

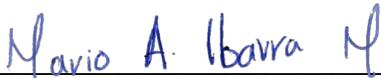
Salamanca, Gto., a 19 de Agosto del 2021.

M. en I. HERIBERTO GUTIÉRREZ MARTIN
COORDINADOR DE ASUNTOS ESCOLARES
P R E S E N T E.-

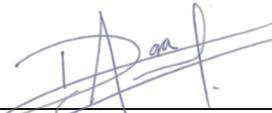
Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno(a) Ing. Oscar Almanza Conejo del **Programa de Maestría en Ingeniería Eléctrica (Instrumentación y Sistemas Digitales)** y cuyo número de **NUA** es: 732838 del cual soy director. El título de la tesis es: "Algoritmo para detectar emociones humanas utilizando señales electroencefalográficas".

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

A T E N T A M E N T E



Dr. Mario Alberto Ibarra Manzano
DIRECTOR DE TESIS
SECRETARIO



Dra. Dora Luz Almanza Ojeda
DIRECTOR DE TESIS



Dr. Sergio Eduardo Ledesma Orozco
PRESIDENTE



Dr. Horacio Rostro González
VOCAL

EmpastadoOAC

Final Audit Report

2021-08-20

| | |
|-----------------|--|
| Created: | 2021-08-19 |
| By: | Mario A Ibarra Manzano (ibarram@ugto.mx) |
| Status: | Signed |
| Transaction ID: | CBJCHBCAABAABZfuMEGX5CgMKsbRMO4PccvBi4mVtk7G |

"EmpastadoOAC" History

-  Document created by Mario A Ibarra Manzano (ibarram@ugto.mx)
2021-08-19 - 10:24:45 PM GMT- IP address: 187.193.52.67
-  Document emailed to Sergio Eduardo Ledesma Orozco (selo@ugto.mx) for signature
2021-08-19 - 10:26:04 PM GMT
-  Email viewed by Sergio Eduardo Ledesma Orozco (selo@ugto.mx)
2021-08-19 - 10:26:46 PM GMT- IP address: 104.47.58.254
-  Document e-signed by Sergio Eduardo Ledesma Orozco (selo@ugto.mx)
Signature Date: 2021-08-19 - 10:29:38 PM GMT - Time Source: server- IP address: 172.83.174.215
-  Document emailed to Horacio Rostro González (hrostrom@ugto.mx) for signature
2021-08-19 - 10:29:40 PM GMT
-  Email viewed by Horacio Rostro González (hrostrom@ugto.mx)
2021-08-20 - 5:58:38 AM GMT- IP address: 46.193.65.53
-  Document e-signed by Horacio Rostro González (hrostrom@ugto.mx)
Signature Date: 2021-08-20 - 6:01:19 AM GMT - Time Source: server- IP address: 46.193.65.53
-  Agreement completed.
2021-08-20 - 6:01:19 AM GMT



UNIVERSIDAD DE GUANAJUATO

**CAMPUS IRAPUATO-SALAMANCA
DIVISIÓN DE INGENIERÍAS**

**“Algoritmo para detectar emociones humanas
utilizando señales electroencefalográficas”**

TESIS PROFESIONAL

QUE PARA OBTENER EL GRADO DE:

**MAESTRÍA EN INGENIERÍA ELÉCTRICA
(INSTRUMENTACIÓN Y SISTEMAS DIGITALES)**

PRESENTA:

ING. OSCAR ALMANZA CONEJO

ASESORES:

**DRA. DORA LUZ ALMANZA OJEDA
DR. MARIO ALBERTO IBARRA MANZANO**

SALAMANCA, GUANAJUATO

SEPTIEMBRE 2021

Dedicatoria

*Para Sarita (†):
Por enseñarme que la vida no se mide
en días sino en risas.*

Agradecimientos institucionales

Este proyecto de tesis se llevó a cabo gracias al apoyo de varias instituciones, las cuáles merecen una mención especial y se les extiende una alta gratitud en la realización de este trabajo.

- Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por el apoyo con fondos para mis estudios de maestría bajo el número de beca **CVU: 1007303**



- A la **División de Ingenierías Campus Irapuato-Salamanca** de la Universidad de Guanajuato por proporcionar las herramientas necesarias para la culminación de mis estudios. A la Dirección de Posgrado en Ingeniería Eléctrica, opción Instrumentación y Sistemas Digitales. Al laboratorio de DSP por el espacio y equipo proporcionado durante la realización de este trabajo de tesis.



Agradecimientos personales

Según Charles Darwin, “las emociones son adaptaciones obtenidas a través de la evolución que permiten que, tanto humanos como animales, maximicen sus posibilidades de supervivencia y reproducción”.

Primero, lo más importante, gracias a mis padres: gracias por tanto apoyo. Gracias por darme la vida, por confiar en mí, y por todos los valores que me inculcaron. Por darme tantas experiencias valiosas y por sobre todo, ser los únicos que siempre por siempre estarán. Pido a la vida tenerlos conmigo muchísimo más tiempo. Gracias también a mi hermana, Ana Karen, por dar la luz necesaria para mi familia.

Gracias a toda mi familia, abuelos(as), tíos(as), primos(as). Gracias por todas y cada una de sus palabras de aliento.

Gracias a todos mis profesores, en especial a la Dra. Dora Luz Almanza Ojeda y al Dr. Mario Alberto Ibarra Manzano por su tan valiosa asesoría durante estos dos años de maestría. Gracias también a mis compañeros de estudio.

Por último, quiero agradecer especialmente a aquellos familiares quienes hoy no están en vida conmigo y mi familia. En especial a mi abuelo Antonio y a mi prima Sarita. Gracias por enseñarme tantas y maravillosas cosas. Prometo recordarles por siempre.

Resumen

Las emociones son estados mentales asociados a cambios repentinos que influyen en el comportamiento, pensamiento y en la salud de las personas. Con el paso de los años, se han desarrollado modelos de reconocimiento emocional con la finalidad de otorgar una mejor calidad de vida en las personas. En la última década, el reconocimiento emocional se ha llevado a cabo mediante el análisis de señales fisiológicas. Las señales fisiológicas modelan el comportamiento en órganos o tejidos en respuesta a alguna excitación. Otra alternativa consiste en la creación de modelos dirigidos a aplicaciones de interfaz cerebro-computadora, los cuales han sido importantes para identificar las emociones en los seres humanos. En este trabajo se propone una clasificación de emociones utilizando señales electroencefalográficas, las cuales se transforman a una representación tiempo-escala. Las wavelets han sido un método comúnmente utilizado para procesar señales electroencefalográficas o electrocardiográficas. Es por ello que en nuestro enfoque se utiliza la transformada continua wavelet para obtener imágenes de escalogramas de las señales electroencefalográficas. Consideramos la representación de tiempo-escala como imágenes que muestran la energía asociada a este tipo de señales. La extracción de características de los escalogramas mediante redes neuronales convolucionales permite proponer dos modelos para clasificar cuatro y ocho emociones. Los resultados de la clasificación fueron evaluados utilizando métricas de desempeño y comparados con trabajos relacionados, siendo superiores a algunos de los existentes en la literatura.

Índice general

| | |
|---|-----------|
| 1. Introducción | 1 |
| 1.1. Objetivos | 3 |
| 1.1.1. Objetivo general | 3 |
| 1.1.2. Objetivos particulares | 3 |
| 1.2. Justificación | 3 |
| 1.3. Antecedentes | 5 |
| 1.4. Organización del documento | 8 |
| 2. Estado del arte | 9 |
| 2.1. Caracterización de las emociones | 12 |
| 2.2. Trabajos relacionados | 12 |
| 2.2.1. Características del modelo y conjunto de datos | 13 |
| 2.2.2. Emociones primitivas | 15 |
| 2.2.3. Emociones: clases discretas | 18 |
| 2.3. La transformada wavelet | 20 |
| 2.3.1. La transformada analítica wavelet | 20 |
| 2.3.2. La transformada continua wavelet | 22 |
| 3. Metodología | 26 |
| 3.1. Conjunto de datos | 27 |
| 3.2. Descripción de la señal | 31 |
| 3.3. Arquitectura de red | 35 |
| 4. Resultados | 39 |
| 4.1. Clasificación en el espacio VA | 41 |

| | |
|--|-----------|
| 4.2. Clasificación en el espacio VAD | 43 |
| 4.3. Discusión | 47 |
| 5. Conclusiones | 51 |
| Bibliografía | 53 |

Índice de figuras

| | |
|--|----|
| 2.1. Espacio de emociones propuesto por Lui, et al., en [1]. | 10 |
| 2.2. Wavelets más comunes. La diferencia principal entre (a), (b) y (c) es la variación del factor β y γ | 22 |
| 2.3. Variaciones del parámetro β y γ para una familia de MWT. [2] | 24 |
| 3.1. Diagrama general de la clasificación de señales EEG. | 27 |
| 3.2. Arreglo de datos adquiridos por cada sujeto de pruebas en la base de datos DEAP [3]. | 28 |
| 3.3. Diagrama del sistema de adquisición 10-20. | 29 |
| 3.4. La relación μ_n/σ_n con respecto al valor y excitación de cada video [3]. | 29 |
| 3.5. Espacio de emociones VA. | 30 |
| 3.6. Espacio de emociones VAD. | 31 |
| 3.7. Componentes de frecuencias en las señales EEG: theta, alpha, beta y gamma. | 32 |
| 3.8. Señal EEG en su representación de escalograma en mapa de color <i>jet</i> | 34 |
| 3.9. Arquitectura de la capa “Inception (3a)”. | 36 |
| 3.10. Arquitectura de red aplicada a la clasificación de emociones en el espacio VA y VAD. | 38 |
| 4.1. Modelos de clasificación propuestos en este trabajo. | 40 |
| 4.2. Desempeño de entrenamiento y validación en el espacio VA. | 41 |
| 4.3. Desempeño de entrenamiento y validación en el espacio VAD. | 44 |
| 4.4. Desempeño de entrenamiento y validación en el espacio VAD añadiendo ruido a las muestras para aumentar el número de ejemplos. | 45 |

Índice de tablas

| | |
|---|----|
| 1.1. Bandas de frecuencia, características y localización en las EEG. . . . | 7 |
| 2.1. Características del modelo y conjunto de datos de trabajos relacionados | 14 |
| 2.2. Características de trabajos con enfoque de emociones primitivas . . . | 17 |
| 2.3. Características de trabajos con enfoque de emociones por clases discretas | 19 |
| 3.1. Sintonización de la red GoogLeNet a nuestro enfoque. | 37 |
| 4.1. Desempeño de entrenamiento y validación en el espacio VA para las últimas cinco iteraciones. | 42 |
| 4.2. Matriz de confusión y métricas de evaluación del modelo en el espacio VA. | 43 |
| 4.3. Matriz de confusión y métricas de evaluación del modelo en el espacio VAD. | 45 |
| 4.4. Desempeño de entrenamiento y validación en el espacio VAD para las últimas cinco iteraciones añadiendo ruido a las muestras para aumen- tar el número de ejemplos. | 46 |
| 4.5. Matriz de confusión y métricas de evaluación en el espacio VAD añadien- do ruido a las muestras para aumentar el número de ejemplos. | 46 |
| 4.6. Comparativa con otras propuestas de clasificación de emociones. . . . | 49 |

Capítulo 1

Introducción

A lo largo de los años los seres humanos han desarrollado diversos lenguajes de comunicación para expresar ideas, emociones y/u opiniones acerca de temas de interés. Las primeras formas de comunicación se llevaron a cabo mediante la expresión de lenguaje corporal, misma que el emisor otorgaba al receptor para que interpretara la idea expresada por el emisor. A partir de entonces, los seres humanos fueron modificando este tipo de lenguaje hasta poder caracterizar su comunicación en un sistema a partir de un conjunto de expresiones, tanto verbales, faciales y/o corporales. De esta manera los seres humanos se han dado a la tarea de investigar y comprender el lenguaje corporal de su emisor, hoy de manera tan automática que pareciera natural, para así poder adaptar su comportamiento a las necesidades o la fluidez de una conversación con base en la recepción e interpretación emocional.

En [4] se presenta que existen también micro-expresiones que se manifiestan de manera involuntaria mediante las expresiones faciales. Estas micro-expresiones tienen una duración muy corta, aproximadamente de 0.25 a 0.333... segundos. Estas micro-expresiones son solamente adquiridas por una cámara de una frecuencia de disparo igual o mayor a 25 cuadros por segundo. También, Ekman, et al., en [5] caracteriza seis emociones como las más importantes, las cuáles son: sorpresa, miedo, ira, disgusto, tristeza y felicidad, pues éstas registran cambios más significativos en algunas zonas faciales tales como la frente, cejas, párpados, mejillas, nariz, labios y la barbilla. Además, Khatri, et al., en [6] establece que las expresiones faciales proveen señales más sensibles sobre la respuesta emocional de las personas. Con lo cual, en [7]

A. Mehrabian cuantifica que el impacto total en la comunicación está representado por tres componentes principales: la parte verbal, la cual tiene un impacto del 7%, 38% la parte vocal y un 55% de manera facial.

Con base en lo anterior, para lograr la mayor comprensión de información, es necesario esclarecer las emociones de los sujetos involucrados en una conversación. En la última década, los trabajos enfocados al reconocimiento emocional a partir del uso de señales electroencefalográficas han demostrado una alta exactitud en los modelos creados a partir de inteligencia artificial. La mayoría de estos modelos están basados en la clasificación de emociones primitivas, es decir, aquella relación que existe entre valor, excitación y dominio emocional que presenta una persona en diferentes situaciones. Los niveles presentes en las emociones primitivas se pueden también utilizar para crear modelos basados en la identificación de emociones más concretas como la tristeza o la felicidad. Para esto, dependiendo el enfoque de la investigación, existen dos métodos de caracterización de emociones. El primero, se realiza con base en un estudio donde las emociones son analizadas previamente y se excita al sujeto para adquirir una emoción específica. El segundo método es mediante la adquisición de los niveles de valor, excitación y dominio para, posteriormente, realizar una discretización de los valores de las emociones primitivas para adquirir emociones específicas.

Con base en lo anterior, en este trabajo se proponen dos modelos de reconocimiento de emociones utilizando señales electroencefalográficas. El primer modelo clasifica, con una exactitud elevada con respecto a la literatura, las siguientes emociones: “tristeza”, “enojo”, “relajación”, y “felicidad”. En el segundo modelo, las emociones son: “tristeza”, “despreocupación”, “miedo”, “enojo”, “protección”, “satisfacción”, “sorpresa”, y “felicidad”. Ambos modelos se derivan de la discretización de las emociones primitivas de un conjunto de datos de señales EEG. El primer modelo está basado en el uso de el valor y la excitación, vectores de datos mutuamente ortogonales que generan un espacio de dos dimensiones para la discretización de cuatro emociones específicas. De manera similar, el segundo modelo está basado en el uso del valor, excitación y dominio, creando así un espacio de tres dimensiones para caracterizar ocho emociones.

1.1. Objetivos

1.1.1. Objetivo general

Diseñar y desarrollar un algoritmo de procesamiento de señales electroencefalográficas (EEG) para obtener un modelo que permita el reconocimiento de emociones humanas mediante una extracción de características a partir de redes neuronales convolucionales.

1.1.2. Objetivos particulares

- Analizar un conjunto de señales electroencefalográficas para modelar diferentes emociones en sujetos de prueba.
- Realizar un algoritmo de procesamiento de señales EEG utilizando técnicas de análisis espectral.
- Aplicar a las señales EEG un método de extracción de características para identificar emociones humanas en diferentes sujetos.
- Desarrollar una arquitectura basada en redes neuronales convolucionales que permita realizar un reconocimiento de emociones humanas.

1.2. Justificación

Conforme a su crecimiento e interacción con otros, los seres humanos han podido interpretar de manera natural las expresiones faciales y el lenguaje corporal de otras personas, así como también el timbre de voz, para deducir de estos las reacciones emotivas que puedan dar pie al desarrollo de una conversación verbal o no verbal. El problema que presenta para las personas al interpretar dichas emociones es que, en muchas ocasiones, los seres humanos tienden a alterar sus emociones con la finalidad de que el receptor adopte otra postura ante tal emoción.

Es claro que las expresiones faciales, así como el timbre de voz, tiendan a ser fácilmente alterados por el emisor, generando así una interpretación errónea por

parte del receptor. He aquí la importancia de un análisis más profundo que permita obtener una emoción más precisa.

En [5] se establecen seis emociones básicas e invariantes con respecto a la cultura: sorpresa, miedo, ira, disgusto, tristeza y felicidad. Estas emociones son las que se identifican de manera natural por las personas en las expresiones faciales, lenguaje corporal y en el habla. Partiendo de lo anterior, es importante detectar emociones que puedan poner en riesgo al sujeto o a otros sujetos en su entorno. Por ejemplo, emociones negativas, como la ira y disgusto, han demostrado una gran influencia en desordenes de estrés post-traumáticos [8]. De igual manera, el mantener un estado emocional estable es importante para la salud. En [9], se establece que las emociones negativas llegan a provocar enfermedades respiratorias y cardiovasculares. Derivado de lo anterior, es importante determinar las emociones reales que experimenta un sujeto. Por ejemplo, derivado de la contingencia provocada por el virus SARS-Cov-2, o COVID-19, la población mundial se vió obligada a un confinamiento con la finalidad de reducir el número de contagios. Durante este confinamiento las personas experimentaron múltiples emociones, en su mayoría negativas. En [10], se establece que el confinamiento por COVID-19 provocó efectos negativos en la salud mental y emocional de los sujetos de prueba, de los cuales una proporción importante se presentó en sujetos con desordenes emocionales y psicosociales.

De manera similar, es importante analizar las emociones reales en las interacciones en redes sociales. Por ejemplo, existen casos a nivel mundial donde usuarios de diferentes plataformas de redes sociales han causado daños a propiedades, sujetos en su entorno e inclusive daños a su persona. En [11] se expone el comportamiento, intimidación, vulnerabilidad, entre otras, a partir del uso de redes sociales para encontrar factores de riesgo individuales y sociales. Así, al conocer las emociones que un sujeto experimenta, se pueden prevenir situaciones de riesgo. Es por ello que un modelo de reconocimiento emocional resulta esencial en el entendimiento del comportamiento humano. Además, son necesarios modelos en los cuáles un mayor número de emociones sean clasificadas con una alta exactitud.

1.3. Antecedentes

Conforme avanzan las investigaciones en el área del estudio comportamiento humano, han surgido numerosas explicaciones acerca de las emociones. En estas investigaciones se abren una y otra vez debates en los cuales científicos encaran la problemática de tratar de definir concretamente “¿Qué es una emoción?”. Sin embargo, otorgar una respuesta final a la pregunta sugiere aún más problemas, por lo que es preferible asociar esta palabra, según R. Roberts en [12], como una experiencia típicamente unificada de estados mentales, tales como la fe, el deseo, una perturbación psicológica y el comportamiento.

Aunque es posible entender mediante señales no fisiológicas (expresiones faciales, reconocimiento del habla, lenguaje corporal, entre otras) la emoción principal por parte del emisor, está claro que siempre existirá una redundancia, en menor o mayor grado, al transmitir la información al(los) receptor(es). Esto último radica en la facilidad del ser humano al alterar sus expresiones corporales y faciales, lo cual conduce a una distorsión de la veracidad en la información. Para esto, Feldman en [13] estableció que las emociones pueden ser expresadas en dimensiones mutuamente ortogonales mediante la excitación y el valor de las emociones. El enfoque del valor hace referencia al grado en el cual los individuos acceden al componente hedónico (placer) con respecto a experiencias afectivas. Mientras que el enfoque de excitación accede, mediante las experiencias, a la componente de excitación.

El enfoque médico del impacto en la salud de los sujetos con respecto a un estado emocional sano tiende a ser de igual manera importante. En [14] se establece que sujetos con altos niveles de salud mental positiva poseen habilidades de autocompación, lo cual se proyecta como una resistencia ante desordenes de psicopatología (trastornos mentales diversos), mismo que puede funcionar como una estrategia de regulación emocional. También, proyecta la posibilidad de que un estado mental positivo no solo impacte sobre la psicopatología, sino que puede incrementar los factores de emociones positivas.

Por otro lado, en [9] se establece que la inhibición crónica de la tristeza y el llanto están ligadas a enfermedades respiratorias como el asma, así como la inhibición crónica de la ira está relacionada con desordenes cardiovasculares tales como la hipertensión. Siendo estos padecimientos unas de las enfermedades más comunes. Por

lo cual, es necesario mantener un estado mental sano para que este no influya en el estado de salud física de una persona.

Con base en lo anterior, es importante conocer las emociones que el ser humano experimenta en algún momento, para esto se han desarrollado diferentes algoritmos de reconocimiento emocional mediante el análisis de señales y/o imágenes no fisiológicas, tales como análisis de expresiones faciales mediante visión por computadora [15–18] o cambios en el timbre de voz mediante procesamiento digital de señales [19–22]. Es importante mencionar que el uso de señales no fisiológicas tiene una desventaja frente a las señales fisiológicas y es que los sujetos de prueba pueden fácilmente manipular su comportamiento al momento de la adquisición de datos. Ejemplo, un sujeto estimulado para sentir ira o tristeza puede fácilmente cambiar ésta por felicidad. De igual manera, el timbre de voz puede ser alterado para expresar otro tipo de emoción.

Es por esto que, para el enfoque de reconocimiento emocional, se han desarrollado algoritmos basados en el uso de señales fisiológicas, las cuales modelan los cambios en una región específica del cuerpo [23], tales como las señales electroencefalográficas (EEG), electrocardiácas (ECG) o electromiográficas (EMG). Mismas que pueden ser procesadas de manera simultánea pues cada una de éstas modelan un comportamiento distinto, dependiendo de su localización corporal [24], partiendo de un objetivo central al llevar a cabo una actividad.

Uno de los métodos más populares para el modelado de la actividad electroencefalográfica es la medición de la diferencia de potencial eléctrico mediante diversos dispositivos de hardware. Siendo uno de los más populares las diademas de electrodos. Éstas realizan un contacto directo sobre el cuero cabelludo del sujeto para poder monitorear y/o adquirir las fluctuaciones en la magnitud del voltaje. Los cambios en la magnitud se derivan del flujo de corriente iónica emitidas por las neuronas en el cerebro, a partir del campo electromagnético generado por alguna actividad cognitiva en los diferentes lóbulos electroencefalográficos.

En las señales EEG existen ondas de frecuencias asociadas a un estado mental específico. Las bandas de interés se presentan brevemente en la Tabla 1.1, donde se describen las bandas de frecuencia de las componentes en las señales EEG, sus características y localización encefalográfica.

A partir de lo descrito anteriormente, en este documento se propone una metodología para el reconocimiento de emociones en dos modelos, uno para clasificar cuatro emociones y otro para ocho. Mediante el uso de un conjunto de datos de libre acceso se emplean señales EEG, en las cuáles el sujeto de pruebas recibió estímulos audiovisuales con respecto a videos musicales que han sido clasificados para excitar al sujeto a diferentes valores de emociones primitivas. Además, mediante un procesamiento de análisis espectral, se obtienen representaciones en el dominio de tiempo-escala de cada señal EEG. Al obtener su distribución tiempo-escala, ésta es almacenada como un archivo de imagen (*.png) y etiquetada a su clase correspondiente dependiendo el modelo. Las representaciones son utilizadas como una base de datos de entrada a un modelo de clasificación. La red utilizada para entrenar y validar el modelo, es una red basada en la arquitectura de las redes neuronales convolucionales (CNN). Cada modelo cuenta con un número de características específico a la salida de cada entrenamiento de red.

Tabla 1.1: Bandas de frecuencia, características y localización en las EEG.

| Bandas de frecuencia de interés en las EEG. | | | |
|---|------------|---|----------------------|
| Nombre | Frecuencia | Características | Lóbulo |
| Theta (θ) | 3 – 7 Hz | Corta duración en tiempo. Hiperventilación. Presente en adultos jóvenes. | Frontal |
| Alpha (α) | 8 – 13 Hz | Baja actividad cerebral | Occipital & Parietal |
| Beta (β) | 14 – 29 Hz | Estado de alerta. Actividad consciente. Actividades motrices y de concentración. | Frontal |
| Gamma (γ) | 30 – 47 Hz | Usualmente presentes en bebés. Presente en sueños profundos en adultos. Baja magnitud | Diversos lóbulos. |

La aportación principal de este trabajo es la transformación de las señales EEG a una distribución de tiempo-escala asociada a su energía. Otra aportación es la caracterización de las emociones a partir de las emociones primitivas. Una tercera aportación es la propuesta de un modelo para clasificar ocho emociones con un alto desempeño. Como cuarta aportación, se demuestra que al aumentar las clases de un

espacio de cuatro a ocho emociones, el modelo no presenta una pérdida significativa de exactitud.

1.4. Organización del documento

En el Capítulo 2 se presentan los trabajos relacionados al reconocimiento emocional, las diferentes técnicas de procesamiento de las señales EEG y algunos resultados con respecto a diferentes modelos de clasificación. El Capítulo 3 presenta el conjunto de datos utilizado, el procesamiento aplicado a las señales EEG, la arquitectura de la red y las clases de salida del modelo de clasificación. En el Capítulo 4 se muestran los resultados del entrenamiento de los modelos de clasificación así como un análisis de las diferentes métricas de evaluación de los modelos entrenados. Por último, en el Capítulo 5 una breve conclusión del trabajo presentado en este documento.

Capítulo 2

Estado del arte

Tras la última década, ha tomado una gran popularidad el análisis de señales EEG debido a que presentan una mayor fiabilidad en comparación con otro tipo de señales. Tal es la importancia de estas señales, que se han usado en trabajos de imaginación motora, como en [25], donde se expone una técnica de análisis de cuaterniones basadas en el uso de las señales EEG para clasificar movimiento de izquierda, derecha o neutro del brazo en sujetos de pruebas. De manera similar, en [26] se presenta el uso de señales electroencefalográficas para clasificar cuando se produce un parpadeo en sujetos a diferentes escenarios como lo son en lectura, estimulación audio-visual y de factores externos diversos. Este último trabajo contiene los datos de más de 2300 parpadeos oculares voluntarios e involuntarios. Existen también modelos enfocados a la medicina, como lo es en [27] donde se proponen modelos de diagnósticos tempranos en la enfermedad de Alzheimer. Aquí, se utilizan características de tiempo-frecuencia, a partir de señales EEG, para realizar un modelo de clasificación de sujetos con enfermedad de Alzheimer y sujetos sanos. El trabajo concluyó que los sujetos con enfermedad de Alzheimer presentaban valores de magnitud menores tanto en la transformación de la señal al dominio de la frecuencia, así como en las características extraídas al analizar los cuatro lóbulos cerebrales.

Los trabajos anteriormente mencionados mantienen una dependencia con respecto a un vector de la magnitud de potencial eléctrico de la señal original EEG. En el caso de los artículos [25, 26], al analizar un movimiento, las frecuencias β que son las encargadas de la actividad motriz del cuerpo, supondría un análisis aún más

profundo de las señales EEG.

El análisis de las emociones tiende a sugerir una mayor fiabilidad con respecto a las posibles emociones a detectar. En [13], Felman, et al., estableció que las emociones pueden ser expresadas mediante las emociones primitivas de valor y excitación, pero además, existe una tercer emoción primitiva llamada dominio. En [28], Verma et al., define el valor como una escala de rangos de la emoción que puede ir de negativo a positivo, siendo la infelicidad o tristeza emociones negativas y la felicidad o el regocijo como positivas. Al rango de estímulo que se genera por una emoción se le conoce como excitación y por último, se le denomina dominio a la escala de rangos en la cual el sujeto siente control sobre sus emociones. Ésta última puede abarcar una escala desde una emoción sumisa hasta una dominante. En la Figura 2.1 se muestra un modelo de tres dimensiones que muestra las emociones primitivas y las diferentes emociones obtenidas a partir de la discretización de emociones primitivas.

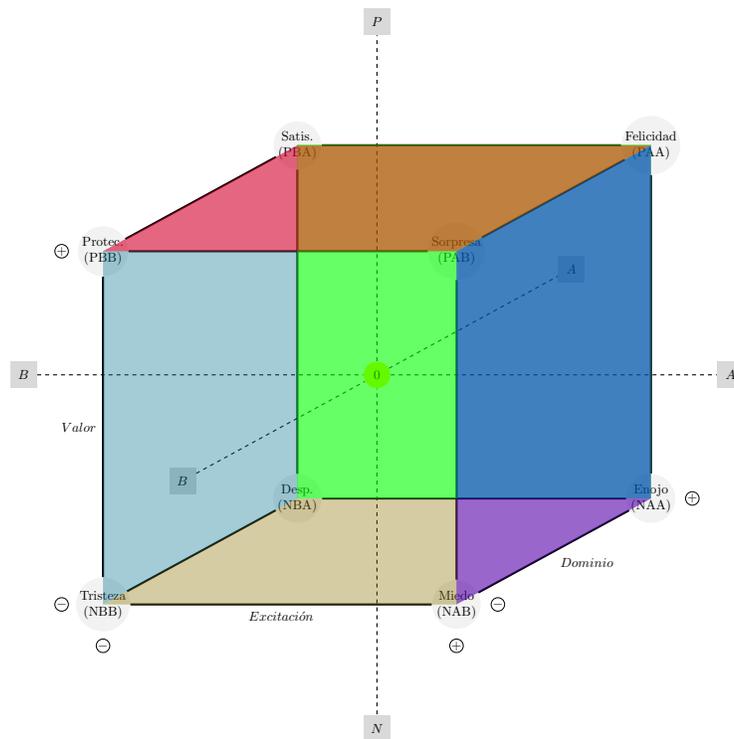


Figura 2.1: Espacio de emociones propuesto por Lui, et al., en [1].

La mayoría de los trabajos orientados al reconocimiento emocional están basados en la clasificación binaria, es decir de un alto o bajo nivel en cada una de las emociones primitivas, obteniendo hasta tres modelos entrenados y cada uno con sus respectivas métricas de evaluación. Esto porque resulta mucho más simple entrenar tres modelos con dos clases cada uno, que un modelo con hasta ocho clases.

Existen modelos basados en el procesamiento de señales EEG siguiendo metodologías diversas como la selección específica de canales de electrodos, para determinar si existen lóbulos o canales que predominen en información sobre otros al obtener una mayor exactitud en la clasificación. También, existen procesos de separación de bandas de frecuencias específicas asociadas a las señales EEG como se muestran en la Tabla 1.1. Estos modelos se basan en una mayor obtención de exactitud al suprimir las bandas que no aportan información relevante en las señales EEG (θ y γ). Así, mediante un proceso de filtrado, se logra obtener la mayor información cognitiva modelada por las ondas α y β , las cuales son frecuencias asociadas al estado de alerta y actividad cognitiva consiente. Si bien las ondas asociadas a una actividad cognitiva significativa, ondas α y β , tienden a mostrar una mayor magnitud en el lóbulo frontal, no se debe descartar la propagación de estas señales inclusive en el lóbulo occipital. Es por esto que en diversas investigaciones se ha optado por realizar un filtro pasa banda, el cual es aplicado directamente sobre cada periodo de la señal discreta, manteniendo así la dependencia temporal y realizando una extracción de características aplicada directamente a las ondas α y β .

Existen también metodologías donde las señales son procesadas mediante análisis espectral utilizando la transformada de fourier (FT), transformada rápida de fourier (FFT) o la transformada corta de fourier (STFT). También, se utiliza la transformada wavelet (WT), transformada continua wavelet (CWT) y la descomposición continua wavelet (CWD). El análisis espectral provee información asociada a la energía de una señal, es por ello que diversos autores utilizan una transformación de dominio para realizar extracciones de características de tipo estadísticas y/o de tiempo-frecuencia. A continuación, se presenta brevemente la discretización de las emociones a partir de las emociones primitivas, así como los nombres asociados a cada una de ellas.

2.1. Caracterización de las emociones

Como se mencionó en la sección anterior, en [5] Ekman, et al., define cinco emociones básicas en su trabajo: felicidad, tristeza, sorpresa, miedo e ira. Además, establece que estas emociones son globales e independientes de la cultura. De manera similar, Tomkins, et al., en [29], caracterizó ocho emociones, las cuáles nombró a partir de una emoción positiva, neutral o negativa. Las emociones positivas las nombró como interés/excitación y placer/alegría, en neutral utilizó los términos sorpresa/susto, y por último, para las negativas empleo las palabras aflicción/angustia, miedo/terror, vergüenza/humillación, desprecio/disgusto y enojo/furia.

A partir de lo anterior, en [1] se propone un modelo de tres dimensiones, mutuamente ortogonales, donde se caracterizan emociones discretas a partir de las emociones de valor, excitación y dominio, como se aprecia en la Figura 2.1. En este espacio, se considera el valor de las emociones en una escala de negativo (N) a positivo (P), la excitación y el dominio de un bajo (B) a alto (A) nivel. Es por esto, que emociones como la tristeza presentan los niveles “NBB”, ya que se compone de un valor negativo, una baja excitación y un bajo dominio. Contrario con respecto a la felicidad donde existe un valor positivo, una alta excitación y un alto dominio (“PAA”). Este modelo es el que se utiliza en nuestra metodología para caracterizar las emociones específicas derivadas de las emociones primitivas.

A continuación, se redactan algunos modelos de reconocimiento emocional existentes en la literatura. También, se presentan características asociadas a sus respectivas metodologías, como la base de datos utilizada, técnicas de extracción de características, entre otras.

2.2. Trabajos relacionados

Existen tres aspectos que conforman a las emociones, y que son de interés al momento de realizar un sistema de reconocimiento emocional. Dependiendo la excitación con la cual sea sometido un sujeto de pruebas, las emociones primitivas poseerán una magnitud específica a una escala determinada. El valor, la excitación y el dominio pueden ser utilizadas para caracterizar emociones específicas.

En la última década, la mayoría de los trabajos relacionados al reconocimiento

emocional ha sido enfocado a la creación de modelos de clasificación de emociones primitivas. Si bien el objetivo general de la mayoría de estos trabajos es lograr una clasificación con base en las emociones primitivas, a partir de señales EEG, es importante diferenciar las metodologías. Por ejemplo, en las siguientes subsecciones se presentan diversos métodos de extracción de características a partir de modelos entrenados mediante algoritmos de aprendizaje de máquina y aprendizaje profundo. También, se realiza un análisis con respecto a las similitudes y diferencias entre los distintos conjuntos de datos utilizados en cada una de las metodologías, así como el alcance de exactitud para cada modelo de clasificación.

2.2.1. Características del modelo y conjunto de datos

El analizar las diferencias entre las investigaciones previas, a nivel del conjunto de datos utilizado, es un aspecto importante en toda investigación. Por ejemplo, en este documento se presenta la Tabla 2.1 donde se analizan diferentes aspectos de las metodologías similares.

La primera diferencia notable en la Tabla 2.1 es que la mayoría de las metodologías, ya sea para clasificación de emociones o emociones primitivas, han sido desarrolladas de manera supervisada. Solamente la metodología propuesta por Gannouni, et al., en [30] utiliza un algoritmo no supervisado. Ahí, se muestra un enfoque con respecto al incremento o decremento en las bandas de interés de las señales EEG en algunos lóbulos cerebrales. Éste, establece que el valor puede ser detectado monitoreando el cambio en la intensidad mostrando un incremento en las señales α en el lóbulo frontal y β en el lóbulo parietal derecho del cerebro. Además, se utiliza una técnica para generar clústers, separando así los ejemplos por clase (la técnica no es proporcionada). Los autores utilizan tres etiquetas para su modelo mediante la umbralización de valores de las emociones primitivas: “positiva”, “neutral” y “negativa”.

Por otra parte, en [31], Cimtay, et al., utilizan una metodología en la cual la clasificación es dependiente de los sujetos de prueba. Lo anterior se traduce en una clasificación para cada sujeto y se determinó un promedio de exactitud. En esta metodología, los autores muestran un modelo en el cual utiliza únicamente ocho canales de una diadema (AF1, AF2, F3, F4, F5, F6, F7, T7, T8). Además, se entrenan tres

modelos a partir de tres conjuntos de datos diferentes; LUMED [32] (Loughborough University Multimodal Emotion Database), SEED-IV [33] (SJTU Emotion EEG Dataset) y DEAP [3] (Database for Emotion Analysis using Physiological Signals).

Tabla 2.1: Características del modelo y conjunto de datos de trabajos relacionados

| Emociones primitivas | | | | | |
|--------------------------------|---------------|--------------------------|------------------------------|-------------------------|---------------------|
| Autor y año | ¿Supervisado? | ¿Dependiente de sujetos? | Conjunto de datos | Número de canales | ¿Fusión de modelos? |
| Zhuang, et al. [34] (2017) | Sí | No | DEAP | 32 | No |
| Gannouni, et al. [30] (2020) | No | No | DEAP | Adaptativo | No |
| Cimtay, et al. [31] (2020) | Sí | Sí | DEAP, SEED [35], LU-MED [32] | 8 por conjunto de datos | No |
| Garg, et al. [36] (2019) | Sí | No | DEAP | 32 | No |
| Garg, et al. [37] (2020) | Sí | No | - | 24 | No |
| Taran, et al. [38] (2019) | Sí | No | DEAP | 32 | No |
| Salama, et al. [39] (2020) | Sí | No | DEAP | 32 | Sí |
| Emociones: clases discretas | | | | | |
| Wang, et al. [40] (2011) | Sí | No | Conjunto de datos propio | 62 | No |
| Alazari, et al. [41] (2018) | Sí | No | DEAP | 22 | No |
| Lin, et al. [42] (2010) | Sí | No | Conjunto de datos propio | 32 | No |
| Murugappan, et al. [43] (2010) | Sí | No | Conjunto de datos propio | 62 | No |
| Yang, et al. [44] (2020) | Sí | No | SEED | 32 | No |
| Khare, et al. [45] (2020) | Sí | No | Conjunto de datos propio | 24 | No |

Al analizar el conjunto de datos utilizados por cada trabajo en la Tabla 2.1, encontramos que en su mayoría, el conjunto de datos DEAP fue el más común. Aunque existen investigaciones que desarrollaron su propio conjunto de datos como en [42] y [43] ya que estos trabajos fueron desarrollados previo a la publicación del conjunto de datos DEAP (2012). De manera similar, Khare, et al., en [45] desarrolló propiamente un conjunto de datos.

Otra de las diferencias más notables entre los trabajos relacionados es el número de electrodos, o canales de señales EEG, utilizados. Por ejemplo, en [37] emplearon una técnica a partir de 24 canales de señales EEG. Wang, et al., en [40], a partir de su propio conjunto de datos, utilizó 62 electrodos de adquisición de señales electroencefalográficas, al igual que en [43]. Alazari, et al., en [41] al usar el conjunto de datos DEAP empleó únicamente 22 canales de información. Por último, en [45], Khare, et al., a partir de su conjunto de datos utilizaron solamente 24 señales EEG.

Existen también metodologías asociadas a la implementación de una fusión de modelos. Dentro de los trabajos relacionados incluidos en la Tabla 2.1 existe solamente una metodología que realiza este enfoque. En [39], Salama, et al., propuso un trabajo basado en la fusión de modelos a partir de tres algoritmos de redes neuronales convolucionales, utilizando señales EEG y reconocimiento facial. Este tipo de metodologías mejoran considerablemente la exactitud de los modelos sacrificando memoria y tiempo de entrenamiento.

En las siguientes dos subsecciones se muestran otros aspectos importantes de los trabajos relacionados expuestos en la Tabla 2.1 como las técnicas de metodología empleadas, el tipo de modelo, la exactitud del modelo, entre otros.

2.2.2. Emociones primitivas

Existen en la literatura modelos enfocados a la clasificación de las emociones primitivas. Este tipo de modelos utilizan dos o más emociones primitivas, donde los espacios, o emociones primitivas, más comunes se basan en el uso de valor-excitación (Valence-Arousal, VA) o valor-excitación-dominio (Valence-Arousal-Dominance, VAD).

Es importante mencionar que los modelos realizados mediante la clasificación de emociones primitivas presentan la cualidad de entrenar un modelo por cada emoción

primitiva. Es decir, si se realiza una metodología para clasificar el espacio VA, será necesario contar con dos modelos de clasificación, uno de ellos para el valor y otro para la clasificación de la excitación. La misma metodología sería aplicable si se pretenden clasificar las emociones primitivas en el espacio VAD, generando así tres modelos de clasificación. En la Tabla 2.2 se muestran trabajos relacionados al reconocimiento emocional de emociones primitivas en el espacio VA o VAD.

Una diferencia altamente notable entre los trabajos incluidos en la Tabla 2.2 es la gran variedad de técnicas de procesamiento de las señales EEG, ya sea por cambio de dominio o extracción de características. Por ejemplo, en [30, 34] ambos trabajos utilizan una extracción de características de tiempo-frecuencia asociadas a la energía de la señal, pero con metodologías de clasificación diferentes. Por ejemplo, en [34], Zhuang, et al., únicamente utiliza los vectores de valor y excitación para clasificar ambos modelos mediante dos etiquetas en cada uno, utilizando un algoritmo de máquina de soporte vectorial (SVM). En cambio, en [30], Gannouni, et al., utiliza el espacio VAD con tres etiquetas por modelo (no muestra técnica de generación de clústers).

Otro método para extracción de características es el que se presenta en [36, 38]. En estos trabajos se emplea una extracción de características estadísticas. Por otro lado, en [31, 37, 39] se realiza un procesamiento aplicado las señales EEG para, posteriormente, utilizar la información de transformación como entrada a sus modelos de clasificación.

Como se puede apreciar en la tabla comparativa de trabajos enfocados a la clasificación de emociones primitivas, la mayoría de las metodologías clasifican solamente las emociones primitivas en el espacio VA. Exceptuando los trabajos en [30, 31, 36]. En [30], como se mencionó en el párrafo anterior, utiliza el espacio VAD. En [36] realizó la clasificación en un espacio VADL (Valor-Excitación-Dominio-Relación) y en [31] no se utiliza un espacio definido ya que, en esta metodología, se implementa un algoritmo donde el modelo únicamente detecta si la emoción es “positiva” o “negativa”.

Diversos autores en sus metodologías emplean de dos a tres etiquetas por modelo de clasificación. Estos modelos se basan en la umbralización de los coeficientes de las emociones primitivas para otorgar la etiqueta “positiva” o “negativa”, o “positiva”,

“neutral” o “negativa”, a cada ejemplo del conjunto de datos.

Tabla 2.2: Características de trabajos con enfoque de emociones primitivas

| Emociones primitivas | | | | | | |
|------------------------------|---|-------------------|--------------------------|----------------------------------|---|--------|
| Autor y año | Técnica/Extracción de características | Espacio emocional | Número de clases | Clasificador/Arquitectura de red | Exactitud (%) | F1 (%) |
| Zhuang, et al. [34] (2017) | Características tiempo-frecuencia | VA | 2 por Dim. | SVM | 70.41 72.10 | - |
| Gannouni, et al. [30] (2020) | Características tiempo-frecuencia | VAD | 3 por Dim. | - | 82.35 79.95 71.14 | - |
| Cimtay, et al. [31] (2020) | Arreglo 2-D usando señales EEG y aumento de datos | - | Positivo & Negativo | Inception-ResnetV2 | 86.5 72.8 81.8 por conjunto de datos | - |
| Garg, et al. [36] (2019) | Características estadísticas | VADL | 2 por Dim. | LSTM | 84.89 83.85 84.37 80.72 | - |
| Garg, et al. [37] (2020) | Imágenes CWD | VA | 2 por Dim / 3 por Dim | CNN | 92.19 61.23 /83.59 55.56 | - |
| Taran, et al. [38] (2019) | Características estadísticas | VA | 2 por Dim. | MC-LS-SVM (Morlet kernel) | 90.63 | 90.64 |
| Salama, et al. [39] (2020) | Señales EEG y reconocimiento facial | VA | 2 por Dim | 3D-CNN | 96.13 96.79 | - |

Además, de la Tabla 2.2 también se puede apreciar que se pueden aplicar metodologías de clasificación mediante el uso del aprendizaje de máquina o aprendizaje profundo. Donde, de los trabajos citados en la tabla comparativa, solamente uno de ellos presenta otras métricas de evaluación del desempeño del modelo. El realizar otras métricas permite al lector interpretar inclusive el equilibrio del conjunto de datos, con respecto al número de clases en el modelo.

2.2.3. Emociones: clases discretas

Los modelos basados en emociones específicas como la felicidad, miedo, ira, etc., tienden a presentar algunas complicaciones comparados con los modelos de emociones primitivas. Ya sea utilizando modelos de aprendizaje de máquina o aprendizaje profundo, estos modelos contienen un mayor número de etiquetas, lo que los hace más susceptibles a pérdidas de exactitud. A continuación, en la Tabla 2.3 se muestran los trabajos relacionados a emociones específicas, mismo enfoque de clasificación que se presenta en este trabajo.

Como se puede apreciar en la Tabla 2.3, en [40–44] utilizan metodologías basadas en la extracción de características de tiempo-frecuencia. De estos trabajos, solamente el propuesto por Alazrai, et al., realiza una combinación de características estadísticas y de tiempo-frecuencia. También, existen modelos similares a nuestra metodología propuesta, como en [45] donde utilizan imágenes de entrada a partir del cálculo de la energía asociada a las señales EEG. La técnica de análisis espectral utilizada en esta metodología se realiza a partir de la pseudo distribución Wigner-Ville (SPWVD). Este último trabajo es aquel con el cual se comparará directamente nuestro enfoque propuesto.

Por otra parte, la mayoría de los trabajos citados realizaron la discretización de los ejemplos contenidos en los conjuntos de datos de manera directa, es decir no realizaron una umbralización de datos empleando las emociones primitivas. Solamente la metodología propuesta por Alazrai, et al., en [41] discretizó los valores de las emociones primitivas VA, esto debido a que en esta metodología se utiliza la base de datos DEAP. Además, los trabajos contenidos en la Tabla 2.3 clasifican únicamente cuatro emociones, mientras que uno de nuestros modelos propuestos clasifica ocho emociones.

Tabla 2.3: Características de trabajos con enfoque de emociones por clases discretas

| Emociones: clases discretas | | | | | | |
|--------------------------------|--|-------------------|------------------|----------------------------------|---------------|--------|
| Autor y año | Técnica/Extracción de características | Espacio emocional | Número de clases | Clasificador/Arquitectura de red | Exactitud (%) | F1 (%) |
| Wang, et al. [40] (2011) | Características de frecuencia | - | 4 | kNN | 66.51 | - |
| Alazari, et al. [41] (2018) | Características estadísticas & tiempo-frecuencia | VA | 4 | SVM | 75.1 | - |
| Lin, et al. [42] (2010) | Características de tiempo-frecuencia | - | 4 | SVM | 82.29 | - |
| Murugappan, et al. [43] (2010) | Características de tiempo-frecuencia | - | 4 | kNN | 83.26 | - |
| Yang, et al. [44] (2020) | Características de tiempo-frecuencia | - | 4 | BiLSTM | 84.21 | 80.37 |
| Khare, et al. [45] (2020) | Imágenes SPWVD | - | 4 | AlexNet | 90.98 | 90.92 |
| | | | | ResNet50 | 91.91 | 91.95 |
| | | | | VGG16 | 92.71 | 92.68 |
| | | | | Arquitectura propia de CNN | 93.01 | 93.02 |

También, en la tabla comparativa, se encuentra que cuatro de las metodologías propuestas están basadas en el uso de aprendizaje de máquina, donde dos de estas están desarrolladas a partir de un SVM y dos mediante un kNN (k-próximos vecinos). Por otro lado, la metodología propuesta en [44] se basa el uso de una arquitectura de memoria bidereccional a corto plazo (BiLSTM). Y, por su parte, Khare, et al.,

en [45] desarrolló tres modelos de clasificación basados en el uso de las CNN. Es destacable también que solamente dos de las metodologías citadas en la Tabla 2.3 contienen otras métricas de evaluación de desempeño del modelo entrenado.

2.3. La transformada wavelet

Numerosas metodologías de análisis espectral han sido desarrolladas a partir de la Transformada de Fourier (TF): espectrograma, pseudo-espectro y la distribución Wigner-Ville, por mencionar algunas. Estas metodologías parten de utilizar una señal senoidal como “onda madre” para realizar el procesamiento de la transformación al dominio de tiempo-frecuencia. Si bien, en la mayoría de los casos, el uso de la TF ha demostrado otorgar buenos resultados, es importante conocer la naturaleza o comportamiento de los fenómenos a modelar mediante un procesamiento digital de señales para elegir correctamente la metodología de transformación de dominio.

Dos de las metodologías más usadas para el procesamiento de las señales EEG son la TF y la Transformada Wavelet (TW). La TW se desarrolló con la finalidad de estudiar la variación en tiempo o frecuencia de una señal $x(t)$ y una de las diferencias principales en el análisis de ambas ondas madre (TF y TW), es el número de oscilaciones. Por ejemplo, en la TW es posible variar el número de oscilaciones contenidas bajo su respectiva envolvente, contrario al uso de la TF donde, a pesar de ser una suma de senos y cosenos, esto mantienen una sola oscilación, en cada periodo de la señal, constante con respecto al tiempo.

A continuación, se presenta la teoría de la transformada wavelet utilizada para la metodología de este trabajo.

2.3.1. La transformada analítica wavelet

La transformada wavelet se origina a partir de la necesidad de estudiar la variación en tiempo o frecuencia de una señal $x(t)$. La TW se define como en (2.1), donde $W_\psi(s, \tau)$ son los coeficientes wavelet, $x(\tau)$ la señal original a transformar, $\psi(t)$ función utilizada para analizar simultáneamente el tiempo y la frecuencia de una señal, s es la escala (dilata o comprime la wavelet) y τ el parámetro de posición de desplazamiento con respecto a t [2].

$$W_\psi(t, s) \equiv \frac{1}{s} \int_{-\infty}^{+\infty} \psi^*\left(\frac{\tau - t}{s}\right) x(\tau) d\tau \quad (2.1)$$

Al utilizar únicamente señales analíticas en la TW, se omiten las frecuencias negativas. Lo anterior se puede expresar de manera matemática como $\Psi(\omega) = 0$ para $\omega \leq 0$. La transformada analítica wavelet, o TAW, se define como en (2.2).

$$W_\psi(t, s) = \frac{1}{2\pi} \int_0^\infty \Psi^*(s\omega) X(\omega) e^{i\omega t} d\omega \quad (2.2)$$

Así, a partir de que la TAW comprende únicamente frecuencias positivas, $\Psi(\omega)$ tiene un pico máximo de amplitud en el dominio de la frecuencia en $\omega = \omega_\psi$ al cual se le denomina “pico de frecuencia”.

Una manera de caracterizar el comportamiento de las wavelets, es mediante las versiones normalizadas de sus derivadas en el dominio de la frecuencia. En (2.3) se define una ecuación general para la n-ésima derivada de la wavelet en el dominio de la frecuencia, a la cual se le conoce como “derivadas adimensionales”.

$$\hat{\Psi}_n(\omega) \equiv \omega^n \frac{\Psi^n(\omega)}{\Psi(\omega)} \quad (2.3)$$

A partir de (2.3) se define la “duración” de la wavelet como se muestra en (2.4).

$$P_\psi \equiv \sqrt{-\tilde{\Psi}_2(\omega_\psi)} = \sqrt{-\omega_\psi^2 \frac{\Psi''(\omega_\psi)}{\Psi(\omega_\psi)}} \quad (2.4)$$

El factor P_ψ se utiliza para determinar el número de oscilaciones comprendidas en la ventana de $\pm P_\psi$ donde el punto central de la ventana se encuentra en ω_ψ . El número de oscilaciones se calcula mediante $\frac{P_\psi}{\pi}$.

A continuación, se presenta la teoría de la transformada wavelet morse (TWM). Ésta última es la que se utiliza para calcular la transformada continua wavelet (TCW) utilizada para la representación de los escalogramas.

2.3.2. La transformada continua wavelet

Una de las variantes de la TW es la TCW. A la representación gráfica del espectro de una señal procesada mediante la TCW se le denomina escalograma. En [46] se describe a un escalograma como la densidad de energía asociada a una frecuencia en específico en un tiempo específico. La TCW utiliza como onda madre la MWT y ha demostrado ser superior a la STFT y FFT debido a su capacidad de procesar y transformar una mayor cantidad de información de la señal original al identificar las variaciones en tiempo-frecuencia a diferentes resoluciones. En la Figura 2.2 se muestran las wavelets más comunes, donde la línea azul representa la envolvente, la línea roja la parte real y la línea amarilla la parte imaginaria de la wavelet.

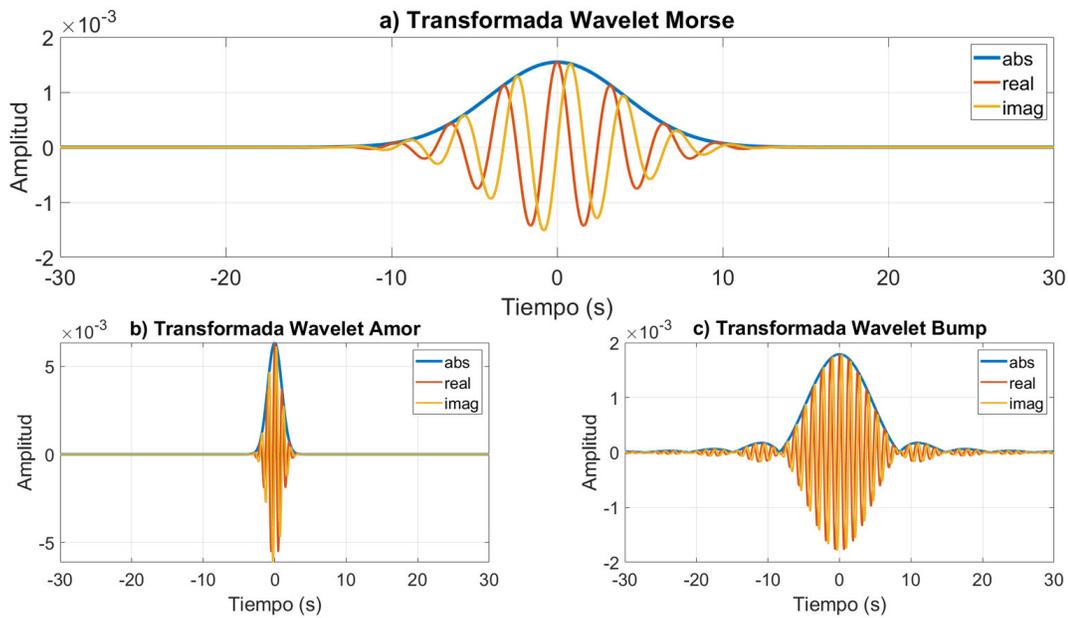


Figura 2.2: Wavelets más comunes. La diferencia principal entre (a), (b) y (c) es la variación del factor β y γ .

La TWM comprende a una familia de wavelets definidas en el dominio de la frecuencia como en (2.5), donde $a_{\beta,\gamma}$ es una constante de normalización, $U(\omega)$ la función escalón unitario, β y γ son parámetros de control de la wavelet. Al variar los parámetros β y γ la TWM cambia de forma.

$$\Psi_{\beta,\gamma}(\omega) = \int_{-\infty}^{+\infty} \psi_{\beta,\gamma}(t) e^{-i\omega t} dt = U(\omega) a_{\beta,\gamma} \omega^\beta e^{-\omega^\gamma} \quad (2.5)$$

La TWM es, en esencia, un conjunto de filtros pasa-banda donde se analizan diferentes conjuntos de frecuencias. Estos filtros se aplican a las variaciones de frecuencia y al conjunto de frecuencias analizadas en cada filtro se les conoce como escalas. Nótese también que el factor “ ψ ” se reemplaza por “ β,γ ”, así el pico de magnitud en el dominio de la frecuencia se localiza en $\omega_{\beta,\gamma} \equiv \left(\frac{\beta}{\gamma}\right)^{1/\gamma}$. Además, el factor $a_{\beta,\gamma} \equiv 2\left(\frac{e\gamma}{\beta}\right)^{\beta/\gamma}$.

Ahora bien, la duración de la wavelet para la TWM se calcula a partir de (2.6).

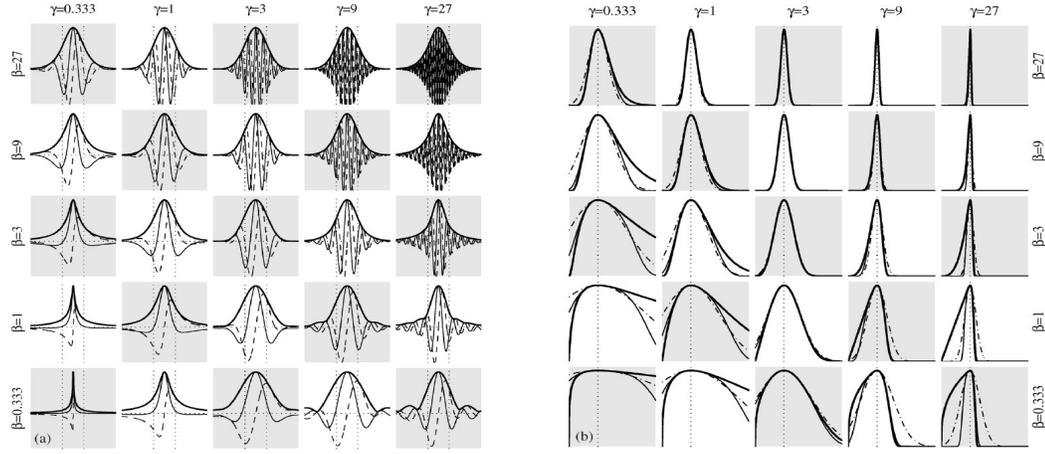
$$P_{\beta,\gamma} \equiv \sqrt{-\tilde{\Psi}_{2;\beta,\gamma}(\omega_{\beta,\gamma})} = \sqrt{\beta\gamma} \quad (2.6)$$

La familia de las MWT se representa gráficamente en la Figura 2.3 propuesta en [2]. En esta gráfica, se puede apreciar una familia de 25 wavelets en el dominio del tiempo ($\psi_{\beta,\gamma}(t)$) en la Figura 2.3a y en el dominio de la frecuencia ($\Psi_{\beta,\gamma}(\gamma)$) en Figura 2.3b.

De la Figura 2.3a, la familia de wavelets en el dominio del tiempo ilustran una línea continua gruesa que representa la envolvente, la línea continua más angosta representa la parte real de la wavelet y la línea discontinua modela la parte imaginaria. Además, la línea de puntos sobre el eje x representa el tiempo y las líneas de puntos sobre el eje y modela el parámetro $\pm P_{\beta,\gamma}$ (duración de la wavelet).

En el dominio de la frecuencia, Figura 2.3b, la familia de MWT muestran una línea continua la cual representa a la wavelet en el dominio de la frecuencia. La línea estrecha continua es una aproximación gaussiana y la línea intermitente punteada representa a una aproximación cuadrática. La línea vertical punteada establece la ubicación del pico máximo de magnitud $\omega_{\beta,\gamma}$.

Nótese que al aumentar simultáneamente el factor β y γ en el dominio del tiempo de la TWM (diagonales de la Figura 2.3a), ésta tiende a mantener constante su ancho de banda y el número de oscilaciones en la wavelet. Esto derivado del cálculo a partir de (2.6) que se muestra en (2.7) como ejemplo. Además, se puede apreciar que la wavelet $\psi_{1/3,1/3}(t)$ se semeja a una delta de Dirac debido a que el factor $P_{\beta,\gamma}/\pi$



(a) Familia MWT en el dominio del tiempo. (b) Familia MWT en el dominio de la frecuencia.

Figura 2.3: Variaciones del parámetro β y γ para una familia de MWT. [2]

es muy pequeño. Lo mismo ocurre en el dominio de la frecuencia para $\Psi_{27,27}(\omega)$. También, analizando las diagonales de la esquina superior izquierda a la esquina inferior derecha de la Figura 2.3b, podemos apreciar que al aumentar los parámetros β y γ el sesgo de la wavelet tiende a variar.

La TCW se desarrolla a partir de la TWM analizando la frecuencia y el tiempo a resoluciones variables. De esta manera se desarrolla la representación de escalograma de una señal discreta.

$$\begin{aligned}
 \frac{P_{27,\frac{1}{3}}}{\pi} &= \frac{3}{\pi} \approx 1 \\
 \frac{P_{9,1}}{\pi} &= \frac{3}{\pi} \approx 1 \\
 &\vdots \\
 \frac{P_{\frac{1}{3},27}}{\pi} &= \frac{3}{\pi} \approx 1
 \end{aligned}
 \tag{2.7}$$

En el Capítulo 3, se presenta la metodología propuesta en este trabajo, donde se describe el conjunto de datos utilizado, la naturaleza y características de la señal

electroencefalográfica, la representación de los escalogramas y la arquitectura de red utilizada para los modelos de clasificación.

Capítulo 3

Metodología

En este trabajo se proponen dos modelos de clasificación de emociones a partir de la discretización de los valores en las emociones primitivas de valor, excitación y dominio. La metodología propuesta consiste en utilizar señales EEG para realizar una transformación de dominio. Como se muestra en la Figura 3.1, el vector de la señal EEG original fue obtenido a partir del conjunto de datos en [3] almacenando información de la diferencia de potencial eléctrico a partir de una excitación en cada sujeto. El vector de la señal EEG es empleado para realizar una transformación mediante análisis espectral utilizando la transformada continua wavelet (TCW) para adquirir una representación de escalograma. Al obtener la representación espectral de la energía asociada a la señal EEG, ésta se almacena como imagen en formato **.png* para ser utilizada como entrada a un clasificador. Esto se realiza para cada vector de señal EEG en el conjunto de datos. La arquitectura del clasificador está basada en la red GoogLeNet [47]. La red GoogLeNet es una red basada en la arquitectura de las redes neuronales convolucionales y en la implementación de una capa llamada “Inception”, que son filtros convolucionales para realizar una extracción de características más robusta. En nuestro trabajo, se emplearon técnicas de transferencia de aprendizaje de los pesos de imageNet. La red se utiliza un conjunto de imágenes de entrenamiento y validación, 80 % y 20 %, respectivamente, para clasificar dos espacios de emociones (VA y VAD).

A continuación, en las siguientes subsecciones se describen de manera detallada cada uno de los módulos anteriormente descritas e ilustradas en la Figura 3.1.

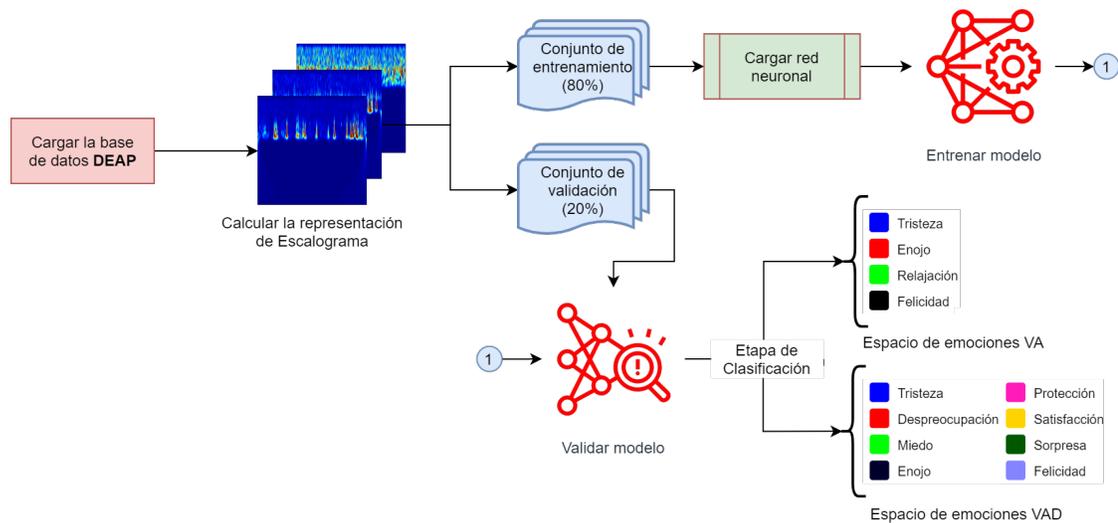


Figura 3.1: Diagrama general de la clasificación de señales EEG.

3.1. Conjunto de datos

En este trabajo utilizamos un enfoque mediante el uso del conjunto de datos *DEAP* (Database for Emotion Analysis using Physiological Signals) [3]. Este conjunto de datos, publicado por Koelstra, et al., en 2012, contiene mediciones de la diferencia de potencial eléctrico de la actividad electroencefalográfica en sujetos de prueba al ser excitados mediante videos musicales.

En el proceso de adquisición de señales EEG del conjunto de datos *DEAP* se utilizó una pantalla para mostrar las diferentes etapas del experimento, las cuáles se mencionan a continuación:

- Muestra de dos segundos con la información de la prueba actual para informar al sujeto de su progreso.
- Una grabación base de cinco segundos de una cruz en pantalla para relajación.
- El video musical de un minuto para generar la excitación.
- Valoración del sujeto para valor, excitación, dominio y relación. La valoración en un rango de 1-9.

Durante las pruebas, cada sujeto fue excitado con 40 diferentes videos musicales mientras usaba una diadema de electrodos para la adquisición de las señales generando un arreglo de datos distribuido como se muestra en la Figura 3.2, donde i representa a los videos musicales, j los canales de electrodos y k la longitud del vector de la señal EEG. Se utilizaron 32 canales en la diadema de electrodos siguiendo el estándar de un *sistema 10-20*. En la Figura 3.3 se muestra un esquema de la diadema donde los electrodos en color gris representan a los utilizados en este conjunto de datos.

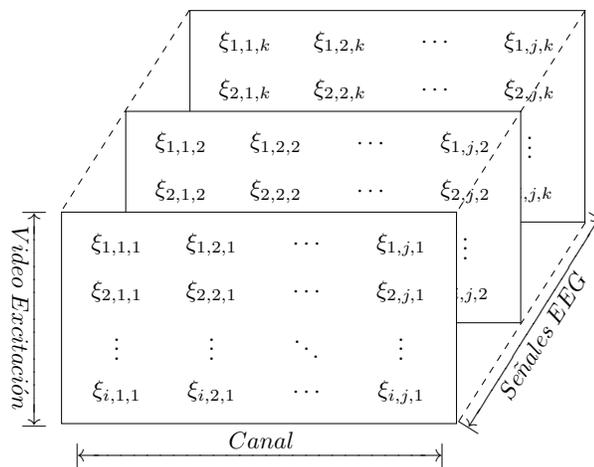


Figura 3.2: Arreglo de datos adquiridos por cada sujeto de pruebas en la base de datos DEAP [3].

El vector original de la señal EEG está muestreado a 128 muestras por periodo de la señal, donde un periodo equivale a 1 segundo. El sujeto de pruebas fue excitado durante 60 segundos, reservando 3 segundos iniciales para adaptar al sujeto previo a la excitación. La totalidad del vector de la señal original corresponde a 8064 muestras y al omitir los 3 segundos de adaptación, el vector se reduce a 7680 muestras. De los 32 sujetos de prueba, las 40 excitaciones y los 32 canales de electrodos, se obtiene una totalidad de 40960 ejemplos de señales EEG.

El arreglo de cada sujeto de pruebas también contiene una matriz de datos de etiquetas. Las etiquetas de este conjunto de datos dependen de la magnitud de las emociones primitivas que el sujeto experimentó y valoró al ver y escuchar cada video. Originalmente, se realizó una valoración de 120 videos por 14 sujetos, posteriormente

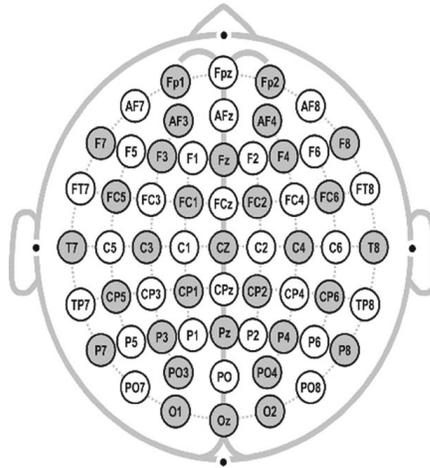


Figura 3.3: Diagrama del sistema de adquisición 10-20.

esto se redujo a 40 videos valorados por 32 sujetos. Las 40 excitaciones, en nuestro modelo etiquetas, fueron seleccionadas mediante los 10 videos más alejados con respecto al centro (aros verdes), en cada cuadrante de la Figura 3.4. Cada aro en la Figura 3.4 representa un video de los 120 valorados, donde la posición de estos se basa en el cálculo de la media entre la desviación estándar (μ_n/σ_n) valorada por cada uno de los 14 sujetos, con respecto al valor y la excitación.

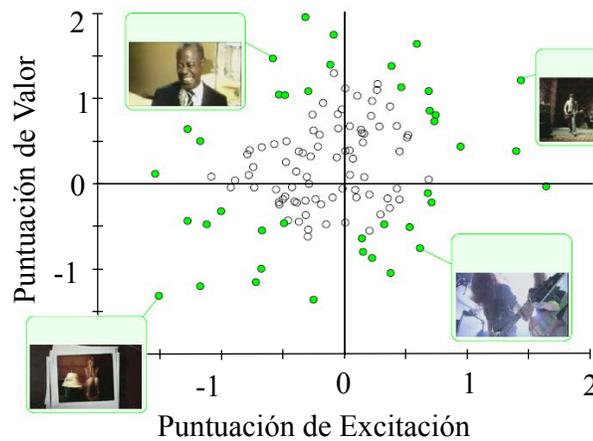


Figura 3.4: La relación μ_n/σ_n con respecto al valor y excitación de cada video [3].

La magnitud para las emociones primitivas se encuentra en el rango de 1 – 9, siendo 1 el valor más bajo de excitación y 9 el más alto. La matriz \mathbb{L}_n en (3.1) representa las etiquetas donde ψ modela el valor, α la excitación, φ el dominio, v la relación e i el video de excitación.

$$\mathbb{L}_n = \begin{pmatrix} \psi_1 & \alpha_1 & \varphi_1 & v_1 \\ \psi_2 & \alpha_2 & \varphi_2 & v_2 \\ \psi_3 & \alpha_3 & \varphi_3 & v_3 \\ \vdots & \vdots & \vdots & \vdots \\ \psi_i & \alpha_i & \varphi_i & v_i \end{pmatrix} \quad (3.1)$$

En nuestro trabajo, utilizamos un umbral de 4.5 para discretizar las emociones primitivas. De esta manera, se proponen dos modelos de clasificación, uno es la representación en el espacio de valor-excitación como en la Figura 3.5. El espacio VA utiliza únicamente las magnitudes de los vectores de valor y excitación para caracterizar cuatro emociones o clases discretas: “tristeza”, “enojo”, “relajación”, y “felicidad”.

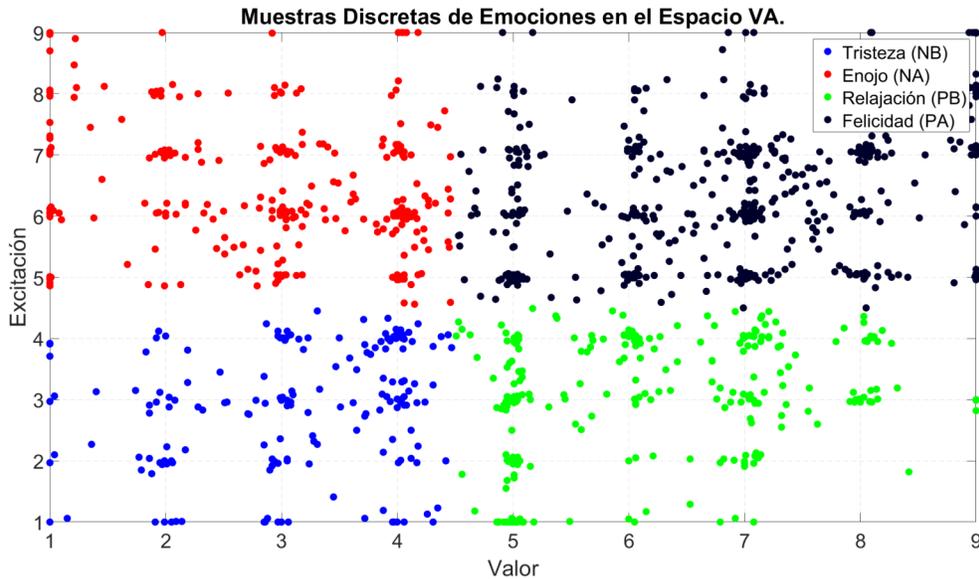


Figura 3.5: Espacio de emociones VA.

El segundo espacio de emociones está modelado por las magnitudes presentes en

los vectores de valor, excitación y dominio, como se ilustra en la Figura 3.6. El espacio *VAD* amplía la dimensión de las emociones discretas al doble, con respecto al espacio *VA*. Las emociones caracterizadas por el espacio *VAD* son: “tristeza”, “despreocupación”, “miedo”, “enojo”, “protección”, “satisfacción”, “sorpresa”, y “felicidad”.

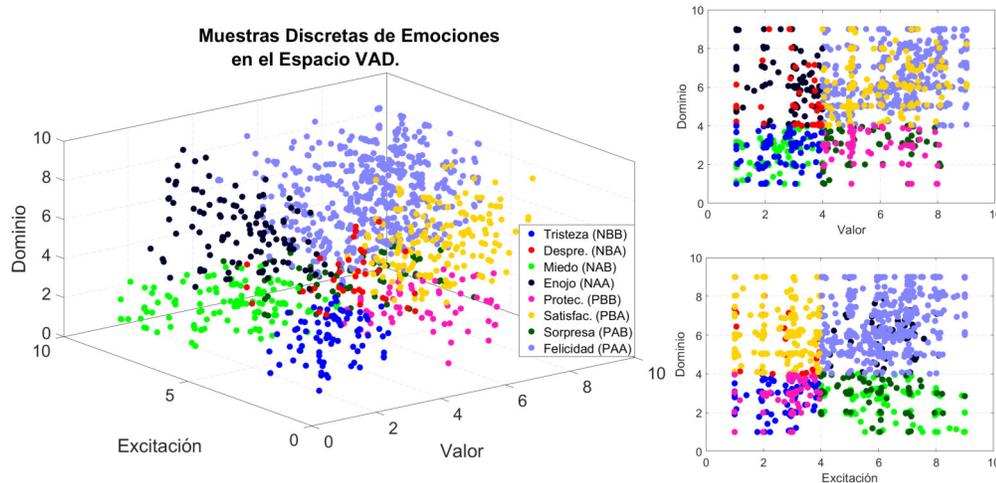


Figura 3.6: Espacio de emociones *VAD*.

3.2. Descripción de la señal

Como se mencionó en el Capítulo 1, para las componentes de frecuencia indicadas en la Tabla 1.1, en cada señal EEG se tienen cuatro frecuencias específicas: theta, alpha, beta y gamma, y están asociadas a ciertos estados mentales. En la Figura 3.7 se aprecia un periodo de una señal EEG aleatoria y sus diferentes componentes de frecuencia.

Algunas investigaciones en el área del reconocimiento emocional proponen una metodología de separación de bandas de interés, [48, 49], de manera que solamente trabajan con las ondas que muestran una mayor actividad encefalográfica (alpha y beta). Lo anterior se realiza mediante un filtro pasa-banda reduciendo al mínimo la información de las bandas theta y gamma. Otro enfoque, como se muestra en [30, 31], se basa en la identificación de regiones en los lóbulos donde se concentra la mayor actividad consciente del sujeto. El lóbulo frontal ha demostrado concentrar la mayor

cantidad de información en características asociadas al estado de alerta, actividades motrices, entre otras.

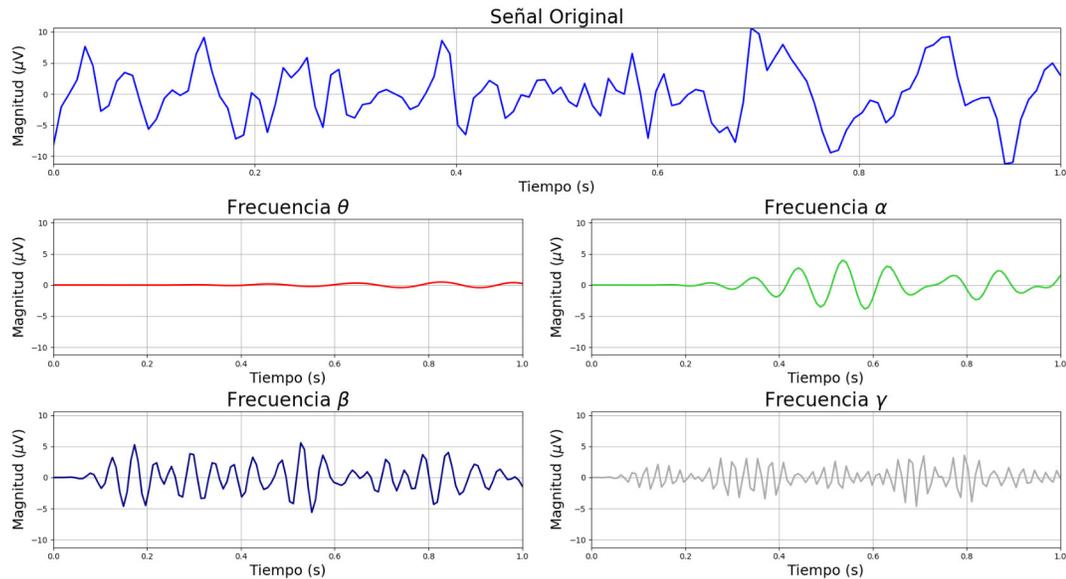


Figura 3.7: Componentes de frecuencias en las señales EEG: theta, alpha, beta y gamma.

Nuestro enfoque es diferente, partimos del principio de si a un sujeto se le excita con miedo, esta emoción estará presente en la señal. De tal manera, es posible omitir un procesamiento de separación de bandas o identificación de lóbulos con una mayor actividad.

Dentro del área del análisis espectral, numerosas metodologías han sido empleadas para clasificar emociones. Una de las más comunes es el uso de wavelets. Las wavelets han demostrado ser altamente eficientes en el tratamiento de señales no estacionarias, contrario a la FFT o STFT. La transformación a un dominio de tiempo-frecuencia está basado en el uso de una onda senoidal, lo cual supone una gran pérdida de términos no transformados de la señal original. Esto debido a que las señales no estacionarias, como lo son las señales EEG, ECG, EMG, etc., no muestran un patrón senoidal o semi-senoidal. Es por ello que en este trabajo se propone una metodología de procesamiento de la señal basada en la transformada wavelet.

Una de las características más significativas de las TW, es el uso de las cajas Heisenberg. Éstas pueden alargar o disminuir su tamaño para adaptarse mejor a

señales como las EEG. También, la TW tiene la característica de mantener una escala pequeña a altas frecuencias y a bajas frecuencias una escala alta, contrario con los dominios de tiempo-frecuencia donde las ventanas son de tamaño constante. Además, en transformaciones basadas en ondas senoidales, la frecuencia es analizada a una misma resolución.

En este trabajo, se utilizó un algoritmo de procesamiento de señales donde se realiza la transformación del vector original de la señal EEG en un escalograma mediante la MWT, es decir, sin emplear una separación de las frecuencias EEG. Para realizar la representación espectral de los escalogramas, se utiliza la ecuación (2.2), como se muestra en el Algoritmo 1 que modela la función wt .

Algoritmo 1 Cálculo de la transformada wavelet a partir de una señal de entrada.

Entrada: señal original ($x(t)$), parámetro β y γ de la wavelet, longitud de la señal (ϕ), $\psi(t)$ tipo de wavelet madre ($\eta = \text{'morse'}$)

Salida: transformada wavelet de la señal original ($W_\psi(t, s)$)

```

1: function COMPUTESCALES( $\beta, \gamma, \phi, \text{'morse'}$ )
2:    $\varrho = 10$ 
3:    $\vartheta = 2^{1/\varrho}$ 
4:    $N = \phi + 2 * \text{floor}(\phi/2)$ 
5:    $[\text{minScl}, \text{maxScl}] \leftarrow \text{frqLimits}(\eta, \phi, \beta, \gamma, \varrho)$ 
6:    $\varsigma \leftarrow \text{max}(\log_2(\text{maxScl}/\text{minScl}), 1/\varrho)$ 
7:    $s = \text{minScl} * \vartheta^{(0:\varsigma\varrho)}$ 
8:    $\omega = 1 : N/2$ 
9:    $\omega = \omega * (2\pi/N)$ 
10:  return  $s, \omega$ 
11: end function
12:  $[s, \omega] \leftarrow \text{COMPUTESCALES}(\beta, \gamma, \phi, \text{'morse'})$ 
13:  $\Psi^*(s\omega) \leftarrow \text{fft}(\psi(t), s, \omega)$ 
14:  $X(\omega) \leftarrow \text{fft}(x(t))$ 
15:  $W_\psi(t, s) \leftarrow \text{ifft}(\Psi^*(s\omega)X(\omega))$ 

```

En el Algoritmo 1 se utilizan los parámetros $x(t)$ la cual representa a nuestra señal EEG, el factor $\beta = 20$ y $\gamma = 3$ de la wavelet, ϕ la longitud del vector de la señal $x(t)$ y 'morse' la wavelet madre utilizada. Se establece el valor $\gamma = 3$ ya que con este valor la envolvente de la wavelet se asemeja lo más posible a una gaussiana y $\beta = 20$ para obtener un número de oscilaciones internas en la wavelet de ≈ 2 , siguiendo

el cálculo de la ecuación (2.7). Además de una duración de $P_{\beta,\gamma} = \sqrt{60} = 7.746$ siguiendo el cálculo de la ecuación (2.6). La función `COMPUTESCALES()` devuelve el cálculo de las escalas y los picos de frecuencia por cada wavelet, donde el factor de incremento de las escalas, ϑ , estará definido por la relación $2^{1/\varrho}$. La variable ϱ representa al número de voces por octava. Una vez obtenida $W_{\psi}(t, s)$, se determina su valor absoluto para realizar la representación de un escalograma, como se aprecia en la Figura 3.8. Además, se utiliza el mapa de color *jet* para aumentar el contraste asociado al mapa de colores en la imagen. Lo anterior no afecta la clasificación debido a que el proceso se realiza para todas las imágenes de escalograma.

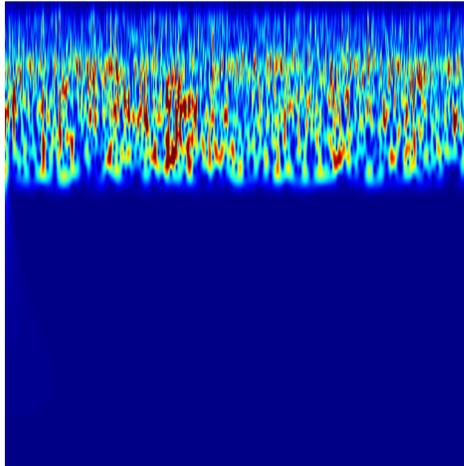


Figura 3.8: Señal EEG en su representación de escalograma en mapa de color *jet*.

Las imágenes de escalograma son almacenadas dando nombre a cada imagen con respecto a su etiqueta, número de sujeto, número de excitación (video) y número de electrodo. Ejemplificando, si la señal del primer sujeto, del primer video y del primer electrodo (FP1) corresponde a la etiqueta “felicidad” (donde “felicidad” = 4 en el espacio VA y “felicidad” = 8 en VAD), la imagen recibirá el nombre de “4_1_1_1.png”, o “8_1_1_1.png”, según el modelo. Así, al momento de leer la imagen para su clasificación, al tomar el primer caracter como la etiqueta, ésta se transforma de texto a entero formando dos vectores de etiquetas equivalente al número de imágenes en el conjunto de entrenamiento y validación.

3.3. Arquitectura de red

Durante la experimentación y sintonización de diferentes modelos de clasificadores, se utilizaron múltiples algoritmos para concluir que el utilizado en esta sección fue el de mejor desempeño. De las diferentes sintonizaciones de red, se utilizó la red GoogLeNet a 100 épocas y una función de activación sigmoide. También, se empleó la misma arquitectura cambiando el número de épocas a 20 y el optimizador a “adam”. En este último entrenamiento se agregaron nuevas capas de desconexión, completamente conectadas, entre otras. De manera similar se utilizó también la red SqueezeNet [50] a 20 épocas y optimizador “adam”. Cada una de las pruebas anteriormente descritas se realizaron únicamente con el modelo a cuatro emociones (espacio VA) y, como se presenta en el siguiente capítulo, fueron menores en exactitud de validación con respecto al modelo presentado en este capítulo.

Al obtener las imágenes de escalograma, éstas son almacenadas para posteriormente ser utilizadas como entrada en el entrenamiento y validación del modelo de clasificación. En nuestro enfoque utilizamos los vectores mutuamente ortogonales, de valor, excitación y dominio del conjunto de datos DEAP [3]. Para realizar la umbralización de etiquetas discretas se utilizó un coeficiente de 4.5. Así, en el espacio VA se obtuvieron cuatro clases y en el espacio VAD ocho. Estas clases representan las emociones de clases discretas para cada modelo entrenado.

La arquitectura de red utilizada para clasificar las emociones está basada en el uso la red propuesta por Szegedy, et al. [47] en el año 2015, llamada GoogLeNet. La red propone el uso de filtros convolucionales de diferentes dimensiones para ser empleados en una sola capa llamada *inception*, Figura 3.9, para extraer características de la imagen de entrada. El tamaño de los filtros convolucionales de la Figura 3.9 se muestran en la Tabla 3.1 como “Inception (3a)”. A continuación, se describe cada uno de los términos en las columnas de la Tabla 3.1.

- Nombre: nombre de la actual capa dentro de la arquitectura.
- Tamaño de filtro/Reducción: se refiere al tamaño de ventana utilizado para los filtros de convolución o max pooling y la reducción de dimensionalidad del arreglo.

- Tamaño de salida: es el resultado de las dimensiones de salida de la imagen (alto \times ancho \times número de mapas de características).
- $\#1 \times 1$, $\#3 \times 3$, $\#5 \times 5$: filtros convoluciones utilizados dentro de la capa inception.
- $\#3 \times 3$ **reduc.**, $\#5 \times 5$ **reduc.**: número de filtros de tamaño 1×1 empleados después de las convoluciones.
- pool proj: filtros 1×1 aplicados después del pooling dentro de la capa inception.

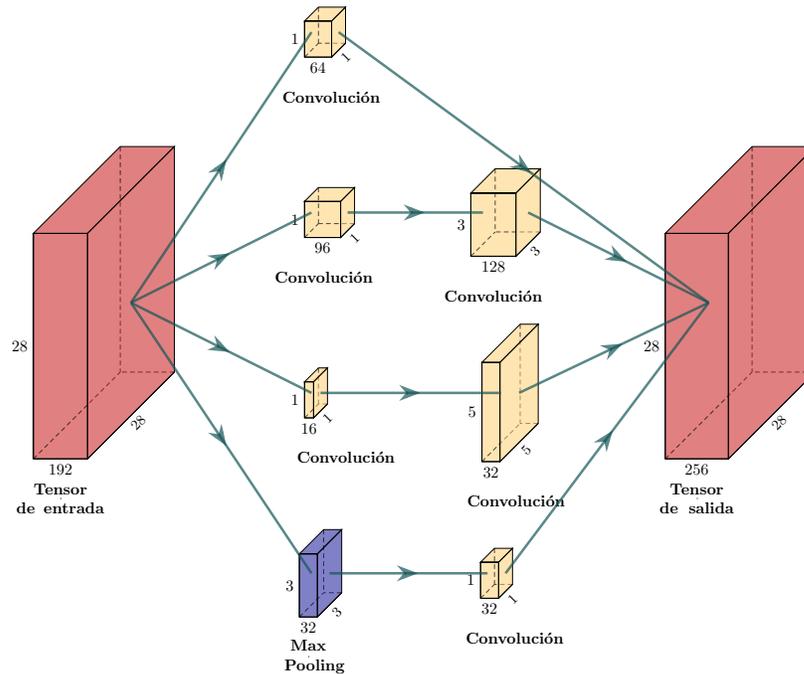


Figura 3.9: Arquitectura de la capa “Inception (3a)”.

Tomando como ejemplo la Figura 3.9, una imagen RGB de tamaño $[228 \times 228 \times 3]$, al entrar a la capa “Inception 3(a)” tendrá un tamaño de $[28 \times 28 \times 192]$. Este tensor, también conocido como arreglo de datos multidimensional, al pasar por la primera capa de filtros convolucionales extraen un número de mapas de características específico mediante tres filtros de tamaño $[1 \times 1 \times k]$, donde k varía según el filtro, y un

filtro “Max Pooling” de tamaño $[3 \times 3 \times 32]$. La segunda capa de filtros convolucionales posee filtros de mayor tamaño para la extracción de características, mientras que el filtro convolucional posterior al filtro “Max Pooling” es de tamaño $[1 \times 1 \times 32]$. El tensor de salida del filtro “Inception 3(a)” será de $[28 \times 28 \times 256]$, donde 256 mapas de características son el resultado de la concatenación de la última capa de filtros convolucionales ($64 + 128 + 32 + 32 = 256$).

Como se puede apreciar en la Tabla 3.1, al adaptar la red a nuestro modelo fue necesario realizar cambios con respecto a la red original. Trabajamos con un dropout (desconexión) al 70% y una capa de salida correspondiente al número de clases por modelo entrenado (nC). También, se modificaron las capas de loss value (valor de pérdida) y predicted labels (etiquetas obtenidas) por una capa fully connected (completamente conectada) y otra capa de clasificación. La capa completamente conectada contiene el número de clases de salida de cada modelo y la capa de clasificación el nombre de las etiquetas. Al realizar estas modificaciones, la arquitectura de red para clasificar emociones en el espacio VA y VAD, se ilustra en la Figura 3.10.

Tabla 3.1: Sintonización de la red GoogLeNet a nuestro enfoque.

| Arquitectura de red | | | | | | | | |
|--------------------------------------|------------------------------|--|----------------|------------------------------|----------------|------------------------------|----------------|-----------|
| Nombre (tipo) | Tamaño de filtro / Reducción | Tamaño de salida | # 1×1 | # 3×3 reduc. | # 3×3 | # 5×5 reduc. | # 5×5 | pool proj |
| Convolución 1 | $7 \times 7/2$ | $112 \times 112 \times 64$ | | | | | | |
| Max Pooling | $3 \times 3/2$ | $56 \times 56 \times 64$ | | | | | | |
| Convolución 2 | $3 \times 3/1$ | $56 \times 56 \times 192$ | | 64 | 192 | | | |
| Max Pooling | $3 \times 3/2$ | $28 \times 28 \times 192$ | | | | | | |
| Inception (3a) | | $28 \times 28 \times 256$ | 64 | 96 | 128 | 16 | 32 | 32 |
| Inception (3b) | | $28 \times 28 \times 480$ | 128 | 128 | 192 | 32 | 96 | 64 |
| Inception (4a) | | $14 \times 14 \times 512$ | 192 | 96 | 208 | 16 | 48 | 64 |
| Inception (4b) | | $14 \times 14 \times 512$ | 160 | 112 | 224 | 24 | 64 | 64 |
| Inception (4c) | | $14 \times 14 \times 512$ | 128 | 128 | 256 | 24 | 64 | 64 |
| Inception (4d) | | $14 \times 14 \times 528$ | 112 | 144 | 288 | 32 | 64 | 64 |
| Inception (4e) | | $14 \times 14 \times 832$ | 256 | 160 | 320 | 32 | 128 | 128 |
| Inception (5a) | | $7 \times 7 \times 832$ | 256 | 160 | 320 | 32 | 128 | 128 |
| Inception (5b) | | $7 \times 7 \times 1024$ | 384 | 192 | 384 | 48 | 128 | 128 |
| avg pool dropout (70%) softmax | $7 \times 7/1$ | $1 \times 1 \times 1024$ $1 \times 1 \times 1024$ $1 \times 1 \times nC$ | | | | | | |

A partir de la Figura 3.10, las imágenes de escalograma de tamaño $[224 \times 224 \times 3]$, son utilizadas como imágenes de entrada al modelo de red, ya sea para entrenamiento o validación. La imagen es procesada en cada capa de la red para extraer un número de características específico por capa de convolución o filtro inception. La salida de la capa “Inception 5(b)” devuelve un arreglo de dimensión $[7 \times 7 \times 1024]$ para, posteriormente, redimensionar el arreglo mediante una capa “avg. pool (agrupamiento promedio)”. El arreglo resultante tendrá un tamaño de $[1 \times 1 \times 1024]$ para ser analizado mediante una capa de dropout al 70%. Por último, el arreglo es procesado y redimensionado mediante una capa de softmax, también conocida como exponencial normalizada, resultando un arreglo de tamaño $[1 \times 1 \times nC]$, donde dependiendo el modelo, nC será igual a cuatro u ocho para el modelo VA y VAD, respectivamente. La capa final de salida determina a qué clase corresponde la imagen de entrada, proveniente del conjunto de imágenes de validación.

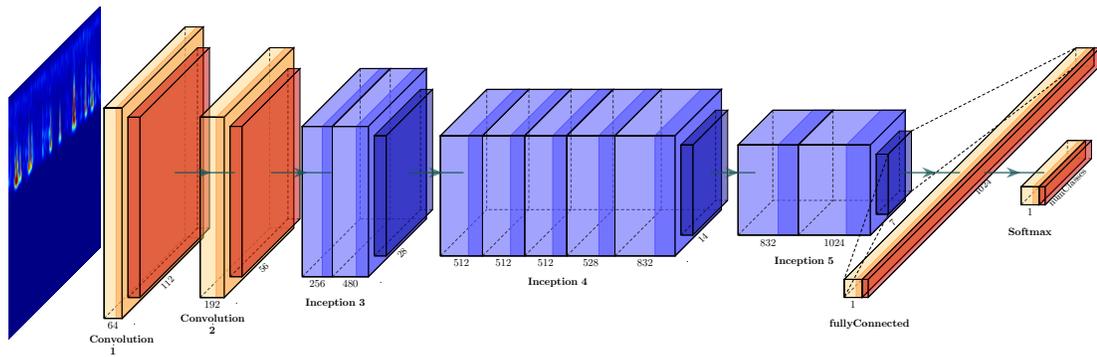


Figura 3.10: Arquitectura de red aplicada a la clasificación de emociones en el espacio VA y VAD.

A continuación, en el capítulo 4 se presentan los resultados obtenidos para la clasificación de emociones en los espacios VA y VAD, al utilizar las imágenes de escalograma como entrada en cada uno de los modelos. Además, se muestran gráficas y tablas de evaluación del desempeño de clasificación.

Capítulo 4

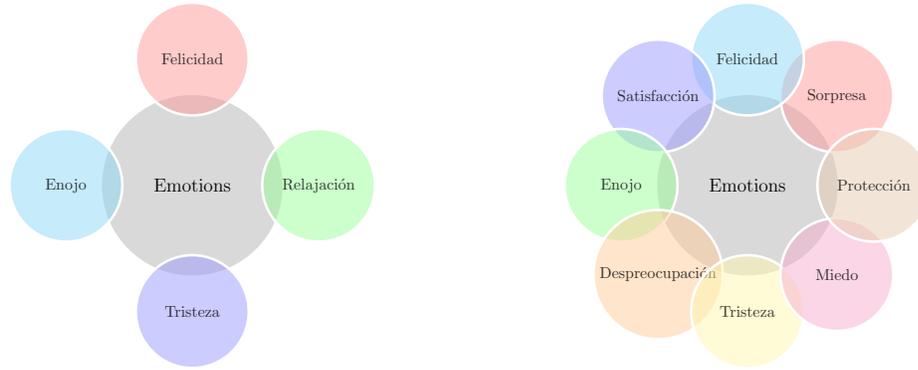
Resultados

En este capítulo se presentan los resultados de los modelos entrenados para este trabajo utilizando la metodología descrita en el Capítulo 3. Como se describió al inicio del Capítulo 3, para obtener los mejores modelos de clasificación, se emplearon algunas sintonizaciones de red. Por ejemplo, en el modelo utilizado mediante la red SqueezeNet se alcanzó una exactitud de validación del 81.60%. Para los modelos realizados a partir de la red GoogLeNet, en el modelo entrenado a 100 épocas y optimizador de sigmoide, la exactitud de validación fue de 78.8%, donde el punto de codo (elbow point) se originó aproximadamente a la época 65. Además, el tiempo de entrenamiento fue de 391 minutos. Del modelo de optimizador “adam” a 20 épocas y etiquetas extra resultó en un modelo de una exactitud de validación de 82.34%. Cada una de estas pruebas se realizaron para el espacio VA y, como se describe en este capítulo, presentaron resultados menores a los resultados finales en nuestra arquitectura de red.

En este trabajo, una de las etapas más importantes fue la realización de un algoritmo para la discretización de emociones primitivas en emociones. Estas emociones, que a su vez resultaron en las etiquetas de las clases de los modelos entrenados, son las que se pueden apreciar en la Figura 4.1a para el espacio VA y la Figura 4.1b en el espacio VAD.

Al obtener las clases discretas, el número de ejemplos en la clase “felicidad” era considerablemente mayor a las demás clases. Debido a ello, se optó por disminuir el desbalance entre clases con la finalidad de evitar un sesgo en la clasificación. Para

disminuir el desbalance, se utilizó la formula presentada en la ecuación (4.1), donde ρ representa un vector de valores aleatorios entre 0 y 1, ζ representa al porcentaje de desbalance ($\zeta = 0.1$ en ambos modelos), χ es la magnitud correspondiente al número de imágenes en la clase menor y κ el vector resultante de imágenes por clase, Γ es el número de imágenes total a utilizar.



(a) Emociones del espacio VA.

(b) Emociones del espacio VAD.

Figura 4.1: Modelos de clasificación propuestos en este trabajo.

$$\begin{aligned} \kappa(n) &= [\rho(n) * \zeta] * \chi + \chi \\ \Gamma &= \sum_{n=0}^{n-1} \kappa(n) \end{aligned} \tag{4.1}$$

De la ecuación (4.1) se obtienen dos valores para Γ : Γ_{VA} y Γ_{VAD} . Γ_{VA} es el número total de imágenes utilizadas en el espacio VA y Γ_{VAD} el número de imágenes utilizada en el espacio VAD. El objetivo principal de reducir el desvalance fue el obtener un conjunto de datos con una variación mínima entre ejemplos por cada clase. De esta manera, en la última columna correspondiente a la Tabla 4.2, Tabla 4.3 y Tabla 4.5 se puede apreciar el número total de ejemplos por clase para cada modelo entrenado. Cada entrenamiento realizado se llevó a cabo utilizando la arquitectura de red presentada en la Figura 3.10. Los conjuntos de entrenamiento y validación corresponden a un 80% y 20%, respectivamente, con respecto al total de ejemplos en Γ_{VA} y Γ_{VAD} . Las imágenes de entrada a la red son de tamaño $[228 \times 228 \times 3]$

en RGB. En la última capa convolucional de la red se reduce la imagen hasta un arreglo de $[7 \times 7 \times 1024]$ donde la tercera dimensión representa el número de mapas de características extraído por la red.

A continuación, en las siguientes secciones se muestran los resultados correspondientes a cada modelo entrenado, así como su matriz de confusión y métricas de evaluación.

4.1. Clasificación en el espacio VA

En este espacio solamente se utilizaron los vectores de valor y excitación (Valence-Arousal, VA) para caracterizar y discretizar las clases que representan a las emociones “tristeza”, “enojo”, “relajación”, y “felicidad”. Como se puede apreciar en la Figura 4.2, donde la exactitud de entrenamiento (línea azul) es $\approx 100\%$ y la exactitud de validación (línea negra) es de 85.16%. La línea inferior roja representa la pérdida de entrenamiento en el modelo. En la Tabla 4.1 se muestran los resultados finales del entrenamiento de este modelo.

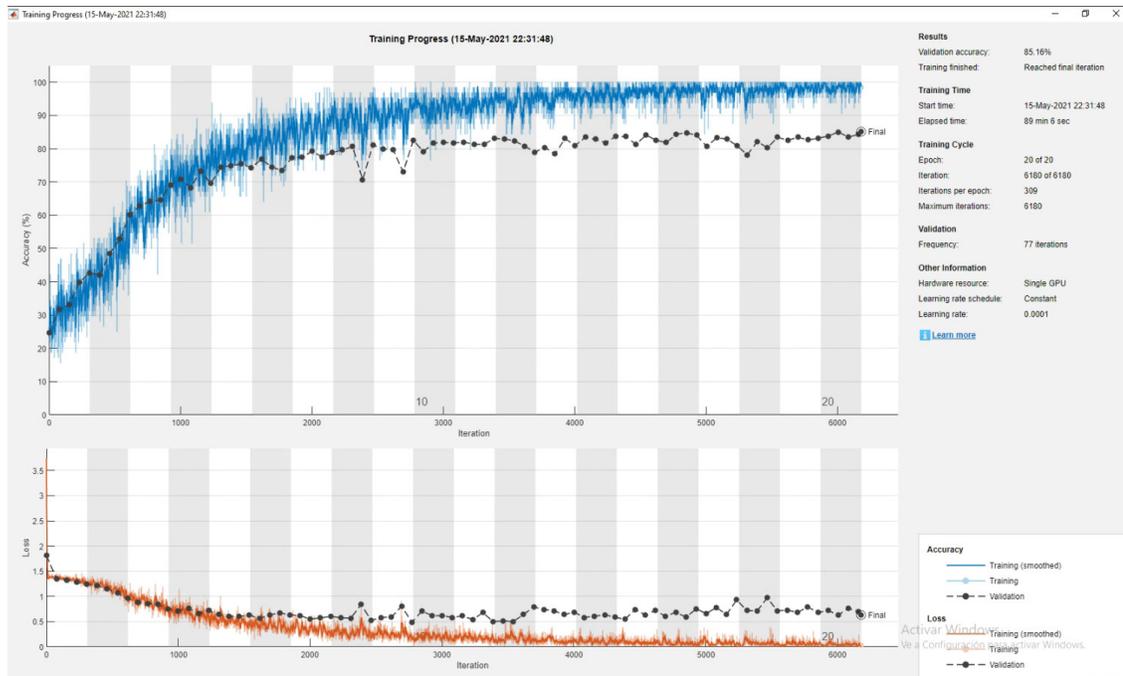


Figura 4.2: Desempeño de entrenamiento y validación en el espacio VA.

Tabla 4.1: Desempeño de entrenamiento y validación en el espacio VA para las últimas cinco iteraciones.

| Últimos cinco resultados de entrenamiento y validación | | | | | | | |
|--|-----------|----------|----------------------------|-------------------------|--------------------------|-----------------------|----------------------|
| Época | Iteración | Tiempo | Exactitud de entrenamiento | Exactitud de validación | Pérdida de entrenamiento | Pérdida de validación | Razón de aprendizaje |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 20 | 6083 | 01:27:28 | 100 % | 83.46 % | 0.0330 | 0.7619 | 1×10^{-4} |
| 20 | 6100 | 01:27:39 | 100 % | - | 0.0113 | - | 1×10^{-4} |
| 20 | 6150 | 01:28:14 | 100 % | - | 0.0055 | - | 1×10^{-4} |
| 20 | 6160 | 01:28:32 | 96.88 % | 84.35 % | 0.0358 | 0.7012 | 1×10^{-4} |
| 20 | 6180 | 01:28:57 | 98.44 % | 85.16 % | 0.0297 | 0.6368 | 1×10^{-4} |

Como se observa en la Tabla 4.1, la exactitud de entrenamiento en la última iteración es de 98.44 % y de validación de 85.16 %. Además, el valor correspondiente a la razón de aprendizaje no cambió conforme el avance de las épocas. También, se aprecia que la exactitud del modelo se incrementó en las últimas cinco iteraciones de entrenamiento. Donde dos de los entrenamientos reportados en la Tabla 4.1 no presentan exactitud y pérdida de validación, esto debido a que las etapas de validación están definidas a cada 72 iteraciones. La frecuencia de validación del modelo (V_{freq}) mantiene la relación de $V_{freq} = \frac{nI_v}{BS}$, donde nI_v es el número total de imágenes para validación y BS el factor de grupo mínimo (mini batch size).

En la Tabla 4.2 se muestra la matriz de confusión correspondiente a la validación del modelo en el espacio VA. También, se pueden apreciar el número de ejemplos utilizados por clase. Además, se reporta el porcentaje de desempeño por clase correspondiente a las métricas de evaluación “precision”, “recall”, y “F1-Score” definidas en las ecuaciones (4.2), (4.3) y (4.4), respectivamente.

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

$$F1 = 2 \times \left(\frac{Precision \cdot Recall}{Precision + Recall} \right) \quad (4.4)$$

Estas métricas se basan en el impacto de los falsos positivos (FP), falsos negativos (FN) y verdaderos positivos (TP) de las etiquetas predichas por los modelos de clasificación. En el estudio del reconocimiento emocional es importante mantener estas métricas debido a que, como se mencionó previamente, el conjunto de datos sobre el cual pueda estar basada una metodología puede presentar un alto desbalance de ejemplos por clase. Así, con el uso de la métrica F1 es posible determinar qué clase contiene más ejemplos y si ésta influye altamente en la exactitud del modelo.

Tabla 4.2: Matriz de confusión y métricas de evaluación del modelo en el espacio VA.

| Matriz de confusión | | | | | |
|---------------------------------------|----------|-------|------------|-----------|-------|
| | Tristeza | Enojo | Relajación | Felicidad | Total |
| Tristeza | 986 | 69 | 58 | 64 | 1177 |
| Enojo | 31 | 1149 | 39 | 57 | 1276 |
| Relajación | 52 | 123 | 1038 | 80 | 1293 |
| Felicidad | 47 | 54 | 62 | 1049 | 1212 |
| Métricas de evaluación del modelo (%) | | | | | |
| Precision | 83.77 | 90.05 | 80.28 | 86.55 | |
| Recall | 88.35 | 82.37 | 86.72 | 83.92 | |
| F1-Score | 86.00 | 86.04 | 83.37 | 85.22 | |

Como se puede apreciar en la Tabla 4.2, la clase con el menor número de ejemplos es “tristeza” con un total de 1177 imágenes utilizadas para la validación. La clase con el mayor número de imágenes fue “relajación”, con 1293 ejemplos. La diferencia entre ambas es de 116, $\approx 10\%$ de la clase “tristeza”.

4.2. Clasificación en el espacio VAD

En este espacio se utilizan los vectores de valor, excitación y dominio, (Valence-Arousal-Dominance, VAD). La discretización y caracterización de los vectores VAD generan ocho clases discretas las cuales son aquellas que se presentan en la Figura 4.1b. En este modelo se obtiene una exactitud de validación de 82.52%, y una exactitud de entrenamiento $\approx 97\%$ con respecto a la Figura 4.3.

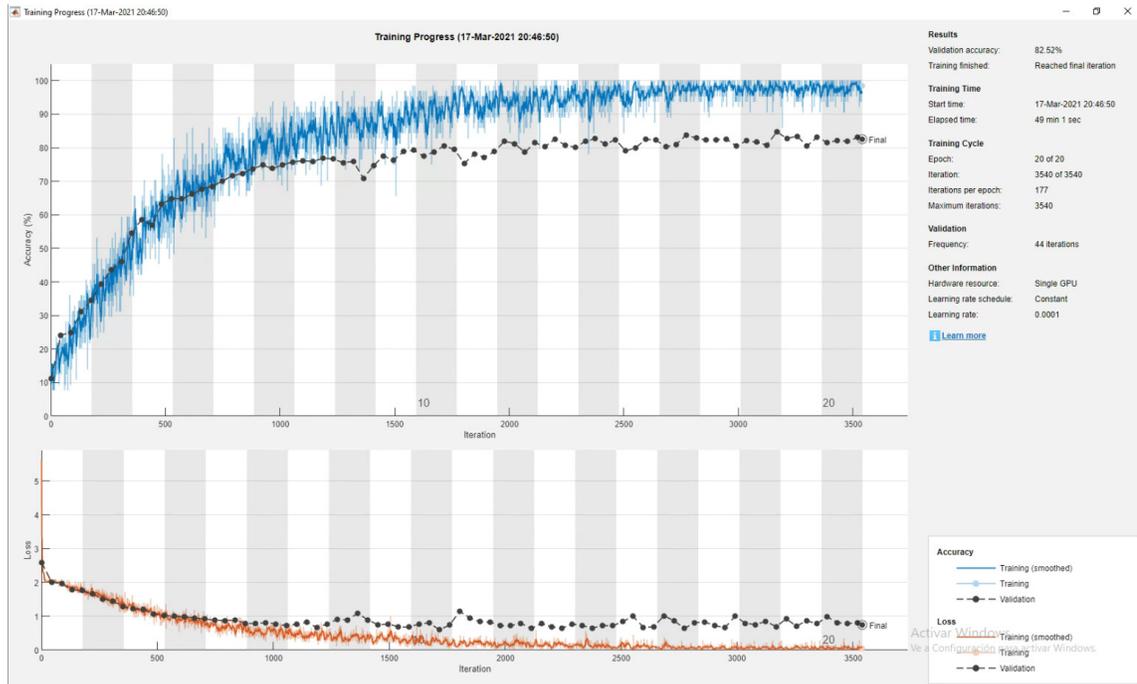


Figura 4.3: Desempeño de entrenamiento y validación en el espacio VAD.

En la Tabla 4.3 se muestran los resultados de la matriz de confusión, el total de ejemplos por clase y las diferentes métricas de evaluación del modelo, con respecto al conjunto de validación de imágenes. Como se aprecia en la tabla de la matriz de confusión, la clase con el menor número de ejemplos fue “enojo”, con un total de 323 imágenes. La clase con el mayor número de imágenes fue “satisfacción”. La diferencia entre ambas es de 47 imágenes. El número de imágenes por clase resulta ser $\approx 30\%$ con respecto al número de imágenes por clase del espacio VA. El número de ejemplos resulta entonces bajo comparado con el espacio VA, para solucionar esto, se llevó a cabo un aumento en el conjunto de datos de imágenes mediante el uso de la adición de ruido normalmente distribuido aplicado a las señales EEG originales.

Al ampliar el número de imágenes por clase, añadiendo el ruido en el espacio VAD, el entrenamiento del modelo aumenta ligeramente tanto en la exactitud de entrenamiento como de validación. En la Figura 4.4 se muestra el desempeño de entrenamiento y validación del modelo y en la Tabla 4.4 los últimos cinco resultados del modelo conforme a sus iteraciones.

Tabla 4.3: Matriz de confusión y métricas de evaluación del modelo en el espacio VAD.

| Matriz de confusión | | | | | | | | | |
|---------------------------------------|----------|---------|-------|-------|---------|-----------|----------|-----------|-------|
| | Tristeza | Despre. | Miedo | Enojo | Protec. | Satisfac. | Sorpresa | Felicidad | Total |
| Tristeza | 277 | 13 | 9 | 9 | 3 | 8 | 8 | 24 | 351 |
| Despre. | 2 | 298 | 9 | 3 | 1 | 12 | 5 | 22 | 352 |
| Miedo | 7 | 12 | 290 | 3 | 6 | 18 | 9 | 19 | 364 |
| Enojo | 3 | 8 | 1 | 283 | 4 | 8 | 1 | 15 | 323 |
| Protec. | 3 | 12 | 21 | 6 | 245 | 13 | 8 | 31 | 339 |
| Satisfac. | 1 | 3 | 5 | 0 | 6 | 344 | 2 | 9 | 370 |
| Sorpresa | 4 | 2 | 15 | 8 | 1 | 6 | 293 | 15 | 344 |
| Felicidad | 5 | 3 | 6 | 4 | 7 | 5 | 9 | 316 | 355 |
| Métricas de evaluación del modelo (%) | | | | | | | | | |
| Pres. | 78.91 | 84.65 | 79.67 | 82.99 | 66.93 | 92.97 | 85.17 | 89.01 | |
| Recall | 84.19 | 84.90 | 79.45 | 89.55 | 89.74 | 83.91 | 85.17 | 70.06 | |
| F1 | 81.47 | 84.77 | 79.56 | 86.14 | 76.68 | 87.75 | 85.17 | 78.41 | |

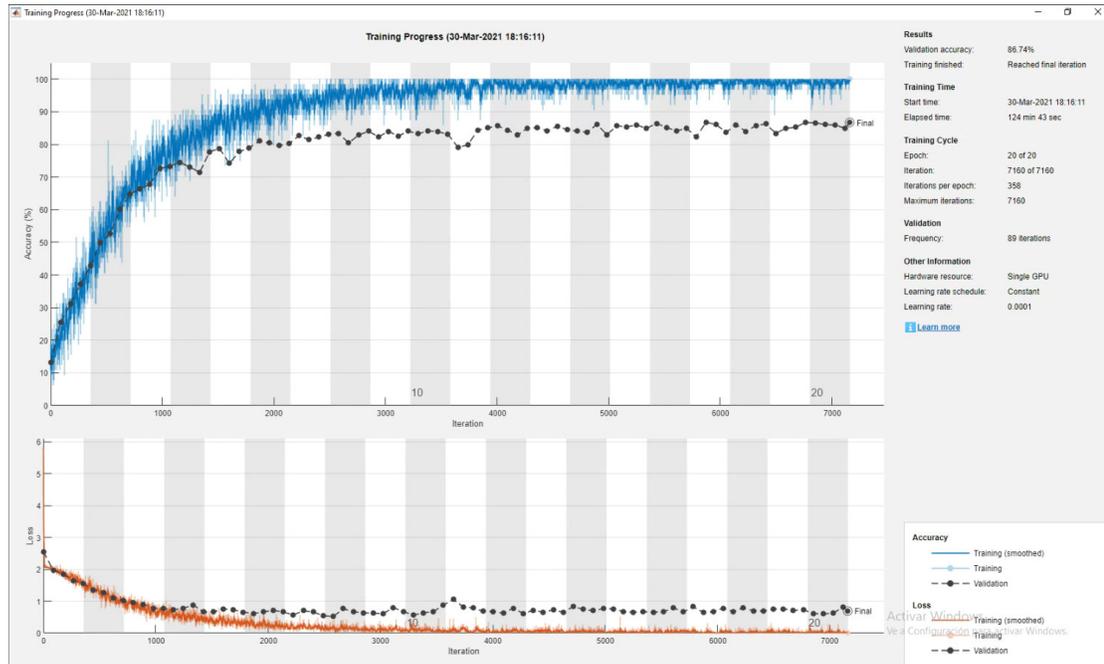


Figura 4.4: Desempeño de entrenamiento y validación en el espacio VAD añadiendo ruido a las muestras para aumentar el número de ejemplos.

A continuación, en la Tabla 4.5 se muestran los resultados de la matriz de confu-

sión, el total de imágenes por clase y las métricas de evaluación en el espacio VAD con ruido añadido a las señales EEG. Donde la clase con el menor número de imágenes fue “tristeza”, con un total de 664 ejemplos, y la más alta “miedo”, con 758 ejemplos. Como se aprecia en la Tabla 4.5, al añadir ruido a las señales mejora en un 4.2 % la exactitud del modelo.

Tabla 4.4: Desempeño de entrenamiento y validación en el espacio VAD para las últimas cinco iteraciones añadiendo ruido a las muestras para aumentar el número de ejemplos.

| Últimos cinco resultados de entrenamiento y validación | | | | | | | |
|--|-----------|----------|----------------------------|-------------------------|--------------------------|-----------------------|----------------------|
| Época | Iteración | Tiempo | Exactitud de entrenamiento | Exactitud de validación | Pérdida de entrenamiento | Pérdida de validación | Razón de aprendizaje |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 20 | 7050 | 02:02:29 | 98.44 % | - | 0.0224 | - | 1×10^{-4} |
| 20 | 7100 | 02:03:12 | 100 % | - | 0.0073 | - | 1×10^{-4} |
| 20 | 7120 | 02:03:47 | 100 % | 84.93 % | 0.0010 | 0.8226 | 1×10^{-4} |
| 20 | 7150 | 02:04:10 | 100 % | - | 0.0013 | - | 1×10^{-4} |
| 20 | 7160 | 02:04:33 | 100 % | 86.74 % | 0.0103 | 0.7001 | 1×10^{-4} |

Tabla 4.5: Matriz de confusión y métricas de evaluación en el espacio VAD añadiendo ruido a las muestras para aumentar el número de ejemplos.

| Matriz de confusión | | | | | | | | | |
|---------------------------------------|----------|---------|-------|-------|---------|-----------|----------|-----------|-------|
| | Tristeza | Despre. | Miedo | Enojo | Protec. | Satisfac. | Sorpresa | Felicidad | Total |
| Tristeza | 629 | 5 | 9 | 0 | 4 | 1 | 2 | 14 | 664 |
| Despre. | 16 | 629 | 16 | 4 | 9 | 3 | 5 | 8 | 690 |
| Miedo | 37 | 3 | 678 | 5 | 19 | 6 | 8 | 2 | 758 |
| Enojo | 55 | 1 | 18 | 577 | 12 | 2 | 17 | 23 | 705 |
| Protec. | 42 | 1 | 11 | 1 | 596 | 2 | 8 | 17 | 678 |
| Satisfac. | 12 | 1 | 25 | 5 | 1 | 673 | 5 | 12 | 734 |
| Sorpresa | 37 | 0 | 3 | 15 | 34 | 0 | 578 | 14 | 681 |
| Felicidad | 31 | 11 | 11 | 6 | 28 | 4 | 12 | 619 | 722 |
| Métricas de evaluación del modelo (%) | | | | | | | | | |
| Pres. | 89.86 | 91.16 | 89.44 | 80.81 | 85.63 | 89.49 | 81.64 | 85.73 | |
| Recall | 73.22 | 92.77 | 84.96 | 92.76 | 79.68 | 97.40 | 91.02 | 87.31 | |
| F1 | 80.70 | 91.96 | 87.15 | 86.38 | 82.55 | 93.28 | 86.07 | 86.51 | |

A continuación, en la sección 4.3 se lleva a cabo un análisis de los resultados obtenidos en las pruebas de clasificación en cada uno de los modelos. Además, se presenta un análisis con respecto al número de imágenes utilizadas en cada modelo, así como la diferencia de imágenes entre cada una de las clases.

4.3. Discusión

La arquitectura de red de clasificación es un modelo basado en la arquitectura GoogLeNet, en la cual se realizaron modificaciones para ser adaptable a dos modelos de clasificación para identificar cuatro y ocho emociones. Los conjuntos de datos en el espacio VA y VAD mostraron resultados altamente competitivos con los que se presentaron en el Capítulo 2, donde con base en nuestro enfoque, la exactitud de nuestros modelos fue de 85.16 % en el espacio VA y 82.52 % y 86.74 % en el espacio VAD, con respecto a un modelo sin aumento de ejemplos mediante ruido añadido y un modelo con aumento de ejemplos.

Al analizar las métricas de evaluación del modelo en el espacio VA, se aprecia que la mejor precisión se obtiene en la clase “enojo” (90.05 %), siendo “relajación” la más débil con un porcentaje de 80.28 %. Al ser “relajación” la clase con la menor precisión se deduce que también fue la clase con el mayor número de falsos positivos (FP). En recall, la clase menor fue “enojo”, con 82.37 %, y “tristeza” la más alta con 88.35 %. Resultando “tristeza” como la clase con la mejor predicción del modelo. Por último, la clase con el menor coeficiente en F1 fue “relajación” con 83.37 % y “enojo” la más alta con 86.04 %. La medida F1 modela el equilibrio entre precisión y recall, por lo que “enojo” es la clase con el menor número de falsos positivos y la mejor predicha en el modelo. Además, en la Tabla 4.2 se muestran el número total de imágenes por clase donde la diferencia máxima (MdI), con respecto a la clase con el menor y mayor número de ejemplos, es de 116, mientras que la diferencia mínima (mdI) fue de 35 imágenes. La diferencia promedio de imágenes por clase (adI) es de 83.

En el modelo VAD la diferencia entre las métricas por clases tiende a ser más amplia. Por ejemplo, la menor precisión de este modelo resulta en la clase “protección” con un porcentaje de 66.93 %, siendo “satisfacción” la mayor con un valor de 92.97 %.

En recall “felicidad” resulta ser la clase con el menor porcentaje y “protección” la mayor con valores de 70.06 % y 89.74 %, respectivamente. El menor valor de F1 entre clases resultó en “protección” con 76.68 % y la mayor en “satisfacción” con 87.75 %. En este modelo, la diferencia entre clases a partir de las métricas de evaluación tiende a ser muy amplia en comparación con el modelo en el espacio VA. Por ejemplo, la diferencia máxima en precisión es de 22.08 %, en recall de 19.68 % y en F1 de 11.07 %. El factor $MdI = 47$, correspondiente a la clase “enojo” y “satisfacción”, $mdI = 16$ (“enojo” y “protección”) y $adI = 30.57$ imágenes. Debido a lo anterior, se amplifica el conjunto de datos sobre el modelo en el espacio VAD, para disminuir la diferencia entre clases. En este modelo, la clase con la mayor precisión fue “despreocupación” con 91.16 % y la menor “enojo” con 80.81 %, por lo cual “despreocupación” contiene el menor número de falsos positivos y “enojo” el mayor de estos. En recall “tristeza” fue la clase con el menor número de ejemplos predichos con 73.22 % y la más alta “despreocupación” con 92.77 %. La clase “tristeza” y “satisfacción” resultaron con valores de 80.70 % y 93.28 %, siendo “tristeza” la más baja y “satisfacción” la más alta en la métrica de F1. Las clases “satisfacción” y “despreocupación” fueron las más altas en la métrica F1 de manera que estas modelan el mejor equilibrio entre precisión y recall.

En la Tabla 4.4 se muestra que de los últimos cinco entrenamientos, donde solo dos de estos fueron validados la exactitud aumentó ligeramente con respecto a sus iteraciones previas. Además, en la Tabla 4.5, se puede apreciar que la diferencia máxima en precisión es de 10.35 %, con respecto a “enojo” y “despreocupación”, en recall de 24.18 % (“tristeza” y “satisfacción”) y en F1 un 12.58 % correspondiente a “tristeza” y “satisfacción”. En este modelo la precisión aumentó y el recall y F1 disminuyó ligeramente. El factor de diferencia de imágenes $MdI = 94$, $mdI = 14$, y $adI = 45.71$ imágenes.

A partir de la exactitud obtenida en cada uno de los modelos, nuestro trabajo compite con los presentados en el estado del arte. En la Tabla 4.6 se muestra una tabla comparativa con respecto diferentes trabajos de reconocimiento emocional. Como se observa en la Tabla 4.6, nuestro trabajo obtiene mejores resultados al clasificar cuatro emociones excepto por la metodología propuesta por Khare, et al., en [45], donde mediante las imágenes de PSWVD, y utilizando arquitecturas conocidas basadas

en las CNN, obtuvieron mejores resultados. Si bien presenta mejores resultados en exactitud y F1, nuestro enfoque desarrolló un modelo con el doble de emociones clasificadas al propuesto por Khare.

Tabla 4.6: Comparativa con otras propuestas de clasificación de emociones.

| Emociones: clases discretas | | |
|--------------------------------|---|---|
| Autor y año | Exactitud (%) | F1 (%) |
| Wang, et al. [40] (2011) | 66.51 | - |
| Alazari, et al. [41] (2018) | 75.1 | - |
| Lin, et al. [42] (2010) | 82.29 | - |
| Murugappan, et al. [43] (2010) | 83.26 | - |
| Yang, et al. [44] (2020) | 84.21 | 80.37 |
| Khare, et al. [45] (2020) | 93.01 | 93.02 |
| Nuestra propuesta | 85.16 (Espacio VA) 86.74 (Espacio VAD) | 85.16 (Espacio VA) 85.16 (Espacio VAD) |

Algunas ventajas de nuestro trabajo es la transformación de las señales EEG a un escalograma. También, la adaptación del algoritmo a otro espacio de representación espectral. Al cambiar la transformación de las señales a otra metodología de análisis espectral, la red es altamente adaptable a los cambios en las imágenes de entrada. Además, nuestro algoritmo permite discretizar las emociones primitivas en emociones específicas para clasificar cuatro y ocho emociones. Por último, este modelo no presenta pérdida significativa de exactitud al aumentar el número de clases.

Algunas desventajas en la metodología propuesta es la adaptación a otro conjunto de datos. Cambiar de base de datos requeriría un análisis a una nueva distribución de datos. Además, al utilizar una red basada en las CNN, es necesario el uso de

imágenes como entrada. Este tipo de clasificador no puede ser adaptado como un modelo de clasificación con mapas de características a la entrada, pero puede ser utilizado como extractor de características para clasificarlas mediante otro algoritmo de clasificación.

Capítulo 5

Conclusiones

Los dos modelos presentados en este trabajo, VA y VAD, mostraron una alta exactitud en la clasificación de emociones. Al comparar nuestro trabajo con los que se proponen en el estado del arte, nuestro enfoque logra una exactitud mayor a la mayoría de los trabajos con los que es comparado. Además, nuestros resultados presentan otras métricas de evaluación de desempeño del modelo, mientras que la mayoría de los autores solo presentan la exactitud.

El procesamiento de las señales electroencefalografías utilizando la transformada wavelet morse para obtener imágenes de representación espectral (tiempo-frecuencia o tiempo-escala), ha sido hasta el momento poco explorado en la literatura. Si bien se puede interpretar como una metodología novedosa, se han presentado buenos resultados en modelos basados en las CNN. Esto propone entonces un enfoque donde las señales no estacionarias sean procesadas mediante algoritmos basados en la WT con buenos resultados.

También, es notable la aportación del modelo utilizando la red GoogLeNet donde la exactitud, y demás métricas de evaluación, presentaron una disminución mínima en los valores de evaluación al aumentar el número de clases.

Por lo anterior, se propone entonces como trabajo a futuro aumentar el número de clases (emociones) mediante la discretización de las emociones primitivas. Además, se propone también una metodología de aumento de datos mediante el uso de técnicas de procesamiento digital de señales como la interpolación y/o extrapolación de los datos de la señal. Esto para aumentar el número de ejemplos en cada clase ya que,

como se describió en la sección 4.3, el aumento de clases supone un decremento considerable en el número de ejemplos por clase. También, analizar la posibilidad de utilizar las CNN únicamente como método de extracción de características para ingresar estos valores a otras redes de tipo LSTM o BiLSTM.

Por último, una perspectiva más de este trabajo es la generación de una base de datos propia utilizando una diadema de electrodos basada en el sistema 10-20. Al crear nuestra base de datos, se propone realizar el experimento en el mismo número de sujetos masculinos y femeninos de diferentes edades.

Bibliografía

- [1] Y. Liu and O. Sourina, “EEG-based subject-dependent emotion recognition algorithm using fractal dimension,” *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2014-January, no. January, pp. 3166–3171, 2014.
- [2] J. M. Lilly and S. C. Olhede, “On the analytic wavelet transform,” *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 4135–4156, 2010.
- [3] S. Koelstra, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “DEAP: A database for emotion analysis; Using physiological signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.
- [4] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, “Recognising spontaneous facial micro-expressions,” in *2011 International Conference on Computer Vision*, pp. 1449–1456, 2011.
- [5] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.
- [6] N. N. Khatri, Z. H. Shah, and S. A. Patel, “Facial Expression Recognition: A Survey,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 149–152, 2014.
- [7] A. Mehrabian, “Communication without words,” *Communication theory*, vol. 6, pp. 193–200, 2008.

-
- [8] C. P. McLean and E. B. Foa, “Emotions and emotion regulation in posttraumatic stress disorder,” *Current Opinion in Psychology*, vol. 14, pp. 72–77, 2017.
- [9] R. Pandey and A. Choubey, “Emotion and health: An overview,” *SIS Journal of Projective Psychology and Mental Health*, vol. 17, pp. 135–152, 2010.
- [10] Achraf Ammar, et al., “Effects of home confinement on mental health and lifestyle behaviours during the COVID-19 outbreak: Insight from the ECLB-COVID19 multicenter study,” *Biology of Sport*, vol. 38, no. 1, pp. 9–21, 2021.
- [11] N. Andalibi and J. Buss, “The human in emotion recognition on social media: Attitudes, outcomes, risks,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Apr. 2020.
- [12] R. C. Roberts, “What an Emotion is: A Sketch,” *The Philosophical Review*, vol. 97, no. 2, pp. 183–209, 1988.
- [13] L. A. Feldman, “Valence Focus and Arousal Focus: Individual Differences in the Structure of Affective Experience,” *Journal of Personality and Social Psychology*, vol. 69, no. 1, pp. 153–166, 1995.
- [14] H. R. Trompetter, E. de Kleine, and E. T. Bohlmeijer, “Why Does Positive Mental Health Buffer Against Psychopathology? An Exploratory Study on Self-Compassion as a Resilience Mechanism and Adaptive Emotion Regulation Strategy,” *Cognitive Therapy and Research*, vol. 41, no. 3, pp. 459–468, 2017.
- [15] J. F. Cohn, “Foundations of human computing: Facial expression and emotion,” in *Proceedings of the 8th International Conference on Multimodal Interfaces, ICMI '06*, (New York, NY, USA), p. 233–238, Association for Computing Machinery, 2006.
- [16] N. Mehendale, “Facial emotion recognition using convolutional neural networks (FERC),” *SN Applied Sciences*, vol. 2, no. 3, 2020.
- [17] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017.

-
- [18] R. K. Madupu, C. Kothapalli, V. Yarra, S. Harika, and C. Z. Basha, “Automatic human emotion recognition system using facial expressions with convolution neural network,” in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1179–1183, 2020.
- [19] E. Frant, F. Frant, I. Ispas, V. Dragomir, M. Dascalu, E. Zoltan, and I. C. Stoica, “Voice Based Emotion Recognition with Convolutional Neural Networks for Companion Robots,” *Romanian Journal of Information Science and Technology*, vol. 20, no. 3, pp. 222–240, 2017.
- [20] Mustaqeem and S. Kwon, “MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach,” *Expert Systems with Applications*, vol. 167, p. 114177, 2021.
- [21] M. Z. Uddin and E. G. Nilsson, “Emotion recognition using speech and neural structured learning to facilitate edge intelligence,” *Engineering Applications of Artificial Intelligence*, vol. 94, no. June, p. 103775, 2020.
- [22] K. Wang, G. Su, L. Liu, and S. Wang, “Wavelet packet analysis for speaker-independent emotion recognition,” *Neurocomputing*, vol. 398, pp. 257–264, 2020.
- [23] S. R. Devasahayam, *Signals and systems in biomedical engineering: signal processing and physiological systems modeling*. Springer Science & Business Media, 2012.
- [24] H. Kantz, J. Kurths, and G. Mayer-Kress, *Nonlinear analysis of physiological data*. Springer Science & Business Media, 2012.
- [25] P. Batres-Mendoza, C. R. Montoro-Sanjose, E. I. Guerra-Hernandez, D. L. Almanza-Ojeda, H. Rostro-Gonzalez, R. J. Romero-Troncoso, and M. A. Ibarra-Manzano, “Quaternion-based signal analysis for motor imagery classification from electroencephalographic signals,” *Sensors*, vol. 16, no. 3, 2016.
- [26] M. Agarwal and R. Sivakumar, “Blink: A fully automated unsupervised algorithm for eye-blink detection in eeg signals,” in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1113–1121, 2019.

-
- [27] N. Kulkarni, “Use of complexity based features in diagnosis of mild alzheimer disease using eeg signals,” *International Journal of Information Technology*, vol. 10, pp. 59–64, 2018.
- [28] G. K. Verma and U. S. Tiwary, “Affect representation and recognition in 3D continuous valence–arousal–dominance space,” *Multimedia Tools and Applications*, vol. 76, no. 2, 2017.
- [29] S. S. Tomkins and R. McCarter, “What and Where Are the Primary Affects? Some Evidence for a Theory.,” *Perceptual and motor skills*, vol. 18, pp. 119–158, 1964.
- [30] S. Gannouni, A. Aledaily, K. Belwafi, and H. Aboalsamh, “Adaptive Emotion Detection Using the Valence–Arousal–Dominance Model and EEG Brain Rhythmic Activity Changes in Relevant Brain Lobes,” *IEEE Access*, vol. 8, pp. 67444–67455, 2020.
- [31] Y. Cimtay and E. Ekmekcioglu, “Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition,” *Sensors (Switzerland)*, vol. 20, no. 7, pp. 1–20, 2020.
- [32] Y. Cimtay and E. Ekmekcioglu, “LUMED Dataset,” 2020.
- [33] W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, “EmotionMeter: A Multimodal Framework for Recognizing Human Emotions,” *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2019.
- [34] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, “Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain,” *BioMed Research International*, vol. 2017, p. 9, 2017.
- [35] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, “Emotionmeter: A multimodal framework for recognizing human emotions,” *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

- [36] A. Garg, A. Kapoor, A. K. Bedi, and R. K. Sunkaria, “Merged LSTM Model for emotion classification using EEG signals,” *2019 International Conference on Data Science and Engineering, ICDSE 2019*, pp. 139–143, 2019.
- [37] D. Garg and G. K. Verma, “Emotion Recognition in Valence-Arousal Space from Multi-channel EEG data and Wavelet based Deep Learning Framework,” *Procedia Computer Science*, vol. 171, no. 2019, pp. 857–867, 2020.
- [38] S. Taran and V. Bajaj, “Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method,” *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 157–165, 2019.
- [39] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, “A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition,” *Egyptian Informatics Journal*, vol. 22, no. 2, 2020.
- [40] X. W. Wang, D. Nie, and B. L. Lu, “EEG-based emotion recognition using frequency domain features and support vector machines,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7062 LNCS, no. PART 1, pp. 734–743, 2011.
- [41] R. Alazrai, R. Homoud, H. Alwanni, and M. I. Daoud, “EEG-based emotion recognition using quadratic time-frequency distribution,” *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–32, 2018.
- [42] Y.-p. Lin, C.-h. Wang, T.-p. Jung, S. Member, T.-l. Wu, S.-k. Jeng, J.-r. Duann, and J.-h. Chen, “EEG-Based Emotion Recognition in Music Listening,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.
- [43] M. Murugappan, N. Ramachandran, and Y. Sazali, “Classification of human emotion from EEG using discrete wavelet transform,” *Journal of Biomedical Science and Engineering*, vol. 03, no. 04, pp. 390–396, 2010.

-
- [44] J. Yang, X. Huang, H. Wu, and X. Yang, “EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network,” *Procedia Computer Science*, vol. 174, no. 2019, pp. 491–504, 2020.
- [45] S. K. Khare and V. Bajaj, “Time-Frequency Representation and Convolutional Neural Network-Based Emotion Recognition,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–9, 2020.
- [46] Ö. Türk and M. S. Özerdem, “Epilepsy detection by using scalogram based convolutional neural network from eeg signals,” *Brain Sciences*, vol. 9, no. 5, 2019.
- [47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 1–9, 2015.
- [48] Z. Liang, S. Oba, and S. Ishii, “An unsupervised EEG decoding system for human emotion,” *Neural Networks*, vol. 116, pp. 257–268, 2019.
- [49] J. Huang, X. Xu, and T. Zhang, “Emotion classification using deep neural networks and emotional patches,” *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, vol. 2017-January, pp. 958–962, 2017.
- [50] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size,” *arXiv:1602.07360*, 2016.