

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA EN SISTEMAS

**DESARROLLO DE UN MODELO DE RECONOCIMIENTO EN
TIEMPO REAL DE GESTOS DE LA MANO USANDO SEÑALES
ELECTROMIOGRÁFICAS, COVARIANZAS Y REDES
NEURONALES ARTIFICIALES**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

JOSÉ DAVID GONZÁLEZ GAVILÁNEZ

jose.gonzalez@epn.edu.ec

DIRECTOR: MARCO E. BENALCÁZAR, PHD

marco.benalcazar@epn.edu.ec

Quito, diciembre del 2019

AVAL

Certifico que el presente trabajo fue desarrollado por José David González Gavilánez, bajo mi supervisión.

Marco Enrique Benalcázar, PhD
DIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo José David González Gavilánez, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

José David González Gavilánez

DEDICATORIA

El presente trabajo dedico a mis padres Diego González y Janeth Gavilánez, quienes fueron mi motivación y guía, gracias a ellos pude concluir mis estudios. Sin olvidar a mis hermanos y familia que siempre estuvieron a mi lado. Mis compañero y amigos que hicieron la universidad mas llevadera y ahora forman parte de mi vida.

AGRADECIMIENTO

A mi familia, por su apoyo incondicional y consejos que permanecerán siempre en mí.

A la Escuela Politécnica Nacional, por brindarme los conocimientos necesarios para lograr este objetivo.

A Marco Benalcázar por la oportunidad de ser parte del proyecto de investigación PIJ-16-13 y por su enseñanza tanto en el ámbito académico como personal.

A mis amigos y compañeros, que siempre estuvieron prestos a compartir sus conocimientos y saben que pueden contar conmigo.

Sin olvidarme de la persona que estuvo presente durante casi toda mi vida académica, por su desinteresado apoyo y amor brindado.

JOSÉ DAVID GONZÁLEZ GAVILÁNEZ

ÍNDICE DE CONTENIDO

AVAL	I
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
ÍNDICE DE CONTENIDO.....	V
ÍNDICE DE FIGURAS	VII
ÍNDICE DE ECUACIONES.....	VIII
ÍNDICE DE TABLAS	VIII
RESUMEN	IX
ABSTRACT	X
1. INTRODUCCIÓN	¡Error! Marcador no definido.
1.1 Pregunta de investigación.....	3
1.2 Objetivo General.....	3
1.3 Objetivos Específicos	3
1.4 Hipótesis.....	3
2. MARCO TEÓRICO	4
2.1. Estado del arte	4
2.2. Sensores para el reconocimiento de gestos	5
2.2.1. Dispositivos basados en visión:	5
2.2.2. Dispositivos no basados en visión:	5
2.3. Unidad motora	8
2.4. Señales electromiográficas.....	9
2.4.1. Características de la señal Semg.....	9
3. METODOLOGÍA	11
3.1. Materiales.....	11
3.2. Conjunto De Datos	12
3.3. Notación.....	13
3.4. Métodos	13
3.4.1 Fase de Diseño	14
3.4.2. Fase de Entrenamiento	22
3.4.3. Fase de evaluación	23

4. <i>RESULTADOS Y DISCUSIÓN</i>	24
4.1. Metodología de Evaluación	24
4.2. Resultados	25
4.2.1. <i>Resultados de clasificación</i>	25
4.2.2. <i>Resultados de reconocimiento</i>	27
4.2.3. <i>Resultados de tiempo</i>	28
4.3. Discusión.....	28
5. <i>CONCLUSIONES Y RECOMENDACIONES</i>	30
6. <i>REFERENCIAS BIBLIOGRÁFICAS</i>	31
7. <i>ANEXOS</i>	34
7.2. Gráficos de las funciones de transferencias.	34
7.3. Resultados de reconocimiento solo utilizando covarianzas utilizando los usuarios de entrenamiento y validación.	¡Error! Marcador no definido.

ÍNDICE DE FIGURAS

Figura 1. Gestos reconocidos por el sistema	2
Figura 2 CyberGlove IIR	6
Figura 3 Wii-moteR.....	6
Figura 4 Funcionamiento sistema WiSee.....	6
Figura 5 Gestos interpretables por una pantalla multi-touch	7
Figura 6 Funcionamiento sistema ultrasónico	7
Figura 7 . Myo Armband	8
Figura 8 Estructura de una unidad motora	8
Figura 9 Señal EMG	9
Figura 10 Myo Armband	11
Figura 11 Sensores del Myo Armband.....	12
Figura 12 Estructura de la metodología utilizada	13
Figura 13 Representación de las ventanas sucesivas.....	14
Figura 14 i) Señal EMG del gesto Fist obtenido por un sensor del Myo Armband ii) Señal EMG filtrada iii) Señal EMG rectificadas	15
Figura 15 Proceso de reconocimiento de una ventana	21
Figura 16 Arquitectura de la red neuronal	22
Figura 17 Solapamiento entre el vector de predicciones y la señal generada	24
Figura 18 Resultados de reconocimiento por usuario (evaluación)	27
Figura 19 Resultados de reconocimiento por usuario (entrenamiento)	27
Figura 20 Resultados obtenidos por usuario función RELU	34
Figura 21 Resultados obtenidos por usuario función ELU.....	34
Figura 22 Resultados obtenidos por usuario función TANH	35
Figura 23 Resultados obtenidos por usuario función log-sigmoid	35
Figura 24 Resultados obtenidos por usuario función softplus	35
Figura 25 Resultados obtenidos por usuario solo utilizando covarianzas, la línea naranja representa la precisión promedio igual a 88,30%.....	1
Figura 26 Tiempo de respuesta promedio por usuario (ms).....	37

ÍNDICE DE ECUACIONES

Ecuación 1 Covarianza	16
Ecuación 2 Valor Absoluto Medio (MAV)	17
Ecuación 3 Longitud de forma de Onda (WL).	18
Ecuación 4 Moda	19
Ecuación 5 Fórmula del factor de solapamiento.....	24
Ecuación 6 Fórmula de la exactitud en la matriz de confusión.	25
Ecuación 7 Fórmula de la sensibilidad en la matriz de confusión.....	25
Ecuación 8 Fórmula de la precisión en la matriz de confusión.	26

ÍNDICE DE TABLAS

Tabla 1. Resultados obtenidos en la fase de entrenamiento por cada usuario.....	23
Tabla 2 Matriz de confusión usuarios de evaluación	26
Tabla 3 Matriz de confusión usuarios de entrenamiento	26
Tabla 4 Tabla comparativa con resultados de trabajos simialres	29

RESUMEN

En la presente investigación, se presenta un modelo de reconocimiento de gestos de la mano humana basado en señales electromiográficas (EMG), covarianzas y redes neuronales artificiales. El modelo propuesto reconoce 5 gestos de la mano: Fist, Fingers Spread, Wave In, Wave Out, Double Tap, y el no gesto que es todo lo que no se reconoce dentro de los 5 gestos.

El modelo se compone de 2 fases: a) fase de diseño y entrenamiento; b) fase de evaluación. La fase de diseño está compuesta por 5 etapas: 1. adquisición de datos, 2. pre-procesamiento, 3. extracción de características, 4. clasificación, y 5. post-procesamiento. Para la adquisición de señales EMG, se utilizó el dispositivo Myo Armband con un enfoque de ventana deslizante; en el pre-procesamiento se empleó rectificación y filtrado de la señal EMG; para la extracción de características, una de las principales diferencias con otras investigaciones es que utilizó principalmente covarianzas; en la clasificación se empleó una red neuronal feed-forward que consta de 3 capas: entrada, oculta, y salida; y finalmente en el post-procesamiento se aplicó: reducción, y eliminación de valores atípicos.

El modelo se entrenó por usuario y requirió de 25 repeticiones por cada uno de los gestos a reconocer. El modelo presentado logra una exactitud de $92,45\% \pm 11,00$, con un tiempo promedio de respuesta de $40,58$ milisegundos $\pm 1,62$. El modelo propuesto funciona en tiempo real sobre un ambiente de bajos recursos computacionales los resultados presentados son superiores a los definidos en la hipótesis.

PALABRAS CLAVE: Reconocimiento de gestos de la mano, señales EMG, tiempo real, Myo Armband, ReLU, Redes Neuronales Artificiales.

ABSTRACT

In this research, a hand gesture recognition model based on electromyographic signals, covariances and artificial neural networks is proposed. The proposed model recognizes five hand gestures: Fist, Fingers Spread, Wave In, Wave Out, Double Tap, and the non-gesture that is all that is not recognized within the 5 gestures. The model consists of 2 phases: design and training phase, and evaluation phase. The design phase is composed of 5 stages: data acquisition, pre-processing, feature extraction, classification, and post-processing. For the acquisition of EMG signals, a sliding window approach is applied, and the Myo Armband is used to measure the EMG of the forearm; in the pre-processing rectification and filtering of the EMG signal is used; for the extraction of characteristics, mainly covariances are used; In the classification a feed-forward neural network is used, consisting of 3 layers: input, hidden, and output; and finally, reduction, and elimination of outliers are applied in the post-processing. The model is trained per user and requires 25 repetitions for each of the gestures to recognize. The model reaches an accuracy of $92.45 \pm 11.00\%$, with an average time response of 40.58 ± 1.62 milliseconds. The proposed model works in real time on an environment with low computational resources, the results presented are higher than the rate established in the hypothesis.

KEYWORDS: Hand gesture recognition EMG signals, real time, Myo Armband, ReLU, Artificial Neural Networks

1. INTRODUCCIÓN

El reconocimiento de gestos de la mano humana tiene como objetivo interpretar y detectar los movimientos de la mano en el instante de tiempo en que estos son generados [1]. El reconocimiento de gestos puede ser aplicado para el desarrollo de sistemas de reconocimiento de lenguaje de señas [2], interacción humano-computadora (HCI) [3], desarrollo de prótesis activas [4], entre otras.

Tomando en cuenta las posibles aplicaciones del reconocimiento de gestos fue requisito indispensable desarrollar un modelo que respondiera con gran precisión y en tiempo real. Para que un modelo sea considerado en tiempo real su tiempo de respuesta debe ser inferior a 300ms [5]. Adicionalmente la portabilidad es una característica imprescindible para aplicaciones como prótesis activas. Esta característica implica que el modelo debía ser capaz de funcionar con bajos recursos computacionales (computador de escritorio con un procesador de 4GHz de velocidad y 16 GB de RAM). En este sentido, la siguiente investigación se presenta en el marco del proyecto de investigación junior PIJ-16-13 “Clasificación de señales electromiográficas del brazo humano usando técnicas de reconocimiento de patrones y machine learning”.

Para la adquisición de datos de entrada en un modelo de reconocimiento de gestos los métodos más comunes son: guantes, brazaletes, cámaras, sensores de ultrasonido, etc. En este proyecto, para obtener la información de entrada para el modelo de reconocimiento de gestos, se utilizó el brazalete Myo Armband que posee 8 sensores, el cual recolecta señales electromiográficas (EMG). Se optó por el uso de señales EMG debido a que son una manifestación eléctrica de la activación muscular. El Myo Armband se utiliza en el antebrazo adaptándose al tamaño de este y según el tiempo de uso o la aplicación no se vuelve incómodo, como en el caso de los guantes [6]; otra ventaja del brazalete es que no es susceptible a los cambios de luz, como en el caso de las cámaras, en las cuales se dificulta su uso, especialmente en escenarios en el que la distancia entre la mano y la cámara varía [7].

Existen diversos métodos para la extracción de características de la señal EMG, los cuales están en el dominio del tiempo (por ejemplo, el valor absoluto medio), frecuencia (por ejemplo, la frecuencia media) o tiempo-frecuencia (por ejemplo, wavelets). En este trabajo, se utilizó como método para la extracción de características la covarianza entre los canales del Myo Armband. La covarianza mide cómo dos variables aleatorias varían entre sí [8]. La

justificación de este método de extracción de características se presenta en la sección de métodos (2.3).

Para la clasificación de las señales EMG se utilizó una red neuronal feed-forward, debido a que este tipo de redes neuronales son aproximadores universales [9] capaces de implementar cualquier frontera de decisión. La arquitectura de la red propuesta constó de tres capas: una capa de entrada, una capa oculta y una capa de salida.

En esta investigación se desarrolló un modelo de reconocimiento de gestos de la mano basado en señales EMG que funciona en tiempo real y con una exactitud superior al sistema propietario de reconocimiento de gestos del Myo Armband. Los gestos que se usaron en esta investigación son los mismos que el sistema propietario del Myo Armband reconoce por defecto (Figura 1): el puño (Fist), la mano con los dedos estirados (Fingers Spread), mano hacia adentro (Wave In), la mano hacia afuera (Wave out), y el doble golpe (Double Tap) [10].

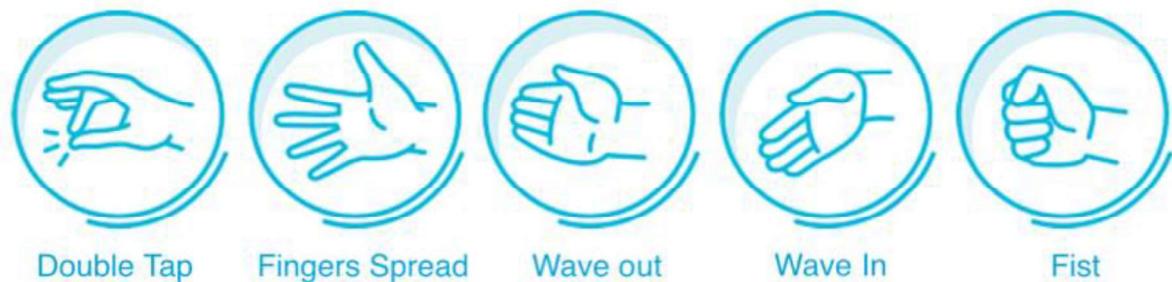


Figura 1. Gestos reconocidos por el sistema

1.1 Pregunta de investigación

¿Cuál es la arquitectura de un modelo de reconocimientos de gestos en tiempo real (300 ms) basado en señales electromiográficas y utilizando redes neuronales artificiales que logre obtener una exactitud de reconocimiento superior al 88% y pueda funcionar con bajos recursos computacionales (computador de escritorio con un procesador de 4GHz de velocidad y 16 GB de RAM)?

1.2 Objetivo General

Desarrollar un sistema de reconocimiento de 6 clases de gestos de la mano humana (Wave In, Wave Out, Fist, Fingers Spread, Double Tap, y el no gesto) que responda en tiempo real y con una exactitud de reconocimiento igual o mayor al 88% mediante el uso de señales EMG, covarianzas y redes neuronales artificiales.

1.3 Objetivos Específicos

- Implementar un sistema de extracción de características que incluya covarianzas y otras características en tiempo y frecuencia de la señal EMG que funcione en tiempo real.
- Reconocer las 6 clases (Wave In, Wave Out, Fist, Fingers Spread, Double Tap, y más el no gesto de la mano humana) de gestos de la mano humana en un tiempo de respuesta inferior a 300ms.
- Evaluar la exactitud de reconocimiento del modelo propuesto que debe ser superior al 88%.
- Evaluar el tiempo de respuesta del modelo propuesto que debe ser inferior a 300ms para que su funcionamiento se considere en tiempo real.

1.4 Hipótesis

Utilizando señales EMG, covarianzas, redes neuronales artificiales, y con ayuda del Myo Armband se podrá desarrollar un sistema de reconocimiento de gestos de la mano en tiempo real, logrando una exactitud de reconocimiento de las 6 clases (Wave In, Wave Out, Fist, Fingers Spread, Double Tap, y más el no gesto de la mano humana) igual o mayor al 88%.

2. MARCO TEÓRICO

2.1. Estado del arte

La señal EMG es la manifestación eléctrica de la activación neuromuscular asociada con un músculo esquelético que se contrae. Esta es una señal controlada por el sistema nervioso y depende de las propiedades anatómicas y fisiológicas de los músculos. A continuación, se presentan los trabajos más relevantes donde se utilizó machine learning, señales EMG y se reconocieron similar cantidad de gestos que la presente investigación.

Varios trabajos de investigación abordan el problema de reconocimiento de gestos en tiempo real. En el trabajo desarrollado por [11], se muestra un porcentaje de exactitud del 96,7% para un sistema de reconocimiento de gestos en tiempo real, con una técnica de votación por mayoría simple, imágenes instantáneas de las señales EMG y redes neuronales convolucionales. Sin embargo, el principal problema de este sistema es que requiere el uso de una estación de trabajo con altas capacidades de cómputo. Por otro lado, el trabajo expuesto en [12] se caracteriza por su propuesta de un sistema de reconocimiento con sensores de bajo costo basado en máquinas de vectores de soporte (SVM) que alcanzaban un 91,06% de exactitud de reconocimiento, empero solo reconoce 4 gestos; y, además, requiere un gran número de repeticiones por cada gesto a reconocer para entrenar el modelo de clasificación. De la misma forma, el sistema implementado por [13] propone un sistema de reconocimiento en tiempo de real de bajo consumo energético, pero que solo depende de 5 repeticiones por cada gesto a reconocer, obteniendo una exactitud de 88%; sin embargo, en este caso el reconocimiento es de solo 3 gestos.

Adicionalmente, cabe destacar los trabajos previos a este proyecto que se realizaron como parte del proyecto PIJ-16-13, en los cuales se desarrolló modelos que funcionan en tiempo real. Estos permitieron observar, entre algunas cosas, que en el trabajo realizado por [14] existe un porcentaje de exactitud promedio del 85.08%, con un enfoque basado en extracción de automática de características con redes neuronales artificiales (RNA). En esta misma línea, el trabajo desarrollado por [15] muestra un porcentaje promedio de exactitud del 90.7%, siendo este modelo aquel que utiliza redes neuronales artificiales para la clasificación de las señales EMG. En el trabajo realizado por [16] obtiene una exactitud de reconocimiento igual a 88.1% y de clasificación igual a 95.79% con un sistema por usuario y que necesita 10 repeticiones para entrenamiento y reconoce los mismos gestos

que el presente trabajo. Por último, el trabajo expuesto por [17] presenta un porcentaje promedio de exactitud de reconocimiento del 87,53%, utilizando SVM.

Tomando en cuenta los resultados de trabajos anteriores desarrollados en el mismo campo se optó por generar un modelo de reconocimiento que supere el 88% de exactitud de reconocimiento que corresponde a un promedio de la exactitud de reconocimiento de los trabajos realizados dentro del proyecto junior PIJ-16-13, utilizando bajos recursos computacionales y que sea capaz de reconocer los 5 gestos, mas el no gesto, en tiempo real.

2.2. Sensores para el reconocimiento de gestos

Existen varios tipos de sensores que son empleados para la adquisición de datos para sistemas de reconocimiento de gestos:

2.2.1. Dispositivos basados en visión:

Los sistemas de reconocimiento de gestos basados en visión dependen de una o varias cámaras para analizar e interpretar el movimiento de las secuencias de video capturadas. De manera similar a los dispositivos de contacto, los dispositivos basados en visión son diversos. Por ejemplo: cámaras infrarrojas, cámaras tradicionales, estero cámaras, cámaras PTZ.

2.2.2. Dispositivos no basados en visión:

Este tipo de dispositivos utilizan acelerómetros, pantallas multi-touch y sensores electromiográficos. Los cuales incluyen diferentes tipos de detectores para el reconocimiento de gestos como:

- a) **Mecánicos:** por ejemplo, Immersionc propone el "CyberGlove IIR" (Figura 2), que es un guante inalámbrico para el reconocimiento de gestos de las manos. Animazooc propone un traje de cuerpo llamado "IGS-190R" para capturar gestos corporales. Este dispositivo se usa generalmente en asociación con otros dispositivos: por ejemplo, el traje de cuerpo del IGS-190 está acoplado con dieciocho dispositivos inerciales (giroscopios) que permiten la detección de movimiento [18].



Figura 2. CyberGlove IIR

- b) **Inerciales:** miden y registran la fuerza específica de un cuerpo, la velocidad angular o la orientación del cuerpo, utilizando acelerómetros, giroscopios o magnetómetros. En acelerómetros tenemos el Wii-moteR (Figura 3). Una desventaja de estos sensores es que sufren de un error acumulado debido a que el sistema se basa a la aceleración con respecto al tiempo para calcular la velocidad y la posición, cualquier error de medición, por pequeño que sea, se acumula con el tiempo [19].



Figura 3. Wii-moteR

- c) **Electromagnéticos:** estos dispositivos miden la variación de un campo magnético artificial para la detección de movimiento (Figura 4). A diferencia de los dispositivos inerciales, los dispositivos electromagnéticos pueden generar algunos problemas de salud debido al electromagnetismo artificial producido [20].

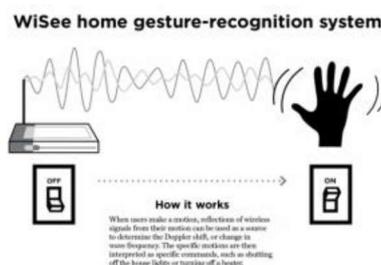


Figura 4. Funcionamiento sistema WiSee

- d) **Pantallas Táctiles:** las pantallas multitáctiles se vuelven cada vez más comunes en nuestra vida (por ejemplo, tablets). [21] propone reconocer las interacciones gestuales multitáctiles utilizando Hidden Markov Model (HMM). Este tipo de sensor son fáciles de implementar ya que las tablets y celulares son comunes y fáciles de conseguir, pero el tipo de gestos que puede reconocer es limitado (Figura 5).

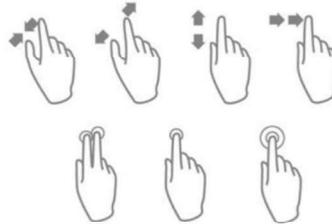


Figura 5. Gestos interpretables por una pantalla multi-touch

- e) **Ultrasónico:** La posición se calcula de acuerdo con el tiempo de propagación/reflexión y la velocidad del sonido. La orientación es entonces triangulada. Estos dispositivos (Figura 6) no son precisos y tienen una resolución baja, pero son útiles para entornos que carecen de luz y tienen interferencia electromagnética o ruido. [22].

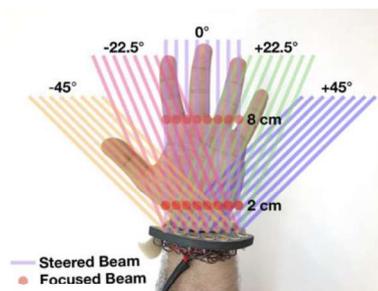


Figura 6. Funcionamiento sistema ultrasónico

- f) **Biomecánicos:** Dispositivos que utilizan técnicas biomecánicas como la electromiografía para medir los parámetros del gesto. Dentro de este tipo se encuentra el Myo Armband (Figura 7), el cual se utilizó para la realización de este proyecto. Este, no requiere métodos invasivos como agujas para obtener la señal y se adapta al tamaño del antebrazo del usuario.



Figura 7. Myo Armband

2.3. Unidad motora

Una unidad motora (UM), como se observa en la Figura 8, está compuesta por una motoneurona y las fibras musculares que esta inerva. La corriente eléctrica que es inducida en esta motoneurona determina el patrón de descarga (disparo) de la unidad motora y, por lo tanto, la actividad de la UM. El número de UMs por músculo en los humanos puede variar desde alrededor de 100 para un músculo pequeño de la mano hasta 1000 o más para los músculos de las extremidades grandes (por ejemplo, piernas) [23]. Las señales EMG se componen esencialmente de potenciales de acción de unidades motoras superpuestas (MUAP) de varias unidades motoras.

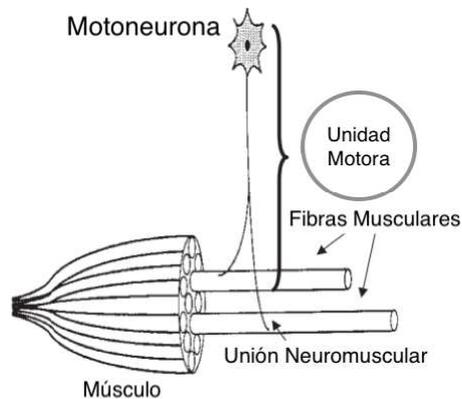


Figura 8. Estructura de una unidad motora

2.4. Señales electromiográficas

La señal electromiográfica (EMG) es una señal eléctrica producida durante la contracción muscular. La EMG depende de las propiedades anatómicas y fisiológicas de los músculos esqueléticos. Existen dos clases de señales EMG: superficiales e intramusculares [24].

Las señales de EMG superficiales e intramusculares se registran mediante electrodos no invasivos y electrodos invasivos, respectivamente. Según [25], la información proporcionada por cualquiera de las dos clases de señales es igualmente valiosa, concluyendo que, no existen diferencias significativas en su capacidad de clasificación. Sin embargo, las señales intramusculares requieren el uso de agujas lo que genera incomodidad al uso.

2.4.1. Características de la señal Ssemg

La información requerida de las señales EMG superficiales (sEMG) debido al ruido y complejidad de estas permanece oculta dentro de las señales de EMG sin procesar (Figura 9). Por esta razón, si estas señales sin procesar se utilizan como la entrada de un sistema de clasificación de sEMG su exactitud de clasificación es baja.

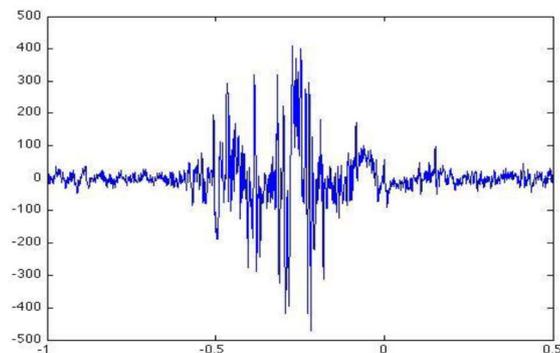


Figura 9. Señal EMG

Existen diferentes características que se pueden extraer de la señal EMG para ser usadas como la entrada de un sistema de clasificación. A continuación, los principales dominios de los cuales se pueden extraer características:

- Dominio del tiempo: Las características del dominio del tiempo son extraídas de la señal EMG en representación de tiempo [26]. Estas características, como el valor absoluto medio (MAV), el cuadrado medio raíz (RMS) y la longitud de onda (WL)

son las más populares en el reconocimiento de señales EMG debido a la alta velocidad de procesamiento en la clasificación.

- Dominio de la frecuencia: Las características del dominio de la frecuencia muestran la densidad de espectro de potencia de señal EMG . Existen algunas características en el dominio de la frecuencia como: la frecuencia media (MNF) y la relación de frecuencia (FR), entre otras. Estas características requirieren más tiempo de procesamiento en comparación con las características del dominio del tiempo. Estas son comúnmente utilizadas en la fatiga muscular y la estimación de la fuerza muscular [27]
- Dominio de tiempo-frecuencia: El análisis de wavelets se basa en utilizar el dominio tiempo-frecuencia, el cual cuantifica los cambios temporales del contenido de frecuencia de las señales no estacionarias sin perder información en el tiempo o la frecuencia [28].

3. METODOLOGÍA

3.1. Materiales

En este apartado, se describe las características del MYO ARMBAND y el conjunto de datos que se emplearon para las diferentes fases de diseño, entrenamiento y testeo del modelo propuesto en esta investigación.

Es un dispositivo portátil (Figura 10), desarrollado por la compañía Talmics Labs, que pesa 94 gramos y tiene un espesor de 1,14 centímetros. Está compuesto de ocho electrodos electromiográficos (EMG), una unidad de medición inercial de 9 ejes (IMU) (giroscopio de 3 ejes, acelerómetro de 3 ejes y magnetómetro de 3 ejes) y un módulo de transmisión bluetooth.

Los electrodos del Myo Armband censan las señales relacionadas con la actividad muscular del antebrazo del usuario y la IMU detecta los movimientos del antebrazo en el espacio tridimensional. Los datos adquiridos se envían, a través del módulo Bluetooth Low Energy (BLE) integrado en el brazalete, a otros dispositivos electrónicos (actuadores, microcontroladores, etc.), que realizan funciones específicas según los datos recibidos y el software instalado [29].

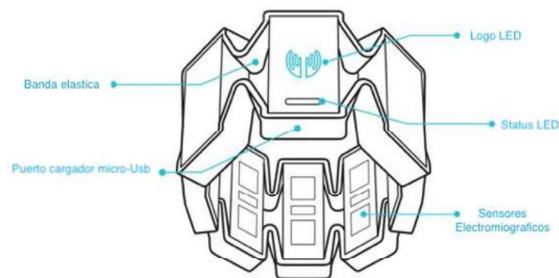


Figura 10. Myo Armband

El rango de potenciales proporcionados por el brazalete Myo está entre -128 y 128 en unidades de activación [29]. Estas unidades de activación son valores enteros de la amplificación de los potenciales medidos por los sensores EMG. El Myo es capaz de extraer datos sEMG a una frecuencia de muestreo de 200Hz.

Este sensor se ha diseñado para funcionar mejor en la parte más ancha del antebrazo, es decir, el antebrazo superior [30]. A diferencia de otros sensores EMG, el brazalete Myo no requiere que el usuario afeite el área alrededor del cual se usará el brazalete, ni tampoco

requiere de geles especiales. Esto permite procedimientos de configuración más fáciles en entornos experimentales o del mundo real.

3.2. Conjunto De Datos

El conjunto de datos usados para el desarrollo de la presente investigación constaron de 120 usuarios, de los cuales se grabó las señales EMG de los 5 gestos mencionados anteriormente. Se recolectó 50 muestras por cada gesto, cada una de las muestras obtenidas consta de 5 segundos de duración. Adicionalmente a los 5 gestos se grabó 10 muestras del gesto de relajación a cada uno de los usuarios. Obteniendo un total de 31.200 muestras recolectadas para el desarrollo del modelo.

Las muestras obtenidas se caracterizan por tener 8 canales que son los que posee el Myo Armband (Figura 11), con valores en el rango de $[-1,1]$, con una frecuencia de muestro de 200 Hz, con una duración de 5 segundos. Es decir, existen 1000 registros por 8 canales dando un total de 8000 valores.



Figura 11. Sensores del Myo Armband

Para este trabajo se dividió aleatoriamente la muestra inicial a 120 usuarios en dos grupos de 60 usuarios cada uno. La muestra del primer grupo sirvió para el desarrollo del modelo planteado y los 60 restantes sirvieron para la evaluación final del modelo.

Así mismo, el primer conjunto de usuarios que se utilizó para el desarrollo fue segmentado en dos grupos: el cincuenta por ciento fue utilizado para la selección del modelo a utilizar y el porcentaje restante para ajustarlo (validación preliminar).

3.3. Notación

En el presente trabajo, las matrices se denotan con letras mayúsculas en negrita (**A**), los vectores se representan con letras minúsculas en negritas (**c**), las constantes se denotan con letras mayúsculas (Y), y por ultimo los índices se denotan con letras minúsculas en cursivas (*i*).

3.4. Métodos

Para el desarrollo del modelo propuesto se utilizó una metodología dividida en dos fases: fase de diseño y entrenamiento, y fase de evaluación o testeo.

Estructura de la metodología utilizada:

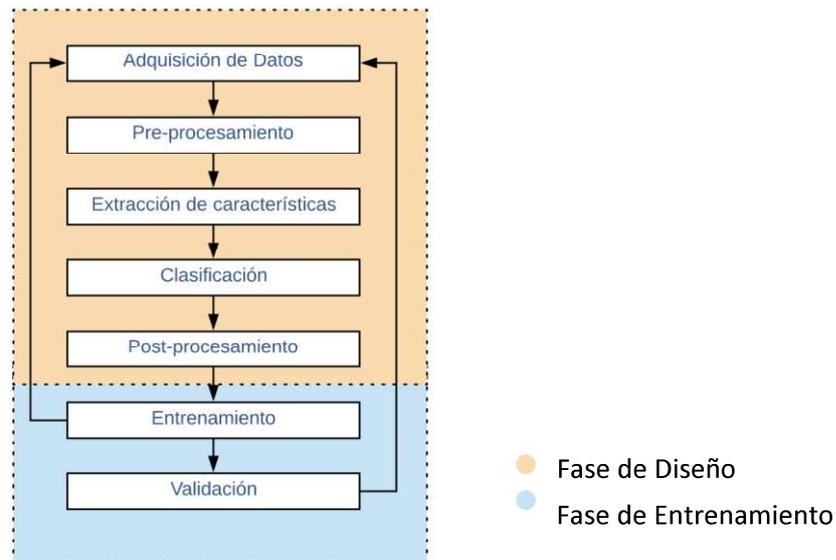


Figura 12. Estructura de la metodología utilizada

3.4.1 Fase de Diseño

3.4.1.1. Adquisición de Datos

El modelo propuesto toma como entrada las señales EMG del brazo obtenidas a través del Myo Armband. Para la adquisición de estos datos se utilizó el concepto de ventana deslizante con longitud M , como se describe en la Figura 13, donde $M=66$ y con un salto entre ventanas de 66 puntos, a su vez esta ventana es dividida en 6 sub-ventanas de longitud $N=11$, con un salto entre sub-ventanas de 11 puntos, cada una de estas se representa con una matriz A_i de tamaño $8 \times N$, donde 8 representa el número de canales que posee el Myo Armband. Cada uno de los elementos de la matriz A_i está en el intervalo $[-1,1]$. Si tomamos en cuenta que cada una de las señales obtenidas, por cada usuario, tiene aproximadamente 1000×8 valores, quiere decir que si dividimos el tamaño de la señal para el tamaño de las sub-ventanas se forman alrededor $1000/11 \approx 90$ sub-ventanas aproximadamente por cada señal.

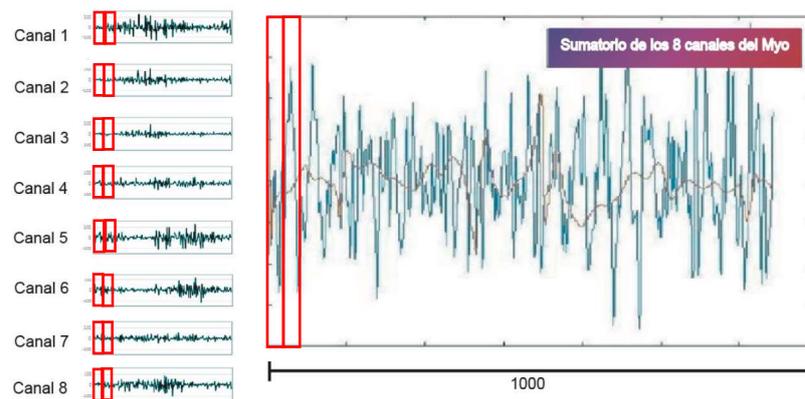


Figura 13. Representación de las ventanas sucesivas

3.4.1.2. Pre-procesamiento

Cada una de las sub-ventanas A_i obtenidas pasaron por un proceso de filtrado y rectificación. Este proceso como se observa en la Figura 14 tiene como objetivo obtener la envolvente de la señal que contiene el gesto a reconocer, para lograrlo se utilizó estas dos técnicas:

- 1) A la sub-ventana \mathbf{A}_i se le aplicó la función valor absoluto para eliminar los valores negativos, una vez culminado este proceso se obtuvo la matriz $\mathbf{B}_i = \text{abs}(\mathbf{A}_i)$, la cual posee valores en el rango de $[0, 1]$.
- 2) En el proceso de filtrado se aplicó un filtro pasa bajos de butterworth de segundo orden a la matriz \mathbf{B}_i y con una frecuencia de corte de 1 Hz, obteniendo la matriz \mathbf{C}_i .

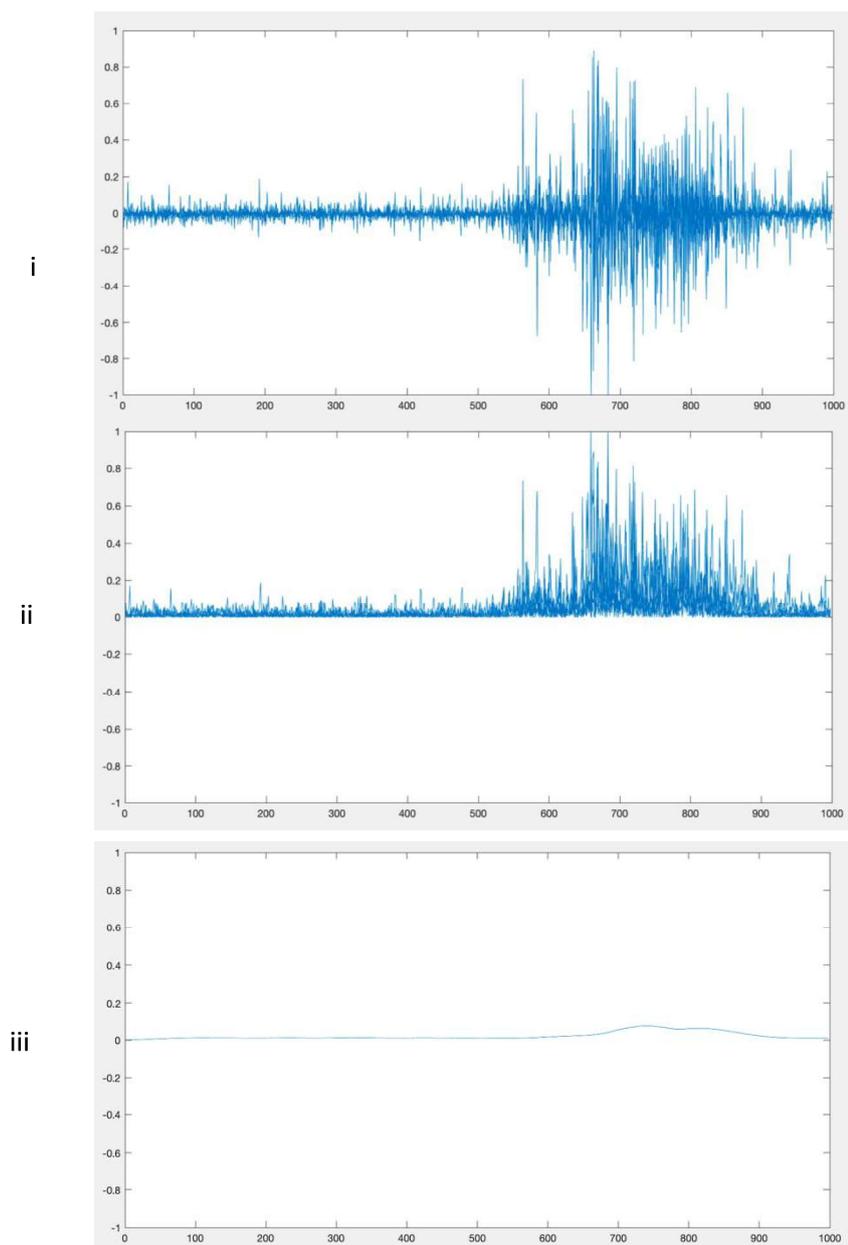


Figura 14. i) Señal EMG del gesto Fist obtenido por un sensor del Myo Armband ii) Señal EMG filtrada iii) Señal EMG rectificadas

3.4.1.3. Extracción de características

El objetivo de esta fase fue caracterizar cada una de las sub-ventanas, para ello se utilizó principalmente la covarianza, la cual permitió determinar como varían los canales del Myo en un determinado gesto y caracterizar matemáticamente esa variación. Como complemento a la covarianza se emplearon funciones que permitieron caracterizar una señal de entrada por su potencia, frecuencia, ancho de banda, forma de onda y para ello se uso las siguientes funciones matemáticas: la potencia de banda, frecuencia media, ancho de banda ocupado, valor absoluto medio, longitud de forma de onda.

El resultado de esta etapa fue un vector r de longitud igual a 68 por cada una de las sub-ventanas C_i . El vector está compuesto por:

COVARIANZA (28 valores)	POTENCIA DE BANDA (8 valores)	FRECUENCIA MEDIA (8 valores)	ANCHO DE BANDA OCUPADO (8 valores)	VALOR ABSOLUTO MEDIO (8 valores)	LONGITUD DE FORMA DE ONDA (8 valores)
-----------------------------------	---	--	--	--	---

Es importante recalcar que solo para la covarianza se empleó la señal resultante de la etapa anterior. Para las demás funciones se usó la señal en bruto, sin ningún pre-procesamiento.

COVARIANZA: mide como dos variables aleatorias varían entre sí a lo largo del tiempo [8]. A diferencia de la varianza, esta puede ser negativa o positiva. Cuando el valor de la covarianza es negativo significa que las variables varían en direcciones opuestas; cuando es positivo implica que las dos variables aleatorias tienden a variar en la misma dirección; y cuando el resultado es cero significa que las variables no tienen ninguna relación entre sí. El objetivo de utilizar esta función es medir cómo varían las señales obtenidas por los 8 sensores del Myo Armband en un determinado gesto, y así discriminar según el grado de variación a qué gesto corresponde una muestra dada.

La covarianza se define por la siguiente ecuación:

$$\sigma(c_{i,j}, c_{j,i}) = E[(c_{i,j} - E[c_{i,j}])(c_{j,i} - E[c_{j,i}])]$$

Ecuación 1. Covarianza

Donde $E[c_{i,j}]$ representa el valor esperado de la variable $c_{i,j}$, también conocido como la esperanza matemática.

Tomando como entrada la sub-ventana \mathbf{C}_i de $8 \times N$, donde N es igual a 11, el resultado aplicando la función de covarianza es una matriz de covarianzas de tamaño 8×8 , la cual es una matriz triangular. Por esta razón, se tomó en cuenta la parte superior colocándola en un vector \mathbf{c} de longitud igual a 28, que posteriormente se concatena con las salidas de las demás funciones de extracción de características para formar el vector de características \mathbf{r} .

POTENCIA DE BANDA: La potencia de banda devuelve la potencia promedio de la señal de entrada. En este caso como \mathbf{A}_i es una matriz, la potencia de banda calcula la potencia promedio en cada fila de forma independiente, retornando así un valor de potencia por cada uno de los canales de la señal, formando un vector \mathbf{q} de longitud igual a 8.

FRECUENCIA MEDIA: Esta función estima la frecuencia media normalizada del espectro de potencia de una señal en el dominio del tiempo. Esta función retorna un vector \mathbf{w} de longitud igual a 8.

ANCHO DE BANDA OCUPADO: Esta función retorna la diferencia de las frecuencias del 99% del ancho de banda de la señal de entrada. En este caso se obtendrá un vector de \mathbf{b} de longitud igual a 8.

VALOR ABSOLUTO MEDIO: se calcula tomando el promedio del valor absoluto de la señal, y es el valor promedio de toda la señal rectificadas [31]. Esta función se aplicó a cada uno de los canales por ello se obtuvo un vector \mathbf{a} de tamaño 8.

$$MAV = \frac{1}{N} \sum_{n=1}^N |d_n|$$

Ecuación 2. Valor Absoluto Medio (MAV)

Donde \mathbf{d} representa una columna de la matriz \mathbf{A}_i .

LONGITUD DE FORMA DE ONDA: Esta función es la longitud acumulativa de la forma de onda en un segmento de tiempo. La longitud de forma de onda está relacionada con la amplitud, frecuencia y tiempo de la forma de onda. Igual que la anterior función se aplicó para cada uno de los canales dando como resultado un vector \mathbf{o} de longitud 8.

$$WL = \sum_{n=1}^N |L_{n+1} - L_n|$$

Ecuación 3. Longitud de forma de Onda (WL).

Donde L_n representa la longitud de la forma de onda en un instante de tiempo n . La longitud de onda es la distancia entre los puntos correspondientes de dos ondas consecutivas.

Todos los vectores resultantes de cada una de las técnicas utilizadas para la extracción de características se concatenan para formar un vector \mathbf{r} de longitud igual a 68, el cual posteriormente se utilizará para la clasificación.

3.4.1.4. Clasificación

En la etapa de clasificación, se optó por utilizar redes neuronales artificiales (RNA) tipo feed-forward. Las redes neuronales artificiales están formadas por elementos de procesamiento simple llamadas "neuronas". El diseño de las RNAs fue motivado por la estructura del cerebro de los mamíferos [32].

Cada neurona en una red puede recibir señales de entrada, procesarlas y enviar una señal de salida. En una red, una neurona está conectada al menos con una neurona, y cada conexión se evalúa mediante un número real, llamado coeficiente de peso, que refleja el grado de importancia de la conexión dada en una red neuronal. En principio, las redes neuronales feed-forward son aproximadores universales, es decir, pueden realizar un mapeo arbitrario de un espacio vectorial a otro espacio vectorial [33].

La principal ventaja de las redes neuronales es que pueden extraer patrones ocultos en los datos. El proceso de "captura" de patrones se denomina "aprendizaje de la red neuronal" o "entrenamiento de la red neuronal". En el formalismo matemático, aprender significa ajustar los pesos de la red de tal manera que cumplan algunas condiciones (por ejemplo, error de entrenamiento y número máximo de épocas). Existe dos tipos principales de procesos de aprendizaje: supervisado y no supervisado. El aprendizaje supervisado (por ejemplo, la red neuronal de alimentación de múltiples capas) significa que la red neuronal conoce la salida deseada, el ajuste de los coeficientes de peso se realiza de tal manera que las salidas calculadas y deseadas estén lo más cerca posible. El entrenamiento no supervisado (por ejemplo, la red de Kohonen) [34] significa que solo se tiene la entrada y para el ajuste de pesos no se conoce la salida deseada.

Específicamente la red utilizada es una red neuronal de aprendizaje supervisado feed-forward, que consta de tres capas: capa de entrada, oculta y de salida. La clasificación toma como entrada el vector r resultante de la etapa anterior, y entrega como resultado las probabilidades que la señal de entrada corresponda a cada uno de los gestos a evaluar. En la fase de entrenamiento se describe la arquitectura de la red.

Una vez obtenidas las probabilidades que la señal de entrada corresponda a cada uno de los gestos, se realiza una predicción basada en un umbral en la cual se evaluó el gesto con mayor probabilidad y si el resultado fue mayor a un umbral $u=0.40$ se le asignó una etiqueta correspondiente al gesto, caso contrario ésta es reemplazada por la etiqueta del no gesto. La utilización de un umbral tiene como objetivo descartar los falsos positivos que se pueden generar de la etapa anterior.

3.4.1.5. Post-Procesamiento

Esta etapa consta de tres fases:

- **Reducción:** Una vez obtenida la etiqueta de las 6 sub-ventanas A_i se procedió a calcular la moda, que es un proceso estadístico, del cual se obtiene la etiqueta que le corresponde a la ventana principal. Cada etiqueta obtenida por cada ventana forma un vector f que describe a que gesto corresponde cada una de las ventanas que forman la señal.

$$Mo = Li + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2} \right) * C$$

Ecuación 4. Moda

Donde:

Li = Límite inferior de la clase modal.

$\Delta 1$ = Es el delta de frecuencia absoluta modal y la frecuencia absoluta pre-modal.

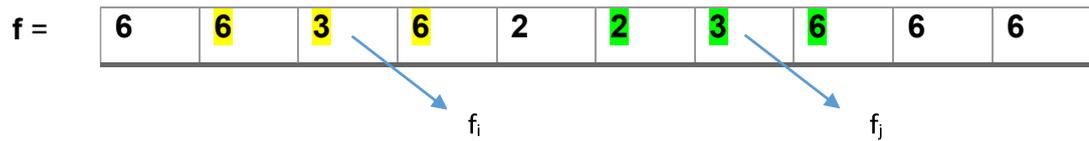
$\Delta 2$ = Es el delta de frecuencia absoluta modal y la frecuencia absoluta post-modal.

C = Amplitud del intervalo modal.

- **Eliminación de valores atípicos:** Este proceso está basado en valores futuros y pasados de la señal. Esta fase verifica dos posibles escenarios:

- (a) Si el elemento f_{i-1} es igual al elemento f_{i+1} , el elemento f_i debe ser igual a f_{i-1} .
- (b) Si los elementos f_{i-1} , f_{i+1} y f_i son diferentes entonces los elementos f_{i-1} , f_i deben ser iguales y tomar el valor de f_{i-1} .

Antes del proceso:



Después del proceso:



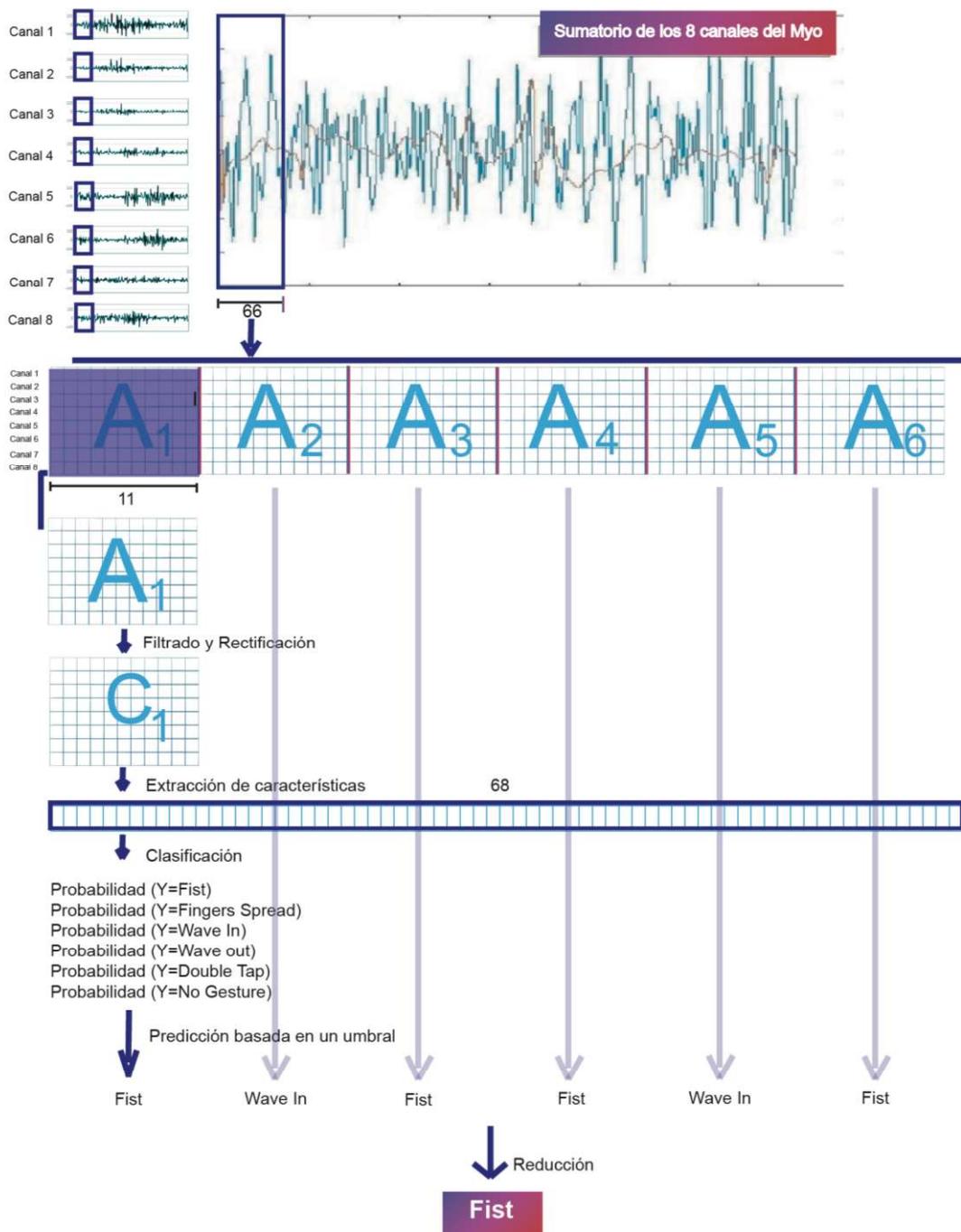


Figura 15. Proceso de reconocimiento de una ventana

3.4.2. Fase de Entrenamiento

El objetivo de esta fase es entrenar el modelo previamente descrito. Es importante señalar que este modelo se entrena por cada usuario. Esta fase tiene dos etapas: la etapa de entrenamiento y la de validación

3.4.2.1. Entrenamiento

Conjunto de datos de entrenamiento

De un total de 60 usuarios que colaboraron para el entrenamiento, los datos de 30 se destinaron para ajustar los parámetros del modelo y los otros 30 para la validación preliminar del modelo. Cada usuario tiene 50 muestras por cada uno de los gestos a evaluar (Fist, Fingers Spread, Wave In, Wave out, y Double Tap) dando un total de 250 muestras por usuario. El modelo propuesto se entrena por usuario, para lo cual, se necesita 25 repeticiones para el entrenamiento por cada uno de los gestos que se necesite reconocer.

Arquitectura de la red

La red neuronal que se utilizó consta de 3 capas: entrada, oculta y salida como se muestra en la figura 16. La capa de entrada presenta 68 nodos que corresponden al vector de características resultante del proceso de extracción de características. La capa oculta presenta 500 nodos, número que se definió a través de un proceso de experimentación utilizando los datos de los 30 usuarios definidos para esta fase. La capa de salida está conformada por 6 nodos que predicen las probabilidades de que la entrada pertenezca a cada clase. Se utilizaron 500 iteraciones para el entrenamiento de la red neuronal.

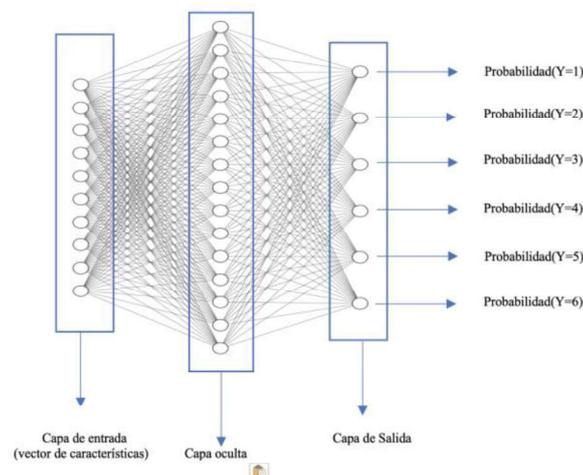


Figura 16. Arquitectura de la red neuronal

Para las neuronas de la capa oculta se utilizó como función de transferencia a ReLU (unidad lineal rectificadora). Esta función se seleccionó después de hacer experimentos con las siguientes funciones log-sig, tanh, softplus, ELU y ReLU. Para el entrenamiento de la red se empleó la regularización mediante decaimiento del valor de los pesos con un factor de regulación igual a $1e-5$.

A continuación, se presentan los resultados obtenidos con las funciones de transferencia y los gráficos obtenidos por cada función se presentan en el Anexo 7.1.

Tabla 1. Resultados obtenidos en la fase de entrenamiento por cada usuario

Nombre de la función de transferencia sigmoideas	Resultado de reconocimiento promedio (conjunto de datos de entrenamiento)
ReLU	94,51%
ELU	92,32%
Tanh	92,85%
log-sigmoid	91,73%
softplus	90,26%

3.4.2.2. Validación preliminar

Para esta etapa, de los 60 usuarios destinados a la fase de entrenamiento, se utilizó los datos de los 30 usuarios restantes que no fueron utilizados en la etapa anterior. El objetivo de esta fase es validar si el modelo obtenido previamente realmente realiza buenas predicciones y descarta que se haya hecho overfitting. En este punto no se realiza ningún cambio en la arquitectura del modelo. Si el modelo obtenido no obtiene más del 88% de exactitud y presenta overfitting, se regresa a la etapa de diseño para reajustar el modelo y así empezar todo el proceso nuevamente.

3.4.3. Fase de evaluación

Para la evaluación se utilizó los datos de los 60 usuarios restantes. El objetivo de esta fase fue evaluar la exactitud de reconocimiento del modelo, teniendo en cuenta que los datos de este grupo de usuarios no se utilizaron para el entrenamiento y validación preliminar del modelo. El resultado que se obtiene en esta fase es el que se reporta en este proyecto.

4. RESULTADOS Y DISCUSIÓN

4.1. Metodología de Evaluación

El conjunto de testeo esta formado por 60 usuarios. Para cada usuario se adquirieron 50 muestras de los 5 gestos a reconocer, donde cada muestra tuvo una duración de 5 segundos a una tasa de muestreo de 200Hz. De las 50 muestras adquiridas para cada gesto 25 fueron destinadas para entrenar modelos individuales y los 25 restantes para testeo. Para cada usuario el modelo fue entrenado utilizando las 25 muestras de entrenamiento por cada gesto a reconocer. Una vez entrenado el modelo se procedió a aplicarlo sobre las 25 muestras de testeo empleando ventanas sucesivas de 66 puntos por cada gesto a reconocer, obteniendo así 2 vectores: un vector conteniendo las etiquetas predichas por el modelo, y el otro vector conteniendo los instantes de tiempo a los que corresponde dichas etiquetas.

Para que el reconocimiento fuese considerado exitoso, todas las etiquetas del vector de respuesta excluyendo la etiqueta del no gesto debe coincidir con la clase correspondiente al gesto y en el instante de tiempo en que este es generado. Para ello se utilizó un factor de solapamiento que se calcula con la siguiente ecuación:

$$p = 2 * \frac{A \cap B}{A + B}$$

Ecuación 5. Fórmula del factor de solapamiento

Donde A es el tiempo de duración de la actividad muscular de la señal generada, B es el tiempo de duración de la actividad muscular en el vector de predicciones (Figura 17). Y ρ es el factor de solapamiento. Para que el reconocimiento sea considerado exitoso este factor no debe superar un umbral $k = 0.25$.

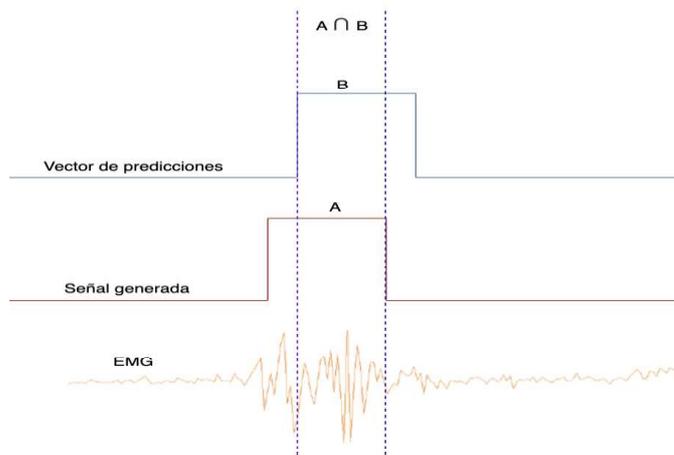


Figura 17. Solapamiento entre el vector de predicciones y la señal generada

La clasificación se considera exitosa cuando la etiqueta que más se repite dentro del vector excluyendo el no gesto es igual al gesto que realizó el usuario. Caso contrario la clasificación es errada.

En cuanto al tiempo de respuesta del modelo, se utilizó un computador de escritorio con un procesador de 4GHz de velocidad y 16 GB de RAM, siendo el tiempo de respuesta de reconocimiento obtenido para este modelo un promedio por cada ventana de 66 puntos. Y para clasificación el tiempo registrado es la suma de los tiempos obtenidos por cada una de las ventanas de 66 puntos.

4.2. Resultados

4.2.1. Resultados de clasificación

Para los 60 usuarios de evaluación se obtuvo una exactitud de clasificación promedio de 95,3% con una desviación estándar de $\pm 9.54\%$ como se observa en la matriz de confusión Tabla 2, y para el conjunto de datos de entrenamiento y validación representada en la Tabla 3 la exactitud de clasificación es del 99,0% con una desviación estándar de $\pm 1.34\%$. Los resultados de esta fase se analizan en la sección de discusión.

Para calcular la exactitud de clasificación del modelo se utilizó la ecuación 6 que es la relación que existe entre la suma de los verdaderos positivos (vp) y verdaderos negativos (vn) sobre la suma de los positivos (p) y los negativos (n).

$$Exactitud = \frac{vp + vn}{p + n}$$

Ecuación 6. Fórmula de la exactitud en la matriz de confusión.

La sensibilidad de la matriz de confusión es igual a la relación entre los verdaderos positivos y los falsos negativos (fn) que se muestra en la ecuación 7.

$$Sensibilidad = \frac{vp}{vp + fn}$$

Ecuación 7. Fórmula de la sensibilidad en la matriz de confusión.

La precisión es calculada por la ecuación 8 que es igual a la relación de los verdaderos positivos con la suma de los verdaderos positivos y los falsos positivos (fp).

$$\text{Precisión} = \frac{vp}{vp + fp}$$

Ecuación 8. Fórmula de la precisión en la matriz de confusión.

Tabla 2. Matriz de confusión usuarios de evaluación

	OBSERVACIONES							%PRECISION
		WAVE IN	WAVE OUT	FIST	FINGERS SPREAD	DOUBLE TAP	NO GESTO	%ERROR
PREDICCIONES	WAVE IN	1454 18,6%	11 0,1%	6 0,1%	2 0,0%	5 0,1%	37 0,5%	96,0% 4,0%
	WAVE OUT	28 0,4%	1466 18,6%	2 0,0%	7 0,1%	8 0,1%	27 0,3%	95,3% 4,7%
	FIST	6 0,1%	4 0,1%	1474 18,9%	10 0,1%	6 0,1%	45 0,6%	95,4% 5,6%
	FINGERS SPREAD	7 0,1%	13 0,2%	4 0,1%	1471 18,9%	9 0,1%	55 0,7%	94,4% 5,6%
	DOUBLE TAP	5 0,1%	6 0,1%	14 0,2%	10 0,1%	1472 18,9%	42 0,5%	95,0% 5,0%
	NO GESTO	0 0,0%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	94 1,2%	100% 0,0%
	%SENSIBILIDAD	96,9%	97,7%	98,3%	98,1%	98,1%	31,3%	95,3%
	%ERROR	3,1%	2,3%	1,7%	1,9%	1,9%	68,7%	4,7%

Tabla 3. Matriz de confusión usuarios de entrenamiento

	OBSERVACIONES						%PRECISION
		WAVE IN	WAVE OUT	FIST	FINGERS SPREAD	DOUBLE TAP	%ERROR
PREDICCIONES	WAVE IN	748 19,9%	5 0,1%	2 0,1%	3 0,1%	7 0,2%	97,8% 2,2%
	WAVE OUT	1 0,0%	739 19,7%	0 0,0%	2 0,1%	1 0,0%	99,5% 0,5%
	FIST	0 0,1%	4 0,1%	739 18,9%	1 0,1%	0 0,1%	99,3% 0,7%
	FINGERS SPREAD	1 0,0%	1 0,0%	4 0,1%	1471 18,9%	0 0,0%	99,2% 0,8%
	DOUBLE TAP	0 0,0%	1 0,0%	5 0,1%	1 0,0%	742 19,8%	99,1% 0,9%
	%SENSIBILIDAD	99,7%	98,5%	98,5%	99,1%	98,9%	99,0%
	%ERROR	0,3%	1,5%	1,5%	0,9%	1,1%	1,0%

4.2.2. Resultados de reconocimiento

Para los 60 usuarios de evaluación o testeo como se observa en la Figura 18 se obtuvo una exactitud de reconocimiento promedio de 92,45% con una desviación estándar de $\pm 11,00\%$ y para el conjunto de datos de entrenamiento y validación como se muestra en la Figura 19 es del 93,18% con una desviación estándar de $\pm 4,96\%$.

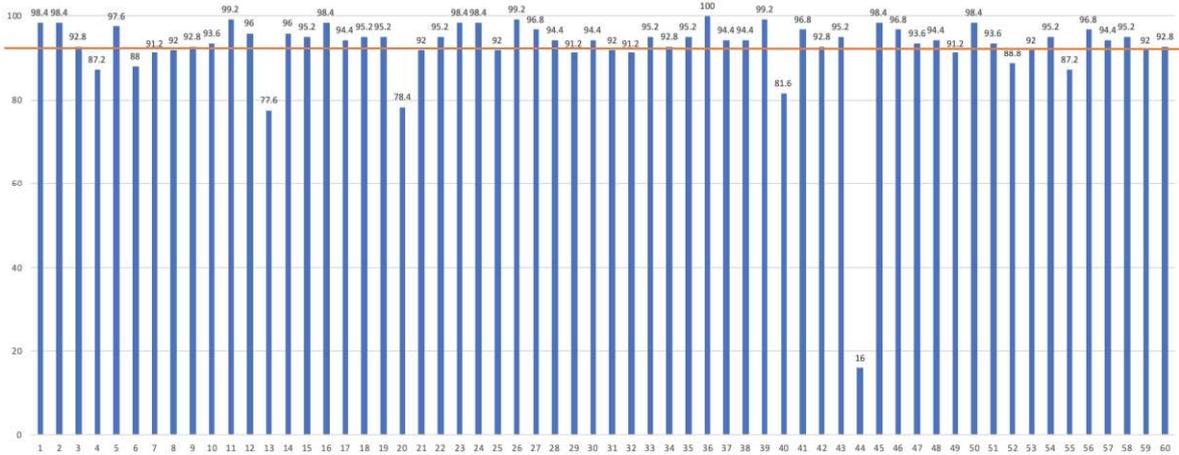


Figura 18. Resultados de reconocimiento por usuario (evaluación)

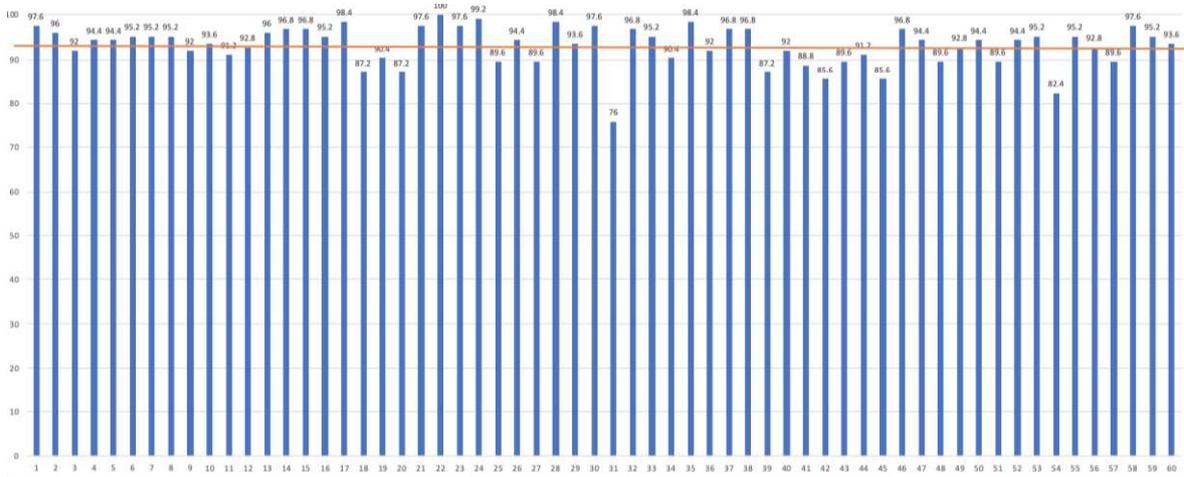


Figura 19. Resultados de reconocimiento por usuario (entrenamiento)

4.2.3. Resultados de tiempo

Para reconocimiento se obtuvo un tiempo de respuesta promedio igual a 40,58 milisegundos y con una desviación estándar igual a $\pm 1,62\%$, como se muestra en el Anexo 7.3. Adicionalmente se calculó el tiempo de respuesta promedio de la clasificación igual a 202,11 milisegundos con una desviación estándar de $\pm 0,008\%$.

4.3. Discusión

Al revisar las matrices de confusión se establece que, de los 5 gestos descritos en el modelo de clasificación, aquellos que tienden a fallar con mayor frecuencia son los gestos *double tap* y *fingers spread*. Al comparar los resultados obtenidos entre los datos de evaluación y entrenamiento se observa que no presenta overfitting ya que existe una diferencia tolerable de 3,7%.

En cuanto a los datos obtenidos para el reconocimiento se puede observar un resultado de reconocimiento promedio de 92,45% con una desviación estándar de $\pm 11,00$. En la Figura 17 se puede observar que el usuario 44 presenta los resultados de reconocimiento más bajos dentro del conjunto de datos analizados, lo cual causó la alta desviación estándar. Esto puede deberse a que cada uno de los usuarios producen señales EMG diferentes o un error en el conjunto de datos utilizado para la presente investigación. Si analizamos los resultados sin tomar en cuenta este usuario se obtiene un resultado promedio de exactitud de reconocimiento de 93,74% con una desviación estándar igual a $\pm 4,58$, que es un resultado similar al obtenido en la etapa de entrenamiento y validación, lo cual indica que no existe overfitting en reconocimiento.

La diferencia que existe entre los resultados de reconocimiento y clasificación está dada por la forma en que se evalúan estas, ya que para clasificación no se toma en cuenta el instante de tiempo en que se genera el gesto, esto hace que predecir el reconocimiento sea una tarea más compleja.

La respuesta promedio del modelo es igual a 40,58 milisegundos con una desviación estándar de $\pm 1,62$, lo cual denota que no existe una diferencia significativa en el tiempo que tarda en responder el modelo de un usuario a otro. El tiempo más alto registrado fue de 45,28 milisegundos el cual es significativamente inferior que el valor definido para tiempo real (300ms).

Al comparar los resultados obtenidos en esta investigación con los de [17], un proyecto de reconocimiento de gestos que se desarrolló como parte del proyecto PIJ-16-13, se destaca que el modelo propuesto supera en 4,92% en exactitud de reconocimiento, y en tiempo de procesamiento es 9,41 veces más rápido utilizando un computador con las mismas características al que se utilizó para este trabajo. En la Tabla 3 se muestra una comparación con modelos similares mencionados en ese trabajo.

Tabla 4. Tabla comparativa con resultados de trabajos similares

<i>Modelo</i>	<i>Resultados de clasificación</i>	<i>Resultados de reconocimiento</i>
Modelo Propuesto	95.3%	92.45%
Modelo 1 [17]	-	87.5 %
Modelo 2 [16]	95.79%	88.1 %
Modelo 3 [15]	90.7%	-
Modelo 4 [13]	88%	-
Modelo 5 [12]	91,06%	-
Myo Armband	83.07%	-

5. CONCLUSIONES Y RECOMENDACIONES

En el presente documento, se ha presentado un modelo de reconocimiento de gestos de la mano humana que trabaja en tiempo real utilizando señales EMG, covarianzas y redes neuronales artificiales. Este modelo reconoce 5 clases de gestos: *fist*, *wave left*, *wave right*, *fingers spread* y *double tap*; así como una sexta clase correspondiente al no gesto. Para entrenar el modelo se requirió de 25 repeticiones por cada gesto y por cada usuario. El modelo responde en un promedio de 40,58 milisegundos, utilizando un computador con bajos recursos computacionales, con una precisión de reconocimiento del 92,45% que es más alta que la tasa de 88% definida en la hipótesis y con una precisión de clasificación promedio del 95,3%.

Cuando se compara el rendimiento del modelo descrito con el sistema de reconocimiento propietario Myo Armband que tiene un 83,07% de exactitud de reconocimiento se determinó que el presente trabajo es un 9.38% más exacto, concluyendo así que un modelo entrenado por usuario genera mejores resultados que un modelo general. El modelo planteado reconoce, clasifica y en menor tiempo de respuesta que los modelos mencionados en este trabajo.

Se encontró que la precisión del modelo de reconocimiento basado en covarianzas (resultados covarianzas Anexo 7.2) mejora cuando se utiliza funciones para analizar las señales como potencia de banda, frecuencia, ancho de banda en el proceso de extracción de características. Estas funciones se deben utilizar con la señal sin pre-procesar para obtener un alto rendimiento del modelo. Para el post-procesamiento la eliminación de valores atípicos es un proceso útil debido a que través de valores futuros y pasados de la señal logra mejorar el reconocimiento.

Para futuros trabajos, se debe considerar utilizar en el proceso de extracción de características las etiquetas de los valores futuros y pasados de la señal para complementar las características ya utilizadas. Además de reducir el tamaño de las sub-ventanas para obtener una mejor observación de la señal, aunque esto aumentaría el tiempo de respuesta del modelo.

6. REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Konar and S. Saha, *Gesture Recognition: Principles, Techniques and Applications*. Clam: Springer International Publishing, 2018
- [2] C. Derr and F. Sahin, "Signer-independent classification of American Sign Language word signs using surface EMG," in *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, 2017.
- [3] L. R. Medsker and L. C. Jain, "RECURRENT NEURAL NETWORKS Design and Applications," Book, 2001.
- [4] D. Farina et al., "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2014.
- [5] S. Benatti et al., "A sub-10mW real-Time implementation for EMG hand gesture recognition based on a multi-core biomedical SoC," in *Proceedings - 2017 7th International Workshop on Advances in Sensors and Interfaces, IWASI 2017*, 2017.
- [6] P. Polygerinos, K. C. Galloway, S. Sanan, M. Herman, and C. J. Walsh, "EMG controlled soft robotic glove for assistance during activities of daily living," in *IEEE International Conference on Rehabilitation Robotics*, 2015.
- [7] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *2014 IEEE International Conference on Image Processing, ICIP 2014*, 2014.
- [8] A. Vargas Sabadías, *Estadística descriptiva e inferencial*. Cuenca: Universidad de Castilla-La Mancha, 1998.
- [9] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, 1991.
- [10] A. Boyali, N. Hashimoto, and O. Matsumoto, "Hand posture and gesture recognition using MYO armband and spectral collaborative representation based classification," in *2015 IEEE 4th Global Conference on Consumer Electronics, GCCE 2015*, 2016.
- [11] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, 2016.

- [12] S. Benatti et al., "EMG-based hand gesture recognition with flexible analog front end," in IEEE 2014 Biomedical Circuits and Systems Conference, BioCAS 2014 - Proceedings, 2014.
- [13] S. Benatti, G. Rovere, J. Bösser, F. Montagna, E. Farella, F. Glaser, P. Schonle, T. Burger, S. Fateh, Q. Huang and L. Benini, "A sub-10mW real-time implementation for EMG hand gesture recognition based on a multi-core biomedical SoC", Advances in Sensors and Interfaces (IWASI), 2017 7th IEEE International Workshop on, 2017.
- [14] Chung Liu, E. A. (2018). Modelo de reconocimiento en tiempo real de gestos de la mano utilizando técnicas de deep learning y señales electromiográficas. 23 hojas. Quito: EPN
- [15] Motoche Macas, C. A. (2018). Reconocimiento de gestos de la mano en tiempo real usando señales electromiográficas y redes neuronales artificiales. 37 hojas. Quito: EPN.
- [16] M. Benalcazar, J. Zea, (2019). "Analysis of Real-Time Hand Gesture Recognition: an LSTM Approach with EMG", Quito: EPN.
- [17] Unapanta Luis, C. A. (2019). Reconocimiento de gestos de la mano en tiempo real basado en señales electromiográficas utilizando myo armband con wavelets y máquinas de vectores de soporte. 47 hojas. Quito: EPN
- [18] N. Y. Y. Kevin, S. Ranganath, and D. Ghosh, "Trajectory modeling in gesture recognition using cybergloves® and magnetic trackers," in IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2004.
- [19] B. Siciliano and O. Khatib, Springer handbook of robotics. 2016.
- [20] S. Yousefi, H. Narui, S. Dayal, S. Ermon and S. Valaee, "A Survey on Behavior Recognition Using WiFi Channel State Information", IEEE Communications Magazine, vol. 55, no. 10, pp. 98-104, 2017.
- [21] S. Weibel, J. Keil and M. Zoellner, "Multi-touch gestural interaction in X3D using hidden Markov models", Proceedings of the 2008 ACM symposium on Virtual reality software and technology - VRST '08, 2008.
- [22] K. Kalgaonkar and B. Raj, "One-handed gesture recognition using ultrasonic Doppler sonar", 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.
- [23] R. Merletti and P. Parker, Electromyography. Hoboken, N. J.: IEEE Press, 2004.

- [24] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan and T. Chang, "Surface Electromyography Signal Processing and Classification Techniques", *Sensors*, vol. 13, no. 9, pp. 12431-12466, 2013.
- [25] Hargrove L., Englehart K., Hudgins B., "A Comparison of Surface and Intramuscular Myoelectric Signal Classification", *Proc. of the 27th Annual Conf. Int.* pag. 5009-5012.
- [26] A. Phinyomark, P. Phukpaaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.
- [27] S. Thongpanja, A. Phinyomark, P. Phukpaaranont, and C. Limsakul, "Mean and Median Frequency of EMG Signal to Determine Muscle Force based on Time-Dependent Power Spectrum," *Elektron. Ir Elektrotehnika*, vol. 19, no. 3, pp. 51–56, Jul. 2013.
- [28] Raez MB, Hussain MS, Mohd-Yasin F (2006) Techniques of EMG signal analysis: detection, processing, classification and applications. *Biol Proced Online*, 8:11–35
- [29] P. Visconti, F. Gaetani, G. Zappatore and P. Primiceri, "Technical Features and Functionalities of Myo Armband: An Overview on Related Literature and Advanced Applications of Myoelectric Armbands Mainly Focused on Arm Prostheses", *International Journal on Smart Sensing and Intelligent Systems*, vol. 11, no. 1, pp. 1-25, 2018.
- [30] I. Thalmic Labs, "Myo Gesture Control Armband," Myo, 2016.
- [31] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette y B. Gosselin, «Deep Learning for Electromyographic Hand Gesture Signal Classification by Leveraging Transfer Learning,» 2018.
- [32] D. Svozil, V. Kvasnicka and J. Pospichal, "Introduction to multi-layer feed-forward neural networks", *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43-62, 1997.
- [33] G.M. Maggiora, D.W. Elrod, R.G. Trenary, Computational neural networks as model-free mapping device, *J. Chem. Inf. Comp. Sci.* 32 (1992) 732-741.
- [34] T. Kohonen, *Self-organisation and Associative Memory*, Springer Verlag, Berlin, 1988.

7. ANEXOS

7.1. Gráficos de las funciones de transferencias.

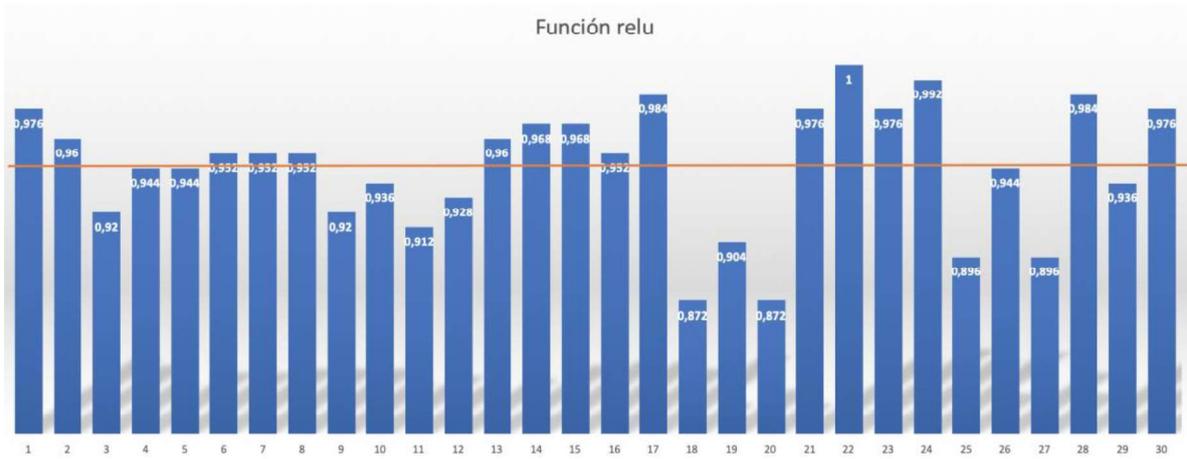


Figura 20. Resultados obtenidos por usuario función RELU

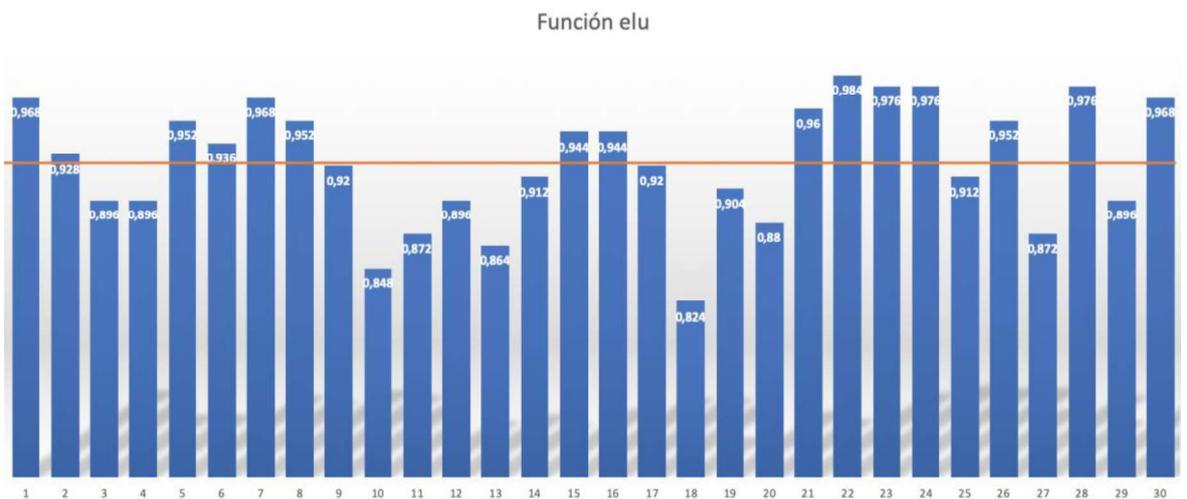


Figura 21. Resultados obtenidos por usuario función ELU

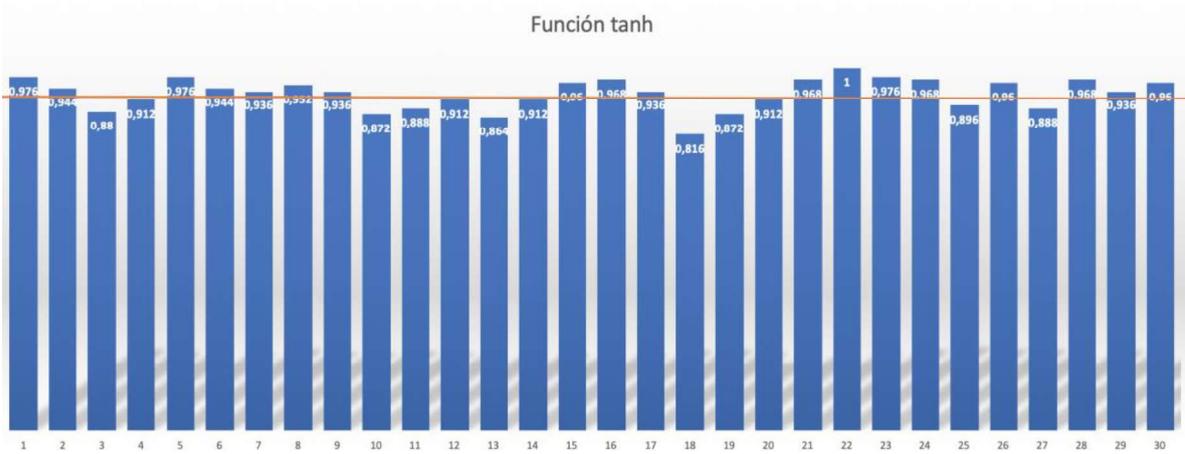


Figura 22. Resultados obtenidos por usuario función TANH

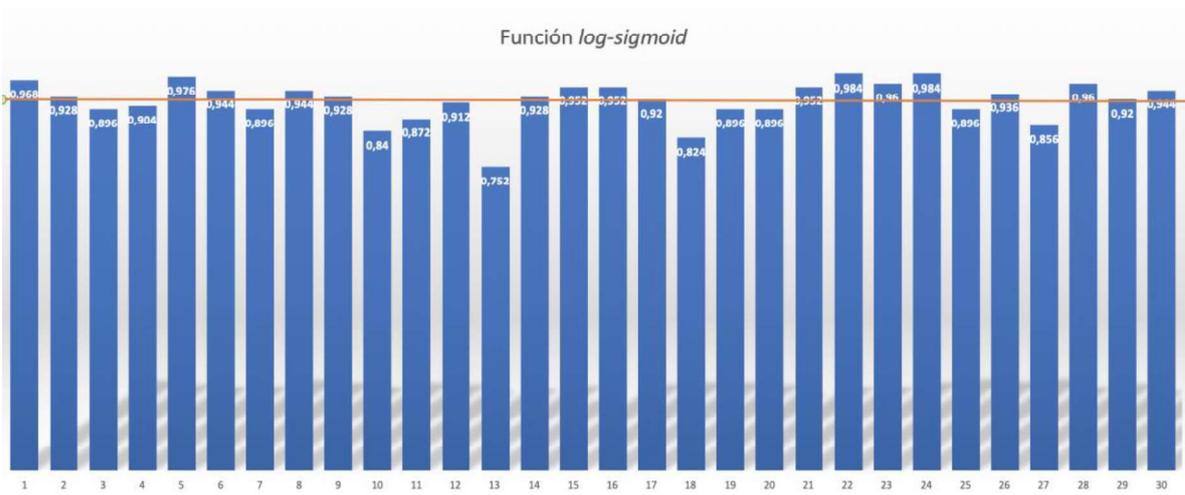


Figura 23. Resultados obtenidos por usuario función log-sigmoid

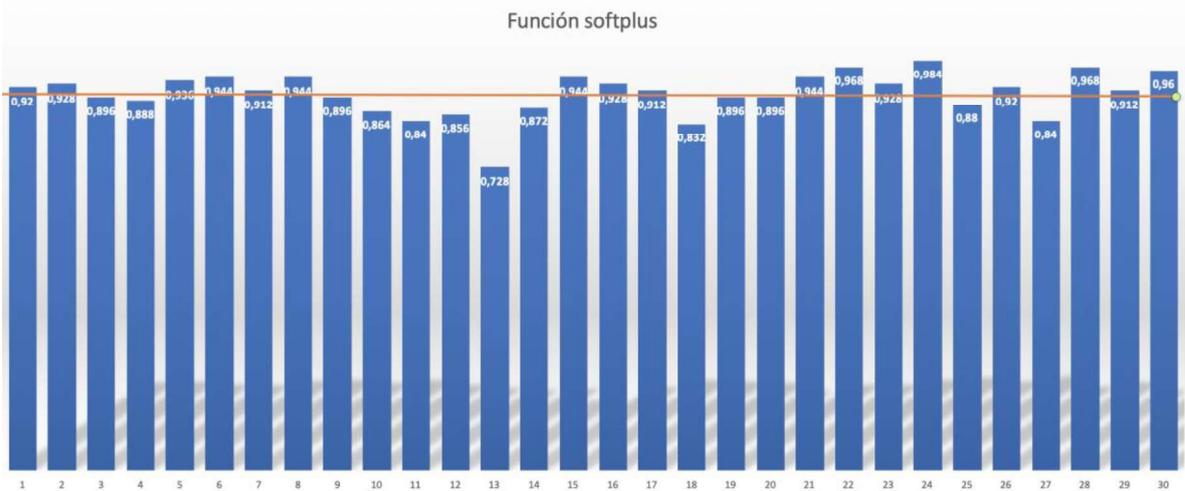


Figura 24. Resultados obtenidos por usuario función softplus

7.2. Resultados de reconocimiento solo utilizando covarianzas utilizando los usuarios de entrenamiento y validación.

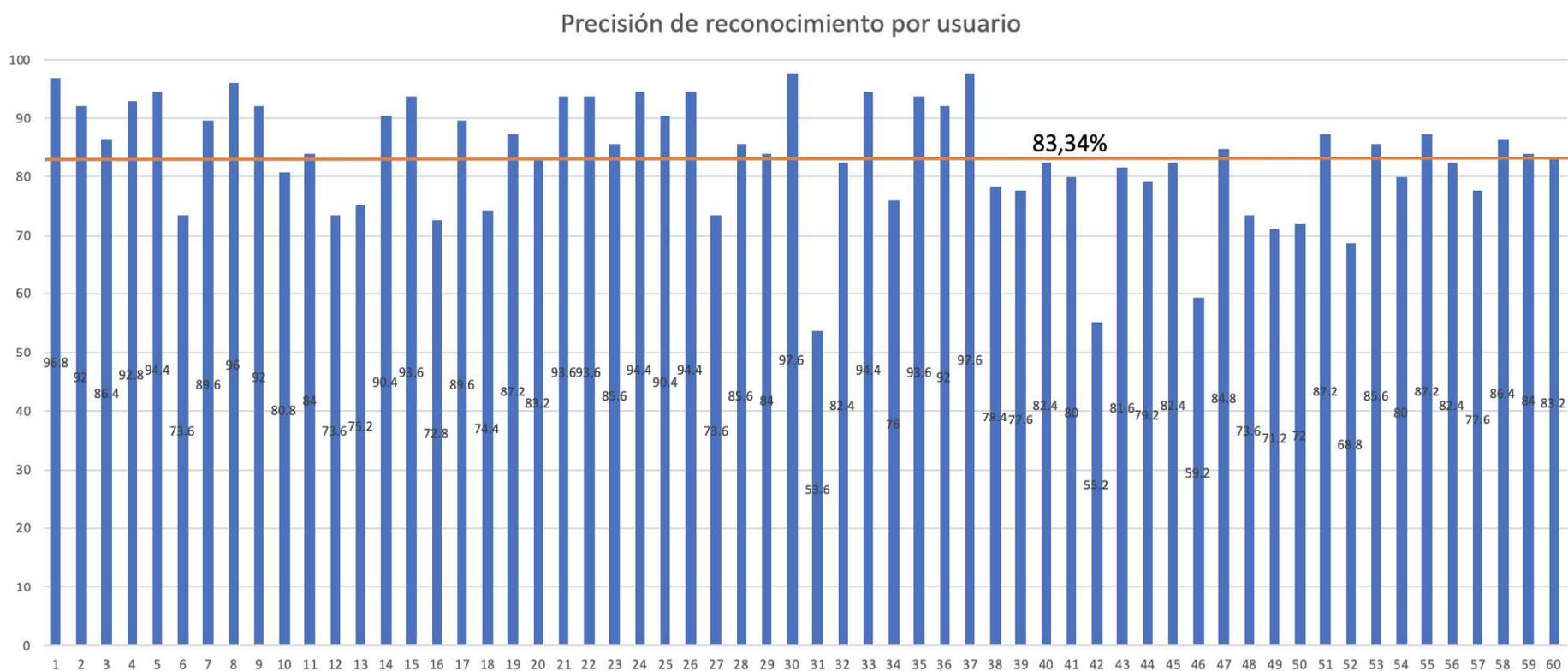


Figura 25. Resultados obtenidos por usuario solo utilizando covarianzas, la línea naranja representa la precisión promedio igual a 88,30%

7.3. Resultados del tiempo de respuesta utilizando los usuarios de evaluación.

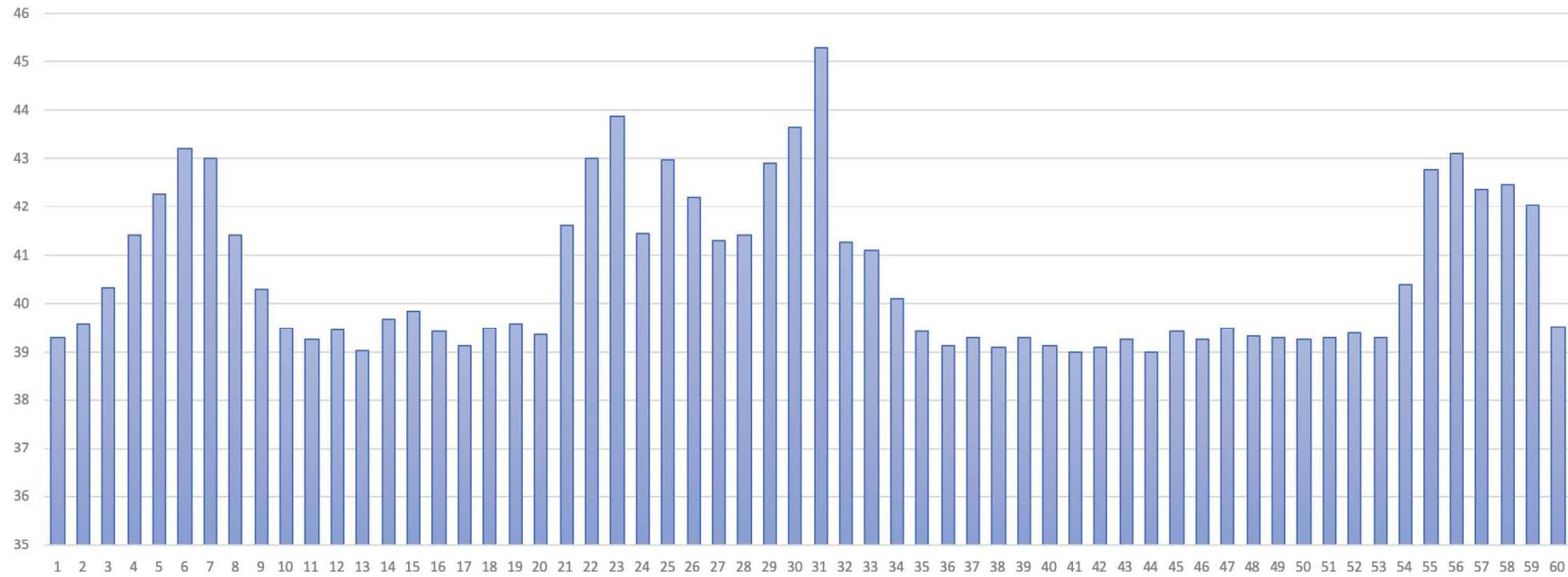


Figura 20. Tiempo de respuesta promedio por usuario (ms)

ORDEN DE EMPASTADO