piel, habitualmente en ambos brazos y piernas y sobre el tórax. Dentro del corazón minúsculos impulsos eléctricos controlan la función del corazón de expansión y contracción del músculo cardíaco para bombear la sangre al cuerpo (Pinzon Duque, 2011). Los electrodos colocados sobre la piel pueden detectar estos impulsos y transmitirlos al electrocardiógrafo por medio de cables (Calero Arevalo, Calle Fuentes,

Ferruzola Campaña, & Torres Panchana, 2010). “Acto seguido, un electrocardiógrafo que incluye una aguja que se mueve arriba y abajo a lo largo de una extensa tira de papel que atraviesa el aparato a un ritmo constante convierte estos impulsos en puntas y depresiones (trazo). Cada ciclo cardíaco o latido se registra como una serie concreta de puntas y depresiones”. (Prieto, s.f.)

El electrógrama (EG). “Es un registro intracardiaco donde los electrodos se han colocado directamente dentro del corazón. La señal EG se usa en equipos implantables tales como marcapasos y desfibriladores”. (Carrión Pérez, 2007).

El electromiograma (EMG). “Las señales electromiográficas son aquellas señales eléctricas producidas durante la contracción y relajación de un músculo” (Córdova Ricapa, 2013)

En el estudio de las señales electromiografías, en especial durante la detección y registro de la señal EMG, existen dos principales fuentes de preocupación que influyen en la fidelidad de la señal. La primera de ellas es la relación señal a ruido y la otra es la distorsión de la señal, lo que significa que la contribución relativa de cualquier componente de frecuencia de la señal no debe ser alterada. (Córdova Ricapa, 2013)

“El EMG es la parte más importante del estudio neurofisiológico. Permite estudiar la actividad eléctrica muscular mediante un electrodo concéntrico de aguja insertado en el músculo que detecta diferencias de potencial extracelular. El estudio del músculo se hace en 2 situaciones: en reposo y en actividad voluntaria” (Gutiérrez-Gutiérrez, Barbosa López, Navacerrada, & Miralles Martínez, 2011).

Está establecido que la amplitud de la señal EMG no tiene un patrón definido, es decir, es al azar, y puede ser representado mediante una función de distribución Gaussiana. La amplitud de las señales EMG varían entre 0 y 10 mV

y estas pueden tener hasta una frecuencia de 10kHz (De Luca Carlo, 2011). “Existen muchos estudios en relación al rango de frecuencia en que se debe estudiar la señal EMG, de modo que dentro de este rango se pueda obtener la mayor cantidad de energía del potencial” (Fridlun Alan & Fowler Stephen, 1978).

El Electroneurograma (ENG). La ENG consiste en el estudio de la conducción de las fibras nerviosas, motoras y sensitivas, evaluándose básicamente la integridad y la función de los nervios periféricos del organismo. En el examen se realiza el registro del potencial eléctrico generado en el músculo (en los estudios de conducción motora) o en el propio nervio (en los estudios de conducción sensitiva) al estimular uno o más puntos de la fibra nerviosa a través de la piel. La información generada permite evaluar parámetros como la velocidad de conducción nerviosa, la latencia o la morfología y la amplitud del potencial evocado motor o sensitivo. Además de los estudios de conducción nerviosa se utilizan otras técnicas como son el estudio de las respuestas tardías (Onda F y Respuesta Refleja H) que exploran los tramos más proximales de los nervios periféricos, la Estimulación Repetitiva, la Electromiografía de fibra simple o aislada (Jitter) a través de la cual se explora la unión neuromuscular o el examen del reflejo del parpadeo o Blink Reflex. (Neurológicas, 2016).

El Electroretinograma (ERG). Permite explorar la función de la retina cuando hay mucha luz o cuando hay poca luz. (Vision, 2018)

Para realizar esta prueba colocamos dos electrodos debajo de los ojos y dos electrodos en el vértice del ojo, en la piel, y sentamos al enfermo frente a una especie de pantalla.

Primero se le hacen unas fotografías con estímulos de luz (flash en el ERG y modelos de rejillas (tablero de ajedrez con cuadrados o rombos) en el P-ERG y

en el m-ERG) con los ojos abiertos, después se ha de estar uso 20 minutos a oscuras, con los ojos cerrados y posteriormente se repiten las fotografías.

El electroretinograma (ERG) registra la respuesta eléctrica de las células de la retina.

Las alteraciones estudiadas más importantes, son:

- Retinosis pigmentaria
- Distrofias de conos y bastones
- Distrofias coriorretinianas
- Enfermedades inflamatorias
- Obstrucción de la arteria y de la vena central de la retina
- Retinopatía diabética
- Desprendimiento de retina
- Ceguera nocturna
- Alteración en la visión de colores

2.2.Marco conceptual

2.2.1. Árboles de decisión

Los árboles de decisión se constituyen como uno de los modelos más utilizados en tareas de clasificación” (Quinlan, 1993). El conocimiento obtenido durante el proceso de construcción del modelo se representa mediante un árbol en el cual cada nodo interno contiene un atributo particular (con un nodo hijo para cada posible valor del atributo) y en el que cada hoja se refiere a una decisión (etiquetada con una de las clases del problema).

Un árbol de decisión puede usarse para clasificar un caso comenzando de su raíz y siguiendo el camino determinado por los valores de los atributos en los

nodos internos hasta que encontremos una hoja del árbol que se refiere a la decisión (clasificación). “Existe una serie de algoritmos desarrollados desde los principios de los 60's para la construcción de árboles de decisión”. CLS (Hunt, 1966) para modelar aprendizaje humano de conceptos, ID3 (Quinlan, 1993) desarrollado con el criterio de ganancia de información para desarrollar sistemas expertos desde ejemplos (casos preclasificados), CART (Breiman, 1984) (Classification And Regression Tree, Arboles de clasificación y regresión) es un sistema recursivo binario de particiones, CHAID (Kass, 1980) (Chi Square Automatic Interaction Detection, Detección de Interacción Automática de Chi Cuadrado) que es un algoritmo recursivo de clasificación no binario, C4.5 (Quinlan, 1993) en él se actualiza (mejora) los sistemas de decisión, etc.

Entre los algoritmos para construir árboles de decisión, el ID3 y el sucesor C4.5, son considerados los más populares para la tarea de clasificación.

Un *árbol de decisión* es una estructura gráfica compuesto por nodos (internos y hojas) y de arcos (ramas). Cada nodo interno está caracterizado por un atributo o variable (con un nodo hijo para cada posible valor del atributo) y en el que cada hoja se refiere a una decisión (etiquetada con una de las clases del problema).

Las características principales de un árbol de decisión son: su construcción sencilla, no necesita determinar de antemano parámetros para su construcción, puede tratar problemas multi-clase de la misma forma en que trabaja con problemas de dos clases y la fácil interpretación de su estructura.

Los árboles de decisión se construyen a partir del método basado en un particionamiento recursivo del conjunto de casos (Hunt, 1966).

La estructura de este método para construir un árbol de decisión a partir de un conjunto T de casos se detalla a continuación.

Sea T el conjunto de casos ya clasificados (datos de entrenamiento) con “ n ” atributos o variables predictoras $\{A_1, \dots, A_n\}$ y una variable clase C con valores c_j , $j = 1, 2, \dots, k$.

Existen tres posibilidades:

1. Si T contiene uno o más casos, y todos pertenecientes a una única clase c_j , el árbol de decisión para T es una hoja etiquetada con la clase c_j .
2. Si T no contiene ningún caso, el árbol de decisión es una hoja, donde la clase asociada a dicha hoja debe ser determinada por información que no pertenece a T . Por ejemplo, la hoja puede etiquetarse de acuerdo a conocimientos base del dominio (conocimiento del problema), como ser la clase mayoritaria.
3. Si T contiene casos pertenecientes a varias clases, La idea es refinar (particionar) el conjunto de casos T en subconjuntos que tiendan a contener casos pertenecientes a una única clase.

Para ello, se selecciona un atributo (o variable) A , que tiene “ m ” posibles valores y T se particiona en los subconjuntos T_1, T_2, \dots, T_m donde T_i contiene todos los casos de T que tienen el valor i -ésimo del atributo seleccionado.

El árbol de decisión para T consiste en que cada nodo interno contiene un atributo (o variable) con una rama para cada valor posible del atributo. El mecanismo de construcción del árbol se aplica recursivamente a cada subconjunto de casos.

En el proceso de *construcción del árbol de decisión*, en cada nodo interno se selecciona un atributo (o variable) que mejor particione o discrimine el

conjunto de casos (subconjuntos que tiendan a contener casos pertenecientes a una única clase, o que consigan nodos más puros) con respecto de un criterio previamente establecido, *criterio de partición*. (el C4.5 y su predecesor el ID3, usan fórmulas basadas en teoría de la información para evaluar la bondad del atributo). El mecanismo de construcción del árbol se aplica recursivamente a cada subconjunto de casos, hasta una *condición de parada*.

Los criterios de partición o división utilizados generalmente se basan en la teoría de la información, esto es, basados en medidas de la impureza de un nodo, específicamente la *entropía* (la cual determina la impureza o incertidumbre de un conjunto de datos).

La bondad de una partición es la disminución de la impureza que se consigue con ella.

Lo que se quiere es buscar particiones que discriminen o que consigan nodos más puros.

Se presenta el criterio de ganancia de información y la proporción de ganancia de información.

Criterio de ganancia de información

El criterio de ganancia de información busca seleccionar el atributo con mayor ganancia de información.

Sea un conjunto de casos T , con múltiples clases $c_j, j = 1, 2, \dots, k$, la *entropía* de T se mide como:

$$I(T) = -\sum_{j=1}^k p_j \log_2 p_j \quad (1)$$

donde p_j es la proporción de casos en T que pertenecen a la clase c_j (o probabilidad de que un caso tomado al azar de T pertenezca a la clase c_j).

Esto es, $I(T)$ es la cantidad de información necesaria para clasificar un caso del conjunto de casos T .

Si el atributo (o variable) A divide el conjunto T en los subconjuntos

T_i , $i = 1, 2, \dots, m$, entonces, la entropía total del sistema de subconjuntos será:

$$I(T, A) = \sum_{i=1}^j P(T_i)I(T_i) \quad (2)$$

donde $I(T_i)$ es la entropía del subconjunto T_i y $P(T_i)$ es la proporción de casos en T que pertenecen a T_i , pueden calcularse, utilizando los tamaños relativos, $n(T_i)$, de los subconjuntos, como:

$$P(T_i) = \frac{n(T_i)}{n(T)}$$

Usando la entropía, la *ganancia de información*, de un atributo A en un conjunto de casos T , se define como:

$$\text{Ganancia}(T, A) = I(T) - I(T, A) \quad (3)$$

donde $I(T)$ es el valor de la entropía a priori antes de realizar la subdivisión y $I(T, A)$ es el valor de la entropía del sistema de subconjuntos generados por la partición según A .

Así, el atributo A seleccionado para determinar la división, será aquel que mayor ganancia obtenga respecto al conjunto T , y se obtendrá a partir de (3).

El criterio de ganancia de información tiene un defecto muy serio y es que presenta una tendencia muy fuerte a favorecer atributos (o variables) con muchos valores (resultados).

Criterio de proporción de ganancia

Esta medida tiene en cuenta tanto la ganancia de información como las probabilidades de los distintos valores del atributo. Dichas probabilidades son

recogidas mediante la denominada información de separación (split information), que no es más que la entropía del conjunto de datos T respecto a los valores del atributo A en consideración, siendo calculada como que representa la información particionada generada al dividir T en “ m ” subconjuntos.

$$I_Separacion(A) = - \sum_{i=1}^m P(T_i) \times \log_2 P(T_i) \quad (4)$$

La información de separación simboliza la información potencial que representa dividir el conjunto de datos, y es usada para compensar la menor ganancia de aquellos test (variables) con pocas salidas (valores). Con ello, tal y como se muestra en (5), la proporción de ganancia es calculada como el cociente entre la ganancia de información (3) y la información de separación (4). Tal cociente expresa la proporción de información útil generada por la división.

$$Porción\ de\ ganancia\ (A) = \frac{Ganancia(T,A)}{I_Separacion(A)} = \frac{I(T)-I(T,A)}{I_Separacion(A)} \quad (5)$$

C4.5 maximiza este criterio de separación, premiando así a aquellos atributos que, aun teniendo una ganancia de información menor, disponen también de menor número de valores para llevar a cabo la clasificación. Sin embargo, si el test incluye pocos valores, la información de separación puede ser cercana a cero, y por tanto el cociente sería inestable. Para evitar tal situación, el criterio selecciona un test que maximice la razón de ganancia, pero obligando a que la ganancia del mismo sea al menos igual a la ganancia media de todos los test examinados. C4.5 ha resultado ser un sistema muy efectivo en la práctica, capaz de ofrecer una representación relativamente simple de los resultados con un bajo coste computacional.

Condición de parada

Cuando se detiene la construcción del árbol de decisión, se construye una hoja a la que se le puede asignar la clase más común de las recogidas por los casos.

Las reglas de parada tratan de predecir si merece la pena seguir construyendo el árbol o no. Ejemplos de este tipo de reglas son: pureza del nodo, cota de profundidad, mínimo de casos.

Árbol de decisión C4.5

El algoritmo del árbol de decisión C4.5 fue propuesto por Quinlan a finales de los años 80 para mejorar las carencias de su predecesor ID3. Desde entonces, ha sido uno de los sistemas clasificadores más referenciados en la bibliografía, principalmente debido a su extremada robustez en un gran número de dominios y su bajo coste computacional.

C4.5 introduce principalmente las siguientes mejoras:

1. Trata eficazmente los valores desconocidos.
2. Maneja los atributos continuos, aplicando una discretización previa.
3. Corrige la tendencia de ID3 a seleccionar los atributos con muchos valores distintos para establecer los test cambiando el criterio de división. Este algoritmo genera árboles de decisión a partir de ejemplos mediante particiones realizadas recursivamente.

En cada nodo, el algoritmo debe decidir cuál atributo o variable elegir para particionar los datos.

Los tipos de pruebas propuestos por Quinlan para C4.5 (Quinlan, 1993) son:

- Si el atributo es discreto, la representación es con un resultado y una rama para cada valor posible de la variable.

- Si el atributo es continuo, se realiza una prueba binaria con resultados $A \leq Z$ y $A > Z$, que comparan el valor de A con el umbral Z . Para calcular Z , se aplica un método, el cual ordena el conjunto de t valores distintos del atributo A presentes en el conjunto de entrenamiento, obteniendo el conjunto de valores $\{a_{i1}, a_{i2}, \dots, a_{it}\}$. Cada par de valores consecutivos aporta un posible umbral teniendo en total $t-1$ umbrales, donde t es como máximo el número de ejemplos. Una vez calculados los umbrales, C4.5 selecciona aquel que maximiza el criterio de separación (proporción de ganancia).

El algoritmo C4.5 incorpora una poda del árbol de decisión una vez que este ha sido inducido. La poda está basada en la aplicación de un test de hipótesis que trata de responder a la pregunta de si merece la pena expandir o no una determinada rama.

Un modelo puede contener varios árboles para distintos atributos de predicción; un árbol varias bifurcaciones, su profundidad y forma está dado por el método de puntuación y del resto de parámetros usados (Barros, 2012).

2.2.2. Naive Bayes

El clasificador Naive Bayes (llamado también clasificador bayesiano simple), es un clasificador bayesiano, esto es, es el que asigna a un objeto u observación en la clase con mayor probabilidad, y supone que todas las variables o atributos son condicionalmente independientes dado el valor de la clase.

Sea A_1, A_2, \dots, A_n , las variables o atributos que permiten predecir el valor de la clase C . La suposición de independencia asumida por el clasificador Naive Bayes da lugar a un modelo de red bayesiana con estructura simple descrita en la

Figura 7. En el existe un único nodo raíz (la clase), y en la que todos los atributos son nodos hoja que tienen como único padre a la variable clase.

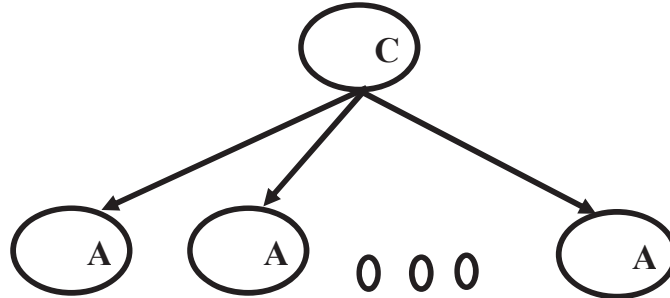


Figura 1 Clasificador Naive Bayes – Independencia asumida. Fuente: <http://www.sc.edu/ccwbayes/docencia/mmcc/docs/t6bayesianos.pdf>

Tomando en cuenta la suposición de que los atributos son condicionalmente independientes dada la clase, la probabilidad de que una observación o conjunto de atributos $\{a_1, a_2, \dots, a_n\}$ pertenezca a la clase “c” es:

$$P(C = c / a_1, a_2, \dots, a_n) = \frac{P(C=c)P(a_1 / C=c)P(a_2 / C=c)\dots P(a_n / C=c)}{P(a_1, a_2, \dots, a_n)} \quad (1)$$

Factorizando

$$P(C = c / a_1, a_2, \dots, a_n) = \frac{P(C=c) \prod_{i=1}^n P(a_i / C=c)}{P(a_1, a_2, \dots, a_n)} \quad (2)$$

Con $c = c_1, c_2, \dots, c_n$

esta ecuación se utilizará para la obtención de las probabilidades en la tarea de clasificación.

Especificación de los parámetros de la red del clasificador Naive Bayes.

La especificación de los parámetros de la red Naive Bayes consiste en estimar los parámetros de las distribuciones de probabilidad que se definen en la red.

Esto es, definida la red Naive Bayes se tendrá que estimar los parámetros de la distribución de probabilidad de la variable clase $P(C)$ y de las distribuciones

condicionales de los atributos dada la clase $P(A_i / C)$, estas distribuciones son representadas en la Figura

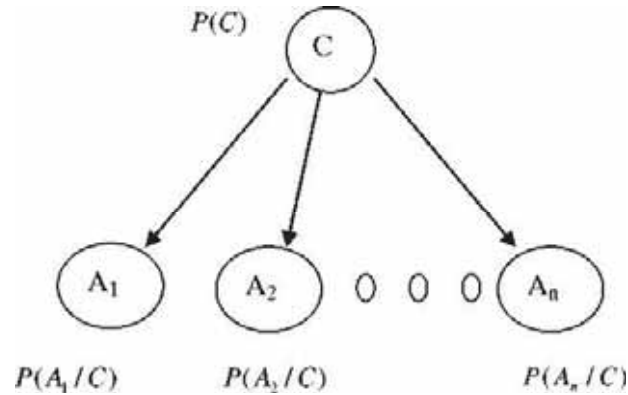


Figura 2 Naive bayes - distribución de probabilidades. Fuente: <https://ccc.inaoep.mx/~esucar/Clases-mgp/pgm06-clasif-2012.pdf>

La estimación de los parámetros va a depender de que las variables o atributos A_i sean discretos o continuos.

Para atributos discretos

La estimación de la distribución de probabilidad de la clase $P(C)$ y de la distribución condicional $P(A_i / C)$ se basa en la frecuencia relativa de ocurrencia de los correspondientes sucesos, que se determinan en el conjunto de datos.

Entonces el valor estimado de probabilidad es:

$$P(a_i / C = c) = \frac{c(a_i, C=c)}{n(C=c)} \quad (3)$$

Donde:

$c(a_i, C = c)$: es el número de casos en que la clase A_i toma el valor a_i y su padre (clase) C toma el valor de c .

$n(C = c)$: es el número de datos del conjunto en que la clase C toma el valor de c .

A pesar de que esta estimación aparenta ser una buena aproximación, para problemas reales, suele suceder que los casos del conjunto de datos no abarcan todas las combinaciones de valores de la variable clase con las variables predictoras, con lo que este tipo de estimación puede llevar a una estimación de parámetros con abundancia de ceros.

Existen varios métodos que intentan solucionar estos problemas. Una de ellos es el estimador basado en la *ley de sucesión de Laplace*, en la cual, en lugar de estimar la probabilidad directamente como,

$$\frac{\text{Casos favorables}}{\text{Casos Totales}}$$

Utilizando como estimación el número que se obtiene al dividir

$$\frac{c(a_i, C = c) + 1}{n(C = c) + r_{A_i}}$$

es decir, el número de casos favorables más uno dividido por el número de casos totales más el número de valores posibles del atributo A denotado por r_{A_i} .

O equivalente

$$\frac{N_{ik} + 1}{N_i + r_i}$$

Donde:

N_{ik} : es el número de casos del conjunto de datos en que la variable A_i toma su k -ésimo valor y la clase C toma su valor c .

N_i : es el número de casos del conjunto de datos en que la variable A_i toma valores y la clase C toma su valor c .

r_i : es el número de valores posibles de la variable A_i

Para atributos continuos

El clasificador Naive Bayes puede ser aplicado también cuando hay variables predictoras continuas, hay dos alternativas:

1. Aplicar previamente un método de discretización de la variable continua.
2. Asumiendo una distribución para cada variable predictora, por lo general gaussiana, con media y varianza estimada de los datos
3. En el segundo caso, el clasificador Naive Bayes supone que la variable predictora en cuestión sigue una distribución normal; por tanto, lo único que se calcula de los datos es la media μ y la desviación típica σ condicionadas a cada valor de la variable clase.

$$4. P(A_i / c) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (4)$$

Evidentemente, esta estimación tiene el inconveniente de que los datos no siempre siguen una distribución normal.

Discretización de variables continuas

En general hay tres razones principales que han llevado a prestar gran atención a la discretización de variables continuas (Dougherty, Kohavi, & Sahami, 1995).

- Muchos algoritmos desarrollados en aprendizaje automático sólo son capaces de aprender desde variables discretas. Debido a que muchas tareas de clasificación de la vida real presentan valores continuos, se hace necesaria una discretización para poder utilizar estos algoritmos.
- Algunos clasificadores, como ocurre con Naive-Bayes, pueden ser utilizados tanto sobre datos discretos como sobre datos continuos, aunque

la correcta discretización de los datos puede llevar a una mejora en su rendimiento.

- Mejora de velocidad en los algoritmos de inducción que utilizan atributos discretos

Los métodos de discretización son desarrollados: usando intervalos de igual ancho, usando intervalos con igual frecuencia, ChiMerge, 1R, discretización usando el método de la entropía, etc.

En los experimentos se emplea la discretización usando intervalos con igual frecuencia con un valor de diez para el número de intervalos, ya que demostró buenos resultados (Kerber, 1992), (Hsu, Huang, & Wong., 2000), (Witten & Frank, 2005), estos estudios fueron realizados en grupos de datos reales de diferentes dominios.

El método de discretización especificado fue implementado con el programa WEKA (Yang & Geoffrey I, 2002).

2.2.3. Secuencia Mínima óptima (SMO)

Es parte de los algoritmos de tipo SVM. Y se basa en actualizar dos parámetros cada vez, considerando los otros parámetros como constantes (Yueguo Luo, 2016). Este clasificador fue diseñado para resolver problemas duales de manera eficiente, el SMO elige dos multiplicadores Lagrange para la optimización simultánea de la función objetivo en cada iteración. (Chih-Hsiang Peng, 2015).

La implementación de support vector machines (SVMs) llamada Sequential Minimal Optimization (SMO): (Joachims, 1998) consiste en una mejora en el algoritmo de entrenamiento de SVMs, de forma tal que este llega a ser 1200 veces más rápido para SVMs lineales y 15 veces más rápido para SVMs

no lineales (Platt, 1998) Las SVMs como método de clasificación son muy populares y ampliamente utilizadas (Kolesov, 2014) debido a su éxito, no solo para clasificar textos sino también para diversos problemas de clasificación. (Kim, 2006) menciona que los clasificadores basados en complejos métodos de aprendizaje como los SVMs. Sequential Minimal Optimization (SMO) es un algoritmo simple que puede resolver rápidamente los grandes problemas de programación cuadrática (QP) de SVM sin utilizar una matriz de almacenamiento y sin utilizar pasos QP de optimización numérica. SMO descompone al problema QP general en sub-problemas QP, utilizando el teorema de Osuna para asegurar la convergencia (Platt, 1998).

2.2.4. Support Vector Machine (SVM)

Las máquinas de soporte vectorial (SVM, de su sigla en inglés Support Vector Machines) son algoritmos que buscan encontrar un hiperplano separador que maximice el margen entre los hiperplanos que se encuentran al borde de cada región a separar. Las regiones a separar corresponden al espacio multidimensional que describe cada una de las clases definidas (Wikstrom, 2005).

Al considerar por ejemplo el problema de separación de dos clases en un espacio de dos dimensiones se puede presentar dos casos. Primero el caso en que las muestras son linealmente separables (figura 9A), donde todas las muestras que corresponden a una misma clase quedan fuera del margen definido por el hiperplano separador, y por consiguiente el error obtenido es cero. Por otro lado, en el caso en que las muestras sean linealmente no separables (figura 9B) se presenta una serie de muestras que se encuentran dentro del margen definido por el hiperplano (Vapnik, 1998).

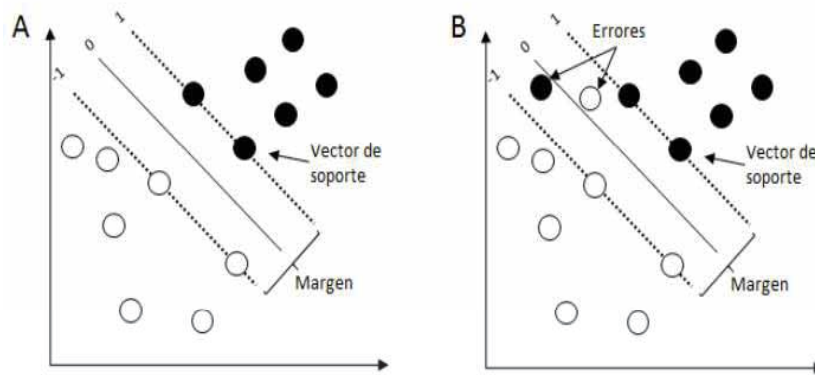


Figura 3 Clasificación binaria con SVM. Fuente: [http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013-Carmona\]%20SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013-Carmona]%20SVM.pdf)

Considerando que frecuentemente los datos no son linealmente separables, el error resultante debe ser minimizado. La expresión a optimizar por el algoritmo esta dado en la siguiente ecuación:

$$\text{Min}_{\vec{w}, b, \xi} \left\{ \frac{1}{2} \|\vec{w}\|^2 + C \sum \xi_i \right\} \quad (1)$$

Sujeto a:

$$y_i(\vec{x}_i \vec{w} + b) + \xi_i - 1 \geq 0, \forall i$$

$$\xi_i > 0, \quad \forall i$$

Donde las variables de decisión son: \vec{w} vector normal al hiperplano separador, b la distancia desde el origen al hiperplano separador, ξ_i variable de holgura que permite que una muestra esté dentro del margen. Los parámetros de entrada son: \vec{x}_i vector de atributos de un objetivo i , y_i clase de pertenencia del objeto i , C penalización de los errores de clasificación.

Las restricciones permiten a cada objeto i estar dentro del margen, permitiendo una holgura. $\xi_i \geq 0$ restringe las variables de holgura a valores mayor o igual a cero, y menor que 1.

Con respecto a la función objetivo a minimizar, la expresión $\frac{1}{2} \|\vec{w}\|^2$ es equivalente a maximizar el margen de separación de clases $2/\|\vec{w}\|$ (donde $\|\vec{w}\|$ es la norma euclidiana de \vec{w}). La minimización de la expresión $C \sum \xi_i$ reduce los casos con holgura ≥ 0 .

Se introduce una transformación del espacio de características con el fin de construir un modelo en los casos que se presentan más muestras respecto al número de dimensiones del espacio de características. Para lo cual se describe a continuación la formulación dual del SVM. Primero se plantea el Lagrangeano de la formulación primal de SVM presentada en la ecuación 3 y posteriormente se verifican las condiciones KKT (Karush - Kuhn - Tucker) (Shawe-Taylor & Cristianin, 2004).

El Lagrangeano (ecuación 2) de la formulación primal de SVM (ecuación 1) es el siguiente:

$$L = \frac{1}{2} \|\vec{w}\|^2 + C \sum \xi_i - \sum \alpha_i \{y_i(\vec{x}_i \cdot \vec{w} + b) + \xi_i - 1\} - \sum \mu_i \xi_i \quad (2)$$

Donde α_i, μ_i son los multiplicadores de Lagrange. Las condiciones KKT de la formulación primal son las siguientes:

$$\frac{\partial L}{\partial w_j} = w_j \sum \alpha_i y_i x_{ij} = 0 \quad (3)$$

$$\frac{\partial L}{\partial b} = - \sum \alpha_i y_i = 0 \quad (4)$$

$$\frac{\partial L}{\partial \xi_i} = C - \xi_i - \mu_i = 0 \quad (5)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) + \xi_i - 1 \geq 0 \quad (6)$$

$$\alpha_i, \xi_i, \mu_i \geq 0 \quad (7)$$

$$\alpha_i \{y_i(\vec{x}_i \cdot \vec{w} + b)\} + \xi_i - 1 = 0 \quad (8)$$

$$\xi_i, \mu_i = 0 \quad (9)$$

Resolviendo las condiciones de las ecuaciones 3 a la 9 se obtiene lo siguiente:

$$w_j = \sum \alpha_i y_i x_{ij} = 0 \quad (10)$$

$$\sum \alpha_i y_i = 0 \quad (11)$$

$$0 < \alpha_i < C \quad (12)$$

$$\mu_i > 0 \quad (13)$$

$$\xi_i \geq 0 \quad (14)$$

Con estas relaciones, el problema dual queda de la siguiente forma:

$$\text{Max}_{\alpha_i} \{ \sum \alpha_i - \sum \alpha_i \alpha_s y_i y_s x_i x_s \} \quad (15)$$

Sujeta a:

$$0 < \alpha_i < C$$

$$\sum \alpha_i y_i = 0$$

Dado que el problema de SVM es un problema de optimización convexo, las condiciones de KKT son necesarias y suficientes para asegurar una solución del problema primal de SVM que cumpla las condiciones KKT de igual manera. Entonces resolver el problema dual de SVM es equivalente a resolver el problema primal.

Cuando la separación lineal de clases no es posible, como ocurre en la figura 10b, se puede realizar una transformación del espacio de origen. Para esto se calculan los productos punto entre vectores de características, utilizando las funciones de Kernel:

$$K(x, z) = \langle \Phi(x), \Phi(z) \rangle \quad \forall x, z \in X \quad (16)$$

Donde \langle , \rangle se define como producto punto, Φ es una función de proyección desde el espacio de origen X hacia el espacio de atributos F (ecuación 17), en el cual se buscan relaciones lineales en este nuevo espacio (figura 4)

$$\Phi: x \mapsto \Phi(x) \in F \quad (17)$$

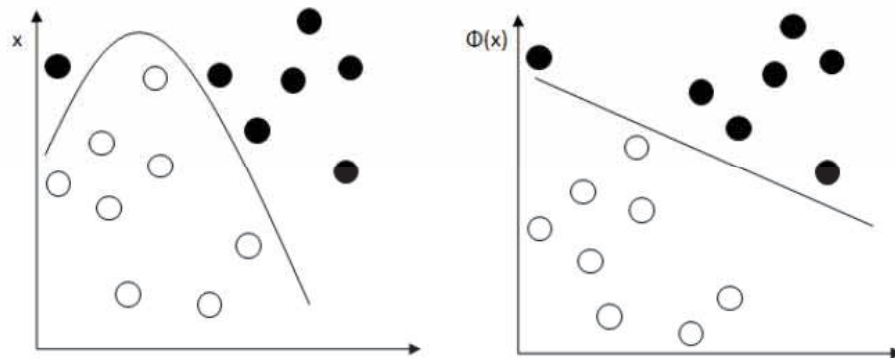


Figura 1: Transformación del espacio original a uno de atributos. Fuente: <https://docplayer.es/1960815-Support-vector-machine.html>

Para que una función sea definida como Mercer kernel debe cumplir con que la matriz de kernel (ecuación 19) sea simétrica y semi-definida positiva (todos sus valores propios son mayores o iguales a cero). La matriz de kernel se expresa como sigue:

$$\begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_l) \\ \vdots & \ddots & \vdots \\ k(x_l, x_1) & \cdots & k(x_l, x_l) \end{pmatrix} \quad (18)$$

Las funciones kernel que existen son diversas, de las cuales las más usadas son:

Lineal: $K(x, z) = \langle x, z \rangle$

Polinomial de grado d : $K(x, z) = (\langle x, z + 1 \rangle)^d, d \in \mathbb{N}$

Gaussiano: $K(x, z) = \frac{-\|x-z\|^2}{2\sigma}, \sigma > 0$

La formulación dual de SVM utilizando una función kernel es la siguiente:

$$\text{Max}_{\alpha_i} \left\{ \sum \alpha_i - \sum \alpha_i \alpha_s y_i y_s K(x_i x_s) \right\}$$

Sujeta a:

$$0 < \alpha_i < C$$

$$\sum \alpha_i y_i = 0$$

Dentro de las ventajas que se pueden mencionar para SVM es que existe una dependencia explícita en los vectores de los márgenes, lo que permite generalizar de mejor forma frente a nuevos objetos, dado que considera el principio de Minimización del Riesgo Estructural. Los vectores de soporte se construyen en base a una función convexa, de modo que se asegura la obtención de un óptimo global y permite la construcción de su formulación dual. Otra gran ventaja es la capacidad de modelar fenómenos no lineales mediante el uso de una transformación del espacio de origen a uno de mayor dimensión (Vapnik, 1998)

Por el otro lado, SVM también presenta limitaciones. Al trabajar con datos numéricos es necesario transformar los atributos nominales a un formato numérico. No existe una función kernel que sea mejor que todas. El uso de distintas funciones de kernel puede determinar diferentes soluciones, por lo que esto se debe determinar para resolver cada problema en particular (Vapnik, 1998).

2.2.5. Análisis Discriminante Lineal (LDA) (Hernández Barajas, 2007)

Supóngase que se tienen G grupos cada uno con $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_g$ observaciones p variadas, además cada grupo posee vector de medias \bar{x}_i , matriz de covarianzas. Si para $\mathbf{i} = 1, 2, \dots, G$ y vector de medias general \bar{x} .

La técnica LDA se basa en los supuestos de normalidad multivariada y de igualdad entre las matrices de varianzas y covarianzas de cada uno de los grupos.

El objetivo del LDA es encontrar un vector \mathbf{a}' de p variables de tal manera que se maximice λ definido por:

$$\lambda = \frac{\mathbf{a}'B\mathbf{a}}{\mathbf{a}'W\mathbf{a}}, \quad (1)$$

donde B y W son dos matrices que representan cada una la variabilidad entre los grupos y la variabilidad dentro de los grupos respectivamente, dadas por;

$$B = \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad (2)$$

$$W = \sum_{i=1}^g (n_i - 1) S_i \quad (3)$$

Los valores de λ que maximizan λ se pueden encontrar por medio de los vectores propios $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_s$ asociados con los valores propios positivos $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_s$ de $W^{-1}B$.

Dónde: $s = \min \{p, g - 1\}$,

De esta manera si $\hat{a} = \hat{e}_1$ entonces \hat{a} se le denomina primer discriminante lineal (LD1), si $\hat{a} = \hat{e}_2$ entonces \hat{a} se le denomina segundo discriminante lineal (LD2) y así hasta $\hat{a} = \hat{e}_s$, en cuyo caso \hat{a} se denomina s-esimo discriminante lineal (LDs).

La regla para clasificar una nueva observación x es asignar a la población o grupo i si se cumple que:

$$\sum_{j=1}^r [\hat{a}'_j(x - \bar{x}_i)]^2 \text{ es mínimo. Donde } r \leq s.$$

2.2.6. Brain Computer Interface – Interfaz Cerebro Computador BCI

La definición de un BCI es: “A brain computer interface is a communication system that does not depend on the brain’s normal output pathways of peripheral nerves and muscles” (Wolpaw & et al., junio, 2000). “La cual fue aprobada en el primer congreso internacional sobre tecnologías BCI, celebrado en junio de 1999 en Rensselaer ville en Nueva York”. Con esta definición podemos entender que una BCI es capaz de traducir los pensamientos de una persona mediante un dispositivo electrónico en acciones sin la intervención física.

Según Arafat (2015), en los últimos años la investigación en interfaces cerebro-máquina se ha extendido por muchos laboratorios de todo el mundo lo que confirma el interés a global en este tipo de tecnologías. Con cada nueva aplicación, la viabilidad y usabilidad de este tipo de interfaces se va asentando paulatinamente y permite esperar en el futuro grandes:

- **Uso extendido de BCIs a personas con discapacidades.** Los implantes de cóclea (oídos biónicos) son ya algo común. Desde hace unos años se viene experimentando también con ojos biónicos, consiguiendo paulatinamente mejores resultados y el control de brazos y piernas usando el pensamiento no es algo raro hoy en día. Para el año 2030 hay previsiones de que este tipo de implantes estén generalizados (Arafat, 2015).
- **Comprensión del funcionamiento del cerebro.** El estudio de esta parte del cuerpo humano está muy poco desarrollado todavía respecto a otras disciplinas. Sin embargo, se espera que, gracias al desarrollo tecnológico que se está experimentando, sea posible alcanzar un alto nivel de interpretación en

las próximas décadas. Esto ayudará a desarrollar interfaces cada vez más eficaces (Arafat, 2015).

Uno de los fines las BCI es dar una mejor condición de vida a las personas con discapacidades, mediante el uso de la tecnología; también sirven para el estudio más profundo de estas discapacidades, “el entrenamiento de interfaces y la evolución del usuario en el manejo del sistema pueden favorecer el estudio de enfermedades y discapacidades, además de proporcionar un análisis minucioso de las mismas y su evolución”. (Moran, 2015).

Las prestaciones de un BCI dependen de la adaptación de la persona al uso de estos dispositivos y puede controlar los patrones que emite su EEG. Para ello es necesario un adecuado entrenamiento que en algunos casos puede llegar a prolongarse durante meses. (Ron-Angevin, 2018).

Los dispositivos BCI siguen un patrón similar de funcionamiento para su estudio, el cual se presenta a continuación:

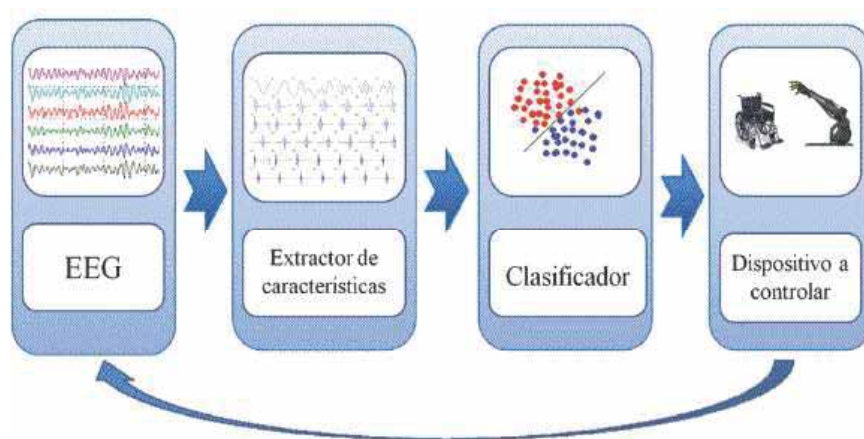


Figura 4 Interfaz Cerebro Computador. Fuente: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-97622013000200006

Registro de la señal. “Se selecciona uno de los varios métodos existentes para registrar la actividad electrofisiológica del cerebro. Éste debe ser seguro para

el usuario, así como entregar registros estables, duraderos y que revelen información útil” (Jaramillo Gonzales, Fundamentos de las interfaces cerebro maquina, 2016).

Procesamiento de la señal. “Se subdivide en dos etapas. En la primera se extraen los rasgos característicos de los registros vinculados con las variaciones en la actividad cognitiva. En la segunda se traducen y clasifican esos rasgos a comandos de control” (Jaramillo Gonzales, Fundamentos de las interfaces cerebro maquina, 2016).

Dispositivo por controlar. “Es el dispositivo al que se destinan los comandos de control para que ejecute la acción deseada por el usuario. Los dispositivos pueden ir de simples a complejos, según la cantidad y complejidad de los comandos de control necesarios” (Jaramillo Gonzales, 2016).

2.2.7. Electroencefalograma

Para medir las señales emitidas por el cerebro existen diferentes métodos y técnicas; todo depende de la colocación y de la profundidad de los electrodos que se ubican en la cabeza, para poder adquirir estas señales.

Electroencefalografía Profunda: “Se utilizan microelectrodos implantados en el interior del cerebro mediante una intervención quirúrgica” (Vaughan T., 1996).

- **Electrocorticograma (ECoG):** utiliza electrodos corticales colocados directamente sobre la corteza cerebral. En este caso también es necesaria una intervención quirúrgica para colocarlos electrodos.
- **Electroencefalograma (EEG):** utiliza electrodos superficiales, colocados sobre la superficie del cuero cabelludo. Es una técnica no invasiva.

Al realizar un registro de la actividad de EEG, podemos identificar una serie de ondas conocidas como ritmos cerebrales. Presentan dos características que las definen:

- **Frecuencia:** hace mención a la periodicidad con la que se repite una onda. Se mide en hercios (Hz). Si una onda presenta una frecuencia de 1 Hz, quiere decir que se repite 1 vez cada segundo.
- **Amplitud:** define la magnitud en el cambio de voltaje (μV) medido entre el punto más alto y el más bajo de la onda. A mayor amplitud, más energía se ha aplicado a la onda.

Por lo tanto, podemos decir que el registro del EEG está definido por una secuencia de oscilaciones en la diferencia de voltaje entre dos electrodos a través del tiempo.

Las ondas registradas en el EEG han sido clasificadas con base en su frecuencia. Las principales bandas de frecuencia, según (Boutros, Galderisi, Pogarell, & Riggio, 2011), son las siguientes:

- Beta (β). “Son ondas que se registran en un rango de frecuencia de 18 a 25 Hz. Se da en momentos de atención, de pánico o de resolución de problemas. Están presentes en regiones centrales y frontales, pero también se han observado en áreas posteriores” (Taywade & Raut, 2012).
- Alfa (α). “Las ondas que la componen tienen una frecuencia que oscila entre 8-12 Hz” (Taywade & Raut, 2012). “Son las ondas cerebrales más abundantes. Aparecen en estados de reposo y de ojos cerrados. Predominan en áreas posteriores y en zonas centrales y anteriores del cerebro” (Sanei & Chambers, 2007).

- Theta (θ). Presentan una frecuencia entre 4-7,5 Hz (Sanei & Chambers, 2007). Se originan en regiones del tálamo. Juegan un papel importante en la infancia, en estados de somnolencia y sueño, en fases 1 y 2 de ondas lentas y en la ejecución de procesos cognitivos.
- Delta (δ). - Se encuentran en un rango de 0,5 a 4 Hz. Aparecen en los primeros años de vida y va desapareciendo con el desarrollo del sistema nervioso. En el adulto, las encontramos durante las fases 3 y 4 del sueño de ondas lentas y en procesos de memorización (Sanei & Chambers, 2007).
- Gamma (γ). - se encuentra en un rango superior a los 30 Hz. Se piensa que reflejan el mecanismo subyacente de la conciencia. La detección de estos ritmos puede tomarse como confirmación de algunas enfermedades cerebrales (Taywade & Raut, 2012). Además, se han asociado con procesos de comunicación y sincronización entre las distintas zonas neuronales durante la realización de actividades mentales complejas.

Sistema de posicionamiento de electrodos

Aunque se conocen varios sistemas de posicionamiento, la disposición de los electrodos en el cuero cabelludo se ajusta a la recomendación, propuesta en 1958 por la Federación Internacional de Sociedades de Electroencefalografía y Neurofisiología Clínica, denominado sistema “diez-veinte” (10-20) y catalogado como el método de posicionamiento estándar (Taywade & Raut, 2012).

“Este sistema se basa en la relación entre la posición de un electrodo y el área subyacente de la corteza cerebral” (Rowan & Tolunsky, 2004). Para ello, “se emplea marcas craneales que son tomadas como puntos de referencia para la localización de electrodos” (Barea Navarro, 2009).

El 10 hace referencia a que las distancias reales entre electrodos adyacentes son el 10%, mientras que el 20 se refiere al 20% de la distancia total delantera/trasera o derecha/izquierda (Barea Navarro, 2009), de manera que la zona que registre cada electrodo sea comparable entre los registros realizados a diferentes individuos (Sanei & Chambers, 2007). Con esto se consigue que haya una división del cráneo y que podamos tomar como referencia algunos puntos para cubrir la totalidad de las regiones cerebrales (Teplan, 2002).

En conjunto, se pueden 19 electrodos superficiales distribuidos por la cabeza y 2 electrodos de referencia en los lóbulos de las orejas, aunque se admite el uso de muchos más electrodos (32, 64, 128, 256...) (Barea Navarro, 2009).

Cada sitio posee una letra para identificar el lóbulo y un número para definir la posición del hemisferio. Por lo tanto, tendríamos la siguiente nomenclatura: F (frontal), C (central), T (temporal), P (posterior) y O (occipital). Además, la “z” hace mención al electrodo que se coloca en la línea media del cráneo (Oostenveld & Praamstra, 2001).

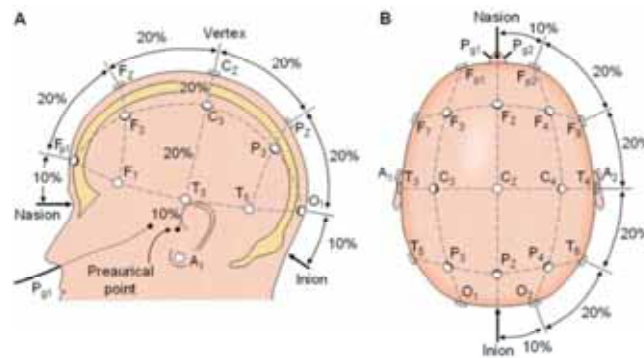


Figura 5 Sistema 10 – 20 para la toma de EEG. Fuente: https://www.researchgate.net/figure/282294960_fig2_Figura-3-Sistema-Internacional-10-20-para-la-colocacion-de-los-electrodos.

2.2.8. Transformadas de Wavelet

La transformada de Wavelet es una técnica matemática que se originó formalmente a principios de los años 80, aunque tuvo sus orígenes en el trabajo propuesto por Fourier en 1807, al utilizar la descomposición ortogonal de señales periódicas. La primera vez que se utilizó la expresión la transformada de Wavelet fue en 1909 en un apéndice de la tesis de A. Haar, donde se expone una alternativa ortogonal para la descomposición de señales, con la desventaja de no ser continuamente diferenciable, limitando sus aplicaciones (Burrus, 1999). En 1980, el físico Grossman y el ingeniero Morlet, desarrollan una función cuadrada integrable, mientras, que, en 1985, S. Mallat encuentra una relación entre los filtros de cuadratura, los algoritmos piramidales y las bases ortonormales.

La mayoría de trabajos enfocados hacia la reducción de ruido presente en señales bioeléctricas como ECG, EMG, EEG, entre otras, exponen unas condiciones fijas para la realización de filtros Wavelet, pero no presentan una metodología clara a la hora de aplicar y evaluar la DWT (Romaniuk, 2000), (Romaniuk, 2001), (Novak, 2000), (Olkkonen, 2002), (Hippenstiel, 2002), (Gotchev, 2003), (Quian, 2002). El proceso de filtrado o eliminación de ruido de fondo, aplicando la transformada de Wavelet, comprende tres partes: cálculo de los coeficientes de Wavelet, umbralización, y reconstrucción de la señal a partir de los coeficientes umbralizados (Burrus, 1999).

Básicamente, lo que hace la transformada de Wavelet es filtrar una señal en el dominio del tiempo mediante filtros paso bajo y paso alto que eliminan ciertas componentes de alta o baja frecuencia de la señal, el procedimiento se repite para las señales resultantes del proceso de filtrado anterior. Por ejemplo, supóngase que se tiene una señal con frecuencias de hasta 1000 Hz, en la primera

etapa de filtrado la señal es dividida en dos partes haciéndola pasar a través de un filtro paso-bajo y un filtro paso-alto con lo cual se obtienen dos versiones diferentes de la misma señal: una que corresponde a las frecuencias entre 0 y 500 Hz. (paso bajo) y otra que corresponde a las frecuencias entre 500-1000 Hz. (paso alto). Posteriormente, se toma cualquiera de las dos versiones (comúnmente la parte del filtro paso bajo) o ambas y se hace nuevamente la misma división (Saltos Torres, 2014). Esta operación se denomina descomposición. (Burrus & Ramesh, 1999).

De esta forma y suponiendo que se ha tomado la parte de la señal correspondiente al filtro paso bajo se tendrían tres conjuntos de datos, cada uno de los cuales corresponde a la misma señal, pero a distintas frecuencias: 0-250 Hz., 250-500 Hz. y 500-1000 Hz (Coomonte Belmonte, 2006). A continuación, se vuelve a tomar la señal correspondiente a la parte del filtrado de paso bajo haciéndola pasar nuevamente por los filtros paso bajo y paso alto, de esta forma ya se tendrían 4 conjuntos de señales correspondientes a las frecuencias 0-125 Hz., 125-250 Hz., 250-500 Hz. y 500-1000Hz. El proceso continúa hasta que la señal se ha descompuesto en un cierto número de niveles predefinidos. Finalmente se cuenta con un grupo de señales que representan la misma señal, pero correspondientes a diferentes bandas de frecuencia.

Con la transformada de Wavelet las altas frecuencias tienen mejor resolución en el tiempo mientras que las bajas frecuencias tienen mejor resolución en el dominio de la frecuencia. Esto significa que una determinada componente de alta frecuencia puede localizarse mejor en el tiempo (con menor error relativo) que una componente de baja frecuencia. Por el contrario, una componente de baja

frecuencia puede localizarse mejor en frecuencia comparado con una componente de alta frecuencia.

La transformada de Wavelet Continua (CWT) fue desarrollada como una técnica alternativa a la STFT como una manera de superar el problema de resolución (Andrade Mora & Navarrete Mejía, 2006). El análisis de la transformada de Wavelet se realiza de manera similar al análisis STFT, en el sentido que la señal es multiplicada por una función (función Wavelet) de manera similar a la función ventana en la STFT, y la transformada se calcula separadamente para distintos segmentos de la señal en el dominio del tiempo (Martinez Aldana, 2009). Sin embargo, existen dos diferencias principales entre la STFT y la CWT:

- a) No se evalúa la transformada de Fourier de las señales ventana y por lo tanto aparecerá un único pico que corresponde a una senoide.
- b) El ancho de la ventana varía a medida que se evalúa la transformada para cada componente del espectro, esto es probablemente la característica más significativa de la transformada Wavelet.

La transformada Wavelet continua se define como sigue:

$$C(\tau, s) = \int_{-\infty}^{\infty} f(t) \psi_{\tau, s}^*(t) dt \quad (1)$$

Donde:

$$\psi_{\tau, s}(t) = \frac{1}{\sqrt{|s|}} \psi\left(\frac{t - \tau}{s}\right) \quad (2)$$

Como se observa en la ecuación anterior, la señal transformada es una función de dos variables, τ y s , los parámetros de traslación y escala respectivamente. $\psi_{\tau, s}(t)$ es la función de transformación que se le denomina

“wavelet madre”, este nombre deriva de dos importantes propiedades del análisis Wavelet:

- El término Wavelet significa “onda pequeña”. La pequeñez se refiere al hecho que esta función (ventana) es de longitud finita (compactamente soportada) y el término onda se refiere a la condición que esta función es de naturaleza oscilatoria. El término madre da a entender que las funciones con diferentes regiones de actuación que se usan en el proceso de transformación provienen de una función principal o wavelet madre. Es decir, la wavelet madre es un prototipo para generar las otras funciones ventanas (Valenzuela Ramírez, 2015).

Traslación	Cambio de escala	Traslación y cambio de escala
$\psi(t-b)$	$\frac{1}{\sqrt{a}} \psi\left(\frac{t}{a}\right)$	$\frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$

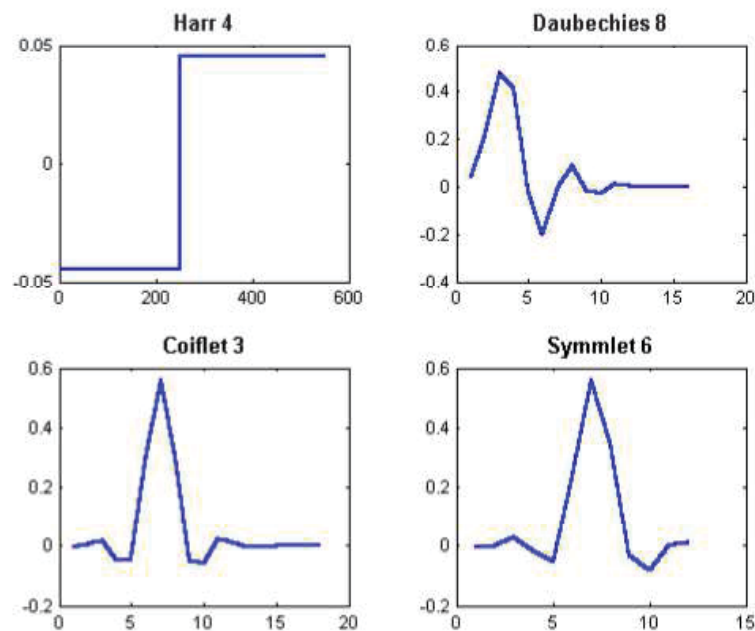


Figura 6 Ejemplo de Wavelet madre. Fuente: <ftp://ftp.unicauca.edu.co/Facultades/.../proyecto/wavelet/serrano.pdf>

Traslación. El término traslación se usa con el mismo sentido que fue usado en la STFT y está relacionado con la localización de la ventana a medida que ésta se desplaza a través de la señal. Obviamente, este término corresponde a la información del tiempo en el dominio transformado. Sin embargo, no se tiene un parámetro que sea la frecuencia como si se tenía antes en el caso de la STFT. En el caso de la transformada de Wavelet se tiene un parámetro de “escala” el que se define como:

$$\text{Escala} = \frac{1}{\text{frecuencia}}$$

Escala. En el análisis de Wavelet el parámetro escala es análogo con el parámetro escala utilizado en los mapas. Tal como en este último caso, las altas escalas corresponden a una visión global no detallada (de la señal) y las bajas escalas corresponden a una vista detallada (Arango Flores, 2007). De igual manera, en términos de frecuencia, las bajas frecuencias (altas escalas) corresponden a una información global de la señal que comúnmente abarca toda la señal, mientras que las altas frecuencias (escalas bajas) corresponden a una información detallada de una característica oculta en la señal que comúnmente dura un tiempo relativamente pequeño (Juarez Segura, 2005).

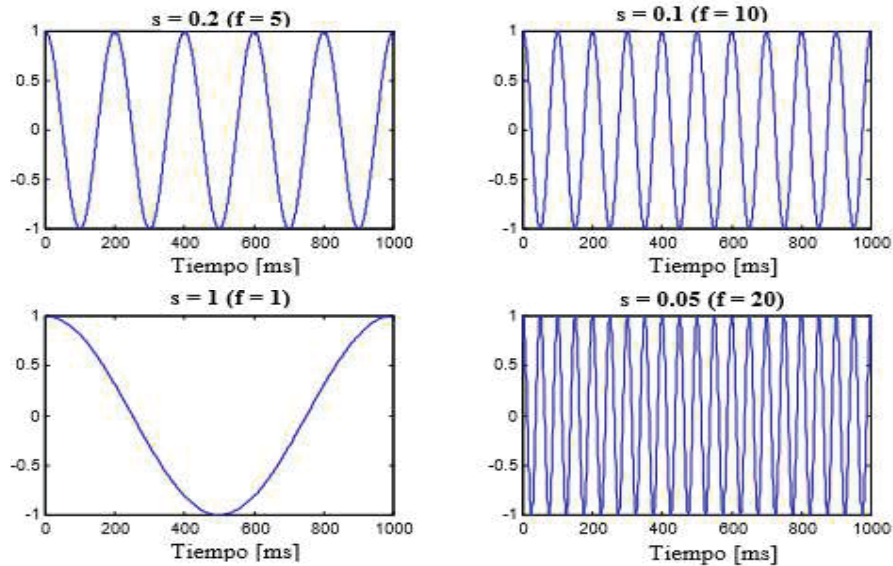


Figura 7 Función coseno en diferentes escalas. Fuente: <ftp://ftp.unicauca.edu.co/Facultades/.../proyecto/wavelet/serrano.pdf>

Conjunción traslación y escala. En señales que corresponden a fenómenos o aplicaciones reales las escalas bajas (altas frecuencias) no tiene una larga duración en la señal, sino que aparecen de tiempo en tiempo como picos o “spikes”. Sin embargo, las altas escalas (bajas frecuencias) comúnmente duran todo el tiempo de duración de la señal.

El escalamiento como operación matemática produce una dilatación o una compresión de una señal; las altas escalas corresponderán a señales dilatadas y las escalas pequeñas corresponden a señales comprimidas. Todas las señales mostradas en la figura nacen de la misma señal coseno, es decir son versiones comprimidas o dilatadas de la misma función. En la figura anterior, para $s = 0.05$ se tiene la menor escala y para $s = 1$ la mayor.

La relación entre la escala y la frecuencia consiste en que las escalas menores corresponden a altas frecuencias y las escalas mayores corresponden a bajas frecuencias (Macri & Benedetti, 2012).

Debido a que la WT incluye información relacionada con el tiempo y la frecuencia, la representación gráfica de esta transformada se realiza en un plano denominado plano tiempo-escala, representado en la Fig. 2.4. Cada celda en esta figura representa un valor de la WT en dicho plano. Es de destacar el hecho que estas celdas tienen un área no nula, lo cual indica que no es posible conocer el valor de un punto particular. Sin tener en cuenta las dimensiones de las celdas, sus áreas, tanto en la STFT como en la WT, son las mismas, y están determinadas por el principio de incertidumbre de Heisenberg.

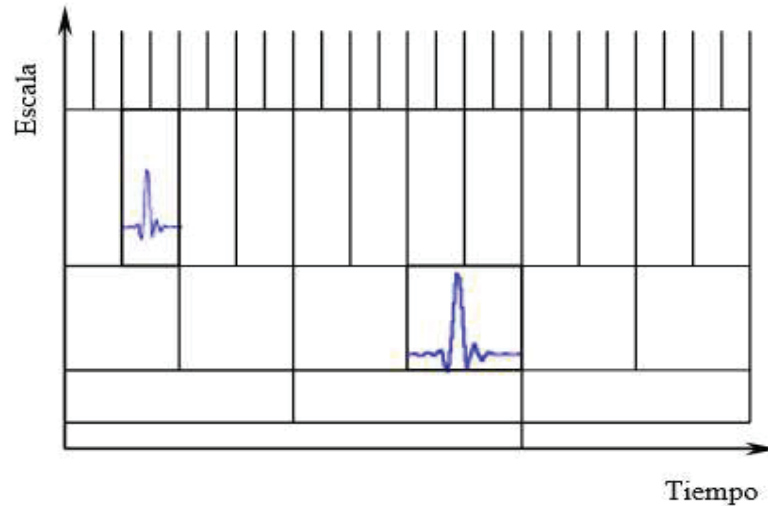


Figura 8 Traslación y escalado. Fuente: <ftp://ftp.unicauca.edu.co/Facultades/.../proyecto/wavelet/serrano.pdf>

Tipos de Transformadas Wavelet

Existen tres tipos de transformada de Wavelet: continua (CWT), semidiscreta (SWT) y discreta (DWT). La diferencia entre ellas radica principalmente en la forma en que los parámetros de desplazamiento y escala son discretizados. A continuación, se describen brevemente estos tres tipos.

Transformada Wavelet continua. (CWT). En este caso los parámetros cambian de forma continua. Esta representación ofrece la máxima libertad en la elección de la wavelet, con la única restricción que satisfaga la condición de media nula. Esta condición permite que la CWT sea invertible en rango. La transformada inversa viene dada por:

$$f(t) = \frac{1}{K_\psi} \iint C(\tau, s) \frac{\psi(\tau, s)}{\tau^2} \cdot d\tau ds$$

Donde ψ satisface la condición de media nula, con K_ψ dada:

$$K_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(w)|^2}{w} \cdot dw < \infty$$

siendo $\Psi(w)$ la transformada Fourier de $\psi(t)$.

El proceso de cálculo de la CWT puede ser descrito en:

1. Tomar una Wavelet madre.
2. Dados dos valores τ y s , calcular un coeficiente $C(\tau, s)$ mediante las ecuaciones (1) y (2), que represente la correlación entre la Wavelet y la sección de la señal bajo análisis. Cuanto mayor sea éste, mayor es la similitud, con lo cual es interesante resaltar que los resultados dependerán por tanto de la forma de la wavelet.
3. Desplazar la Wavelet en el sentido positivo del eje temporal, y repetir los pasos anteriores hasta que se haya cubierto la totalidad de la señal.

4. Escalar la Wavelet en el tiempo, y repetir los pasos 1 a 3.

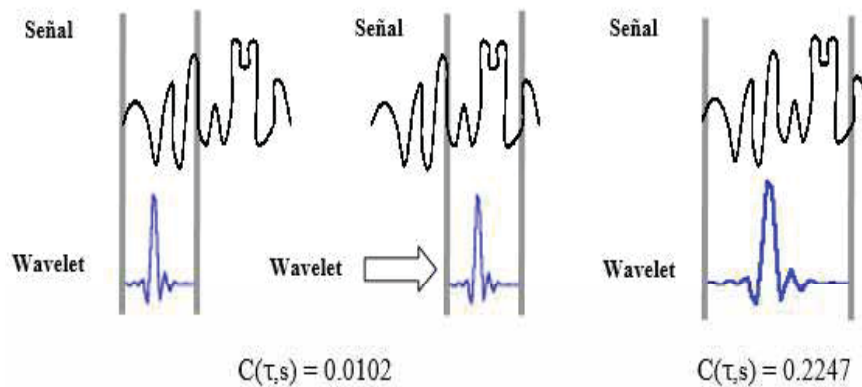


Figura 9 Transformada de Wavelet continua. Fuente: <ftp://ftp.unicauca.edu.co/Facultades/.../proyecto/wavelet/serrano.pdf>

Wavelet semi discreta (SWT). En la práctica, es más conveniente considerar la WT en algunos valores discretos de a y b . Por ejemplo, la escala diádica corresponde a la definición de los parámetros $a = 2^j$, $b = 2^j k$, con $(j, k) \in \mathbb{Z}^2$ denominándose transformada de Wavelet semidiscreta (SWT).

La transformada será reversible si cumple:

$$A\|f\|^2 \leq \sum_{a,b} |\langle f, \psi(\tau, s) \rangle|^2 \leq B\|f\|^2$$

donde A y B son dos constantes positivas y $f(t)$ sigue siendo una función continua.

Wavelet Discreta (DW). Sea la señal a analizar $f[n]$ una función discreta. En este caso la transformada de Wavelet de esta señal viene dada por:

$$C[j, k] = \sum_{n \in \mathbb{Z}} f[n] \psi_{j,k}[n]$$

donde $\psi_{j,k}$ es una Wavelet discreta definida como:

$$\Psi_{j,k}[n] = 2^{-\frac{j}{2}} \cdot \psi[2^{-j}n - k]$$

Los parámetros τ , s están definidos según la escala diádica, de manera que $\tau = 2^j$, $s = 2^j k$, con la diferencia respecto a la SWT que la señal bajo estudio es discreta (De Castro Fernandez, 2002). La transformada inversa se define de forma similar como:

$$f[n] = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C[j, k] \cdot \Psi_{j,k}[n]$$

Si las constantes indicadas son $A = B = 1$, entonces la transformada es ortogonal. Estas wavelets pueden ser construidas mediante un análisis multiresolución.

Por otra parte, realizado el análisis, es posible reconstruir a discreción las componentes relevantes de la señal y caracterizar así diversos fenómenos de interés. En la sección siguiente, volveremos sobre esta descomposición, profundizando su significado.

A diferencia de las Series de Fourier Locales, al análisis se realiza por octavas o rangos de frecuencia que duplican su dimensión hacia las altas frecuencias, a la vez que se reduce el rango temporal de localización. Es posible entonces localizar por medio del espectro, tanto fenómenos locales como patrones de autosimilaridad, a distintas escalas.

Más aún, el apropiado truncamiento de las series de Wavelets, realizado en cada nivel j , no desnaturaliza o destruye la señal. En suma, con un espectro finito podemos representar eficientemente la textura y fenomenología temporal de la señal, clasificada por octavas.

Por supuesto, existen también desventajas. Por un lado, la discretización no conserva ciertas importantes propiedades de la Transformada de Wavelet Continua. Particularmente, la de invariancia respecto de las traslaciones.

Serán el tipo de transformada a utilizar en este trabajo, debido a su facilidad de implementación en ordenadores.

2.3. Antecedentes empíricos de la investigación

Oikonomou Vangelis P. et al. (2017), “*A comparison study on EEG signal processing techniques using motor imaginary EEG data*”. Trabajo presentado en el Instituto de Ingeniería Eléctrica y Electrónica (IEEE).

Conclusiones:

Se realizó una comparación sobre la capacidad discriminativa entre las características relacionadas con CSP (patrones espaciales comunes) y las características espectrales. Nuestros resultados indican que las características espectrales de la señal son superiores a las características CSP (densidad espectral de potencia) si se utiliza el clasificador apropiado. El uso de métodos de estimación de PSD nos proporciona un conjunto de características de alta dimensionalidad. El clasificador debe tener en cuenta la propiedad anterior para lograr su mayor rendimiento. Este hecho es evidente cuando comparamos el rendimiento entre el LDA (análisis discriminante lineal) y el SVM (máquina de soporte de vectores) en el caso en el que el conjunto de características de entrada es el PSD.

Comentario:

De este documento se considera los resultados de los algoritmos de clasificación LDA y SVM para compararlos con el LDA y SMO de nuestro estudio.

Según Yue Zang. et al. (2017), “*Extracting Error-Related Potentials from Motion Imagination EEG in Noninvasive Brain-Computer Interface*”. Trabajo presentado en el Instituto de Ingeniería Eléctrica y Electrónica (IEEE).

Conclusiones:

En este documento, mostramos un paradigma experimental con retroalimentación visual para el MI (movimiento imaginario) de la mano izquierda y derecha y completamos dos sesiones en línea. Luego, la señal de EEG (electroencefalograma) sin procesar, se clasificó usando el clasificador LDA (análisis discriminante lineal) después del preprocesamiento. Nuestra contribución es doble en la parte de la mano izquierda-derecha MI. Primero, en el dominio del tiempo, se estudió la clasificación de la imaginación del movimiento de la mano izquierda y derecha, la precisión promedio fue de 68.72% y se dibujó la curva AUC (área bajo la curva). Segundo, en el dominio de la frecuencia, se mostraron el análisis de poder discriminante y las topografías del cuero cabelludo que muestran las modulaciones del patrón cerebral de la sesión en línea. Los resultados muestran que las características se concentran principalmente en la banda de frecuencia Mu (8-13Hz) y Beta (14-30Hz) en los canales FC3, FC4, C3, C4, CP3 y CP4. Además, diseñamos un paradigma experimental de muestra para encontrar los potenciales relacionados con el error y los ErrPs (error potencial). Los resultados del análisis mostraron que la curva de ErrPs se extrajo con éxito, que tiene una cresta aparente entre 300 ms y 400 ms con un valor de 4uV aproximadamente.

Comentario:

En esta investigación se utilizan los mismos pasos para el proceso de análisis de las señales de electroencefalogramas y los resultados de son comparados con el algoritmo de clasificación LDA.

Según Yuhan ye et al. (2017) “*The Design of Multi-task Simulation Manipulator Based on Motor Imagery EEG*”. Trabajo presentado en el Instituto de Ingeniería Eléctrica y Electrónica (IEEE).

Conclusiones:

Para abordar el problema de las acciones de control limitadas y la baja precisión en el manipulador controlado por EEG (electroencefalograma), se desarrolla un manipulador multitarea basado en la combinación del algoritmo EEG de imágenes motora y el mecanismo de activación / desactivación. En el diseño de BCI (interfaz cerebro computador), se utiliza un algoritmo de clasificación de dos categorías, que incluye el filtro de paso de multibanda, CSP (patrón espacial común) para la extracción de características, información mutua para la selección de características y LDA (análisis discriminante lineal) para la clasificación. Como resultado, las señales de EEG se clasificaron en dos categorías, imágenes motoras a la izquierda y a la derecha, y la precisión de la prueba es de alrededor del 70%, lo que garantiza que podamos controlar el manipulador con un bajo índice de errores. Con el EEG traducido a los comandos de la operación de activación / desactivación, se pueden utilizar dos tipos de señales de control para cambiar y seleccionar diferentes tareas, y posteriormente, el manipulador realizará las tareas correspondientes. En la prueba de eficiencia, encontramos que el sistema de manipulación de tareas múltiples con mecanismo de activación / desactivación tiene un requerimiento de tiempo real bajo, y el tiempo empleado en la selección de tareas es aceptable cuando el número de tareas no es demasiado. Entonces, no hay duda de que tiene dos ventajas importantes:

- 1) Debido a que la tecnología actual solo puede distinguir algunos tipos de señales de EEG, el manipulador no puede obtener suficientes señales de control para realizar tareas

complejas. Para abordar este problema, el mecanismo de activación / desactivación puede usarse para seleccionar más tareas solo por dos tipos de señales de EEG.

2) La mejora del tipo de reconocimiento llevará a la reducción de la precisión del reconocimiento EEG. Por lo tanto, el manipulador multitarea con control EEG directo enfrentará el problema de una alta tasa de error, lo que hace que el manipulador no pueda completar una serie de acciones. En aplicaciones prácticas, el error del manipulador probablemente causará muchos problemas.

En este documento, el manipulador multitarea que proponemos es un manipulador controlado indirectamente a través del mecanismo de activación / desactivación. Inevitablemente sufre las deficiencias de la inflexibilidad, la falta de libertad y el largo tiempo de respuesta. Por lo tanto, el control directo de la señal de EEG sigue siendo el foco de la investigación de BCI.

Comentario:

En investigación sigue los mismos pasos para el análisis de señales de electroencefalogramas planteados en esta investigación, la diferencia radica en el uso de CPS en la obtención de vectores característicos y en nuestro estudio se utiliza la transformada discreta de Wavelet, los resultados de la clasificación del LDA se comparan con nuestro estudio.

CAPÍTULO III

METODOLOGÍA

4.1. Ámbito de estudio: localización política y geográfica

La base de datos se encuentra en un repositorio público a nivel mundial y todo investigador puede acceder a esta para poder realizar estudios utilizando diferentes métodos para mejorar o comparar sus resultados.

4.2. Tipo y nivel de la Investigación

El tipo de investigación es descriptivo, según Hernández Sampieri et al. (2014) “se busca especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis.”

El diseño de investigación es no experimental transversal, según Hernandez Sampieri et al. (2014) “Recolectan datos en un solo momento, en un tiempo único. Su Propósito es describir variables y analizar su incidencia e interrelación en un momento dado.”

4.3. Unidad de Análisis

Las unidades de análisis para la presente investigación están definidas en los datos de las señales de electroencefalogramas de los electrodos C3, Cz y C4 relacionadas a las señales sensorio-motoras.

4.4. Técnicas de recolección de información

El conjunto de datos fue obtenido del II BCI Competition. El archivo descargado está compuesto por sesiones que tienen una duración de 9 segundos cada una y grabadas a 128Hz. Se tiene 280 sesiones, de las cuales 140 se utilizan para entrenamiento y 140 para prueba.

4.5. Técnicas de análisis e interpretación de la información

Para el análisis de la información de los algoritmos de clasificación se utilizan dos fases. La primera es el entrenamiento en la cual la maquina aprende a diferenciar los movimientos hacia la derecha o hacia la izquierda. Y la segunda fase de prueba en la cual a partir del aprendizaje realizado en la primera fase la maquina ya puede distinguir y clasificar los movimientos imaginarios. Para el análisis e interpretación de la información se utiliza la matriz de confusión generada en la etapa de prueba, de esta matriz se realizan los cálculos de precisión, exhaustividad y exactitud y la curva ROC. Con estos resultados se realiza la comparación de resultados de cada algoritmo de clasificación y se determina cual es el más eficaz.

4.6. Técnicas para medir la eficacia de los algoritmos de clasificación

Existen diferentes métricas para poder para poder determinar si la clasificación cumple el objetivo indicado. Una de ellas es la matriz de confusión.

Tabla 1: Matriz de confusión

Clase Real	Clase Predictiva	
	Predicción Positiva	Predicción Negativa
Clase Positiva	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Clase Negativa	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Fuente: Propia

VP (Verdaderos positivos): instancias correctamente reconocidas por el sistema.

FN (Falsos negativos): instancias que son positivas y que el sistema no las reconoce como tales.

FP (Falsos positivos): instancias que son negativas pero el sistema dice que no lo es.

VN (Verdaderos negativos): instancias que son negativas y correctamente reconocidas como tales.

Donde N es el número del conjunto de datos de entrenamiento, además está dada por la siguiente formula:

El número de instancias clasificadas correctamente es la suma de la diagonal de la matriz y el resto están clasificadas de forma incorrecta (Corso, 2009).

Una obtenida la matriz de confusión podemos trabajar con las siguientes métricas: Precisión, cobertura, Exactitud, Valor – F y G – Media.

Precisión: La precisión, en cambio, mide el número de términos correctamente reconocidos respecto al total de términos predichos, sean estos verdaderos o falsos términos. En este caso, la precisión está midiendo la pureza o el grado en que son todos los que están.

Cobertura (Recall): la cobertura mide la proporción de términos correctamente reconocidos respecto al total de términos reales, dicho de otro modo, mide en qué grado están todos los que son:

La cobertura y la precisión mantienen una relación inversa, es decir cuando aumenta la cobertura del modelo generado disminuye la precisión y viceversa cuando disminuye la cobertura aumenta la precisión

Exactitud (Accuracy): Es el porcentaje de aciertos del modelo.

Curvas ROC (Receiver Operating Characteristic): Es una representación gráfica de la **sensibilidad** de los verdaderos positivos ($VP/(VP+FN)$) vs. la **especificidad** de los falsos positivos ($FP/(FP+TN)$).

“El punto (0,1) se llama una clasificación perfecta. La línea diagonal que divide el espacio de la ROC en áreas de la clasificación buena o mala” (Martinez Romero, 2010). “Los puntos por encima de la línea diagonal indican buenos resultados de clasificación, mientras que los puntos por debajo de la línea indican resultados equivocados (aunque el método de predicción puede simplemente ser invertido para conseguir puntos por encima de la línea)” (Santamaria Ruiz, 2010).

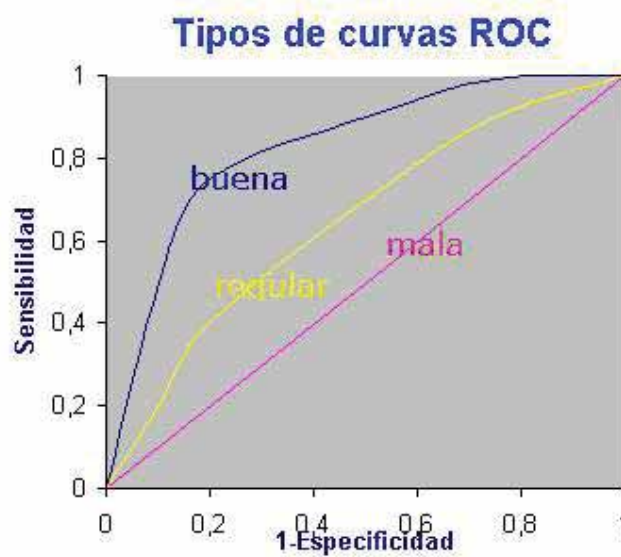


Figura 10 Tipos de curvas ROC. Fuente: http://www.hrc.es/bioest/roc_1.html

CAPÍTULO IV

DESARROLLO DEL PROYECTO

4.1. Desarrollo del proyecto

Para el procesamiento, análisis, interpretación y discusión de resultados se siguió los pasos de la figura 12, donde se muestra la secuencia del procesamiento de señales de electroencefalogramas.

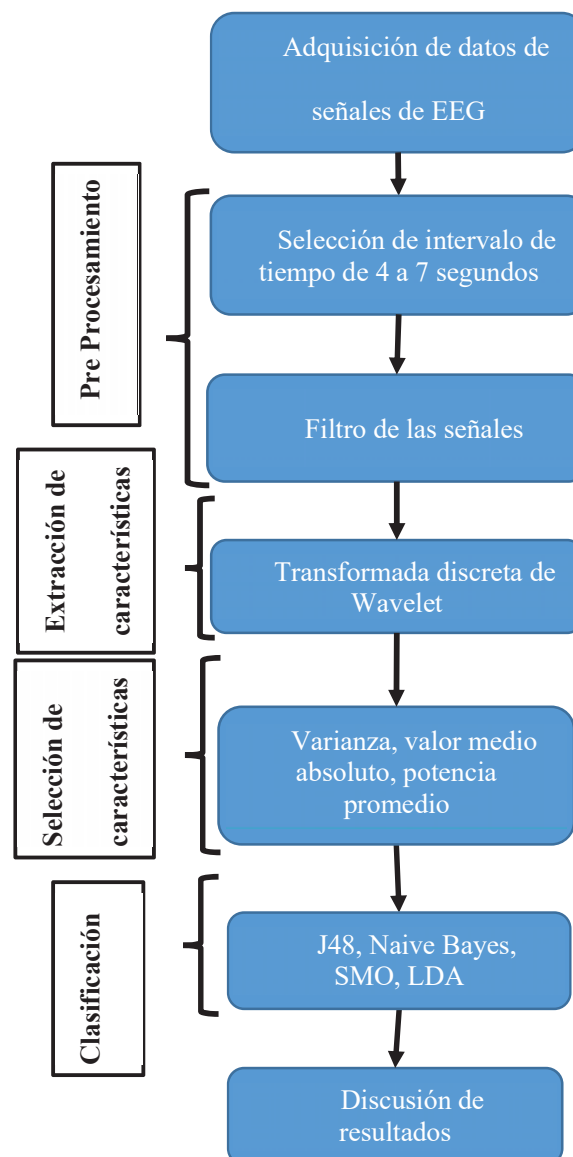


Figura 11 Esquema de desarrollo del proyecto de investigación. Fuente: propia

4.1.1. Adquisición de datos

Los datos fueron obtenidos del sitio Web “Berlin Brain-Computer Interface” de Alemania. Estos datos fueron usados en el concurso de BCI Competition II también llamado BCI Competition 2003, data set III y tiene la siguiente descripción:

El conjunto de datos fue proporcionado por el Departamento de Informática Medica, Instituto de Ingeniería Biomédica, de la universidad de tecnología de Graz (Austria).

El conjunto de datos fue grabado a una mujer de 25 años. El sujeto se encuentra sentado en una silla con reposabrazos. El experimento consistía en moverse un cursor e imaginar su movimiento hacia la derecha o izquierda. La aparición del cursor de movimiento era aleatorio.

El experimento consiste en 7 ejecuciones con 40 ensayos cada uno. Cada ensayo se realizó el mismo día con minutos de diferencia entre ellos. Se obtuvieron 280 ensayos con una duración de 9 segundos cada uno.

En cada sesión se tuvieron 2 segundos donde no se mostraba nada en la pantalla; luego un pitido indicaba que la sesión iba a empezar y durante un segundo mostraba una cruz “+”. Entre el segundo tres y nueve se muestra una flecha hacia la izquierda o derecha, en este momento el sujeto del experimento debe imaginar el movimiento de la mano correspondiente a la flecha (clase 1 izquierda, clase 2 derecha). Figura Nro. 13.

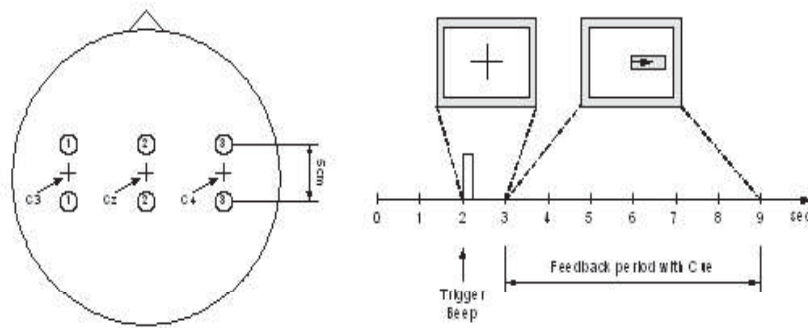


Figura 12 Esquema de captura de datos EEG. Fuente: <http://www.bci.de/competition/ii/>

La grabación de los EEG fue realizada utilizando un amplificador G-tec y electrodos Ag/AgCl. Los electrodos se colocaron en los canales C3, Cz y C4 y fueron grabados a 128 Hz. Estos datos fueron filtrados entre 0.5 y 30 Hz. (BCI Competition II, 2003), (Pfurtscheller, 2003).

4.1.2. Pre procesamiento de datos

El conjunto de datos del BCI Competition se encuentra en formato del programa MatLab con el nombre de dataset_BCIcomp1.mat; cuenta con tablas distribuidas de la siguiente manera:

- x_train con 140 sesiones de 9 segundos cada uno para entrenamiento.
- y_train con 140 etiquetas correspondientes a cada sesión de entrenamiento.
- x_test con 140 sesiones correspondientes al test.

Cada uno de las sesiones tiene 1152 registros correspondientes a los 9 segundos de grabación a 128Hz.

Para el experimento se consideró únicamente los electrodos C3 y C4 que son los más representativos en cuanto corresponde a imaginación motora de ambas manos.

Una vez cargado los valores del conjunto de datos en MatLab, se muestra los canales C3 y C4 de la señal del electroencefalograma con los datos obtenidos en cada una de las sesiones (figura 14).

C3	C4
0.0103	0.0625
-0.0239	0.0342
-0.0098	0.0659
0.0488	0.1230
-0.0381	0.0669
-0.0815	-0.0151
-0.0957	-0.0718
-0.0708	-0.0527
-0.0967	-0.0854
-0.1528	-0.1050
-0.1812	-0.0942
-0.1411	-0.0718
-0.0693	-0.0542
-0.0547	-0.0801
-0.0112	-0.0239
0.0073	-0.0571
-0.0327	-0.0303
-0.0317	-0.0254
-0.0083	0.0103
-0.0132	0.0005
-0.0159	-0.0151

Figura 13 Datos de los canales C3 y C4 de a sesión 1. Fuente: Propia

Utilizando el Signar Analyzer de MatLab se busca determinar de manera visual los tiempos en los que hay mayor influencia de las señales para saber cuál es la intención de movimiento imaginario.

Para ello se abre el archivo que contiene las 144 sesiones de electroencefalogramas y se especifica la frecuencia y el tiempo de grabado de estas. (Figura 15)

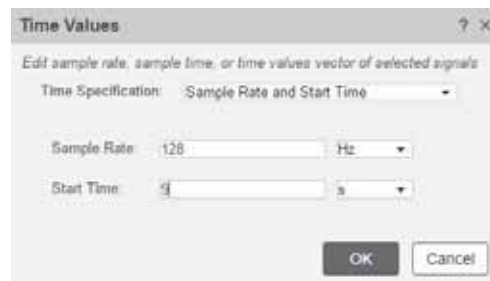


Figura 14 Determinando la frecuencia y tiempo de la señal. Fuente: Propia

Una vez abierto el archivo se procede a ver la gráfica de la señal, con su respectivo espectrograma. Y se nota que los colores más intensos son en la parte del medio, por lo que se decide tomar el intervalo de entre 4 y 7 segundos de la señal para este estudio.

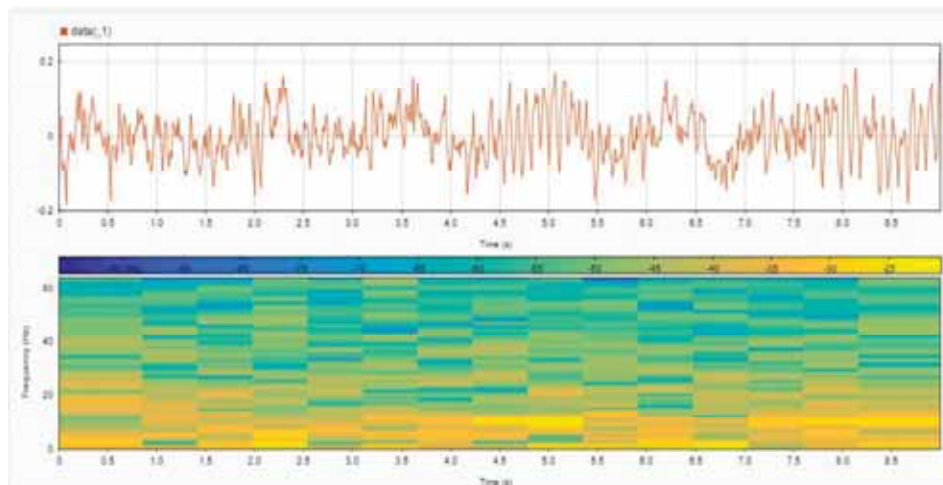


Figura 15 Señales Canal C3. Fuente: Propia

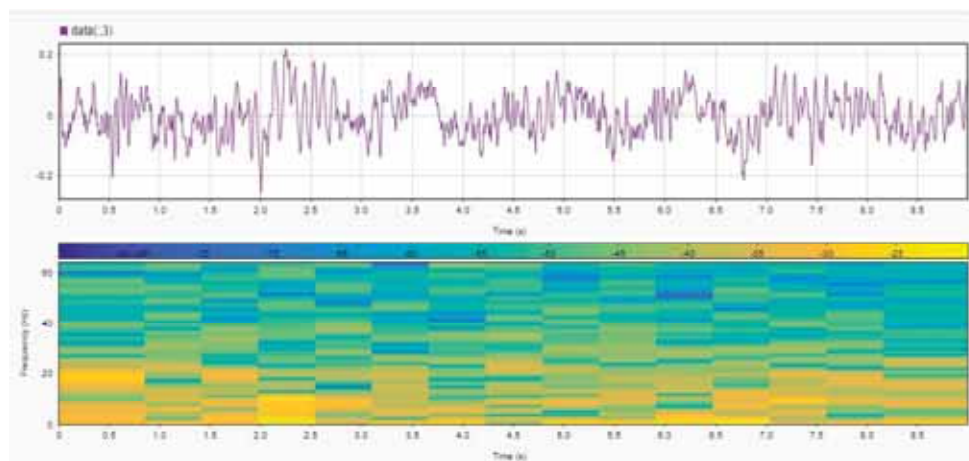


Figura 16 Señales del Canal C4. Fuente: Propia

Después se visualiza la señal en este intervalo de 4 a 7 segundos en los canales C3 y C4, (Figura 18 y 19) Donde se visualiza en la parte superior la señal entre 4 y 7 segundos, en la parte del medio la frecuencia a la cual se va a trabajar la señal; en este caso a 64 Hz. Esta frecuencia se utiliza en la transformada de Discreta de

Wavelet. En la parte inferior se visualiza el total de la señal con la parte resaltada del intervalo de 4 a 7 segundos

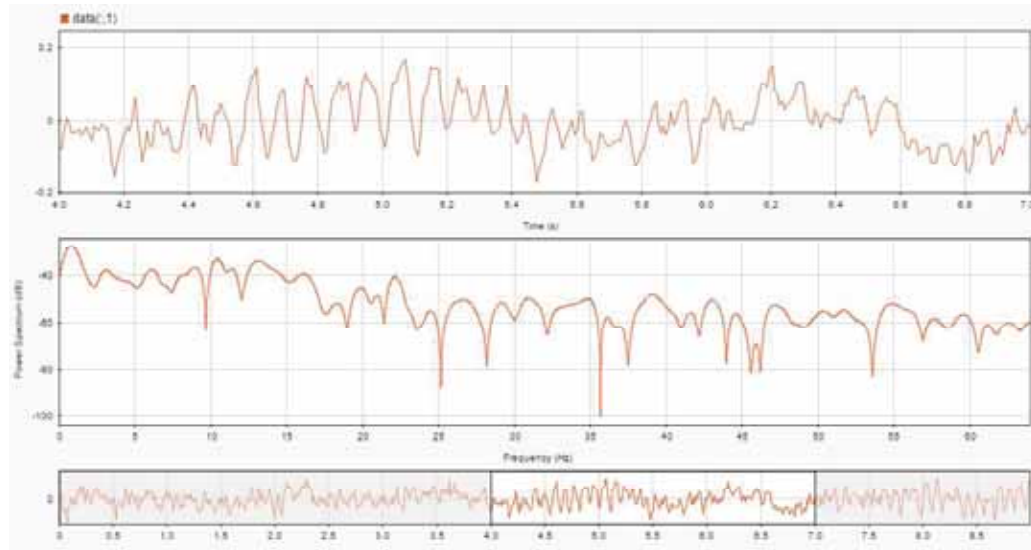


Figura 17 Canal C3 en el intervalo de 4 a 7 segundos. Fuente: Propia

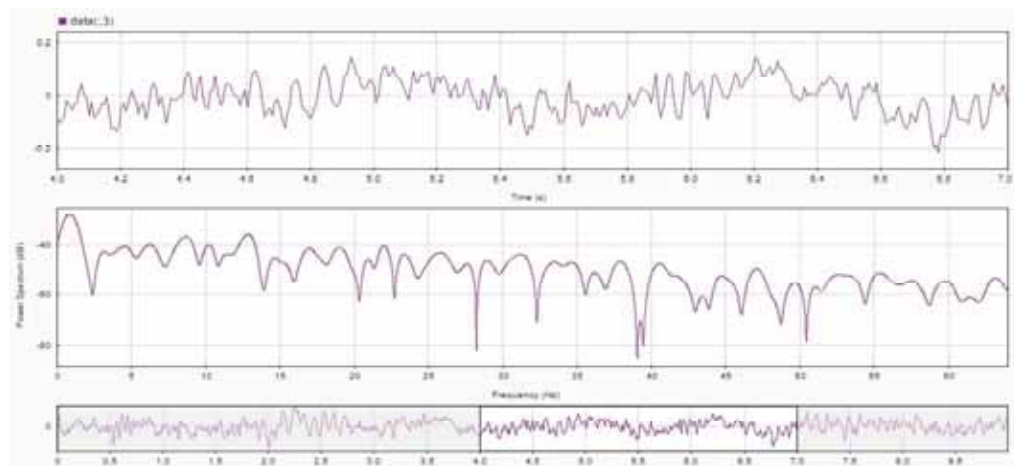


Figura 18 Canal C4 en el intervalo de 4 a 7 segundos. Fuente: Propia

4.1.3. Extracción de características

Al visualizar como cambia el espectro de onda se comprobó que la frecuencia se encontraba entre 0 y 64 Hz, para luego aplicar la Transformada Discreta de Wavelet; donde se han utilizado dos filtros de señal; pasa bajos y pasa altos, para descomponer la señal del electroencefalograma.

La salida del filtro pasa bajos corresponde a los coeficientes de aproximación y los datos de salida del filtro pasa altos corresponde a los coeficientes de los detalles. Cada señal de salida de la transformada de Wavelet tiene una representación para su ancho de banda en función frecuencia y tiempo. Por eso se considera que la Transformada Discreta Wavelet es temporo frecuencial.

Las bandas de frecuencias con las que trabajaremos se encuentran en la banda alfa (8 – 13 Hz) y en la banda beta (18-25 Hz). Entonces la distribución de frecuencias mediante la Transformada Discreta de Wavelet para nuestro estudio se muestra en la figura 23:

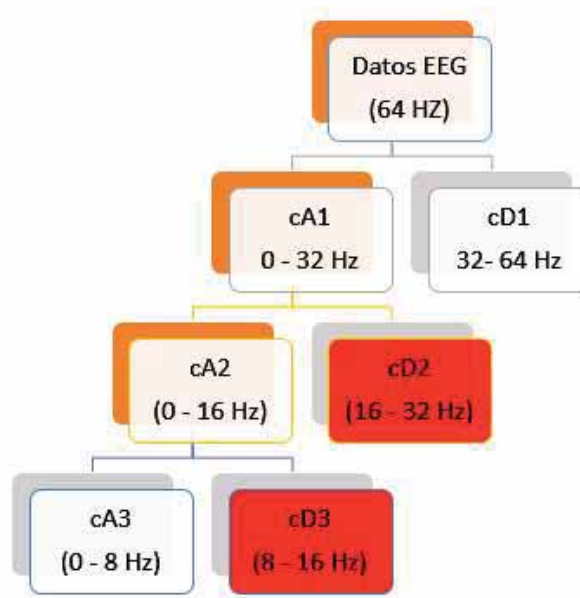


Figura 19 Descomposición de la señal EEG mediante la transformada discreta de Wavelet. Fuente: Propia

De la figura podemos ver que al descomponer la señal en tres niveles obtenemos los rangos específicos para la banda alfa y beta encargadas de los movimientos motores, la banda alfa se encuentra dentro de cD3 (coeficiente de detalle 3) y la banda beta se encuentra en cD2 (coeficiente de detalle 2). En a transformada Discreta de Wavelet se utiliza la familia Daubichies 3 y así obtener la descomposición y las aproximaciones de la señal de electroencefalograma.

4.1.4. Selección de características

Una vez obtenido los detalles cD2 y cD3 para los canales C3 y C4, se procede a extraer los vectores característicos con los siguientes parámetros estadísticos y sus respectivas combinaciones.

Varianza (V). La varianza es el cuadrado del valor medio de la desviación de la señal de cada canal y se define como:

$$V = \frac{1}{N-1} \sum_{n=1}^N |x[n]^2|$$

Valor Medio Absoluto (VMA). se define como la sumatoria de los valores absoluto de cada uno de los vectores canal entre el número de elemento del canal seleccionado y se denota por:

$$PA = \frac{1}{N} \sum_{n=1}^N x_n$$

Potencia Promedio (P). Se calculó como la normalización de cada vector elevado al cuadrado, dividida por la longitud del vector.

$$P_x = \frac{1}{L} \sum_{i=1}^L (x_i)^2$$

Obteniéndose 7 combinaciones y por consiguiente siete diferentes archivos almacenados con extensión .csv. Descritos de la siguiente manera

- **Potencia Promedio (P)** con 4 columnas de vectores característicos
- **Promedio Absoluto (MAV)** con 4 columnas de vectores característicos
- **Varianza (V)** con 4 columnas de vectores característicos

- **Potencia Promedio (P) – Promedio absoluto (MAV)** con 8 columnas de vectores característicos
- **Potencia Promedio (P) – Varianza (V)** con 8 columnas de vectores característicos
- **Promedio absoluto (MAV) – Varianza (V)** con 8 columnas de vectores característicos.
- **Potencia Promedio (P) - Promedio absoluto (MAV) – Varianza (V)** con 8 columnas de vectores característicos.

4.1.5. Clasificación

Una vez obtenido los vectores característicos, estos se proceden a unirlos con las etiquetas de clases, que representan los movimientos imaginarios a derecha e izquierda.

Se procede a clasificar cada uno de los 7 archivos con extensión .csv mediante el software Weka.

Se seleccionaron los algoritmos de clasificación J48, Naive bayes, SMO y LDA, Estos algoritmos de clasificación son los más utilizados en la literatura revisada sobre clasificación de señales de electroencefalogramas para movimiento imaginario. Teniendo un predominio en los estudios el uso de los clasificadores del LDA y SMO.

Para cada uno de los algoritmos de clasificación se ejecuta la etapa de entrenamiento (Training), validación Cruzada (Cross Validation – 10 folds) y Evaluación (Test).

RESULTADOS Y DISCUSION

5.1. Presentación de resultados

En este ítem se presenta los resultados de comparar algoritmos de clasificación para señales de electroencefalogramas. Los resultados se muestran en las siguientes tablas.

Tabla 2: Resultados del algoritmo J48

Algoritmo de Clasificación	Selección de características	Precisión	Cobertura	Exactitud	Área ROC
J48	Potencia Promedio (P)	81.54%	75.71%	79.29%	0.763
	Promedio Absoluto (MAV)	85.45%	67.14%	77.86%	0.826
	Varianza (V)	78.13%	71.43%	75.71%	0.727
	P - MAV	85.71%	68.57%	78.57%	0.803
	P - V	78.13%	71.43%	75.71%	0.731
	MAV - V	85.71%	68.57%	78.57%	0.803
	P - MAV - V	85.71%	68.57%	78.57%	0.803

Fuente: propia

En la tabla 2 se puede ver que para el algoritmo de clasificación J48 se utilizaron como vectores de selección de características los estadísticos descritos en el ítem. 4.1.4.; obteniéndose los resultados de precisión, cobertura, exactitud y la curva de ROC, siendo la exactitud aquel valor que se aproxima más al valor deseado y el utilizado para determinar la eficiencia del algoritmo, obteniendo un porcentaje mayor en la Potencia promedio con un 79.29%.

Tabla 3: Resultados del algoritmo Naive Bayes

Algoritmo de Clasificación	Selección de características	Precisión	Cobertura	Exactitud	Área ROC
Naive Bayes	Potencia Promedio (P)	86.89%	75.71%	82.14%	0.763
	Promedio Absoluto (MAV)	84.62%	78.57%	82.14%	0.826
	Varianza (V)	86.89%	75.71%	82.14%	0.727
	P - MAV	84.38%	77.14%	81.43%	0.803
	P - V	86.89%	75.71%	82.14%	0.731
	MAV - V	84.38%	77.14%	81.43%	0.803
	P - MAV - V	85.48%	75.71%	81.43%	0.803

Fuente: propia

En la tabla 3 se puede ver que para el algoritmo de clasificación Naive Bayes se utilizaron como vectores de selección de características los estadísticos descritos en el ítem. 4.1.4.; obteniéndose los resultados de precisión, cobertura, exactitud y la curva de ROC, siendo la exactitud aquel valor que se aproxima más al valor deseado y el utilizado para determinar la eficiencia del algoritmo, obteniendo un porcentaje mayor en la Potencia promedio, promedio Absoluto, varianza y la combinación de potencia promedio y varianza, con un 82.14%. siendo su porcentaje de clasificación mayor al de J48.

Tabla 4: Resultados del algoritmo SMO

Algoritmo de Clasificación	Selección de características	Precisión	Cobertura	Exactitud	Área ROC
SMO	Potencia Promedio (P)	87.88%	82.86%	85.71%	0.763
	Promedio Absoluto (MAV)	90.48%	81.43%	86.43%	0.826
	Varianza (V)	85.29%	82.86%	84.29%	0.727
	P - MAV	89.39%	84.29%	87.14%	0.803
	P - V	85.29%	82.86%	84.29%	0.731
	MAV - V	89.39%	84.29%	87.14%	0.803
	P - MAV - V	92.06%	82.86%	87.86%	0.803

Fuente: propia

En la tabla 4 se puede ver que para el algoritmo de clasificación SMO se utilizaron como vectores de selección de características los estadísticos descritos en el ítem. 4.1.4.; obteniéndose los resultados de precisión, cobertura, exactitud y la curva de ROC, siendo la exactitud aquel valor que se aproxima más al valor deseado y el utilizado para determinar la eficiencia del algoritmo, obteniendo un porcentaje mayor en la combinación de la potencia promedio – Promedio absoluto - varianza, con un 87.86%. siendo su porcentaje de clasificación mayor al de J48 y Naive Bayes.

Tabla 5: Resultados del algoritmo LDA

Algoritmo de Clasificación	Selección de características	Precisión	Cobertura	Exactitud	Área ROC
LDA	Potencia Promedio (P)	92.19%	84.29%	88.57%	0.763
	Promedio Absoluto (MAV)	92.19%	84.29%	88.57%	0.826
	Varianza (V)	92.19%	84.29%	88.57%	0.727
	P - MAV	89.06%	81.43%	85.71%	0.803
	P - V	90.63%	82.86%	87.14%	0.731
	MAV - V	87.69%	81.43%	85.00%	0.803
	P - MAV - V	86.36%	81.43%	84.29%	0.803

Fuente: propia

En la tabla 5 se puede ver que para el algoritmo de clasificación LDA se utilizaron como vectores de selección de características los estadísticos descritos en el ítem. 4.1.4.; obteniéndose los resultados de precisión, cobertura, exactitud y la curva de ROC, siendo la exactitud aquel valor que se aproxima más al valor deseado y el utilizado para determinar la eficiencia del algoritmo, obteniendo un porcentaje mayor en la potencia promedio, el promedio absoluto y la varianza, con un 88.57%. siendo su porcentaje de clasificación mayor al de J48 y Naive Bayes y sus resultados muy cercanos a los de SMO.

Tabla 6: Resumen de resultados

Algoritmo de Clasificación	Selección de características	Precisión	Cobertura	Exactitud	Área ROC
LDA	Potencia Promedio (P)	92.19%	84.29%	88.57%	0.763
	Promedio Absoluto (MAV)	92.19%	84.29%	88.57%	0.826
	Varianza (V)	92.19%	84.29%	88.57%	0.727
SMO	P - MAV - V	92.06%	82.86%	87.86%	0.803
Naive Bayes	P - V	86.89%	75.71%	82.14%	0.731
J48	Potencia Promedio (P)	81.54%	75.71%	79.29%	0.763

Fuente: Propia

En la tabla 6 se observa el resumen de los máximos resultados obtenidos por cada uno de los algoritmos de clasificación. Donde el LDA es el que obtiene el mayor resultado de exactitud con un 88.57% seguido de SMO con un 87.86%, Naive Bayes con un 82.14% y J48 con un 79.29%.

5.2. Discusión de resultados

- Oikonomou Vangelis P. et al. (2017), en su trabajo “*A comparison study on EEG signal processing techniques using motor imaginary*”. Utiliza el conjunto de datos B de Graz, considera a los patrones espaciales comunes y al método de Welch para la estimación del espectro como características relevantes para aplicar LDA y SVM como clasificadores. Obteniendo un valor promedio de 0.59 en el estadístico Kappa. Nuestro estudio obtuvo en la clasificación del LDA un valor del estadístico Kappa de 0.7143. Siendo este resultado mejor al presentado por (Oikonomou, Georgiadis, Liaros, &

Kompatsiaris) ya que el estadístico Kappa muestra la concordancia entre dos observaciones y por nuestro resultado podemos decir que se tiene mayor concordancia en las observaciones realizada en nuestros resultados.

- Según Yue Zang, et al. (2017), en su trabajo “*Extracting Error-Related Potentials from Motion Imagination EEG in Noninvasive Brain-Computer Interface*”. Usa para la clasificación el algoritmo LDA a partir del error relativo potencial de 69.14%. Nuestra investigación supera estos resultados en todos los algoritmos planteados como J48 con un 79.29%, Naive Bayes con un 82.14%, SMO con un 87.26% y LDA con un 88.57%. Esto hará posible que se pueden construir nuevos interfaces cerebro computador más eficientes para que ayudan a personas con discapacidades motoras.
- Según Yuhan ye et al. (2017) en su trabajo “*The Design of Multi-task Simulation Manipulator Based on Motor Imagery EEG*”. Se considera como clasificador al LDA mediante el patrón espacial común y la información mutua para la selección de características, obteniendo un porcentaje promedio de 68.33%. Nuestra investigación supera estos resultados en todos los algoritmos planteados como J48 con un 79.29%, Naive Bayes con un 82.14%, SMO con un 87.26% y LDA con un 88.57%. Esto hará posible que se pueden construir nuevos interfaces cerebro computador más eficientes para que ayudan a personas con discapacidades motoras.

CONSTRUCCION DEL PROTOTIPPO

6.1.Construcción del prototipo de software interactivo cerebro computador

Para la construcción del prototipo de software interactivo se utilizó la metodología de programación extrema (XP) que consta de las siguientes etapas

6.1.1. Planificación.

En esta etapa se desarrollan las pautas para la construcción del prototipo de software interactivo cerebro computador. De los datos obtenidos anteriormente se decide utilizar los datos resultantes de la potencia promedio ya que es el que obtiene uno de los porcentajes mayores, con un 88.57% de exactitud.

También se realizan las siguientes actividades.

Establecimientos de historia

Tabla 7 Historia

Numero de historia	Nombre de la historia
1	Potencia promedio para el algoritmo LDA

Fuente: propia

Se define la tarea a realizar

Tabla 8 Tarea 1

Numero de tarea: 1	Numero de Historia: 1
Nombre de la tarea	Diseñar una interfaz de la potencia promedio para LDA
Tipo de tarea: desarrollo:	Puntos estimados: 10
Fecha de Inicio: 10/05/2019	Fecha Fin: 10/05/19
Responsable	Carlos Ramón Quispe Onofre
Descripción: Se diseña una interfaz en la cual el usuario busca la dirección de los datos de training y test de movimiento imaginario para luego solicitar que estos se procesen. Para luego visualizar los datos y resultados estadísticos del Test de movimiento imaginario hacia la derecha o izquierda.	
Fuente: Propia	

Para la presente tarea se realizará una sola iteración

6.1.2. Diseño:

En esta etapa se ven las tarjetas CRC (Clase-Responsabilidad-Colaboración) y el prototipo a implementarse

Tabla 9 Tarjeta CRC

Tarjeta CRC Usuario - (Clase-Responsabilidad-Colaboración)	
Responsabilidades	Colaboradores
Ubicar archivos	
Procesar datos	Weka.jar
	Clase LDA
Fuente: Propia	

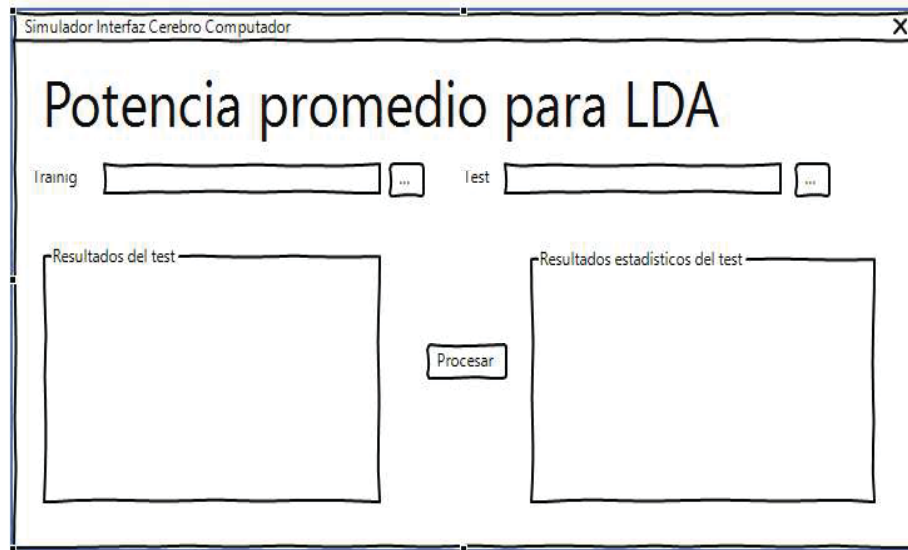


Figura 20 Prototipo de Simulador de Interfaz cerebro computador

6.1.3. Codificación.

Para la comprobación de los resultados se implementó un prototipo interactivo del clasificador LDA a partir de la librería Weka.jar, la clase LDA.java escrita por Eibe Frank de la universidad de Waikato.



Figura 21 Software de clasificación de LDA

Primeramente, se busca la ubicación de los archivos de training (entrenamiento) y test (prueba). De ahí, se presiona el botón procesar, y se obtiene la tabla de resultados de test y el modelo generado por el clasificador LDA.

```

readerTraining = new BufferedReader(new FileReader(RutaTraining));
ArffReader arff = new ArffReader(readerTraining);
Instances dataTrainig = arff.getData();
dataTrainig.setClassIndex(dataTrainig.numAttributes() - 1);

readerTest = new BufferedReader(new FileReader(RutaTest));
ArffReader arffTest = new ArffReader(readerTest);
Instances dataTest = arffTest.getData();
dataTest.setClassIndex(dataTest.numAttributes() - 1);

LDA lda = new LDA();
lda.buildClassifier(dataTrainig);
Evaluation eval = new Evaluation(dataTrainig);
eval.evaluateModel(lda, dataTest);

```

Figura 22 Código para clasificar el algoritmo LDA

Se recuperan los archivos de extensión .arff que son propios del software Weka y se almacenan en las instancias de datos de dataTraining y dataTest.

Luego se instancia la clase LDA, que es el encargado de construir el clasificador de entrenamiento y crea un modelo predictivo. A partir del modelo predictivo se procede a evaluar el Test. Obteniendo de esta manera los mismos resultados de exactitud que los obtenidos con el software Weka. Con esta aplicación se puede lograr construir a futuro interfaces cerebro computador para datos que se procesen en tiempo real.

6.1.4. Pruebas.

Al culminar la codificación se procede a la realización de pruebas unitarias de acuerdo a la siguiente tabla.

Tabla 10 Caso de prueba

Caso de prueba	
Numero de prueba: 1	Numero de historia: 1
Nombre de historia: Potencia promedio para el algoritmo LDA	
Condiciones de ejecución: se debe ejecutar el programa desde el IDE de NetBeans	
Entrada/Pasos de Ejecución: el usuario selecciona las rutas de Training y Test. Luego hacen clic en Procesar.	
Resultado esperado: Muestra la tabla de datos de Test y el sumario de los resultados obtenidos.	
Evaluación de prueba: La prueba finalizo correctamente.	
Fuente: Propia	

6.1.5. Lanzamiento.

Una vez realizadas las pruebas satisfactoriamente se procede a la etapa de puesta en funcionamiento del software.

Conclusiones

- La comparación de los algoritmos de clasificación seleccionados para el análisis de señales de electroencefalogramas para movimiento imaginario a partir de tres estadísticos como son la potencia promedio, promedio absoluto y varianza; determinó que el algoritmo del LDA como el más eficiente con un 88.57% seguido del clasificador SMO con un 87.86%.
- Se construyó un prototipo de software, para comprobar los resultados obtenidos con el Clasificador LDA para la potencia promedio, obteniendo los mismos resultados que con el software Weka.

Recomendaciones

1. Crear un repositorio de datos para movimiento imaginario propio de la universidad, que sirva para futuras investigaciones.
2. Descomponer en más niveles la transformada discreta de Wavelet para ver si se logra una mejor clasificación de las señales de electroencefalogramas.
3. Aplicar algoritmos de clasificación como redes neuronales y máquina de soporte de vectores en sus diferentes variantes para obtener mejores resultados a la presente investigación.
4. Realizar diferentes métodos de extracción de características o la combinación de estos métodos como pueden ser: Transformada discreta de wavelet, Transformada de Fourier, análisis de componentes principales, entre otros. Con el fin de obtener mejores resultados a la presente investigación.
5. Utilizar otros vectores característicos como podría ser valor medio ponderado, potencia de emisión, desviación estándar entre otros, también con el objetivo de lograr mejores resultados que los obtenidos en la presente investigación.
6. Construir una interfaz cerebro – computador a partir de los resultados obtenidos en la presente investigación.

Bibliografía

- Alam Iztac, H. C. (2014). Adquisición no-invasiva de señales de EEG y procesamiento de señales motrices. Mexico.
- Andrade Mora, S. M., & Navarrete Mejía, M. G. (2006). *Análisis y tratamiento de la señal electrocardiográfica para la detección de parámetros de normalidad bajo la plataforma labview "ADPAN-ECG"*. Bucaramanga - Venezuela. Obtenido de http://cienciagora.com.co/imgs2012/imagenes/Tesis_Miguel_Ingenieria_Electronica_UPB.pdf
- Arafat, I. (2015). Brain-Computer Interface: Past, Present & Future. *International Islamic University Chittagong*. Obtenido de http://www.academia.edu/1365518/Brain_Computer_Interface_Past_Present_and_Future
- Arango Flores, G. A. (2007). *Clasificación de fallas en motores electricos utilizando señales de vibración*. Pereira - Colombia. Obtenido de <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/1149/621313A662c.pdf?sequence=1>
- Barea Navarro, R. (2009). Electroencefalografía. Madrid, España: Universidad de Alcalá.
- Barros, R. C. (2012). A Survey of Evolutionary Algorithms for Decision-Tree Induction. *Systems, Man, and Cybernetics. IEEE Transactions*, 291-312.
- BCI Competition II*. (2003). Obtenido de <http://www.bbc.de/competition/ii/>
- Boutros, N., Galderisi, S., Pogarell, O., & Riggio, S. (2011). *Standard electroencephalography in clinical psychiatry: a practical handbook*. Oxford, UK.
- Breiman, L. F. (1984). *Classification and Regression Trees*. Wadsworth International Group.

- Burrus, S., & Ramesh, A. G. (1999). *Introduction to wavelets and Wavelet Transforms. Electrical and Computer Engineering Department, Rice University*. Prentice Hall.
- Calero Arevalo, J., Calle Fuentes, P., Ferruzola Campaña, I., & Torres Panchana, E. (2010). *Análisis, diseño e implementación de un tele -alertador cardíaco con el sistema labview*. Guayaquil - Ecuador. Obtenido de <http://repositorio.ucsg.edu.ec/bitstream/3317/8565/1/T-UCSG-PRE-TEC-ITEL-232.pdf>
- Camacho, J., & Manian, V. (2016). Real-time single channel EEG motor imagery based Brain Computer Interface. En *2016 World Automation Congress (WAC)* (págs. 1-6). Rio Grande: IEEE. doi:10.1109/WAC.2016.7582973
- Carrilo-de-la-peña, M., & Galdo-Alvarez, S. y.-B. (2008). Equivalent is not equal: Primary motor cortex (MI) activation during motor imagery and execution of sequential movements. *Brain Research*, 1226, 134-143.
- Carrión Pérez, P. A. (2007). *Procesado de señales biomédicas*. Cuenca: Ediciones de la Universidad de Castilla - La Mancha.
- Chih-Hsiang Peng, T.-W. K.-C.-F.-J. (2015). Trainable and Low-Cost SMO Pattern Classifier Implemented via MCMC and SFBS Technologies. En *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS*, (págs. 23:2295-2305).
- Coomonte Belmonte, r. (2006). *Sistema de reconocimiento de personas mediante su patron de iris basado en la transformada de wavelet*. Madrid - España. Obtenido de http://oa.upm.es/1329/1/PFC_RAFAEL_COOMONTE_BELMONTE.pdf
- Córdova Ricapa, F. (2013). *Desarrollo De Un Sistema Para La Evaluacion De La Actividad Muscular Mediante Electrodo De Superficie*. Lima.

- Corso, C. (2009). Aplicación de algoritmos de clasificación supervisada usando Weka. Universidad Tecnológica Nacional, Facultad Regional Córdoba. Obtenido de http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cy nthia/CNIT_2009_Aplicacion_Algoritmos_Weka.pdf
- Das, P., Sadhu, A. K., Konar, A., Bhattacharya, B. S., & Nagar, A. K. (2015). Adaptive Parameterized AdaBoost Algorithm with application in EEG Motor Imagery Classification. En *2015 International Joint Conference on Neural Networks (IJCNN)* (págs. 1-8). Killarney. doi:10.1109/IJCNN.2015.7280835
- De Castro Fernandez, R. M. (2002). *Análisis de la teoría de ondículas orientadas a las aplicaciones en ingeniería eléctrica: Fundamentos*. Madrid - España. Obtenido de https://www.uta.cl/hdiaz/Documents/Fundamentos_Wavelets.pdf
- De Luca Carlo, J. (2011). *Surface Electromyography: Detection and Recording*. Delsys. Obtenido de URL:http://www.delsys.com/Attachments_pdf/WP_SEMGintro.pdf
- Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. En *In Proceedings of the 12th International Conference on Machine Learning* (págs. 194-202).
- Fauzi, H., Shapiyai, M. I., Setiawan, N. A., & Mustafa, J. J. (2017). Channel Selection for Common Spatial Pattern Based on Energy Calculation of Motor Imagery EEG Signal. En *International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)* (págs. 33-39). Yogyakarta. doi:10.1109/ICCEREC.2017.8226692
- Fridlun Alan, J., & Fowler Stephen, C. (1978). An eight-channel computer-controlled scanning electromyography.
- Gutiérrez-Gutiérrez, G., Barbosa López, C., Navacerrada, F., & Miralles Martínez, A. (Octubre de 2011). Utilidad del electromiograma en el diagnóstico de las

- miopatías inflamatorias. *Reumatología Clínica*, 8(4), 195-200.
doi:doi:10.1016/j.reuma.2011.10.012
- Hayashi, T., Yokoyama, H., & Wada, I. N. (2017). Prediction of individual finger movements for motor execution and imagery: An EEG study. En *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (págs. 3020-3023). Banff, AB. doi:10.1109/SMC.2017.8123088
- Hernández Barajas, F. (2007). *Comparación entre análisis discriminante no métrico y regresión logística multinomial*. Medellín: Universidad de Colombia.
- Hernandez Sampieri, R. F. (2014). *Metodología de la investigación*. México D.F. : McGRAW-HILL / INTERAMERICANA EDITORES, S.A. .
- Hosni, S., Gadallah, M., & M., B. S. (2008). "Classification of EEG Signals Using Different Feature Extraction Techniques for Mental Task. *Report of Ain shams University, El Cairo*.
- Hsu, C.-N., Huang, H.-J., & Wong., T.-T. (2000). Why discretization works for naive bayesian classifiers. En *In ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*, (págs. 399-406). San Francisco, CA, EEUU: Morgan Kaufmann Publishers Inc.
- Hunt, E. B. (1966). *Experiments in Induction*. New York: Academic Press.
- Jaramillo Gonzales, A. (2016). *Fundamentos de las interfaces cerebro maquina*. Mexico: comunicaciones Libres.
- Jaramillo Gonzales, A., & Hermosillo Valadez, J. (2015). Interfaces cerebro-máquina, un encuentro entre neurociencias y robótica. *Inventio. La genesis de la cultura universitaria en Morelios*, 29-36.
- Jeannerod, M., & Frak, V. (1999). Mental imaging of motor activity in humans. *Curr Opin Neurobiol*, 9(6), 735-739.

- Joachims, T. (1998).). Text categorization with support vector machines: Learning with many relevant features. Springer Berlin: Heidelberg.
- Juarez Segura, D. (2005). *Eliminación del ruido en señales radar utilizando Wavelets*. Mexico. Obtenido de <http://docplayer.es/48415504-Instituto-politecnico-nacional.html>
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. En *Applied Statistics* (Vol. 2, págs. 119-127.).
- Kerber, R. (1992). . Chimerge: discretization for numeric attributes. En *In Proceedings of National conference on artificial intelligence* (págs. 123-128). Menko Park: AAAI Press.
- Kim, S.-B. e. (2006). Some effective techniques for naive bayes text classification. En *Knowledge and Data Engineering, IEEE* (págs. 1457 - 1466.). IEEE.
- Kolesov, A. e. (2014). On Multilabel Classification Methods of Incompletely Labeled Biomedical Text Data. . En *Computational and Mathematical Methods in Medicine*.
- Kraeutner, S., Gionfriddo, A., & Bardouille, C. y. (2014). Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations. *Brain Research*, 1588, 81-89.
- Macri, M., & Benedetti, M. (2012). Análisis multirresolución del motor trifásico de inducción sometido a huecos de tensión. *Ingeniare. Revista chilena de ingeniería*, 20(1), 66-78. Obtenido de <https://scielo.conicyt.cl/pdf/ingeniare/v20n1/art07.pdf>
- Martinez Aldana, I. (2009). *Detección de fallos en rodamientos mediante análisis de energía y lógica difusa*. Madrid - España. Obtenido de https://e-archivo.uc3m.es/bitstream/handle/10016/8369/PFC_Ivan_Martinez_Aldana.pdf?sequence=1

- Martinez Romero, C. I. (2010). Predicción de fugas de clientes en una compañía de seguros utilizando redes neuronales artificiales evolutivas. Valparaiso - Chile. Obtenido de http://opac.pucv.cl/pucv_txt/txt-0000/UCF0318_01.pdf
- Minguez, J. (2014). *Tecnología de Interfaz Cerebro - Computador*. España: Universidad de Zaragoza.
- Moran, A. (2015). *Diseño de interfaces cerebro-máquina controlados mediante registros de EEG*. Madrid-España: Universidad Autónoma de Madrid.
- Müller-Putz, G. R. (2011). Brisk movement imagination for the non-invasive control of neuroprostheses: a first attempt. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 4219-4222.
- Neurológicas, I. I. (enero de 2016). <http://www.iensa.es/especialidades/neurofisiologia/item/26-electroneurografia-eng/26-electroneurografia-eng>.
- Oikonomou, V. P., Georgiadis, K., Liaros, G., & Kompatsiaris, S. N. (2017). Sparse Kernel Machines for motor imagery EEG classification. En *IEEE Engineering in Medicine and Biology Society* (págs. 207-210). doi:10.1109/CBMS.2017.113
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high resolution EEG and ERP measurements. En *Clinical Neurophysiology* (págs. 713-719).
- Park, B., Kim, H. J., Lee, & Min, B. (2017). Motor imagery EEG classification with optimal subset of wavelet based common spatial pattern and kernel extreme learning machine. En *9th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (págs. 2863-2866). Seogwipo. doi:doi: 10.1109/EMBC.2017.8037454

- Pattnaik, S., Dash, M., & Sabut, S. K. (2016). DWT-based feature extraction and classification for motor imaginary EEG signals. En *2016 International Conference on Systems in Medicine and Biology (ICSMB)* (págs. 186-201). Kharagpur: IEEE. doi:10.1109/ICSMB.2016.7915118
- Pérez Corrales, P. (2016). Primera carrera de drones controlados por la mente. *tendencias tecnológicas 21*.
- Pfurtscheller, G. (2003). *Graz description of data set III of BCIcompetition 2003*. Obtenido de <http://www.bbc.de/competition/ii>
- Pinzon Duque, M. C. (2011). *Análisis de señal del impulso cardíaco para el mejoramiento del diagnóstico de patologías del corazón*. Perreira, colombia. Obtenido de <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/1873/6161207547P661.pdf;jsessionid=3062CC5AB16378B009A3A9E3BE489262?sequence=1>
- Pires, G. T.-B. (2011). Playing Tetris with non-invasive BCI. *2nd International Conference on Serious Games and Applications for Health*, 1–6.
- Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector.
- Prieto, A. (s.f.). *Electrocardiograma*. Obtenido de http://www.cenetec.gob.mx/cd_inter/guiapaciente/electrocardiograma.pdf.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Mateo.: Morgan Kaufmann.
- Reuderink, B. N. (2009). Affective Pacman: A frustrating game for brain–computer interface experiments. *Intelligent technologies for interactive entertainment. Lecture Notes of the Institute for Computer Sciences. Social Informatics and Telecommunications Engineering, 9*, 221-227.

- Robinson, N., Thomas, K. P., & Vinod, A. P. (2017). Canonical correlation analysis of EEG for classification of motor imagery. En *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (págs. 2317-2321). Banff: IEEE. doi:10.1109/SMC.2017.8122967
- Robledo Carmona, J. M., Jiménez Navarro, M., & Robledo Carmona, L. (2014). *ELECTROCARDIOGRAMA*. España: Servicio de Cardiología del Hospital Virgen de la Victoria.
- Rodriguez, M., & Llanos, C. y. (2009). The kinematics of motor imagery: Comparing the dynamics of real and virtual movements. *Neuropsychologia*, 47(2), 489-496.
- Ron-Angevin, R. &.-L. (2018). *Problemática de las Interfaces Cerebro-Computador*. Obtenido de https://www.researchgate.net/publication/238623170_Problematica_de_las_Interfases_Cerebro-Computador
- Rowan, J., & Tolunsky, E. (2004). *Conceptos básicos sobre EEG con mini-atlas*. Madrid, España: Elsevier.
- Saltos Torres, M. G. (2014). *Análisis de señales sísmicas del volcán Cotopaxi mediante las transformadas de Wavelet y Fourier*. Ecuador. Obtenido de <http://repositorio.espe.edu.ec/bitstream/21000/9642/1/T-ESPE-048135.pdf>
- Sanei, S., & Chambers, J. (2007). *EEG Signal Processing*. Cardiff, UK: Ltda.
- Santamaria Ruiz, W. (2010). *Modelo de detección de fraude basado en el descubrimiento simbólico de reglas de clasificación extraídas de una red neuronal*. Colombia. Obtenido de <http://www.bdigital.unal.edu.co/3086/1/299742.2010.pdf>
- Shawe-Taylor, J., & Cristianin, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge: Press.

- Shedeed, H. A.-S. (2013). rain EEG signal processing for controlling a robotic arm. *Computer Engineering & Systems (ICCES), 2013 8th International Conference*, 152-157.
- Shenoy, H. V., Vinod, A. P., & Guan, C. (2015). Shrinkage estimator based regularization for EEG motor imagery classification. En *10th International Conference on Information, Communications and Signal Processing (ICICS)* (págs. 1-5). Singapore. doi:10.1109/ICICS.2015.7459836
- Tangemann, ., M.-R. (2009). Playing Pinball with Non-Invasive BCI. In *Advances in Neural Information Processing Systems. MA: MIT Press 21*, 1641–1648.
- Taywade, S., & Raut, R. (2012). A Review: EEG signal analysis with different methodologies. . *Proceeding on a National Conference on Innovative Paradigms in Engineering and Techonology*, 29-31.
- Teplan, M. (2002). Fundamentals of EEG measurements. En *Measurements sciences review* (págs. 1-10).
- Thomas, K. P., Robinson, N., & Vinod, A. P. (2017). EEG-based motor imagery classification using subject-specific spatio-spectral features,. En *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (págs. 2302-2307). Banff: IEEE. doi:10.1109/SMC.2017.8122964
- Toro, J. (1984). POTENCIALES EVOCADOS. *Acta Médica Colombiana Vol. 9 No. 6*, 319-324. Obtenido de <http://actamedicacolombiana.com/anexo/articulos/06-1984-03.htm>
- Tu, W. &. (2009). Three-Class Motor Imagery EEG Data by Combining Wavelet Packet Decomposition and Common Spatial Pattern. *International Conference on Intelligent HumanMachine Systems and Cybernetics, Vol No.1*, 189-191.

- Tyagi, A., & Nehra, V. (2017). Time frequency analysis of non-stationary motor imagery EEG signals. En *International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)* (págs. 44-50). Gurgaon. doi:10.1109/IC3TSN.2017.8284448
- Valenzuela Ramírez, M. A. (2015). *Construcción de un mecanismo de procesamiento de patrones temporales aplicado al reconocimiento de voz*. Chile. Obtenido de <http://repositorio.uchile.cl/bitstream/handle/2250/135331/Construccion-de-un-mecanismo-de-procesamiento-de-patrones.pdf?isAllowed=y&sequence=1>
- Van de Laar, B. G.-O. (2013). Experiencing BCI Control in a Popular Computer Game. . *IEEE Transactions on Computational Intelligence and AI in Game*, 176 – 184.
- Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley and Sons,.
- Vaughan T., W. J. (1996). EEG-Based communication: Prospects and Problems. *IEEE Trans. Rehab. Eng., vol. 4, , 425-430*.
- Vision, A. (13 de 03 de 2018). *Admira Vision*. Obtenido de Electroretinograma: <https://www.admiravision.es/es/articulos/pruebas-diagnosticas/articulo/electroretinograma#.XHRXH-hKjIU>
- Wang, B., Jun, L., Bai, J., Peng, L., & Li, Y. a. (2005). EEG Recognition Based on Multiple Types of Information by Using Wavelet Packet Transform and Neural Networks. *Proceedings of the IEEE 27th Annual Conference Engineering in Medicine and Biology*, 5377-5380.
- Wikstrom, G. (2005). *Data Classification using Support Vector Machines*. Estocolmo: Stockholm University.
- Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques. Segunda Edición*. Elsevier.

- Wolpaw, J., & et al. (junio, 2000). Brain-computer interface technology: A review of the first international meeting. *IEEE Trans. Rehab. Eng Vol. 8,* 164-173.
- Yang, Y., & Geoffrey I, W. (2002). A comparative study of discretization methods for naïve bayes classifiers. En *In Proceedings of PKAW 2002: The 2001 Pacific Rim Knowledge Acquisition Workshop* (págs. 159-173).
- Ye, Y., Yang, C., Li, X., Ju, Z., & Li, Z. (2017). *The design of multi-task simulation manipulator based on motor imagery EEG*. Banff, Canada: IEEE. doi:10.1109/SMC.2017.8123135
- Yueguo Luo, Z. X. (2016). Classification noise detection based SMO algorithm. En *Optik- International Journal for Light and Electron Optics*, (págs. 7021-7029).
- Zela Ari, L., & Mamani Chambi, W. (2015). *Hipertensión Arterial y Estilo de Vida en Adultos Mayores de la Micro Red de Salud Revolución San Román, Juliaca - 2015*. Juliaca - Perú. Obtenido de http://repositorio.upeu.edu.pe/bitstream/handle/UPEU/240/Luzmila_Tesis_bachiller_2015.pdf?sequence=1
- Zhang, Y., Chen, W., & Wang, J. Z. (2007). Extracting error-related potentials from motion imagination EEG in noninvasive brain-computer interface. En *2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)* (págs. 768-773). Ningbo: IEEE. doi:10.1109/ICCIS.2017.8274876