



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

DESARROLLO DE MODELO PARA LA IDENTIFICACIÓN DE
SOMNOLENCIA BASADO EN REDES NEURONALES
CONVOLUCIONALES Y PROCESAMIENTO DE SEÑALES DE
ELECTROENCEFALOGRAFÍA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
MECÁNICO

LUIS FELIPE GUARDA BRÄUNING

PROFESOR GUÍA:

ENRIQUE LÓPEZ DROGUETT

MIEMBROS DE LA COMISIÓN:

VIVIANA MERUANE NARANJO

EDUARDO SALAMANCA HENRÍQUEZ

SANTIAGO DE CHILE

2018

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL MECÁNICO
POR: LUIS FELIPE GUARDA BRÄUNING
FECHA: 2018
PROF. GUÍA: ENRIQUE LÓPEZ DROGUETT

La confiabilidad en todo proceso de ingeniería es fundamental para asegurar su correcto y seguro desarrollo. Por este motivo se vuelve imperante ampliar las fronteras de análisis en confiabilidad, dejando de tener como eje central el análisis de equipos técnicos, apuntando de manera paralela al estudio de confiabilidad del operador de dichos equipos.

Las nuevas herramientas de aprendizaje de máquinas permiten a una computadora analizar información en base a algoritmos para aprender a identificar características concretas de su entorno. Dentro de estas nuevas herramientas, las redes neuronales convolucionales (CNN) se presentan como una de las técnicas más robustas para el reconocimiento de imágenes. Mediante esta técnica es posible determinar patrones de frecuencia en señales de electroencefalografía (EEG) para determinar diversos estados anímicos de un individuo.

El objetivo general de la memoria consiste en desarrollar un modelo para la detección del estado de somnolencia de un operador en base a redes neuronales convolucionales utilizando señales de electroencefalografía. Los objetivos específicos son:

- Desarrollar, comprender y manejar un modelo de detección de sujetos con ojos abiertos y cerrados en base a CNN, a través del procesamiento de señales de electroencefalografía.
- Desarrollar, comprender y manejar un modelo de detección de sujetos en estado de somnolencia utilizando CNN a través del procesamiento de señales de electroencefalografía.
- Comparar la clasificación de sujetos en estado de somnolencia realizada con redes neuronales convencionales (NN), *Support Vector Machines* (SVM) y *Random Forest* (RF).

Para llevar a cabo los objetivos planteados, se utilizó la “base de datos de somnolencia multimodalidad ULg”. Con ella, en primer lugar, se generaron seis conjuntos de datos a evaluar, con sus respectivas etiquetas. Luego se elaboró un clasificador utilizando CNN, capaz de clasificar entre sujetos con y sin somnolencia, con el cual además se encontró el conjunto de datos que entregaba el mayor desempeño para el modelo creado. Posteriormente se procedió a comparar el desempeño de la red elaborada con el conjunto de datos óptimo generado con NN, SVM y RF.

Luego de comparar el rendimiento entregado por cada modelo, se obtuvo que las redes neuronales convolucionales representan el método más a fin para realizar la clasificación de sujetos bajo estado somnolencia, entregando una exactitud de clasificación superior al 86%.

Tabla de contenido

| | |
|--|----|
| 1. Introducción | 1 |
| 1.1 Objetivo general..... | 2 |
| 1.2 Objetivos específicos | 2 |
| 1.3 Alcances..... | 2 |
| 2. Antecedentes | 3 |
| 2.1 Fatiga mental..... | 3 |
| 2.1.1 Riesgos de la fatiga | 3 |
| 2.1.2 Estadísticas de fatiga mental | 4 |
| 2.1.3 Nivel de somnolencia percibido por el individuo | 5 |
| 2.2 Electroencefalografía..... | 5 |
| 2.2.1 Ondas cerebrales..... | 7 |
| 2.3 Análisis de componentes independientes | 7 |
| 2.4 Transformada de Fourier | 8 |
| 2.5 Espectrogramas | 8 |
| 2.6 <i>Support Vector Machines</i> | 10 |
| 2.7 <i>Random Forest</i> | 11 |
| 2.8 Redes neuronales | 11 |
| 2.8.1 Redes neuronales convencionales | 11 |
| 2.8.2 Redes neuronales convolucionales | 15 |
| 2.8.3 Entrenamiento de una red neuronal | 17 |
| 2.9 Métricas de rendimiento para tareas de clasificación | 19 |
| 3. Metodología | 21 |
| 3.1 Prefactibilidad de análisis de espectrogramas de EEG mediante CNN..... | 21 |
| 3.1.1 Preprocesamiento de las señales..... | 21 |
| 3.1.2 Implementación de CNN para análisis de espectrogramas de EEG..... | 22 |
| 3.1.3 Resultados obtenidos | 23 |
| 3.1.4 Análisis de resultados | 24 |
| 3.2 Detección de sujetos bajo estado de fatiga | 24 |
| 3.2.1 Base de datos | 24 |
| 3.2.2 Preprocesamiento de datos y selección de etiquetas | 25 |

| | | |
|-------|---|----|
| 3.2.2 | Implementación de redes neuronales convolucionales..... | 28 |
| 3.2.3 | Implementación redes neuronales convencionales..... | 30 |
| 3.2.4 | Implementación <i>Support Vector Machines</i> | 32 |
| 3.2.5 | Implementación <i>Random Forest</i> | 32 |
| 3.3 | Recursos..... | 33 |
| 4. | Resultados | 34 |
| 4.1 | Resultados red neuronal convolucional para cada conjunto de datos de somnolencia... | 34 |
| 4.2 | Comparación de resultados por modelo para espectrogramas de trece segundos correspondientes a los canales F_z - P_z | 36 |
| 5. | Análisis de resultados..... | 37 |
| 6. | Conclusiones | 39 |
| 7. | Bibliografía..... | 40 |

1. Introducción

Debido a la automatización y al desarrollo tecnológico en los ambientes de trabajo, es cada vez más común la ejecución de tareas monótonas que requieren altos niveles de atención. Cuando estas tareas desempeñan un papel importante en el mantenimiento y deben garantizar la seguridad en entornos de trabajo como salas de control, es imperativo que los operadores conserven un nivel de alerta adecuado. En este contexto la detección temprana de la somnolencia se ha vuelto de vital importancia para garantizar el desarrollo correcto y seguro de las tareas mencionadas.

En este contexto, se define la confiabilidad humana como la probabilidad de que una persona (1) realice correctamente una acción requerida por el tiempo requerido y (2) que no realice ninguna actividad extraña que pueda degradar el sistema [1].

El problema radica en que no es posible determinar si un operador es capaz de realizar una tarea crítica sólo analizando su aparente nivel de rendimiento en ella, puesto que este puede realizar un proceso mental compensatorio para que su desempeño en la tarea parezca normal.

Bajo este contexto, la somnolencia se define como un estado de conciencia con oscilaciones entre el sueño y la vigilia, y un deseo irresistible de dormir, acompañado de pesadez y/o torpeza. Esto puede causar tiempos de reacción lentos y una reducción de la vigilancia. La somnolencia es también utilizada indistintamente bajo el termino de fatiga [2].

Usando señales de electroencefalograma (capturadas desde varios electrodos ubicados en el cuero cabelludo), es posible registrar variaciones en el potencial eléctrico del cerebro de un individuo, donde cada uno de ellos proporciona información específica sobre el estado mental de un sujeto. En particular, los canales frontales y parietales proporcionan información sobre el nivel de somnolencia de un individuo, que se representa en variaciones de las ondas alfa y theta [3].

Por otro lado, las Redes Neuronales Convolucionales (CNN) son una de las técnicas de aprendizaje automático más efectivas para el reconocimiento de imágenes debido a su gran capacidad para extraer características y patrones jerárquicos esenciales de las imágenes [4].

De este modo, este trabajo de investigación propone la utilización de redes neuronales convolucionales para detectar automáticamente si un operador se encuentra en estado de somnolencia mediante el uso de los espectrogramas de las señales de electroencefalografía de cada sujeto.

1.1 Objetivo general

Desarrollar un modelo para la detección del estado de somnolencia de un operador en base a redes neuronales convolucionales utilizando señales de electroencefalografía.

1.2 Objetivos específicos

- Desarrollar, comprender y manejar un modelo de detección de sujetos con ojos abiertos y cerrados en base a CNN, a través del procesamiento de señales de electroencefalografía.
- Desarrollar, comprender y manejar un modelo de detección de sujetos en estado de somnolencia utilizando CNN a través del procesamiento de señales de electroencefalografía.
- Comparar la clasificación de sujetos en estado de somnolencia realizada con redes neuronales convencionales, *Support Vector Machines* y *Random Forest*.

1.3 Alcances

Como primer alcance, este trabajo de título propone realizar un análisis de prefactibilidad de análisis de espectrogramas de EEG mediante el uso de redes neuronales convolucionales, realizando un clasificador de sujetos con ojos abiertos y ojos cerrados, utilizando los espectrogramas respectivos.

Como segundo alcance, se propone la realización de un modelo computacional utilizando redes neuronales convolucionales, el cual mediante señales de electroencefalografía sea capaz de distinguir si un sujeto se encuentra o no bajo estado de somnolencia, determinando de qué manera debe ser presentada la base de datos para realizar la clasificación.

Finalmente se deberá comparar el modelo de redes neuronales convolucionales realizado con algoritmos de redes neuronales convolucionales, *Support Vector Machines* y *Random Forest* utilizando la base de datos previamente encontrada.

Cabe destacar que los datos utilizados para llevar a cabo el trabajo de título fueron proporcionados por el profesor guía.

2. Antecedentes

2.1 Fatiga mental

La fatiga mental se describe como un cambio en los estados tanto psicológicos como fisiológicos que un individuo experimenta durante el desarrollo de una actividad cognitiva que demanda alta concentración durante un prolongado período de tiempo [5]. Estos cambios se manifiestan principalmente como una baja en el desempeño cognitivo y psicomotor del sujeto. Es decir, un deterioro en la eficiencia durante la realización de una labor. A consecuencia de esto, la fatiga mental está vinculada a un estado de disminución de alerta, donde gradualmente predomina el cansancio y la falta de energía.

Desde esta perspectiva, la fatiga se muestra como un indicador de un problema implícito: la pérdida de recursos básicos de un individuo tales como reaccionar en forma oportuna y apropiadamente ante una emergencia o un imprevisto [6].

2.1.1 Riesgos de la fatiga

Los métodos utilizados en la literatura para medir la fatiga mental han mostrado ser subjetivos [3], y el auto reporte de este mismo estado mental lo es más aún. Contar con un método objetivo y reproducible permitiría identificar sujetos en riesgo de posibles accidentes.

Una dificultad en la detección del nivel de fatiga de un sujeto es que no es posible determinar si un operador es capaz de realizar una tarea crítica solo analizando su aparente nivel de rendimiento en ella, puesto que este puede realizar un proceso mental compensatorio para que su desempeño en la tarea parezca normal [7].

A medida que una persona padece fatiga mental, aparecen los siguientes síntomas:

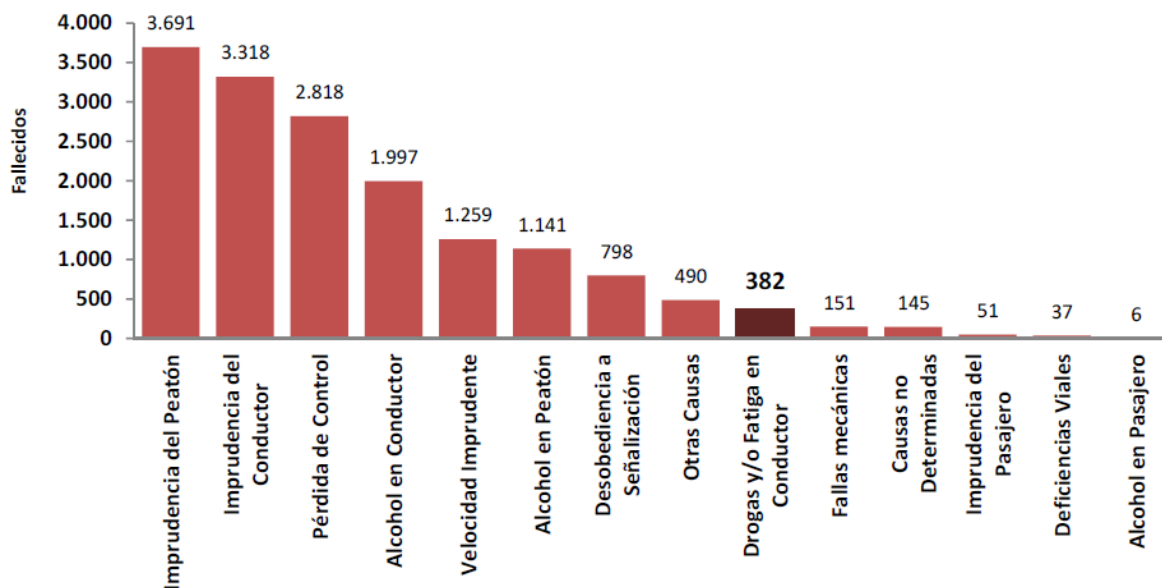
- Movimientos lentos o torpes
- Disminución de la velocidad motora de reacción
- Aparición de visión borrosa o doble
- Dificultad para concentrarse o permanecer alerta
- Dificultad para recordar

Esto implica que aquel individuo que bajo los efectos de la fatiga trabaje en rubros donde es necesario un alto nivel de atención, puede generar un accidente tal como un choque automovilístico.

2.1.2 Estadísticas de fatiga mental

Al analizar el problema de la fatiga mental en una tarea como la conducción, la cual requiere realizar una acción por un determinado periodo de tiempo sin realizar ninguna actividad extraña que pueda degradar el sistema (confiabilidad humana [1]), se puede encontrar que entre el año 2004 y 2013 se registraron un total de 4.645 accidentes de tránsito atribuidos directamente a fatiga mental o al uso de drogas (ambas causas aparecen en la misma categoría), entregando un saldo de 382 fallecidos, 1.367 heridos graves, 980 menos graves, 5.574 heridos leves, lo cual equivale a un total de 7.921 lesionados [8]. Estas cifras dejan en evidencia que la fatiga mental y/o el uso de drogas son la novena causa de muerte en accidentes de tránsito en Chile entre los años 2004 y 2013. En la Tabla 2.1 se puede observar el ranking de fallecidos por causa de accidentes de tránsito en Chile.

Tabla 2.1: Ranking de fallecidos por causa de accidentes de tránsito (2004-2013)



2.1.3 Nivel de somnolencia percibido por el individuo

De manera de poder establecer una relación entre el nivel real de somnolencia percibido por un individuo y los datos fisiológicos que pueden ser extraídos de este para estudiar su nivel de somnolencia, y como manera de evitar la detección de eventos de fatiga donde el sujeto se auto percibe alerta, existe la “Escala de Somnolencia Karolinska” (Karolinska Sleepiness Scale, KSS). Esta escala mide el nivel subjetivo de somnolencia de un individuo a determinada hora del día. En esta escala, la cual va del 1 al 9, los sujetos indican qué nivel refleja mejor el estado psicofísico experimentado en los últimos 10 minutos [9].

En la Tabla 2.2 se muestra la descripción de cada nivel de la escala.

Tabla 2.2: Escala de Somnolencia Karolinska [9].

| Índice | Estado psicofísico |
|---------------|--|
| 1. | Extremadamente alerta |
| 2. | Muy alerta |
| 3. | Alerta |
| 4. | Algo alerta |
| 5. | Ni alerta ni somnoliento |
| 6. | Algunos signos de somnolencia |
| 7. | Sueño, pero no hay esfuerzo para mantenerse despierto |
| 8. | Sueño, algo de esfuerzo para mantenerse despierto |
| 9. | Mucho sueño, gran esfuerzo para mantenerse despierto, luchando contra el sueño |

2.2 Electroencefalografía

La electroencefalografía (EEG) es un método electrofisiológico capaz de registrar la actividad bioeléctrica generada por las neuronas corticales del cerebro y su variación en el tiempo. La EEG se realiza colocando pequeños electrodos de disco sobre el cuero cabelludo de la cabeza o utilizando una tapa de electrodo. Los electrodos se encuentran conectados a un electroencefalograma, el cual amplifica los potenciales de las ondas cerebrales y su actividad [10].

Para asegurar la reproducibilidad estandarizada de la toma de señales de EEG, se utiliza el sistema internacional 10-20 [11]. Este sistema se basa en la relación entre la ubicación de un electrodo y el área subyacente de la corteza cerebral. En la Ilustración 2.1 se observa la disposición convencional del sistema internacional 10-20.

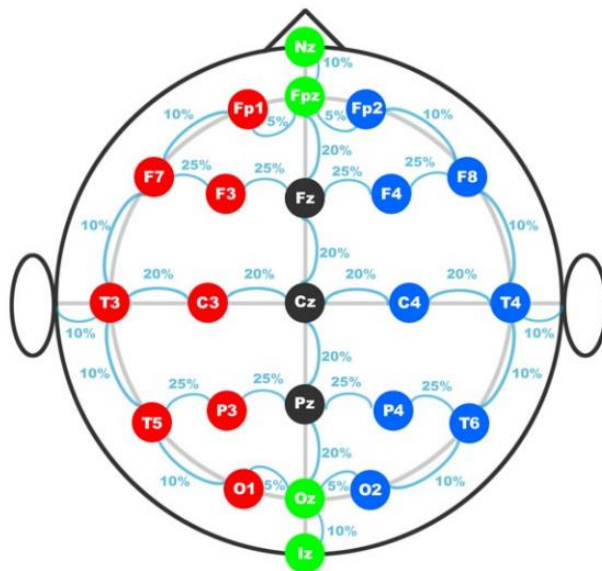


Ilustración 2.1: Sistema internacional de colocación de electrodos 10-20 [10].

Cada sitio de la corteza cerebral posee una letra para identificar el lóbulo correspondiente y un número para identificar la ubicación del hemisferio. En la Tabla 2.3 se muestra la correspondencia de cada carácter del sistema 10-20.

Tabla 2.3: Nomenclatura sistema 10/20 [10].

| Electrodo | Lóbulo |
|-----------|-----------|
| F | Frontal |
| T | Temporal |
| C | Central* |
| P | Parietal |
| O | Occipital |

No existe un lóbulo central, por lo que la letra “C” es utilizada solo con un propósito de identificación. La “Z” (de cero en inglés) hace referencia a los electrodos ubicados en la línea central. Los números pares representan los electrodos fijados en el hemisferio derecho, mientras que los impares hacen referencia a los electrodos fijados en el hemisferio izquierdo.

Se utilizan dos puntos de referencia anatómicos para el posicionamiento de los electrodos. En primer lugar “Nz”, el cual es el punto deprimido entre los ojos, “Iz”, el cual es el punto más bajo del cráneo en la parte posterior de la cabeza y que generalmente se nota por una protuberancia prominente.

2.2.1 Ondas cerebrales

Las señales EEG se pueden clasificar en cuatro segmentos principales dependiendo de su banda de frecuencia. Además, cada banda de frecuencia se relaciona de manera biyectiva con diferentes estados de consciencia, como estado de alerta, concentración intensa, somnolencia, etc. Las bandas de frecuencia se pueden clasificar como [10]:

- **Delta (0.5-4Hz):** su aparición se produce principalmente durante el sueño profundo. Hemisferio derecho totalmente activo.
- **Theta (4-8Hz):** este tipo de ondas están caracterizadas por un estado de somnolencia del individuo con consciencia reducida.
- **Alpha (8-13Hz):** las ondas alfa representan un estado de escasa actividad cerebral y relajación tanto física como mental, aunque consciente del entorno, por ejemplo, al descansar luego de realizar una tarea.
- **Beta (13-20Hz):** las ondas de este tipo se emiten cuando estamos en un estado consciente o alerta. Las ondas beta denotan una actividad mental intensa.

2.3 Análisis de componentes independientes

El análisis de componentes independientes (ICA) es un método computacional que permite obtener las señales originales de un proceso a partir de una mezcla de señales, como es el caso de la electroencefalografía [12].

El algoritmo utilizado para realizar el análisis de componentes independientes asume en primer lugar que las fuentes originales son temporalmente independientes, y que la mezcla es lineal, sin retardos e invariante en el tiempo. De esta manera el problema se plantea como: Sea $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]$ un conjunto de N fuentes independientes, y sea $x(t) = [x_1(t), x(t), \dots, x_M(t)]$ el conjunto de M señales observadas en los sensores. Utilizando ambos supuestos, las señales $x(t)$ pueden ser expresadas como:

$$x(t) = A * s(t)$$

Donde A es la matriz de mezcla. ICA intenta obtener un conjunto de señales $y(t) = [y_1(t), y_2(t), \dots, y_M(t)]$ que resulte ser una estimación de las señales $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]$ a partir de las señales observadas:

$$y(t) = B * x(t)$$

Donde B es la matriz de recuperación, la cual se calcula maximizando una función de contraste $i(y)$, la cual es una medida de independencia de las señales $y(t)$.

Mediante esta técnica, es posible limpiar las señales de electroencefalografía, obteniendo como resultado registros más precisos de cada canal o fuente.

2.4 Transformada de Fourier

La transformada de Fourier es una transformación matemática la cual permite la transformación de una señal $f(t)$ con dominio en el tiempo en una función $F(w)$, la cual se encuentra definida en el dominio de la frecuencia. Además, el teorema de Fourier afirma que cualquier señal periódica puede ser modelada como la suma de sinusoides de diferentes frecuencias y amplitudes, donde a cada una de ellas se les puede aplicar la transformada de Fourier, debido a la propiedad de linealidad que posee la misma [13].

Matemáticamente la transformada se expresa como:

$$F(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt$$

Donde $F(w)$ representa la señal en el dominio de la frecuencia, $f(t)$ es la señal en el dominio del tiempo, t es el tiempo, w la frecuencia y j representa el número imaginario.

En caso de que la señal a tratar no cumpla con el requisito de ser periódica, se puede realizar la transformación de la señal utilizando un intervalo de tiempo auto impuesto, asumiendo que dentro de este la señal se comporta de manera periódica. De esta manera se puede deslizar dicho intervalo de tiempo a través de la señal repitiendo el proceso según corresponde. Este método se conoce como transformada corta de Fourier (STFT).

2.5 Espectrogramas

Los espectrogramas son una representación visual de la transformada corta de Fourier (STFT), donde el eje X y el eje Y representan la escala temporal de la señal y la frecuencia

respectivamente, mientras que la escala de color de la imagen representa la amplitud de la señal en ese punto.

La principal ventaja de los espectrogramas es su capacidad de incluir el tiempo en el análisis espectral que realiza la STFT, lo cual permite solucionar el problema de analizar las frecuencias y el dominio en el que se presentan. En la Ilustración 2.2 y en la Ilustración 2.3 se puede observar una señal de EEG en el dominio del tiempo y el espectrograma de una señal de EEG respectivamente [13].

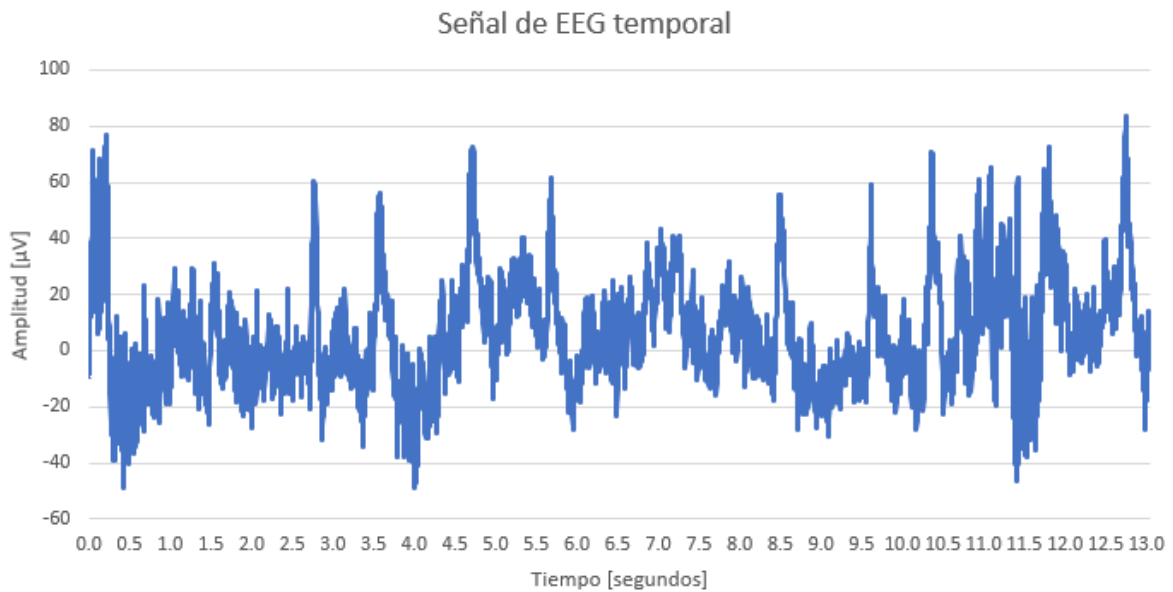


Ilustración 2.2: Señal de electroencefalografía en el dominio del tiempo.

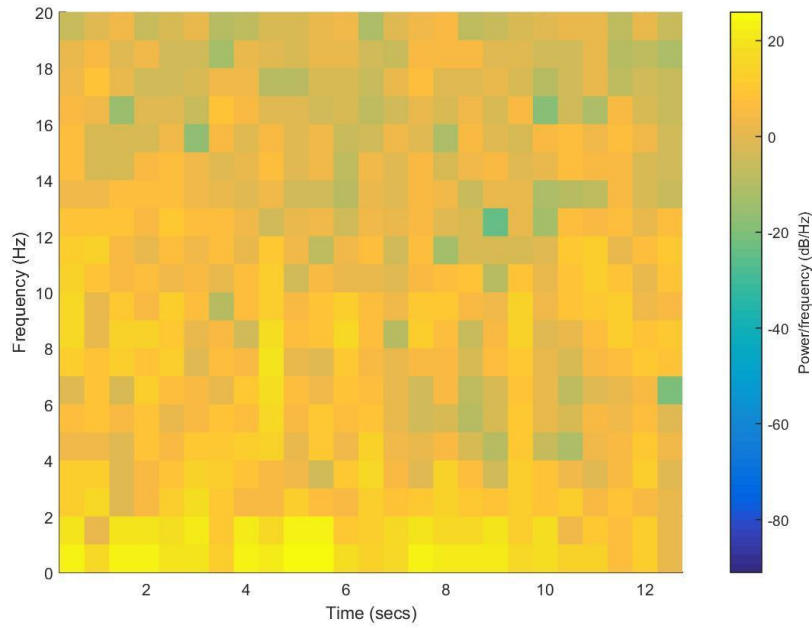


Ilustración 2.3: Espectrograma de una señal de EEG.

2.6 Support Vector Machines

Support Vector Machines (SVM) es un método de clasificación el cual se basa en minimizar el error cuadrático de la función de clasificación, la cual corresponde al hiperplano que separa de mejor manera los datos disponibles. En la Ilustración 2.4 se observa el hiperplano que separa dos clases (cuadrados y círculos).

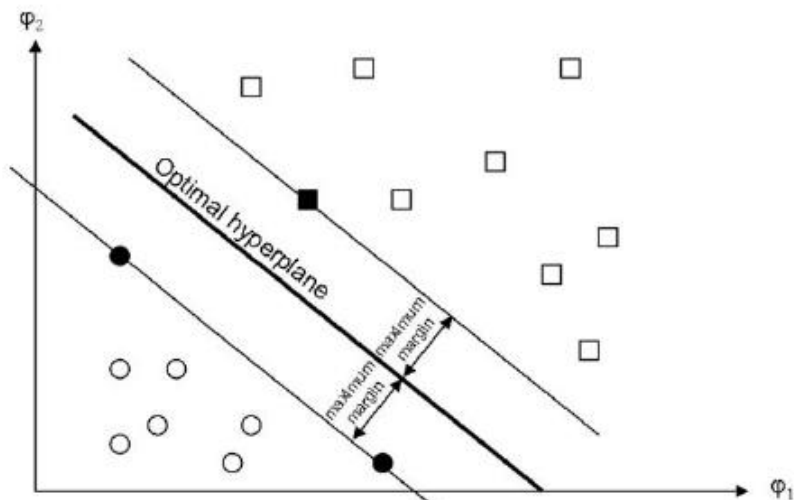


Ilustración 2.4: Clasificación mediante SVM. Las figuras en negra representan los hiperplanos canónicos [14].

El funcionamiento de SVM se basa en maximizar la distancia entre el hiperplano óptimo y los hiperplanos canónicos, los cuales corresponden a los bordes de cada clase. Para este objetivo se minimiza la norma euclidiana de w , que corresponde a los coeficientes que definen el hiperplano. Este problema se puede formular matemáticamente como [14]:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{s.a. } y_i(w^T * x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N$$

$$\xi_i \geq 0, \quad i = 1, \dots, N$$

Donde $x_i \in \mathbb{R}^M$, $y_i \in \{-1, 1\}$, y $\xi_i (i = 1, \dots, N)$ son variables de holgura que permiten la ocurrencia de errores, los cuales son penalizados en la función objetivo. Dicha penalización está representada por el parámetro C .

2.7 Random Forest

Random Forest es un algoritmo de clasificación el cual consiste en una colección de árboles estructurados paralelamente en distintos niveles de profundidad, donde cada uno de los árboles de un mismo nivel realiza una clasificación de la base de datos de manera independiente, utilizando un mismo criterio. De esta forma, cada árbol entrega un voto acerca de la clase de cada elemento del conjunto de datos. Este proceso se repite en cada nivel de profundidad utilizando diferentes criterios de clasificación, hasta que finalmente la clasificación de cada elemento se realiza de acuerdo con el número de votos que posee cada elemento [15].

2.8 Redes neuronales

2.8.1 Redes neuronales convencionales

Las redes neuronales convencionales (NN) son un modelo computacional que consta de un grupo interconectado de nodos (neuronas), cada uno de las cuales se encuentra conectado a muchos otros, intentando emular así el funcionamiento de una red neuronal

biológica. Las NN son ampliamente utilizadas para resolver problemas de clasificación de datos.

La neurona es la estructura básica de una red neuronal. Cada neurona al recibir información, de una neurona anterior o del exterior, realiza una función de suma de todos los valores de entrada x_i multiplicados por su peso w_i . El resultado de este cálculo es ingresado a una función de activación, de modo que la señal debe sobrepasar un límite antes de propagarse. En la Ilustración 2.5 se observa el funcionamiento básico de una neurona [4].

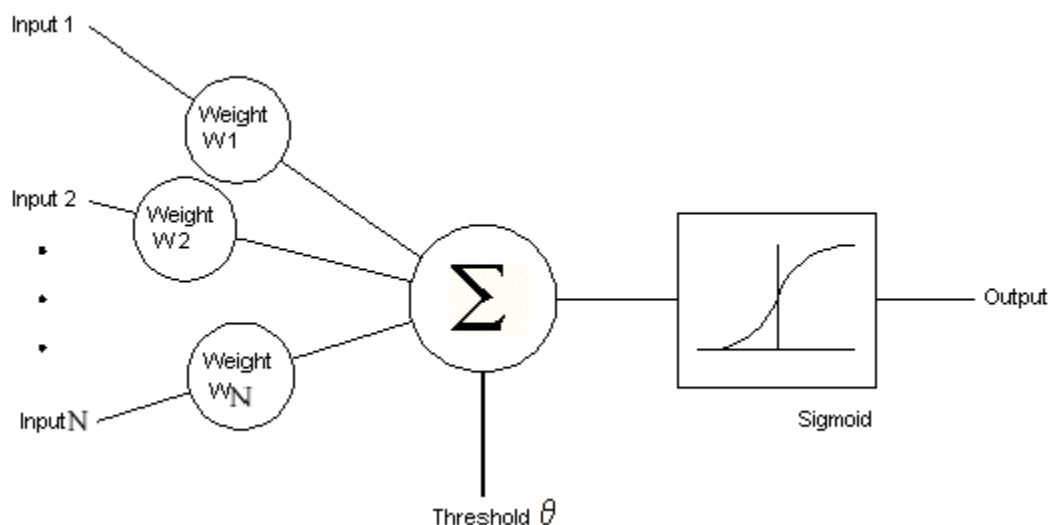


Ilustración 2.5: Funcionamiento básico de una neurona artificial [31].

2.8.1.1 Arquitectura de una red neuronal

La arquitectura básica de una red neuronal consta de tres capas. La primera de ellas es la capa de entrada (*Input Layer*), en la cual se recibe la información que se desea procesar y se realiza una primera evaluación de esta. Posteriormente, las neuronas de la capa de entrada envían su información a cada una de las neuronas de la segunda capa o también llamada capa oculta (*Hidden Layer*), de las cuales puede existir más de una dependiendo de la complejidad del problema que se desea abordar. Finalmente, luego de la o las capas ocultas se encuentra una capa final (*Output Layer*), la cual entrega los resultados de la clasificación realizada por la red. En la Ilustración 2.6 se muestra la arquitectura básica de una red neuronal totalmente conectada [4].

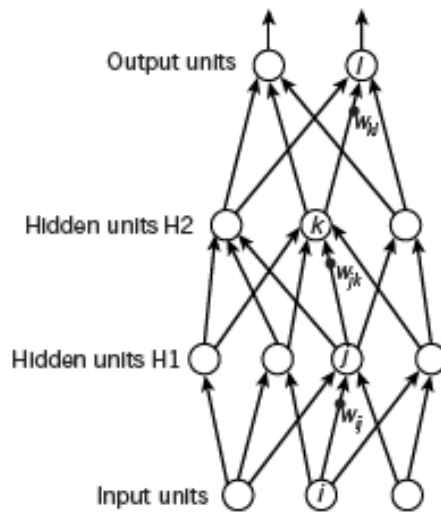


Ilustración 2.6: Arquitectura de una red neuronal totalmente conectada [4].

2.8.1.2 Función de activación

El cálculo realizado por cada neurona se puede escribir de manera abreviada como:

$$f(x_i, w_i) = \phi\left(\sum_{i \in I} (x_i * w_i)\right)$$

Donde ϕ representa la función de activación. Existen diversos tipos de funciones de activación, sin embargo, las más comunes son las siguientes [16]:

Linear $\phi(x)=x$

Binaria $\phi(x) = \begin{cases} x^2, & x < 0 \\ x^3, & x \geq 0 \end{cases}$

Sigmoidea $\phi(x) = \frac{1}{1+e^{-x}}$

Tangente hiperbólica $\phi(x) = \tanh x$

Rectificación de unidad linear (ReLU) $\phi(x) = \max(0, x)$

2.8.1.3 Aprendizaje supervisado de una red neuronal

La etapa de aprendizaje de una red neuronal utiliza el método de *backpropagation* (propagación hacia atrás). En este se ingresa un conjunto de datos (*input*) categorizados según su clase. De esta manera, este *input* se propaga desde la primera capa hasta la capa final, pasando por cada una de las capas ocultas. El *output* obtenido por la capa final es comparado con la salida deseada (definida por las clases), obteniendo así un error para cada una de las salidas, el cual puede expresarse como:

$$J(w) = - \sum_i y_i \log(p_i)$$

Donde y_i corresponde al *output* i-ésimo deseado, p_i a la predicción i-ésima y la función $J(w)$ se representa sobre un espacio multidimensional de dimensión igual al número de pesos de la red [17].

Posteriormente, el error es propagado hacia atrás (*backpropagation*), partiendo desde la capa de salida, pasando por cada una de las capas ocultas hasta llegar a la capa de entrada. La búsqueda del conjunto de pesos que minimiza el error $J(w)$ se realiza mediante la siguiente ecuación:

$$\Delta w_j = -\alpha \frac{\partial J(w)}{\partial w_j}$$

Con w_{ij} el peso de la neurona ij que se busca optimizar y α el hiper parámetro de la red correspondiente a la velocidad de aprendizaje.

Al realizar esto, cada una de las capas intermedias recibe solo una porción del error total, el cual corresponde a la contribución relativa de la neurona correspondiente.

Este proceso es repetido hasta que el peso de cada neurona corresponde al peso que produce el menor error de salida, el cual es obtenido mediante la optimización de la función de error.

2.8.2 Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) son muy similares a las redes neuronales convencionales. Estas están compuestas por neuronas, cada una de las cuales posee pesos, y es activada mediante una función de activación, además de poseer una función de error a la salida de la última capa.

La principal diferencia de este tipo de red neuronal es que asume explícitamente que los datos de entrada (*input*) son imágenes, lo cual permite reducir la cantidad de parámetros de una red.

Generalmente, las CNN están construidas mediante tres diferentes tipos de capas, las cuales son:

1. **Capa convolucional:** la cual permite obtener un mapeo de las características de la imagen.
2. **Capa de *pooling*:** la cual permite reducir la dimensionalidad al seleccionar las características más comunes.
3. **Capa clasificadora:** la cual actúa de manera idéntica a la de una red neuronal convencional, clasificando así los *inputs* entregados.

2.8.2.1 Capa convolucional

En esta primera fase, se realiza una convolución entre la imagen de entrada (*input*) y una función de filtro (también llamado kernel), la cual busca extraer características específicas de la imagen. De esta manera se logra filtrar la imagen con una membrana previamente entrenada. Esta técnica permite que ciertas características se vuelvan más dominantes en la imagen de salida, debido a que poseen un peso más elevado en los píxeles que los representan.

La convolución realizada entre las funciones mencionadas es abordada numéricamente como un producto matricial, lo cual permite trabajar con entradas de tamaño variable [17]. En la Ilustración 2.7 se ilustra el funcionamiento de una capa convolucional a modo de ejemplo.

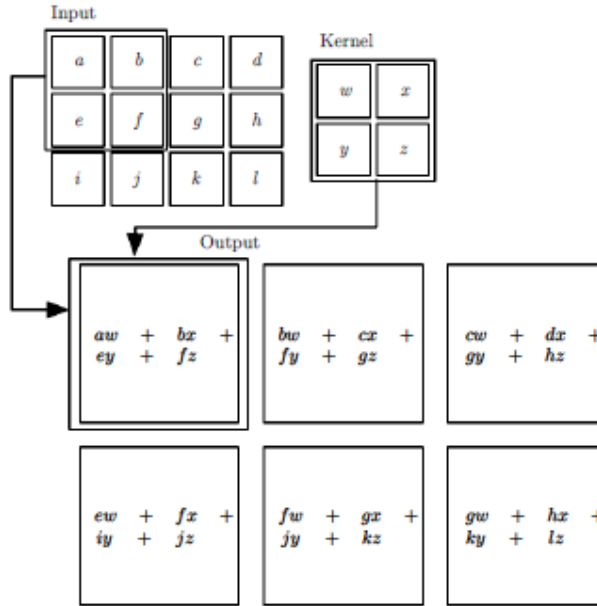


Ilustración 2.7: Ejemplo de convolución para una imagen 2D [17].

2.8.2.2 Capa de pooling

Generalmente la capa de *pooling* se encuentra después de la capa de convolución, y la función de esta es realizar una reducción de muestreo, lo cual implica paralelamente una pérdida de información. Sin embargo, esta pérdida de información es beneficiosa para la red, debido a que conduce a una menor sobrecarga de cálculo para las capas sucesoras de la red.

Usualmente se suele usar la función *max-pooling* en esta capa. Esta función encuentra el valor máximo entre una ventana de muestra, guardando solo este valor para la siguiente capa [17]. En la Ilustración 2.8 se puede observar gráficamente la operación que realiza la función *max-pooling*.

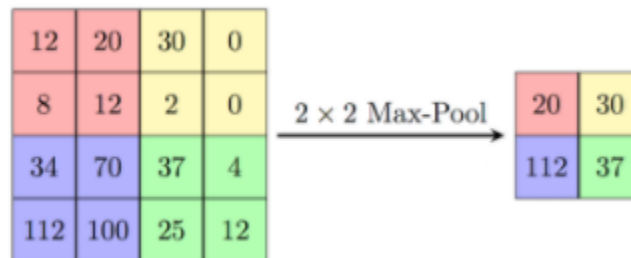


Ilustración 2.8: Diagrama ilustrativo de la función *max-pooling* [17].

2.8.2.3 Capa clasificadora

Luego de las capas de convolución y de *pooling*, se utilizan capas de neuronas completamente conectadas, en las que cada pixel de la imagen se considera como una neurona separada. Esta capa funciona de manera idéntica a la de una red neuronal convencional, sin embargo, gracias a la reducción de dimensionalidad y extracción de características predominantes realizadas por las dos capas previas, esta capa es capaz de realizar una clasificación más exacta.

2.8.3 Entrenamiento de una red neuronal

Para realizar el entrenamiento de una red neuronal en primer lugar se debe separar el conjunto de datos a utilizar en dos grupos, los cuales son entrenamiento (*train*) y prueba (*test*). Los datos de entrenamiento son utilizados para lograr que la red aprenda a clasificar la tarea deseada, lo cual es realizado mediante la asignación de pesos (w_k) a cada una de las neuronas que componen la red mediante la técnica de *backpropagation*. De esta manera la red en primera instancia aprende a clasificar al conjunto de datos de entrenamiento. Posteriormente se debe alimentar la red con el conjunto de datos de prueba, los cuales son desconocidos por el grafo, de modo de evaluar el desempeño presentado por el algoritmo.

Un problema recurrente al realizar el entrenamiento de una red es el sobreajuste (*overfitting*) de esta a los datos de entrenamiento, es decir, la red aprende a clasificar de manera acertada los datos con los que fue entrenada, sin embargo, pierde generalidad a la hora de evaluar datos desconocidos (datos de prueba). La consecuencia más obvia del sobreajuste es el desempeño deficiente presentado por la red (lo que equivale a un mayor error para los datos de prueba o una menor exactitud de clasificación), sin embargo este problema también implica un mayor requerimiento de información innecesaria sobre cada dato, lo cual se traduce en un mayor costo y tiempo de entrenamiento [18].

Para evitar este problema existen principalmente dos métodos, *Dropout* y *regularización L2*.

2.8.3.1 Dropout

Esta técnica consiste en “apagar” cada uno de los nodos de la red de forma individual en cada una de las etapas de entrenamiento con probabilidad $1 - p$ (con p la probabilidad de mantener encendido el nodo). De esta manera, en cada etapa del entrenamiento, la red cuenta con una cantidad reducida de nodos (determinada por p). Al entrenar la red en cada iteración con solo una fracción de los nodos totales, la red se ve forzada a generar una forma más robusta de realizar la clasificación, logrando una mayor generalización para datos nuevos, y así evitando el sobreajuste [19].

Cuando se evalúa el desempeño de la red con los datos de prueba no se aplica *Dropout* a los nodos, buscando obtener así el máximo desempeño de la red. El valor estándar de p que suele utilizarse es de 0.5, sin embargo, este puede variar.

En la Ilustración 2.9 se puede observar gráficamente la consecuencia del *Dropout* en una red neuronal artificial.

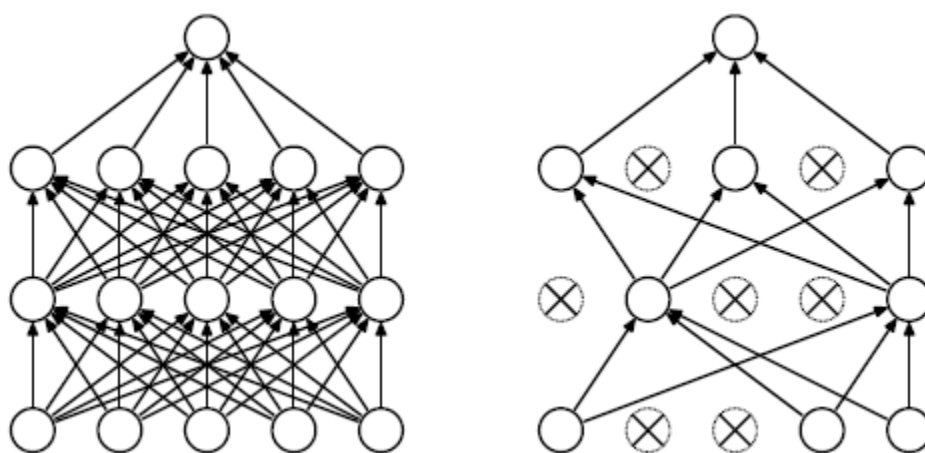


Ilustración 2.9: Izquierda: ejemplo de red neuronal artificial. Derecha: ejemplo de una red neuronal convencional luego de aplicar Dropout. Los nodos cruzados fueron apagados debido al Dropout [19].

2.8.3.2 Regularización L2

La regularización L2 busca evitar el sobreajuste de la red incorporando una penalización a la función de error que busca optimizar el algoritmo, de esta manera se previene que pesos individuales puedan tener demasiada influencia sobre la clasificación final. La penalización impuesta corresponde a la sumatoria de los cuadrados de cada uno de los pesos de la red multiplicados por un factor β de ajuste, el cual suele tener un orden entre 0,1 y 0,00001. Matemáticamente la regularización puede escribirse como [20]:

$$J'(w) = J(w) + \beta \sum_{j=i}^n w_j^2$$

Con $J'(w)$ la nueva función de costo, $J(w)$ la función de costo original, β el factor de ajuste de la regularización y w los pesos de cada una de las neuronas de la red.

2.9 Métricas de rendimiento para tareas de clasificación

Con el objetivo medir el rendimiento de un clasificador se puede utilizar una matriz de confusión, la cual da cuenta de cada uno de los aciertos y desaciertos realizados por el clasificador mediante la contabilización de cuatro variables, las cuales se describen a continuación [21]:

- **Verdaderos positivos (VP):** elementos clasificados como verdaderos y que realmente son verdaderos.
- **Verdaderos negativos (VN):** elementos clasificados como negativos y que realmente son negativos.
- **Falsos positivos (FP):** elemento clasificado como positivo y que realmente es negativo.
- **Falsos negativos (FN):** elemento clasificado como negativo y que realmente es positivo.

En la Tabla 2.4 se muestra la estructura de una matriz de confusión para una clasificación binaria.

Tabla 2.4: Diagrama matriz de confusión

| | Clasificados como positivos | Clasificados como negativos |
|------------------|-----------------------------|-----------------------------|
| Reales positivos | Verdaderos Positivos | Falsos Negativos |
| Reales negativos | Falsos Positivos | Verdaderos Negativos |

A partir de la matriz de confusión se pueden definir las siguientes métricas de desempeño:

- **Exactitud:** se define como la tasa de aciertos totales que obtiene un clasificador, y da cuenta de la eficacia general del clasificador:

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN}$$

- **Sensibilidad:** se define como la tasa entre los verdaderos positivos y las muestras que efectivamente son positivas, y da cuenta de la eficacia del clasificador para detectar muestras que efectivamente son positivas:

$$Sensibilidad = \frac{VP}{VP + FN}$$

- **Especificidad:** se define como la tasa entre los verdaderos negativos y las muestras que efectivamente son negativas, y da cuenta de la eficacia del clasificador para detectar muestras que efectivamente son negativas:

$$Especificidad = \frac{VN}{FP + VN}$$

- **Precisión:** se define como la tasa entre los verdaderos positivos y el número de muestras predichas como positivas, y da cuenta de la coincidencia entre las etiquetas efectivas con las muestras positivas entregadas por el clasificador:

$$Precisión = \frac{VP}{VP + FP}$$

- **Puntaje F1:** se define como el medio armónico entre la sensibilidad y la especificidad, y da cuenta de la relación entre las etiquetas efectivamente positivas y las entregadas por el clasificador:

$$Puntaje F1 = \frac{2 * VP}{2 * VP + FP + FN}$$

3. Metodología

El presente capítulo da cuenta del procedimiento realizado para llevar a cabo la detección de somnolencia, sin embargo, previo a esto se expone un análisis de prefactibilidad de análisis de espectrogramas de EEG mediante CNN, el cual se realizó mediante la clasificación de señales de EEG de sujetos con ojos abiertos y sujetos con ojos cerrados.

3.1 Prefactibilidad de análisis de espectrogramas de EEG mediante CNN

De modo de establecer la factibilidad del reconocimiento de espectrogramas de señales de electroencefalografías mediante la utilización de redes neuronales convolucionales, se construyó un clasificador de señales de EEG pertenecientes a sujetos con los ojos cerrados y sujetos con los ojos abiertos.

Este problema de clasificación se encuentra ampliamente estudiado en la literatura, por lo cual se sabe que existe un aumento significativo en la amplitud de las ondas alfa del cerebro al cerrar los ojos [22]. Teniendo en cuenta este factor, se procedió a generar un clasificador, siguiendo la metodología expuesta a continuación.

3.1.1 Preprocesamiento de las señales

Para llevar a cabo el clasificador, se contó con una base de datos de 95 registros de señales de EEG de sujetos en estado de relajación, de los cuales 58 correspondían a sujetos con los ojos abiertos y 37 a sujetos con los ojos cerrados. Las señales fueron tomadas utilizando un equipo de electroencefalografía de 20 canales siguiendo el sistema internacional 10/20. La frecuencia de muestreo del equipo era de 200 Hz, y las señales poseían una duración promedio de 5 minutos.

En primer lugar, se realizó un pre procesamiento a la base de datos utilizando el software de Matlab “EEGLAB”. Utilizando EEGLAB se reconstruyó cada señal utilizando el método de análisis de componentes independientes. Esto permitió que, mediante un análisis visual, se pudiesen eliminar canales contaminados con el ruido producido por eventos como el parpadeo de los ojos [23]. En la Ilustración 3.1: se puede observar la visualización de un canal contaminado por el parpadeo de los ojos en el software, donde el color rojo indica una alta amplitud de las ondas captadas por los electrodos correspondientes.

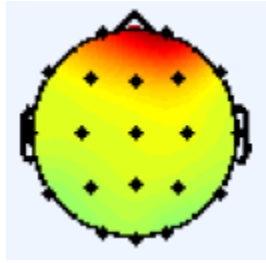


Ilustración 3.1: Visualización canal contaminado por el parpadeo de los ojos mediante EEGLAB.

Posteriormente, con el objetivo de aumentar el tamaño de la base de datos disponibles para generar el modelo deseado, cada señal fue segmentada en ventanas de 5 segundos, los cuales contienen toda la información referente a cada clase [24].

Con la base de datos segmentada, se alcanzó un total de 4.118 datos, de los cuales 2.600 correspondían a sujetos con los ojos abiertos y 1.518 a sujetos con los ojos cerrados.

Luego se procedió a obtener el espectrograma del canal *O2* para cada uno de los 4.118 datos segmentados, utilizando el software Matlab. Los espectrogramas fueron obtenidos utilizando una ventana de 256 puntos, una superposición del 50% y solo se consideraron las frecuencias de 0 a 40 Hz [25]. Además, cada imagen fue obtenida en una dimensión de 96x96 píxeles, y a cada espectrograma le fue eliminada tanto la barra de color como la etiqueta de los ejes, debido a que estos elementos solo representan ruido para el clasificador.

Una vez obtenidas las imágenes a utilizar en la CNN, estas fueron convertidas a escala de grises.

3.1.2 Implementación de CNN para análisis de espectrogramas de EEG

Finalmente, se construyó una CNN utilizando una de las arquitecturas propuesta por David Verstraete [26]. La arquitectura utilizada constó de 11 capas, de las cuales 6 correspondían a capas convolucionales, 3 a capas de *pooling* y 2 capas totalmente conectadas.

El detalle de la arquitectura utilizada se muestra a continuación, donde las *C* corresponden a las capas convolucionales, las *P* corresponden a las capas de *pooling* y *FC* corresponde a las capas totalmente conectadas. El número precedente a cada una de las letras corresponde a la cantidad de filtros aplicados en cada capa, mientras que los números entre corchetes corresponden al tamaño de los filtros o la cantidad de neuronas de cada capa según corresponde.

Arquitectura utilizada:

Input[96x96]-32C[3x3]-32C[3x3]-32P[2x2]-64C[3x3]-64C[3x3]-64P[2x2]-128C[3x3]-28C[3x3]-128P[2x2]-FC[100]-FC [100]

Para el entrenamiento de la red, la base de datos se dividió en un 90% y un 10%, para el entrenamiento y la prueba respectivamente. Además, la red fue entrenada mediante la utilización de lotes de sesenta y cuatro datos cada uno, para de esta forma acelerar la velocidad de entrenamiento de la red [27].

Paralelamente, para evitar el sobreajuste de la red se utilizó un *Dropout* de 0,5 y regularización L2 con parámetro β de 0,001.

En la Tabla 3.1 se muestran los hiper parámetros utilizados para entrenar la red.

Tabla 3.1: Hiper parámetros utilizados para la red.

| Hiper parámetro | Valor |
|--------------------------|--------|
| Tasa de aprendizaje | 0,0001 |
| Función de activación | ReLU |
| Tamaño de los lotes | 128 |
| <i>Dropout</i> | 0,5 |
| Beta (regularización L2) | 0,001 |
| Épocas | 100 |

3.1.3 Resultados obtenidos

En la Tabla 3.2 se muestra el resumen de los resultados obtenidos para clasificación de sujetos con ojos abiertos y sujetos con ojos cerrados mediante la utilización de redes neuronales convolucionales, utilizando como dato de entrada los espectrogramas del canal O2.

Tabla 3.2: Resumen de resultados para la clasificación de sujetos con ojos abiertos y sujetos con ojos cerrados mediante CNN.

| Exactitud | Puntaje | | | |
|-----------|---------|--------------|---------------|-----------|
| | F1 | Sensibilidad | Especificidad | Precisión |
| 98,54% | 98,85% | 99,61% | 96,71% | 98,10% |

3.1.4 Análisis de resultados

Al observar las métricas expuestas en la Tabla 3.2, es evidente notar el alto desempeño alcanzado por el clasificador realizado. Con esto se evidencia no solo la factibilidad de realizar un algoritmo de clasificación de señales de EEG mediante el uso de espectrogramas, si no también que el uso de esta técnica presenta un rendimiento superior al de otros algoritmos utilizados para realizar el mismo problema de clasificación [28].

Además, esto revela que los espectrogramas son capaces de aportar por si solos información referente al estado de un individuo, presentándose como una alternativa a la extracción de características de una señal.

3.2 Detección de sujetos bajo estado de fatiga

3.2.1 Base de datos

Para lograr llevar a cabo el modelo de detección de somnolencia, se utilizó la base de datos de somnolencia multimodalidad ULg (*ULg Multimodality Drowsiness Database*), también denominada *DROZY*. Esta base datos contiene varios tipos de datos relacionados con la fatiga y/o somnolencia como imágenes y señales entre otros, y su objetivo es servir como herramienta para desarrollar y evaluar sistemas de monitoreo de fatiga.

La base de datos *DROZY* fue obtenida a partir de 14 sujetos jóvenes y sanos (11 mujeres y 3 hombres), los cuales realizaron tres pruebas de vigilancia psicomotora (PVT) sucesivas de 10 minutos de duración en condiciones de privación de sueño inducido mediante un estado de vigilia prolongada. Cabe destacar que previo al primer experimento los sujetos tuvieron una noche de sueño normal. En la Ilustración 3.2 se puede observar el cronograma utilizado para la recolección de la base de datos *DROZY*.

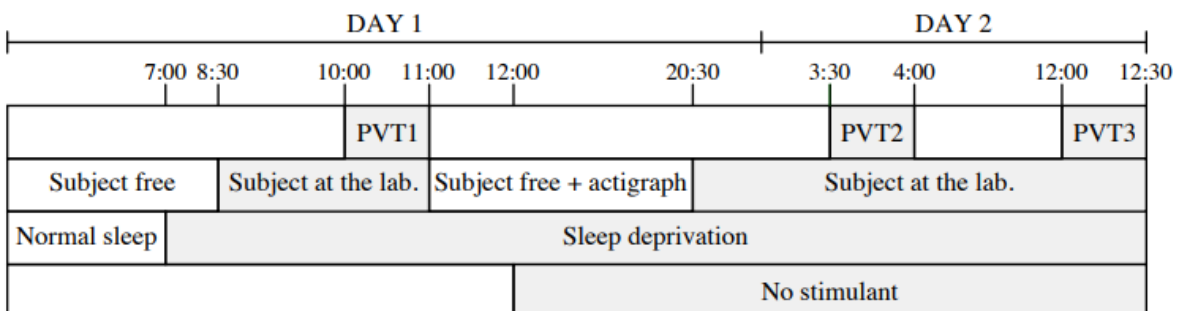


Ilustración 3.2: Resumen del cronograma de recolección de datos de la base de datos *DROZY* [29].

Para cada sujeto y prueba de vigilancia psicomotora, la base de datos contiene los siguientes datos perfectamente sincronizados en el tiempo [29]:

- **Puntuaciones de la escala de somnolencia Karolinska (KSS):** al comienzo de cada PVT cada sujeto autoevaluó su nivel subjetivo de somnolencia mediante la KSS.
- **Datos PVT:** tiempos de estímulo y reacción para cada sujeto y prueba.
- **Señales de polisomnografía (PSG):** cinco canales de EEG (Fz, Cz, Pz, C3 y C4), dos canales (horizontal y vertical) de electrooculografía (EOG), electrocardiograma (ECG) y electromiografía (EMG), todos muestreados a 512 Hz.
- **Imágenes de sensor Kinect v2:** intensidad de infrarrojo cercano e imágenes de rango con un tamaño de 512x424 píxeles, escala de grises y valores de distancia de 16 bits y velocidad de cuadro de 30 cuadros por segundo.
- **Videos del sensor Kinect v2:** videos de intensidad de infrarrojo cercano (1 por PVT) con un tamaño de 512x424 píxeles y valores de escala de grises de 8 bits.
- **Señales de rostro (obtenidas manualmente):** Anotaciones 2D (en píxeles) y 3D (en milímetros) de 68 marcas de rostro para 720 marcos seleccionados a mano.
- **Señales de rostro (obtenidas automáticamente):** anotaciones 2D y 3D de 68 puntos de referencia de rostro para todos los marcos, resultantes del procesamiento automático utilizando modelos locales limitados específicos de tema.

Para el desarrollo del clasificador deseado, se utilizó únicamente las señales de EEG y las puntuaciones de la escala de somnolencia Karolinska.

3.2.2 Preprocesamiento de datos y selección de etiquetas

De los cinco canales de EEG disponibles en la base de datos, se decidió utilizar solo los canales *Fz* y *Pz* para la generación del modelo, debido a la relación existente entre la fatiga mental y el incremento de amplitud de las ondas theta en el lóbulo frontal del cerebro y el aumento de amplitud de las ondas alfa en el lóbulo parietal del cerebro bajo esta condición [2].

Posterior a esta elección se procedió a generar las clases con las cuales se generarían los modelos de clasificación deseados. Para esto se hizo uso de las puntuaciones de la escala de somnolencia Karolinska indicadas por cada sujeto antes de cada PVT. De esta manera y siguiendo la descripción de cada índice de la escala, se definió de manera preliminar como

estado base (sin somnolencia) a los sujetos que presentaban un índice KSS menor o igual a 4, y como estado de somnolencia a los sujetos que presentaban un índice KSS mayor o igual a 7 [9].

Sin embargo, debido a que el índice KSS es un indicador subjetivo señalado por cada sujeto antes de cada prueba, y con el objetivo de generar clases lo más distantes entre ellas (a modo de evitar el solapamiento de clases), se eliminaron todos los datos correspondientes al PVT 2, dado que como se muestra en Tabla 3.3, la media de las puntuaciones de dicha prueba no corresponden a ninguna de las clases previamente expuestas.

Tabla 3.3: Promedio y desviación estándar del índice KSS por PVT.

| | PVT 1 | PVT 2 | PVT 3 |
|--------------------------------|-------|-------|-------|
| Promedio KSS | 3,2 | 5,6 | 7,2 |
| Desviación Estándar KSS | 1,3 | 1,7 | 1,3 |

Finalmente se definió el estado base con todos los sujetos del PVT 1 que presentaran un KSS menor o igual a 4 y el estado de somnolencia con todos los sujetos del PVT 3 que presentaran un KSS mayor o igual a 7. Cabe destacar que en ningún caso existieron sujetos con un índice KSS correspondiente al estado de somnolencia en el PVT 1 o viceversa.

Una vez que la base de datos fue acotada y las clases definidas, se procedió a segmentar las señales de los canales Fz y Pz , con el objetivo de aumentar el número de datos disponibles. En primer lugar, cada canal fue segmentado en archivos de trece segundos de largo, dado que esa es la duración mínima que debe poseer una señal de EEG para contener toda la información referente a la fatiga mental del individuo [2]. Luego de realizar esta segmentación se obtuvieron un total de 460 archivos para cada clase, constituyendo así un total de 920 archivos para el banco de datos de las señales de trece segundos.

Paralelamente se segmentaron las señales en archivos de cinco segundos de largo, de modo de poder evaluar la relevancia del largo de la señal con la cantidad de archivos disponibles para realizar la clasificación. Con este largo de señal se obtuvo un total de 1200 archivos para cada clase, obteniendo de esta manera 2400 archivos para el banco de datos de las señales de cinco segundos.

Posteriormente, se procedió a generar los espectrogramas de cada uno de los archivos (señales segmentadas) para cada uno de los bancos de datos generados previamente, mediante el uso del software Matlab 2016. Los espectrogramas fueron obtenidos utilizando una ventana de 512 puntos, una superposición del 50% y solo se consideraron las frecuencias de 0 a 20 Hz, debido a que las ondas cerebrales relevantes para el problema a solucionar se encuentran en este rango [2]. Además, cada imagen fue obtenida en una dimensión de 96x96 píxeles, y a cada espectrograma le fue eliminada tanto la barra de color como la etiqueta de los ejes, debido a que estos elementos solo representan ruido para el clasificador. En la

Ilustración 3.4 se puede un observar un espectrograma utilizada, al cual se le removieron los ejes y la barra de colores

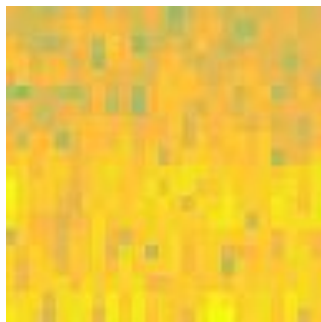


Ilustración 3.4: Ejemplo de espectrograma utilizado con dimensiones de 96x96 pixeles, sin barra de colores ni etiquetas en los ejes.

Luego cada uno de los espectrogramas generados fue convertido desde formato RGB a escala de grises, de modo que la matriz representativa de cada imagen posea una dimensión de dos y no de tres. En la Ilustración 3.3 se puede observar un espectrograma luego de ser convertido a escala de grises.



Finalmente, luego de pre procesar toda la base de datos y de generar las etiquetas
Ilustración 3.3: Ejemplo de espectrograma de EEG en escala de grises.

correspondientes, se obtuvieron seis conjuntos de espectrogramas (todos en escala de grises) a usar para lograr realizar la detección de sujetos bajo estado de fatiga, los cuales se describen a continuación:

- 1. Espectrogramas canal Fz duración 13 segundos:** conjunto de espectrogramas del canal Fz con una duración de trece segundos y un rango de frecuencias 0-20 Hz, cada uno de los cuales con una dimensión de 96x96x1 pixeles.
- 2. Espectrogramas canal Pz duración 13 segundos:** conjunto de espectrogramas del canal Pz con una duración de trece segundos y un rango de frecuencias 0-20 Hz, cada uno de los cuales con una dimensión de 96x96x1 pixeles.
- 3. Espectrogramas canales Fz - Pz duración 13 segundos:** conjunto de espectrogramas de los canales Fz y Pz con una duración de trece segundos y un rango de frecuencias 0-20 Hz. Para este conjunto de datos, se trabajó con imágenes en tres dimensiones,

donde existían dos niveles de profundidad, cada uno de los cuales correspondía a los espectrogramas de los canales Fz y Pz respectivamente, teniendo una dimensión de $96 \times 96 \times 2$ píxeles.

4. **Espectrogramas canal Fz duración 5 segundos:** conjunto de espectrogramas del canal Fz con una duración de cinco segundos y un rango de frecuencias 0-20 Hz, cada uno de los cuales con una dimensión de $96 \times 96 \times 1$ píxeles.
5. **Espectrogramas canal Pz duración 5 segundos:** conjunto de espectrogramas del canal Pz con una duración de cinco segundos y un rango de frecuencias 0-20 Hz, cada uno de los cuales con una dimensión de $96 \times 96 \times 1$ píxeles.
6. **Espectrogramas canales Fz - Pz duración 5 segundos:** conjunto de espectrogramas de los canales Fz y Pz con una duración de cinco segundos y un rango de frecuencias 0-20 Hz. Para este conjunto de datos, se trabajó con imágenes en tres dimensiones, donde existían dos niveles de profundidad, cada uno de los cuales correspondía a los espectrogramas de los canales Fz y Pz respectivamente, teniendo una dimensión de $96 \times 96 \times 2$ píxeles.

En la Ilustración 3.5 se observa un ejemplo de una imagen 3D utilizada, donde el espectrograma del frente corresponde al canal Fz y el de atrás al Pz .

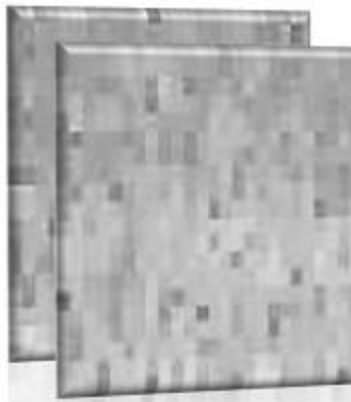


Ilustración 3.5: Ejemplo de imagen 3D, conteniendo los espectrogramas de los canales Fz y Pz .

3.2.2 Implementación de redes neuronales convolucionales

Una vez definidos los conjuntos de datos a utilizar, se dio paso a la construcción de un modelo de redes neuronales convolucionales, con el cual se buscó encontrar que conjunto de espectrogramas entregaba el mejor desempeño de clasificación.

El detalle de la arquitectura utilizada para crear la red neuronal convolucional se muestra a continuación, donde las C corresponden a las capas convolucionales, las P

corresponden a las capas de *pooling* y *FC* corresponde a la capa totalmente conectadas. El número precedente a cada una de las letras corresponde a la cantidad de filtros aplicados en cada capa, mientras que los números entre corchetes corresponden al tamaño de los filtros o la cantidad de neuronas de cada capa según corresponde. La letra Z correspondiente a la capa de entrada posee un valor de uno para los conjuntos de datos que utilizan un solo espectrograma (F_z o P_z), mientras que posee un valor de dos para los conjuntos de datos que utilizan los espectrogramas de canales F_z y P_z como entrada.

Arquitectura utilizada [30]:

Entrada[96x96xZ]-32C[3x3]-32P[2x2]-64C[3x3]-64P[2x2]-128C[3x3]-128P[2x2]-FC[512]-Salida [2]

En la Ilustración 3.6 puede observar gráficamente la arquitectura de la red utilizada.

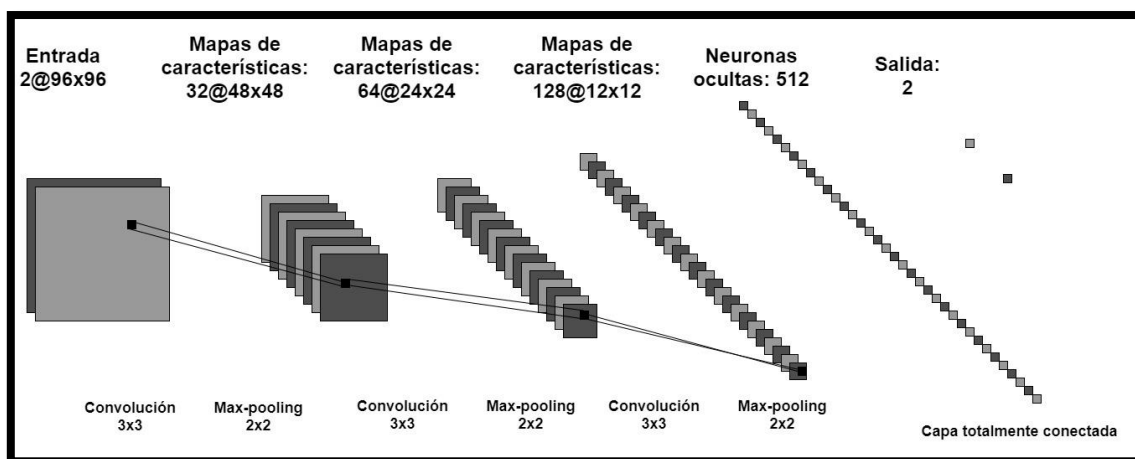


Ilustración 3.6: Diagrama de la arquitectura de la red convolucional utilizada.

Cabe destacar, que si bien la arquitectura utilizada fue tomada de la literatura correspondiente al tema [30], la forma en que se le presentan los datos a la red fue cambiada, ya que en la literatura se entrega a la red un solo espectrograma que contiene la información de 32 canales de EEG, mientras que la red utilizada recibe como entrada dos espectrogramas, donde cada uno de los cuales corresponde a un solo canal (F_z y P_z).

Por otro lado, para el entrenamiento de la red, la base de datos se dividió en un 80% y un 20%, para el entrenamiento y la prueba respectivamente. Además, la red fue entrenada mediante la utilización de lotes de sesenta y cuatro datos cada uno, para de esta forma acelerar la velocidad de entrenamiento de la red [30].

Paralelamente, para evitar el sobreajuste de la red se utilizó un *Dropout* de 0,5 y regularización L2 con parámetro β de 0,001.

En la Tabla 3.4 se muestra el resumen de los hiper parámetros utilizados para entrenar la red. Cada uno de estos fue obtenido mediante diferentes iteraciones, donde se buscó conseguir el mayor desempeño de clasificación posible.

Tabla 3.4: Hiper parámetros de la red neuronal convolucional para la detección de fatiga.

| Hiper parámetro | Valor |
|--------------------------|--------|
| Tasa de aprendizaje | 0,0001 |
| Función de activación | ReLU |
| Tamaño de los lotes | 64 |
| Dropout | 0,5 |
| Beta (regularización L2) | 0,001 |
| Épocas de entrenamiento | 1.000 |

Mediante la utilización de la red descrita, se procedió a evaluar su desempeño mediante el uso de cada uno de los conjuntos de datos previamente descritos. Debido a la escasa cantidad de datos disponible en cada uno de los conjuntos de datos, y de modo que los resultados obtenidos no estuviesen sesgados, cada conjunto de datos fue evaluado cinco veces (variando en cada iteración los datos correspondientes a entrenamiento y prueba), de modo de obtener un resultado promedio de clasificación asociado a una distribución estándar.

Además, para determinar el desempeño de la red, y dado que este varía en la iteración de una época a otra, el algoritmo fue programado para guardar los pesos de la red en la época que presentara la mayor exactitud dentro de las 1.000 épocas utilizadas. De esta manera fue posible obtener siempre el modelo (pesos de la red) que presentara el mejor desempeño para cada uno de los conjuntos de datos utilizados.

El mejor desempeño de clasificación obtenido fue mediante la utilización de los espectrogramas de los canales $Fz-Pz$ con duración de trece segundos. Con este conjunto de datos se procedió a realizar una comparación utilizando redes neuronales convencionales, SVM y *Random Forest*. A continuación, se explica la metodología utilizada para cada uno de estos algoritmos.

3.2.3 Implementación redes neuronales convencionales

El primer algoritmo con el cual se procedió a comparar el desempeño de la CNN fue con una red neuronal convencional, debido a las similitudes obvias que presentan ambos métodos.

De modo de poder utilizar el conjunto de datos correspondientes a los espectrogramas de los canales Fz y Pz con una duración de trece segundos, y dado que las redes neuronales convencionales no permiten como argumento de entrada imágenes si no que solo vectores, se procedió a remodelar la matriz que contenía los espectrogramas como un vector, concatenando de manera vertical cada columna de la matriz, obteniendo así un vector con la misma cantidad de puntos que la matriz de tres dimensiones previa al remodelado. De esta forma cada vector contenía los mismos dieciocho mil cuatrocientos treinta y dos puntos (lo cual equivale al producto de noventa y seis por noventa y seis por dos) correspondientes a cada matriz original.

Para poder cuantificar el efecto de la reducción de dimensionalidad generado por la CNN, realizado mediante capas convolucionales y de *pooling*, se creó una red neuronal convencional con la misma arquitectura de capa totalmente conectada que la red convolucional previamente realizada. De esta manera, se obtuvo una red neuronal de tres capas, con una capa de entrada con dieciocho mil cuatrocientos treinta y dos neuronas, una capa oculta de quinientas doce neuronas (al igual que en la capa totalmente conectada de la CNN) y una capa de salida de dos neuronas (correspondientes al estado base y estado de fatiga).

Al igual que en la CNN, la red neuronal convencional fue entrenada mediante lotes, y se le aplico regularización L2 y *Dropout* de modo de evitar el sobre ajuste.

En la Tabla 3.5 se muestra el resumen de los hiper parámetros utilizados para entrenar la red. Cada uno de estos fue obtenido mediante diferentes iteraciones, donde se buscó conseguir el mayor desempeño de clasificación posible.

Tabla 3.5: Hiper parámetros de la red neuronal convencional para la detección de fatiga.

| Hiper parámetro | Valor |
|--------------------------|--------|
| Tasa de aprendizaje | 0,0001 |
| Función de activación | ReLU |
| Tamaño de los lotes | 64 |
| Dropout | 0,5 |
| Beta (regularización L2) | 0,0001 |
| Épocas de entrenamiento | 10.000 |

Finalmente se utilizó la red descrita para evaluar cinco veces el conjunto de datos correspondiente, obteniendo así una media para cada métrica de desempeño, cada una asociada a una desviación estándar.

3.2.4 Implementación *Support Vector Machines*

Para la implementación de *Support Vector Machines*, y al igual que en la implementación de redes neuronales convencionales se utilizaron los espectrogramas en forma de vector, debido a que SVM solo acepta datos de entrada en forma de vectores.

Para la selección de los hiper parámetros óptimos del algoritmo, se procedió a evaluar tres tipos de kernel, gaussiano, tangente hiperbólica y lineal, cada uno de los cuales además fue probado con distintos valores para las constantes que les corresponden (grilla de búsqueda).

En la Tabla 3.6 se muestran los hiper parámetros obtenidos que entregaron el desempeño más óptimo para el clasificador de SVM.

Tabla 3.6: *Hiper parámetros utilizados para la clasificación de fatiga mediante SVM.*

| Hiper parámetro | Tipo/Valor |
|-----------------------|------------|
| Kernel utilizado | Lineal |
| Penalización de coste | 0,1 |

Finalmente se utilizó el clasificador SVM con los hiper parámetros mostrados para evaluar cinco veces el conjunto de datos correspondiente, obteniendo así una media para cada métrica de desempeño, cada una asociada a una desviación estándar.

3.2.5 Implementación *Random Forest*

Al igual que en la implementación de SVM, se utilizaron los espectrogramas en forma de vectores para el entrenamiento del clasificador *Random Forest*, y la selección de los hiper parámetros se realizó mediante una grilla de búsqueda, con la cual se buscó optimizar el número de estimadores del algoritmo, las características máximas a evaluar por cada estimador y la profundidad máxima del algoritmo.

En la Tabla 3.7 se muestran los hiper parámetros obtenidos que entregaron el desempeño más óptimo para el clasificador *Random Forest*.

Tabla 3.7: Hiper parámetros utilizados para la clasificación de fatiga mediante *Random Forest*.

| Hiper parámetro | Valor |
|-------------------------|-------|
| Numero de estimadores | 1.024 |
| Características máximas | 136 |
| Profundidad máxima | 25 |

Finalmente se utilizó el clasificador *Random Forest* con los hiper parámetros mostrados para evaluar cinco veces el conjunto de datos correspondiente, obteniendo así una media para cada métrica de desempeño, cada una asociada a una desviación estándar.

3.3 Recursos

Las especificaciones de la computadora utilizada para llevar a cabo ambos modelos son: CPU Core i7-6700K 4.2 GHz con 32 GB RAM y GPU Tesla K20. Por otro lado, todos los clasificadores fueron realizados mediante el software Python.

4. Resultados

4.1 Resultados red neuronal convolucional para cada conjunto de datos de somnolencia

Como se describió en el capítulo 3.2.2, la red neuronal convolucional realizada fue probada con seis diferentes conjuntos de datos, de modo de identificar cuál de estos conjuntos resulta más óptimo para la identificación de sujetos bajo estado de somnolencia.

En la Tabla 4.1 se muestran los valores medios obtenidos para cada métrica de desempeño en la detección de sujetos bajo estado de somnolencia para cada uno de los seis conjuntos de datos.

Tabla 4.1: Resumen métricas de desempeño para los distintos conjuntos de datos creados utilizando CNN.

| Largo señal | Canales utilizados | Exactitud | | Puntaje F1 | | Sensibilidad | |
|----------------|--------------------|-----------|---------|------------|---------|--------------|---------|
| Trece segundos | Fz | 79,66% | ± 1,12% | 79,48% | ± 2,90% | 78,38% | ± 5,38% |
| | Pz | 64,78% | ± 3,67% | 66,22% | ± 3,33% | 69,08% | ± 4,42% |
| | Fz-Pz | 86,74% | ± 1,81% | 87,09% | ± 1,99% | 88,42% | ± 5,06% |
| Cinco segundos | Fz | 73,37% | ± 0,85% | 73,12% | ± 1,64% | 71,85% | ± 6,16% |
| | Pz | 63,81% | ± 3,13% | 73,12% | ± 4,25% | 64,42% | ± 5,81% |
| | Fz-Pz | 77,98% | ± 0,27% | 76,47% | ± 2,01% | 75,32% | ± 5,88% |

| Largo señal | Canales utilizados | Especificidad | | Precisión | |
|----------------|--------------------|---------------|---------|-----------|---------|
| Trece segundos | Fz | 80,67% | ± 4,67% | 79,11% | ± 3,06% |
| | Pz | 60,69% | ± 7,64% | 63,95% | ± 5,92% |
| | Fz-Pz | 84,88% | ± 3,19% | 85,96% | ± 1,41% |
| Cinco segundos | Fz | 74,77% | ± 7,34% | 74,95% | ± 3,35% |
| | Pz | 62,98% | ± 5,54% | 62,24% | ± 3,59% |
| | Fz-Pz | 80,19% | ± 5,16% | 78,09% | ± 2,96% |

Como se puede observar, el conjunto de datos correspondiente a los espectrogramas de trece segundos de los canales $Fz-Pz$ presenta el mejor desempeño para cada una de las métricas utilizadas. En la Ilustración 4.1 y en la Ilustración 4.2 se observan las curvas de costo y exactitud respectivamente para los datos de entrenamiento como para los de prueba.

Costo de clasificación red neuronal convolucional a través de las épocas para conjunto de datos de trece segundos y canales $Fz-Pz$.

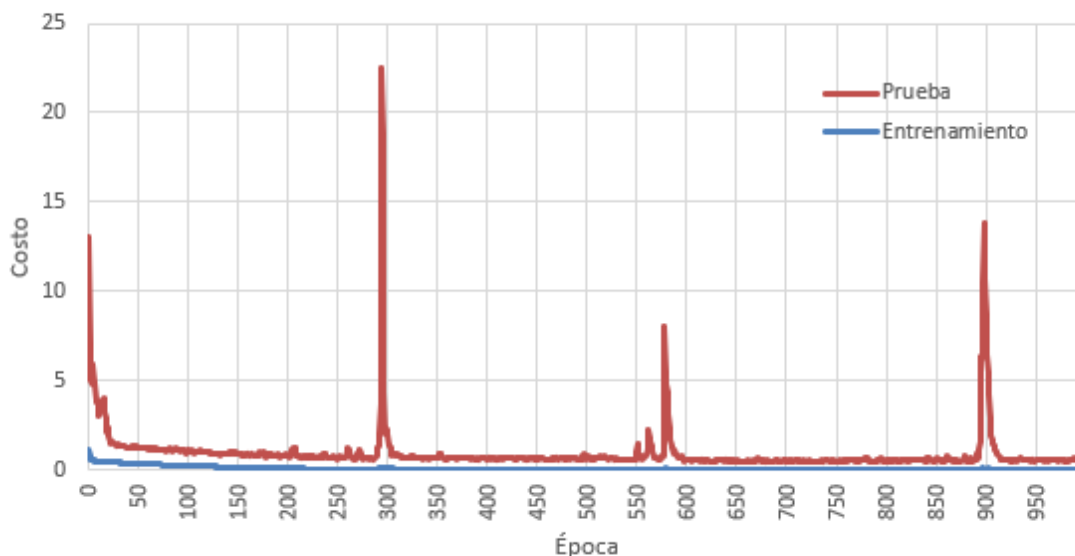


Ilustración 4.1: Costo de clasificación red neuronal convolucional a través de las épocas para conjunto de datos de trece segundos y canales $Fz-Pz$.

Exactitud de clasificación red neuronal convolucional a través de las épocas para conjunto de datos de trece segundos y canales $Fz-Pz$.

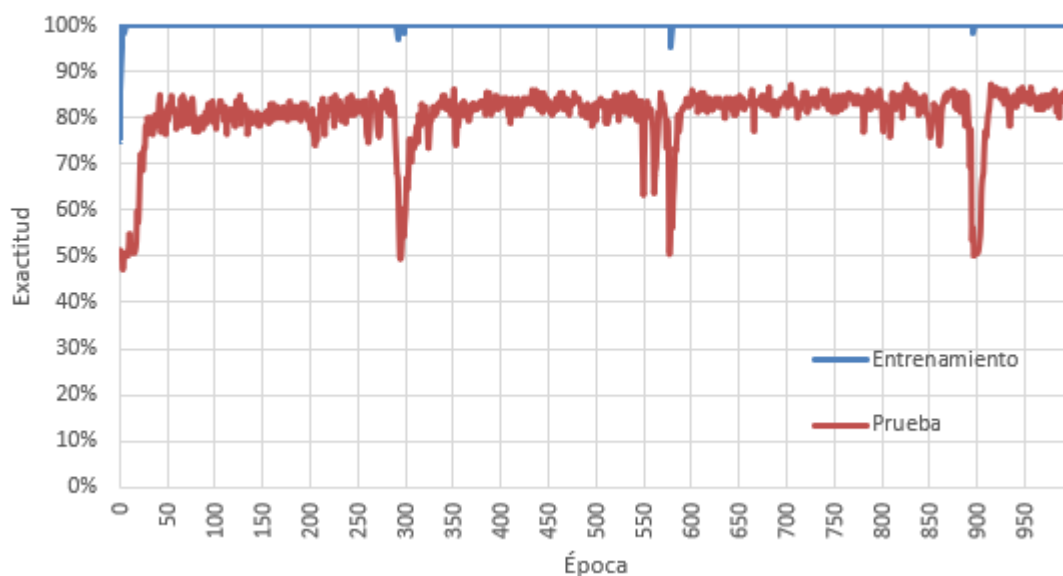


Ilustración 4.2: Exactitud de clasificación red neuronal convolucional a través de las épocas para conjunto de datos de trece segundos y canales $Fz-Pz$.

4.2 Comparación de resultados por modelo para espectrogramas de trece segundos correspondientes a los canales *Fz-Pz*

En la Tabla 4.2, se presentan los valores medios obtenidos para cada métrica de desempeño en la detección de sujetos bajo estado de somnolencia para los modelos de redes neuronales convoluciones (CNN), redes neuronales convencionales (NN), Support Vector Machines (SVM) y Random Forest (RF) utilizando el conjunto de datos de espectrogramas de trece segundos correspondientes a los canales *Fz-Pz*, conjunto con el cual se alcanzó el mayor desempeño.

Tabla 4.2: Resumen comparativo de resultados para cada uno de los modelos de clasificación de sujetos bajo estado de fatiga utilizados.

| Modelo | Exactitud | | Puntaje F1 | | Sensibilidad | |
|---------------|------------------|---------|-------------------|---------|---------------------|---------|
| <i>CNN</i> | 86,74% | ± 1,81% | 87,08% | ± 1,99% | 88,42% | ± 5,06% |
| <i>NN</i> | 74,45% | ± 2,55% | 75,77% | ± 3,02% | 79,35% | ± 6,23% |
| <i>SVM</i> | 73,03% | ± 2,06% | 72,18% | ± 1,50% | 73,10% | ± 2,68% |
| <i>RF</i> | 76,19% | ± 2,56% | 77,28% | ± 3,76% | 80,56% | ± 3,76% |

| Modelo | Especificidad | | Precisión | |
|---------------|----------------------|---------|------------------|---------|
| <i>CNN</i> | 84,88% | ± 3,19% | 85,96% | ± 1,41% |
| <i>NN</i> | 69,49% | ± 5,67% | 72,78% | ± 3,44% |
| <i>SVM</i> | 72,79% | ± 4,31% | 71,41% | ± 3,11% |
| <i>RF</i> | 72,14% | ± 4,83% | 74,55% | ± 5,19% |

5. Análisis de resultados

En primera instancia, al analizar los resultados expuestos en la Tabla 4.1, se puede observar que tanto para las señales de trece como de cinco segundos, la utilización conjunta de los canales Fz - Pz entrega el mejor desempeño para todas las métricas evaluadas, y del mismo modo el canal Fz entrega un mejor rendimiento que el canal Pz . De esto se puede deducir en primera instancia que la presencia de somnolencia en un individuo provoca cambios más relevantes en las ondas theta que en las ondas alfa del cerebro, debido a la relación existente entre las onda alfa y theta con los canales Fz y Pz respectivamente. Sin embargo, los cambios existentes en presencia de somnolencia en el canal Pz también entregan información acerca del estado mental del individuo, lo cual se hace evidente al evaluar los resultados obtenidos al utilizar ambos canales [2].

Paralelamente, se pudo comprobar que es necesario que las señales a utilizar posean una duración mínima de trece segundos [2], debido que señales con una duración inferior a esta, no poseen toda la información referente al nivel de cansancio del individuo. Este factor se evidencia no solo en las métricas expuestas en la Tabla 4.1, donde se observa que para señales de trece segundos se obtienen mejores resultados por canal en comparación a las señales de cinco segundos de duración, sino también al evaluar la cantidad de archivos disponibles para cada conjunto de datos. Los conjuntos de datos con señales de trece segundos poseen en total 920 archivos, mientras que los conjuntos de datos con señales de cinco segundos poseen en total 2.400 archivos. Al poseer más del doble datos, los modelos con señales de cinco segundos debiesen tener una mayor capacidad de generalización, debido a que las a que las CNN presentan resultados un aumento en su rendimiento al aumentar el tamaño de las bases de datos [4], sin embargo, esto evidentemente no ocurre.

Por otro lado, al observar la Ilustración 4.1 y en la Ilustración 4.2, se puede observar que tanto el costo como la exactitud de la CNN construida no convergen a través de las épocas para los datos de prueba, sin embargo, en los puntos en que el costo aumenta súbitamente (épocas 300, 575 y 900 aprox.), la exactitud de clasificación decae, mientras que para el resto de los puntos a pesar de que la exactitud oscila, el costo se mantiene constante dentro de un margen. Este hecho indica que a lo largo de las épocas la red no cae en un sobreajuste excesivo [19]. Sin embargo, se debe mencionar que la red igualmente posee un leve sobreajuste a los datos de entrenamiento, el cual se evidencia al observar en la Ilustración 4.2 la diferencia de exactitud entre los datos de entrenamiento y prueba, sin embargo, esta diferencia es atribuible al tamaño de la base de datos, por lo cual en caso de contar con una base de datos de mayor tamaño, la brecha entre entrenamiento y prueba debiese ser cada vez menor.

Del mismo modo, al analizar los resultados de las Tabla 4.1 y Tabla 4.2, podemos deducir que para realizar la clasificación de señales no estacionarias, como es el caso de las

señales de electroencefalografía, mediante espectrogramas, es imperante el uso de bases de datos de mayor tamaño que la utilizada, debido que al contar con una reducida cantidad de ejemplos, los modelos obtenidos no poseen una capacidad de generalización fija, hecho observable en las desviaciones estándar para cada una de las métricas calculadas, las cuales van desde el 1% en el mejor de los casos, hasta el 7%, en el peor escenario.

Al analizar los resultados expuestos en la Tabla 4.2, se puede observar que el modelo que entrega el mejor desempeño en todas las métricas corresponde a la CNN, seguido por RF, NN y finalmente SVM. Con estos resultados, podemos comprobar que la reducción de dimensionalidad realizada mediante las capas de *pooling* de la CNN es esencial para poder generar una clasificación en base a imágenes.

Para cuantificar el efecto de las capas de *pooling* y de convolución, podemos comparar el desempeño de la CNN con el de la NN realizada, dado que esta última posee la misma arquitectura que para las capas completamente conectadas que la CNN. De este modo, se puede observar mediante el uso de la Tabla 4.2, que la CNN obtiene resultados un 11% superiores a los de la NN en promedio, diferencia que se explica solo por la presencia de las capas convolucionales y de *pooling*, ya que como se mencionó en el capítulo 3.2.3, ambos modelos poseen la misma arquitectura, a excepción de las capas mencionadas.

La diferencia existente en los resultados de la Tabla 4.2 entre la CNN y los otros modelos expuesto nos permite concluir que es necesario realizar una reducción de dimensionalidad si se desea utilizar espectrogramas como datos de entrada para realizar la clasificación de sujetos bajo estado de somnolencia.

Finalmente, al comparar los resultados obtenidos por NN, SVM y RF en la Tabla 4.2, se puede observar que RF es el método más a fin para realizar este tipo de clasificación, alcanzando un 80,56% de sensibilidad en la detección de sujetos bajo estado de fatiga (tasa de aciertos en la detección de sujetos bajo estado de fatiga). Lo sorprendente de este hecho, es que matemáticamente el algoritmo de clasificación utilizada por RF (árboles de decisión) es el más simple en comparación a NN y SVM. Por lo cual, bajo este contexto RF se podría presentar como un método alternativo para la detección de sujetos bajo estado de fatiga si se utilizase algún método de reducción de dimensionalidad previo a su uso.

6. Conclusiones

Es posible realizar la clasificación de señales de electroencefalografía mediante sus espectrogramas utilizando redes neuronales convolucionales.

Detectar cuando un individuo presenta un estado somnolencia, el cual puede poner en riesgo su propia seguridad y el de todo un proceso es una necesidad no satisfecha actualmente, la cual requiere solución en el corto plazo. El presente proyecto de título presenta una metodología que permite detectar la presencia de fatiga en un operador mediante el uso de una técnica no invasiva como lo es la electroencefalografía, y el uso de los espectrogramas de las señales obtenidas, obteniendo resultados superiores al 85% de exactitud.

El índice KSS se presenta como un indicador valido para la generación de etiquetas de sujetos bajo estado de somnolencia, sin embargo, su uso debe ser acompañado por una correcta planificación del cronograma de la toma de datos.

Para lograr detectar fatiga mediante el uso de señales de EEG, estas deben poseer una duración mínima de trece segundos. Señales con una duración inferior a esta no contienen toda la información necesaria acerca del estado mental del sujeto.

Las variaciones de amplitud de las ondas theta en el canal Fz , representan el mejor indicador individual del nivel de somnolencia del individuo a evaluar, mientras que las variaciones de amplitud de las ondas alfa en el canal Pz también aportan información al estado mental del sujeto pero de forma reducida, por lo cual, es óptimo utilizar los espectrogramas de ambos canales para llevar a cabo el modelo de clasificación.

El número de muestras disponibles para llevar a cabo un clasificador de señales de electroencefalografía mediante técnicas de aprendizaje de máquinas es una limitante a la hora de optimizar el desempeño de estos.

Las redes neuronales convolucionales representan el algoritmo de clasificación más óptimo para la clasificación de sujeto bajo estado de fatiga mediante el uso de espectrogramas en comparación a redes neuronales convencionales, *Support Vector Machines* y *Random Forest*, lo cual es atribuible a la reducción de dimensionalidad de los datos de entrada que realizan las CNN.

Se logró comprender y manejar a cabalidad el algoritmo de clasificación de redes neuronales convolucionales.

7. Bibliografía

- [1] A. D. Swain y H. E. Guttman, «Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications,» Washington, D.C., 1983.
- [2] L. J. Trejo, K. Kubitz y L. D. Montgomery, «EEG-Based Estimation and Classification of Mental Fatigue,» *Psychology*, pp. 572-589, 2015.
- [3] W. y. B. A. J. Dittner, «The assessment of fatigue: A practical guide for clinicians and researchers,» *Journal of Psychosomatic Research*, vol. 56, n° 2, pp. 157-170, 2004.
- [4] Y. LeCun, Y. Bengio y G. Hinton, «Deep Learning Review,» *Nature*, vol. 521, n° 14539, pp. 436-444, 2015.
- [5] T. Marek, «Mental fatigue at work and pain perception,» *Work & Stress*, vol. 2, n° 2, pp. 133-137, 1988.
- [6] A. C. d. Seguridad, «Boletín Técnico de Ergonomía: Antecedentes sobre fatiga en la conducción de vehículos,» Asociación Chilena de Seguridad, Santiago de Chile, 2007.
- [7] S. V. y. T. S. D. Wylie, «Study of commercial vehicle driver rest periods and recovery of performance in an operational environment,» de *Managing Fatigue in Transportation. Proceedings of the 3rd Fatigue in Transportation Conference*, Fremantle, 1998.
- [8] C. N. d. S. d. Tránsito, «Informe de drogas y/o fatiga en la conducción,» Comisión Nacional de Seguridad de Tránsito, Santiago de Chile, 2013.
- [9] T. Åkerstedt y M. Gillberg, «STOP, THAT and One Hundred Other Sleep Scales,» *International Journal of Neuroscience*, vol. 52, pp. 29-37, 1990.
- [10] J. R. Daube y D. Rubin, *Clinical Neurophysiology*, Oxford: OXFORD, 2009.
- [11] T. C. Technologies, «Cortical Functions,» 1929-1958, Hong-Kong, 2012.
- [12] A. Labarga, «Análisis de Componentes Independientes en el Estudio de Potenciales,» España, 2002.
- [13] N. Instruments, «<http://www.ni.com>,» [En línea].

- [14] M. V.M., «Can SVM be used for automatic EEG detection of drowsiness during car driving?,» *Safety Science*, vol. 47, pp. 115-124, 2009.
- [15] L. Breiman, «Random Forests,» *Machine Learning*, vol. 45, n° 1, pp. 5-32, 2001.
- [16] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer-Verlag, 2006.
- [17] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*, Cambridge, Massachusetts: MIT Press, 2016.
- [18] D. M. Hawkins, «The problem of overfitting,» *Journal of Chemical Information and Modeling*, vol. 44, n° 1, pp. 1-12, 2003.
- [19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, «Dropout: A Simple Way to Prevent Neural Networks from,» *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.
- [20] E. Phaisangittisagul, «An Analysis of the Regularization between L2 and Dropout in Single Hidden Layer Neural Network,» de *7th International Conference on Intelligent Systems, Modelling and Simulation*, Bangkok, 2016.
- [21] M. Sokolova y G. Lapalme, «A systematic analysis of performance measures for classification tasks,» *Information Processing and Management*, vol. 45, pp. 427-437, 2009.
- [22] R. J. Barry, A. Clarke, S. Johnstone, C. Magee y J. Rusby, «EEG differences between eyes-closed and eyes-open resting conditions,» *Clinical Neurophysiology*, vol. 118, p. 2765–2773, 2007.
- [23] D. J. S. Salazar, «Análisis de componentes independientes aplicado al estudio de la actividad cerebral,» Bogota, 2014.
- [24] S. Valipour, A. Shaligram y G. Kulkarni, «Spectral analysis of EEG signal for detection of alpha rhythm with open and closed eyes,» *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 3, n° 6, 2013.
- [25] E. Ben-Simon, I. Klovatch Podlipsky, H. Okon-Singer y T. Hendler, «The dark side of the alpha rhythm: fMRI evidence for induced alpha modulation during complete darkness,» *European Journal of Neuroscience*, vol. 37, n° 5, pp. 1-9, 2012.

- [26] D. Verstraete, A. Ferrada, E. López Droguett, V. Meruane y M. Modarres, «Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image,» *Shock and Vibration*, vol. 2017, n° 5067651, p. 17, 2017.
- [27] A. S. Berahas, J. Nocedal y M. Takác, «A Multi-Batch L-BFGS Method for Machine,» de *30th Conference on Neural Information Processing Systems*, Barcelona, 2016.
- [28] M. Kaur, «EEG SIGNALS DURING EYES-OPEN & EYES-CLOSED STATE,» Chandigarh, 2017.
- [29] Q. Massoz, T. Langohr, C. François y J. G. Verly, «The ULg Multimodality Drowsiness Database (called DROZY),» Liège, 2016.
- [30] Y.-C. Hung, Y.-K. Wang, M. Prasad y C.-T. Lin, «Brain Dynamic States Analysis based on 3D Convolutional Neural Network,» de *International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, 2017.
- [31] A. O’Riordan, «<http://www.cs.ucc.ie>,» 2005. [En línea].
- [32] A. Dittner, S. Wessely y R. Brown, «The assessment of fatigue: A practical guide for clinicians and researchers,» *Journal of Psychosomatic Research*, vol. 56, n° 2, pp. 157-170, 2004.