

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA EN SISTEMAS

**RECONOCIMIENTO DE GESTOS DE LA MANO EN TIEMPO REAL
USANDO SEÑALES ELECTROMIOGRÁFICAS Y REDES
NEURONALES ARTIFICIALES**

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

CRISTHIAN ALBERTO MOTOCHE MACAS

cristhian.motoche@epn.edu.ec

DIRECTOR: MARCO E. BENALCÁZAR, PHD

marco.benalcazar@epn.edu.ec

Quito, Mayo 2018

AVAL

Certifico que el presente trabajo fue desarrollado por Cristhian Alberto Motoche Macas, bajo mi supervisión.

Marco Enrique Benalcázar, PhD

DECLARACIÓN DE AUTORÍA

Yo, Cristhian Alberto Motoche Macas, 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.

Cristhian Alberto Motoche Macas

DEDICATORIA

Dedico este trabajo a mis padres, hermanos, y a los integrantes del grupo de investigación del proyecto junior PIJ-16-13.

AGRADECIMIENTO

Le agradezco a mis padres y hermanos por su apoyo y aliento, al Dr. Marco Benalcázar por su guía en el desarrollo del proyecto y al grupo de investigación del proyecto junior PIJ-16-13 por el conocimiento compartido.

ÍNDICE DE CONTENIDO

AVAL	II
DECLARACIÓN DE AUTORÍA.....	III
DEDICATORIA.....	IV
AGRADECIMIENTO.....	V
ÍNDICE DE CONTENIDO.....	VI
RESUMEN	VIII
ABSTRACT	IX
1. INTRODUCCIÓN.....	1
1.1. Pregunta de investigación	2
1.2. Objetivo General	3
1.3. Objetivos Específicos.....	3
1.4. Alcance	3
1.5. Marco Teórico	4
1.5.1. Sensores para el reconocimiento de gestos de la mano	4
1.5.2. Electromiografía.....	6
1.5.2.1. Anatomía del músculo esquelético	6
1.5.2.2. Fisiología del músculo esquelético	7
1.5.3. Revisión del estado del arte	8
2. MATERIALES Y MÉTODOS	9
2.1 Materiales	9
2.1.1. Myo Armband.....	9
2.1.2. Conjunto de datos.....	10
2.2. Métodos	10
2.2.1. Notación.....	11
2.2.2. Adquisición de datos	11
2.2.3. Preprocesamiento.....	11
2.2.4. Extracción de características	13
2.2.5. Clasificación.....	16
2.2.6. Postprocesamiento	17
3. RESULTADOS Y DISCUSIÓN	17
3.1 Metodología de Evaluación	17
3.2 Resultados	18
3.2.1. Modelo propuesto	18
3.2.2. Modelo basado solo en los valores de la señal preprocesada.....	20

3.2.3.	Modelo basado solo en el resultado del conjunto de funciones	21
3.2.4.	Modelo generalizado	22
3.3	Discusión	23
4.	CONCLUSIONES	24
5.	REFERENCIAS BIBLIOGRÁFICAS	25
6.	ANEXOS	27
	ANEXO I	27
	ANEXO II	27
	ANEXO III	28

RESUMEN

En el presente trabajo, se propone un modelo de reconocimiento de gestos de la mano humana basado en señales electromiográficas superficiales. El modelo propuesto reconoce cinco gestos estáticos de la mano: *Wave In*, *Wave Out*, *Fist*, *Double Tap*, y *Fingers Spread*. Para este trabajo se aplica un enfoque de ventana deslizante utilizando una ventana principal y una sub-ventana. La sub-ventana es utilizada para observar segmentos de la señal vista a través de la ventana principal. El modelo propuesto está compuesto de cinco bloques: adquisición de datos, preprocesamiento, extracción de características, clasificación y postprocesamiento. Para la adquisición de datos, se aplica el enfoque de ventana deslizante y se utiliza el dispositivo *Myo Armband* para medir la señal electromiográfica superficial del antebrazo. Para el preprocesamiento, se aplica rectificación, filtrado, y detección de la actividad muscular a la señal observada a través de la ventana principal. Para la extracción de características, se genera un vector de características utilizando los valores de la señal preprocesada y los resultados de la aplicación de un conjunto de funciones. Para clasificación, se utiliza una red neuronal *feedforward* de tres capas para etiquetar a cada segmento de la señal electromiográfica vista a través de una sub-ventana. Finalmente, para el postprocesamiento se aplica una votación por mayoría simple para, en base a las etiquetas de las observaciones de las sub-ventanas, asignar la etiqueta correspondiente a la señal observada a través de la ventana principal. El modelo propuesto responde en un promedio de 29.38 ms (tiempo real) y con una exactitud de reconocimiento de 90.7%, que es superior a la tasa definida estado del arte.

PALABRAS CLAVE: Aprendizaje de Máquina, Electromiografía, Procesamiento de Señales, Reconocimiento de Gestos, Redes Neuronales Artificiales.

ABSTRACT

In this paper, a hand gesture recognition model based on superficial electromyographic signals is proposed. The proposed model recognizes five static hand gestures: *Wave In*, *Wave Out*, *Fist*, *Double Tap*, and *Fingers Spread*. For this work, a sliding window approach is applied using a main window and a sub-window. The sub-window is used to observe segments of the signal seen through the main window. The proposed model is composed of five blocks: data acquisition, preprocessing, feature extraction, classification and postprocessing. For data acquisition, a sliding window approach is applied, and the *Myo Armband* is used to measure the superficial electromyographic signal of the forearm. For preprocessing, a function for rectification, filtering and muscle activity detection is applied to the signal seen through the main window observation. For feature extraction, a feature vector is generated using the preprocessed signal values and the results from the application of a bag of functions. For classification, a feedforward neuronal network of three layers is used to classify every segment of the signal seen through a sub-window. Finally, for postprocessing a simple majority voting, based on the labels from the sub-window observations, is applied to assign the corresponding label to the signal seen through the main window. The model responds at an average of 29.38 ms (real time) with a recognition accuracy of 90.7%, which higher than the rate established in the state-of-the-art.

KEYWORDS: Artificial Neural Networks, Electromyography, Gesture Recognition, Machine Learning, Signal Processing.

1. INTRODUCCIÓN

El reconocimiento de gestos de la mano consiste en identificar el instante de tiempo y la clase correspondiente a un movimiento de la mano a partir de los datos medidos durante la realización de dicho movimiento [1]. Los modelos de reconocimiento de gestos de la mano tienen múltiples aplicaciones en diferentes campos científicos y tecnológicos, por ejemplo: implementación de interfaces humano-computador (IHC), control de prótesis activas, manipulación de imágenes digitales durante procedimientos médicos, fortalecimiento de la educación del lenguaje de señas para sordomudos, mapeo de movimientos de la mano para interactuar con objetos y ambientes virtuales en aplicaciones de realidad virtual [2].

Un modelo apropiado para este tipo de aplicaciones debe responder con una alta precisión de reconocimiento y usualmente en tiempo real. Para que un modelo de reconocimiento de gestos funcione en tiempo real, su tiempo de respuesta debe ser menor a 300 ms [3]. Adicionalmente, para algunas aplicaciones particulares (como las IHC) el modelo también debe ser portable, lo que significa que es capaz de funcionar en un computador o *hardware* de recursos limitados de memoria *RAM* y de procesamiento.

Los modelos de reconocimiento de gestos de la mano comúnmente utilizan sensores como guantes instrumentados, cámaras a color, cámaras de profundidad, y sensores electromiográficos para adquirir los datos de entrada para el modelo de reconocimiento [4] [5] [6]. En el presente trabajo se utiliza sensores electromiográficos para el modelo propuesto. De acuerdo con la literatura científica, el estado del arte en cuanto a precisión de reconocimiento es de alrededor del 85% para los modelos que utilizan sensores electromiográficos para el reconocimiento de gestos de la mano [3]. Por esta razón, en el presente trabajo el objetivo general es desarrollar un modelo de reconocimiento de gestos de la mano que alcance una precisión de reconocimiento superior al 85%. Adicionalmente, la respuesta del modelo debe ser en tiempo real y debe ser capaz de ejecutarse en un computador o *hardware* de recursos limitados de memoria *RAM* y de procesamiento.

Machine Learning es un marco de trabajo que se puede utilizar para resolver el problema de reconocimiento de gestos de la mano humana basado en señales electromiográficas superficiales (sEMG, por sus siglas en inglés *superficial electromyography*). Los clasificadores más comúnmente utilizados para el reconocimiento de gestos de la mano incluyen: *Support Vector Machines* [7], *Artificial Neuronal Networks* [8] [9], *Deep Convolutional Neuronal Networks* [10], y *k Nearest Neighbors* [11] [12]. Las características o descriptores comúnmente utilizados para el reconocimiento de gestos de la mano están

definidos en los siguientes dominios: tiempo (e. g., Valor Absoluto Medio y *Zero Crossing*), frecuencia (e. g., Frecuencia Media e Histograma de Frecuencias) y tiempo-frecuencia (e. g., *Wavelets*). Los modelos basados en estos clasificadores y dominios de características presentan una alta precisión de reconocimiento y responden en tiempo real, pero también tienen desventajas. Las desventajas más comunes incluyen: predicción de pocas clases [3], demanda de muchos ejemplos de cada clase para entrenar el modelo [13] y demanda de recursos computacionales elevados [10]. Por lo tanto, el reconocimiento de gestos de la mano sigue siendo un problema que está abierto para nuevas investigaciones.

En este trabajo, se presenta un modelo de reconocimiento de gestos de la mano humana basado en señales sEMG que responde en tiempo real, logra una precisión de reconocimiento que sobrepasa el estado del arte y funciona en una computadora con recursos limitados de memoria *RAM* y de procesamiento. El modelo propuesto sigue un enfoque de ventana deslizante usando una ventana principal y una sub-ventana. El modelo se compone de los siguientes bloques: adquisición de datos, preprocesamiento, extracción de características, clasificación y postprocesamiento. Para la adquisición de datos, se aplica el enfoque de ventana deslizante y se utiliza el dispositivo *Myo Armband* para medir la señal sEMG del antebrazo. Para el preprocesamiento, se rectifica, filtra y detecta la actividad muscular a la señal electromiográfica vista a través de la ventana principal. Para la extracción de características, se genera un vector de características concatenando los valores de la señal preprocesada con los resultados de aplicar un conjunto de funciones a la señal observada a través de la sub-ventana. Para la clasificación de cada segmento de la señal visto a través de la sub-ventana se utiliza una red neuronal *feedforward* de tres capas, con una función de transferencia sigmoidea en la capa oculta. Finalmente, para el postprocesamiento se aplica una votación por mayoría simple, en base a las etiquetas de las sub-ventanas, para asignar la etiqueta de la señal vista a través de la ventana principal.

Siguiendo esta introducción en lo restante de esta sección se presenta la pregunta de investigación, el objetivo general, los objetivos específicos, el alcance y el marco teórico del presente proyecto. El resto de este documento está organizado en tres secciones. En la sección 2, se describe los materiales y métodos empleados en este trabajo. En la sección 3, se presenta el método de evaluación y los resultados obtenidos. Finalmente, en la sección 4, se presenta las conclusiones y el esquema para trabajo futuro.

1.1. Pregunta de investigación

¿Cuál es la estructura de un modelo de reconocimiento de gestos de la mano humana basado en señales electromiográficas superficiales y empleando una red neuronal artificial

para alcanzar una exactitud de reconocimiento mayor a la definida en el estado del arte (85%), responder en tiempo real (menos de 300 ms) y ser capaz de funcionar en un computador o *hardware* con recursos limitados de memoria *RAM* y de procesamiento?

1.2. Objetivo General

Desarrollar un sistema de reconocimiento de gestos estáticos de la mano humana que responda en tiempo real y con una exactitud de reconocimiento mayor a 85% mediante el uso de señales sEMG y redes neuronales artificiales.

1.3. Objetivos Específicos

- a) Desarrollar un modelo de reconocimiento de gestos que cumpla con las siguientes características:
 - Funcionamiento en tiempo real (reconocimiento de un gesto en menos de 300 ms).
 - Exactitud de reconocimiento mayor o igual al 85% en los datos de prueba.
 - Ejecución en un computador de escritorio que tenga como mínimo un procesador de 8 núcleos, un disco duro de 150GB de espacio, velocidad de procesamiento de 4 GHz y memoria *RAM* de 16 GB.
 - Reconocimiento de 5 gestos estáticos de la mano humana basado en un número de ejemplos de entrenamiento que se encuentre en el orden de las unidades.
- b) Validar que el modelo propuesto cumpla con lo descrito anteriormente.

1.4. Alcance

En el presente trabajo de investigación se utilizará el sensor comercial *Myo Armband* para la medición de la señal sEMG del antebrazo de cada usuario. Se desarrollará un modelo de reconocimiento de gestos de la mano humana. El modelo se entrenará y adaptará individualmente a cada usuario. El modelo necesitará de la grabación de 5 repeticiones de cada clase para ser entrenado. El tiempo de respuesta del modelo deberá ser menor a 300 ms para ser considerado un modelo en tiempo real. Adicionalmente, el modelo propuesto debe conseguir una precisión de reconocimiento superior al 85%.

El modelo de reconocimiento será capaz de reconocer cinco gestos de la mano: mano con los dedos juntos inclinada hacia adentro, mano con los dedos juntos inclinada hacia afuera, puño cerrado, doble toque entre el dedo pulgar y el dedo medio y mano con los dedos

extendidos. Estos gestos se denotarán por los siguientes nombres en inglés: *Wave In*, *Wave Out*, *Fist*, *Double Tap*, y *Fingers Spread*, respectivamente.

1.5. Marco Teórico

1.5.1. Sensores para el reconocimiento de gestos de la mano

Los sensores más comunes empleados en el reconocimiento de gestos de la mano humana incluyen: cámaras a color, guantes instrumentados, cámaras de profundidad, guantes con marcadores de color, sensores ultrasónicos, y sensores de señales electromiográficas [4] [5] [6]. En la **Figura 1** se puede observar algunos ejemplos de estos sensores.

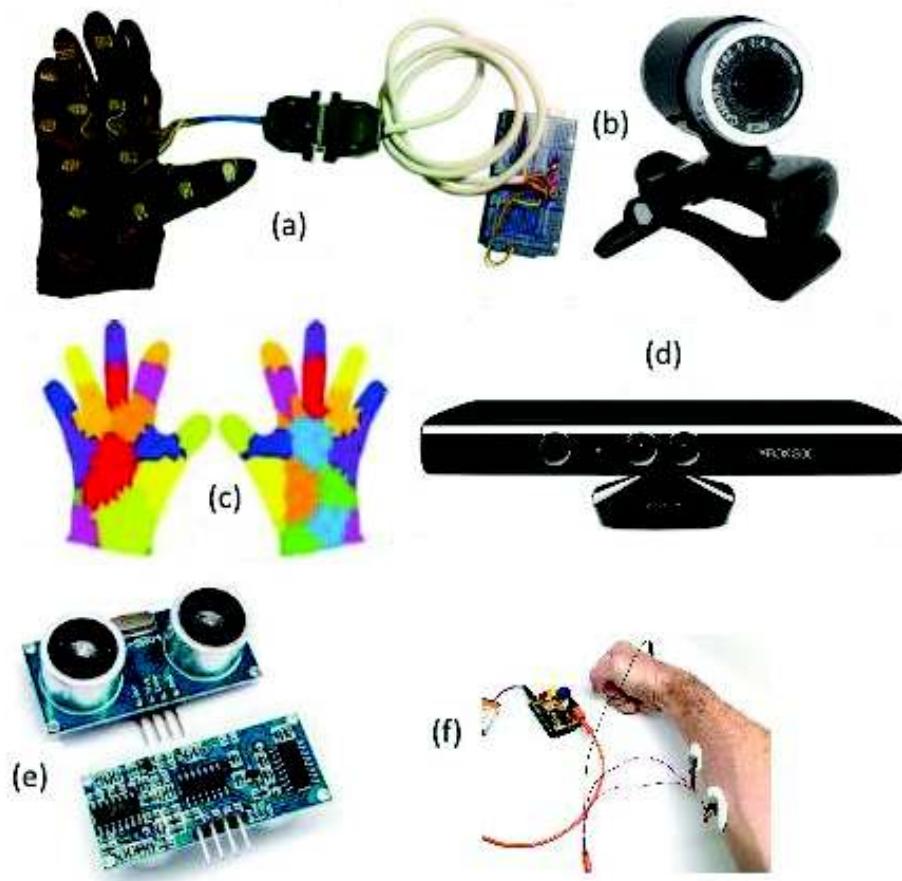


Figura 1. (a) Guantes instrumentados. (b) Cámara web estándar. (c) Guantes con marcadores de color. (d) Cámara de profundidad (Dispositivo Kinect). (e) Sensores de ultrasonido. (f) Electrodos que miden señales electromiográficas ubicados en el brazo.

En la **Tabla 1**, se muestra de forma resumida las ventajas y desventajas que han sido descritas por [4] [5] [6] [9] y [14] sobre estos sensores.

Sensor	Ventajas	Desventajas
Guantes instrumentados (Figura 1 (a))	Miden un número alto de grados de libertad. Obtienen información de la mano y de los dedos directamente.	Son incómodos para un uso prolongado. No se adaptan a la variación de tamaño de la mano de diferentes usuarios.
Cámaras a color (Figura 1 (b))	No requiere estar adherido a las extremidades. Son de bajo costo. Por ejemplo, las cámaras web estándar.	Alta sensibilidad a las variaciones de luz (color, sombra, brillo), posición y distancia entre el usuario y la cámara. Son difíciles de usar en modelos de reconocimiento portables.
Guantes con marcas de colores y cámaras (Figura 1 (c))	Se pueden utilizar con cámaras web estándar y son de bajo costo. Son más confortables que los guantes instrumentados.	Presentan los mismos problemas que las cámaras a color.
Cámaras de profundidad (Microsoft Kinect) Figura 1 (d))	No son afectadas por factores como la iluminación, sombra o color.	Son de costo elevado. Son difíciles de usar en modelos de reconocimiento portables.
Sensores de ultrasonido (Figura 1 (e))	Son de bajo costo. No requieren estar adheridos al cuerpo del usuario.	Sufren del efecto cross-talking. Este efecto se da cuando la onda de un sensor es captada por otro sensor o cuando un sensor capta su propia onda de un disparo previo producto de dispersiones. Son difíciles de usar en modelos de reconocimiento portables.
Sensores electromiográficos (Figura 1 (f))	Su señal no es afectada por variaciones de luz. Se puede medir su señal en personas que han sufrido amputación de su mano.	La señal medida contiene ruido de diferentes fuentes, por ejemplo: ruido del dispositivo de medición, ruido por radiación electromagnética, y ruido por el movimiento del artefacto.

Tabla 1. Descripción de los sensores para reconocimiento de gestos de la mano humana.

En el presente trabajo, se optó por el uso de sensores electromiográficos superficiales para medir la señal electromiográfica que es la entrada para el modelo propuesto. Los sensores electromiográficos son de bajo costo, no son afectados por variaciones del tamaño de la mano de los usuarios y se pueden utilizar en sistemas de reconocimiento portables. La señal de estos sensores no es afectada por las variaciones de luz del ambiente o la posición y orientación de la mano. Adicionalmente, estos sensores pueden ser utilizados por períodos prolongados sin causar incomodidad al usuario.

1.5.2. Electromiografía

La electromiografía (EMG por sus siglas en inglés *elecromyography*) es el estudio de la función muscular por medio de las señales eléctricas que emanan los músculos esqueléticos [15]. Las señales EMG se producen por variaciones fisiológicas en el estado de las membranas de las fibras musculares [16] de los músculos esqueléticos.

1.5.2.1. Anatomía del músculo esquelético

Los músculos esqueléticos están compuestos por células musculares (fibras musculares), vasos sanguíneos, fibras nerviosas y tejido conectivo. A diferencia de los músculos cardiacos y suaves, los músculos esqueléticos se contraen por medio de la señalización del sistema nervioso [17] que es transmitida por las motoneuronas. La **Figura 2** ilustra la anatomía del músculo esquelético.

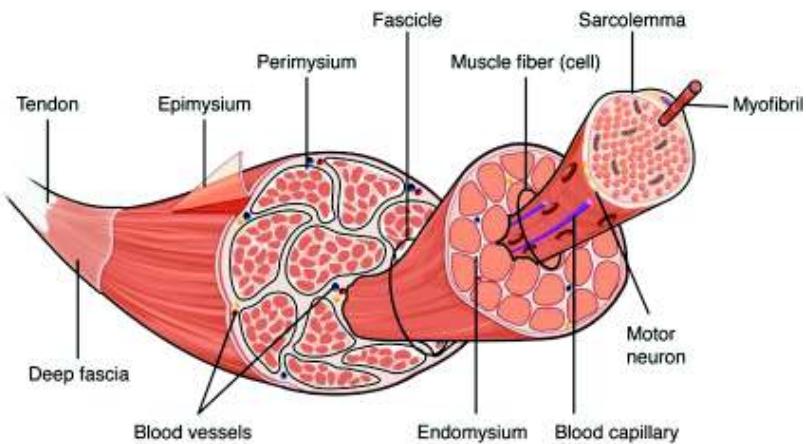


Figura 2. Anatomía de los músculos esqueléticos. Imagen obtenida de [17].

Los músculos están compuestos por fascículos. Los fascículos están compuestos de fibras musculares. Cada fibra muscular está compuesta de miofibrillas que a su vez están formadas de miofilamentos. Las motoneuronas que forman parte del sistema nervioso motor tienen terminaciones en las fibras musculares. El núcleo y dendritas de las motoneuronas están ubicados en la espina dorsal. El axón de la motoneurona se prolonga

desde la espina dorsal hasta el músculo. Las terminaciones de la motoneurona forman una unión neuromuscular con las fibras musculares [17].

Las fibras musculares son células elongadas y cilíndricas que están cubiertas por una membrana plasmática llamada sarcolema. La unidad funcional de las fibras es el sarcómero que es una disposición organizada de los miofilamentos contráctiles, actina y miosina, junto con otras proteínas de soporte. Varios sarcómeros conectados entre sí forman los filamentos longitudinales llamados miofibrillas que se extienden por la fibra y se adhiere al sarcolema en su extremo [17] [18]. La contracción muscular involucra el acortamiento longitudinal individual de los sarcómeros.

1.5.2.2. *Fisiología del músculo esquelético*

El movimiento de las extremidades se controla mediante señales eléctricas que viajan entre los músculos y el sistema nervioso periférico y central [18]. La contracción de los músculos esqueléticos es el resultado de la contracción de varias unidades motoras. Una unidad motora es un conjunto de varias fibras musculares inervadas por una única motoneurona [19]. Una fibra muscular es activada por los impulsos provenientes de una motoneurona, lo cual provoca la despolarización de la fibra. En condiciones normales, el sarcolema de las fibras musculares en reposo mantiene un potencial de aproximadamente -80 mV entre el ambiente intracelular y el extracelular. La despolarización provoca un cambio de -80 mV hasta aproximadamente +30 mV que es restaurado en la fase de repolarización [16]. La inervación en la fibra muscular provoca que dos potenciales de acción intracelulares se propaguen desde la unión neuromuscular hasta los tendones [16]. Un potencial de acción es la forma de onda detectada durante la despolarización.

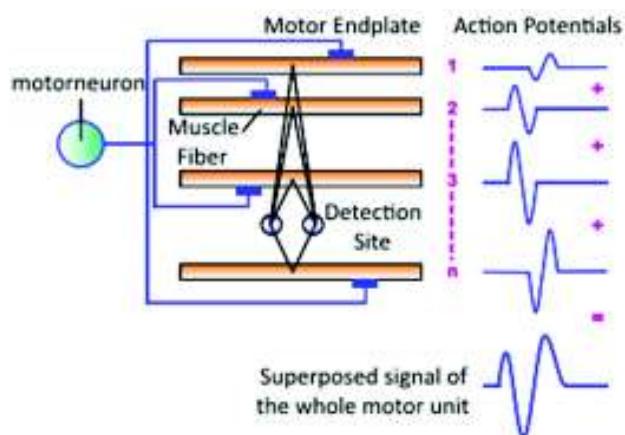


Figura 3. Generación de los potenciales de acción y el potencial de acción de unidad motora. Imagen obtenida de [16].

Las fibras musculares de una unidad motora se solapan y la suma de potenciales de acción produce un potencial de acción de unidad motora (**Figura 3**) [15]. Cuando un músculo se mantiene contraído, la secuencia resultante de potenciales de acción de unidad motora forma un tren de potenciales de acción de unidad motora (MUAPT, por sus siglas en inglés Motor Unit Action Potential Train). Una señal electromiográfica es la suma algebraica de todos los MUAPTs de todas las unidades motoras activas dentro de un área de captación de un electrodo [15]. La formación de la EMG basada en la suma de MUAPTs se ilustra en la **Figura 4**.

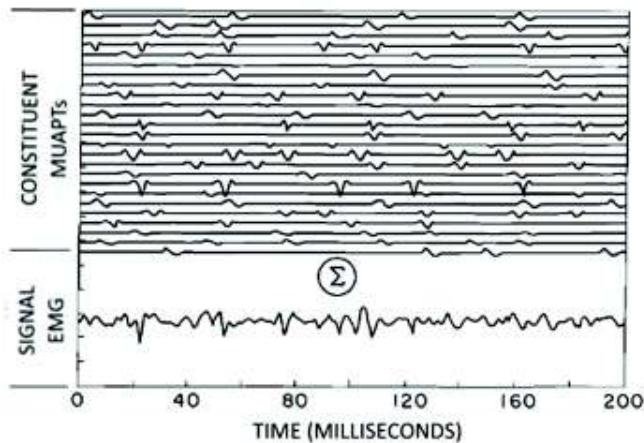


Figura 4. Señal EMG formada por la suma de MUAPTs. Imagen obtenida de [15].

1.5.3. Revisión del estado del arte

Para el presente trabajo se realizó una revisión no exhaustiva del estado del arte de modelos de reconocimiento de gestos de la mano humana. Se limitó a los modelos de reconocimiento en tiempo real, basados en señales sEMGs y que emplean técnicas de Machine Learning para resolver el problema. No se limitó por el año de publicación, pero la revisión del estado del arte consistió en los trabajos más recientes que cumplen las condiciones anteriores.

Los clasificadores más comúnmente utilizados para el reconocimiento de gestos de la mano incluyen: *Support Vector Machines* [7], *Artificial Neuronal Networks* [8] [9], *Deep Convolutional Neuronal Networks* [10], y *k Nearest Neighbors* [11] [12]. Las características o descriptores comúnmente utilizados para el reconocimiento de gestos de la mano están definidos en los siguientes dominios: tiempo (e. g., Valor Absoluto Medio y Zero Crossing), frecuencia (e. g., Frecuencia Media e Histograma de Frecuencias) y tiempo-frecuencia (e. g., Wavelets). Los modelos basados en estos clasificadores y dominios de características presentan una alta precisión de reconocimiento y responden en tiempo real, pero también tienen desventajas que se analizan a continuación.

En [3] se propone un sistema de bajo consumo energético que solo depende de 5 repeticiones de cada clase a reconocer y obtiene una exactitud de 88% en el reconocimiento de 3 clases. El problema de este sistema es la poca cantidad de clases que el sistema puede reconocer. El modelo desarrollado por [10] utiliza *High-Density sEMG*, responde en tiempo real y presenta una tasa de exactitud de reconocimiento de 84.6% utilizando una ventana que abarca solo una “imagen instantánea sEMG”. Cuando se incrementa el tamaño de la ventana a 149 imágenes y aplicando una técnica de votación por mayoría simple logran un 99.6% de exactitud. Sin embargo, el principal problema de este sistema es que requiere el uso de una estación de trabajo con altas capacidades de cómputo al trabajar con *High-Density sEMG*. Por otro lado, en [13] proponen un sistema de reconocimiento que alcanza un 91.7% de exactitud de reconocimiento. El problema en este modelo es que se requiere 10 repeticiones de cada clase para su entrenamiento. A pesar de las soluciones propuestas, el reconocimiento de gestos de la mano sigue siendo un problema que está abierto para nuevas investigaciones.

2. MATERIALES Y MÉTODOS

En esta sección, se describe los materiales y métodos empleados en el presente trabajo. En la subsección de materiales se describirá las características y ventajas del dispositivo *Myo Armband* y los datos utilizados para entrenamiento, validación y testeo. En la subsección de métodos se detallarán los cinco bloques que conforman al modelo propuesto.

2.1 Materiales

2.1.1. Myo Armband

En este trabajo se utilizó el dispositivo *Myo Armband* de *Thalmics*, el cual está ilustrado en la **Figura 5** (a), porque proporciona un kit de desarrollo de software abierto, tiene bajo costo, puede ser expandido de 19 a 34 cm, y tan solo pesa 93 gramos. El *Myo Armband* es un dispositivo inalámbrico que incluye los siguientes componentes: 8 sensores electromiográficos superficiales de acero inoxidable, una unidad de medición inercial de 9 ejes, una conexión *Bluetooth* de baja energía, una batería de ion litio recargable y un retroalimentador haptico (vibración) [20]. El *Myo* mide las señales sEMG a una frecuencia de 200 Hz y representa cada medición con un entero de 8 bits [21]. Los canales del *Myo* están ilustrados en la **Figura 5** (b). Además, el *Myo* está equipado con un software privado que reconoce cinco gestos: *Fist*, *Wave In*, *Wave Out*, *Fingers Spread*, y *Double Tap*, que están ilustrados en la **Figura 5** (c).

2.1.2. Conjunto de datos

Para este trabajo, se utilizó los datos de 10 voluntarios saludables que fueron utilizados previamente en [11] y [12] para entrenamiento, validación y testeo. Se escogió este conjunto de datos para poder comparar los resultados del modelo propuesto con los de otros modelos. El conjunto de datos utilizado en este trabajo contiene un conjunto para entrenamiento y otro conjunto para testeo. El conjunto de entrenamiento consta de cinco repeticiones de los cinco gestos ilustrados en la **Figura 5** (c) grabadas durante dos segundos cada una. Adicionalmente, el conjunto de entrenamiento incluye cinco medidas de las señales sEMG con la mano en la posición de relajación grabadas durante dos segundos cada una. El conjunto de entrenamiento se utilizó para entrenamiento y validación preliminar del modelo aplicando validación cruzada. El conjunto de datos para testeo consiste en 30 repeticiones grabadas durante 5 segundos únicamente de los cinco gestos ilustrados en la **Figura 5** (c). En cada repetición, del conjunto de entrenamiento y testeo, el voluntario comienza con su brazo en una posición de relajación, en un momento aleatorio de la grabación el voluntario realiza el gesto y retorna su brazo a la posición de relajación hasta el final de la grabación.

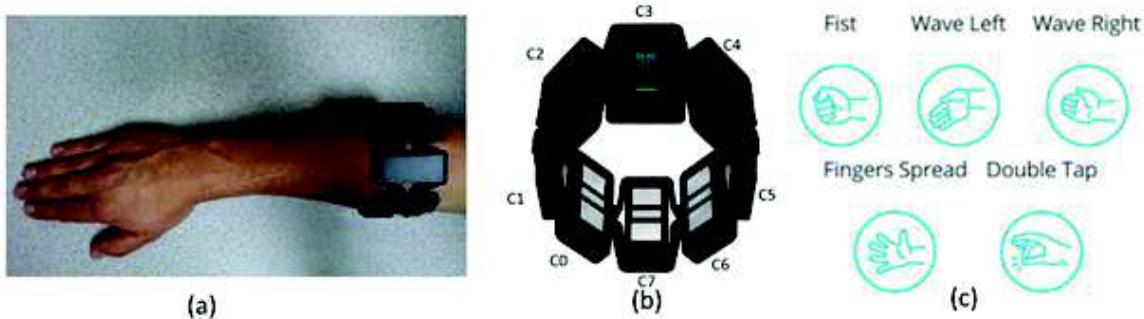


Figura 5. (a) Dispositivo Myo Armband (b) Myo Armband y sus ocho canales. (c) Gestos que actualmente reconoce el software privado del Myo Armband.

2.2. Métodos

El modelo propuesto para el reconocimiento de gestos de la mano sigue el flujo ilustrado en la **Figura 6**. El proceso consta de cinco bloques: adquisición de datos, preprocesamiento, extracción de características, clasificación y postprocesamiento. En el modelo propuesto se aplica un enfoque de ventana deslizante y se aplican estos cinco bloques a cada segmento de la señal vista a través de una ventana principal.

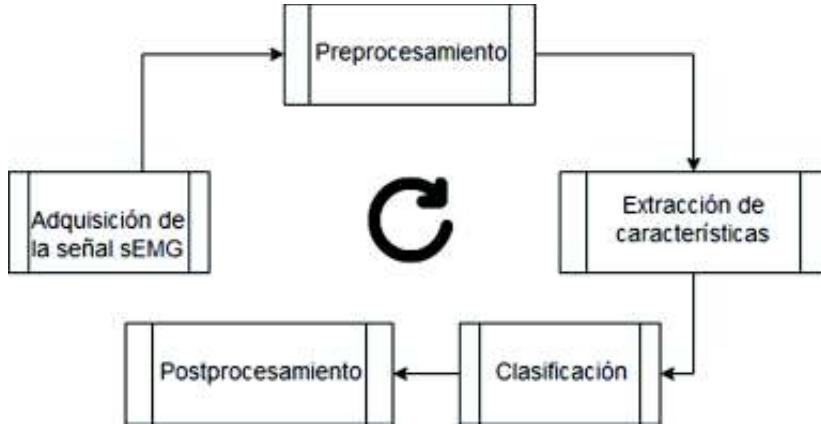


Figura 6. Bloques del modelo propuesto para el reconocimiento de gesto.

2.2.1. Notación

En este documento, las matrices son denotadas con letras mayúsculas en negrita, por ejemplo \mathbf{A} . Los vectores son denotados con letras minúsculas en negrita, por ejemplo \mathbf{x} . Las constantes son denotadas con letras mayúsculas (e. g., N) y los índices son denotados con letras minúsculas en cursiva (e. g., i).

2.2.2. Adquisición de datos

En este bloque, se aplicó el enfoque de ventana deslizante utilizando una ventana principal de longitud N . Se representó a la señal sEMG adquirida con el *Myo Armband* y vistas a través de la ventana principal con una matriz \mathbf{A} de tamaño $N \times 8$, donde 8 es el número de canales del *Myo Armband*. El valor $A_{i,j}$ representa la medida en el instante i por el sensor j , donde $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, 8$, respectivamente. Cada elemento de la matriz \mathbf{A} está en el rango de $[-1, 1]$. Para generar el vector de características para entrenar el modelo, se utilizó una ventana principal MW_{train} de longitud $N_{train} = 400$. La longitud de la ventana principal MW_{train} corresponde a la cantidad de puntos medidos en cada repetición del conjunto de entrenamiento. Para la validación preliminar y el testeo final del modelo, se utilizó una ventana principal MW_{test} de longitud $N_{test} = 200$, con un paso de 20 puntos entre dos ventanas consecutivas.

2.2.3. Preprocesamiento

Las señales sEMG pueden ser modeladas como un proceso estocástico no-estacionario [12]. Esto significa que la distribución de probabilidad de las sEMG cambia con el tiempo. Sin embargo, se puede reducir la no-estacionaridad de las sEMG suavizando sus valores. La idea de este proceso es reducir los cambios de la distribución de probabilidad de la señal sEMG en el tiempo [16]. En este trabajo, para suavizar la señal sEMG se aplica rectificación y filtrado. El preprocesamiento comienza con la rectificación de la señal

utilizando la función valor absoluto. Después, se filtra aplicando un *filtro pasa bajos de Butterworth* ψ de cuarto orden y con una de frecuencia de corte 5 Hz (i. e., frecuencia de normalización de $0.05\pi \text{ rad/sample}$) [11].

Para el conjunto de entrenamiento se aplica una función de segmentación de actividad muscular Φ a la observación de la ventana principal. La función Φ retorna el índice inicial y final que contiene la actividad muscular dentro de la observación de MW_{train} . Esta función es utilizada para remover la cabeza y cola que hacen referencia a la posición de relajación de la mano para cada repetición en el conjunto de entrenamiento [12]. En la **Figura 7** (a) se puede ver las señales sin preprocesar y en la **Figura 7** (b) se ilustra las señales preprocesadas de una repetición del conjunto de entrenamiento de un voluntario.

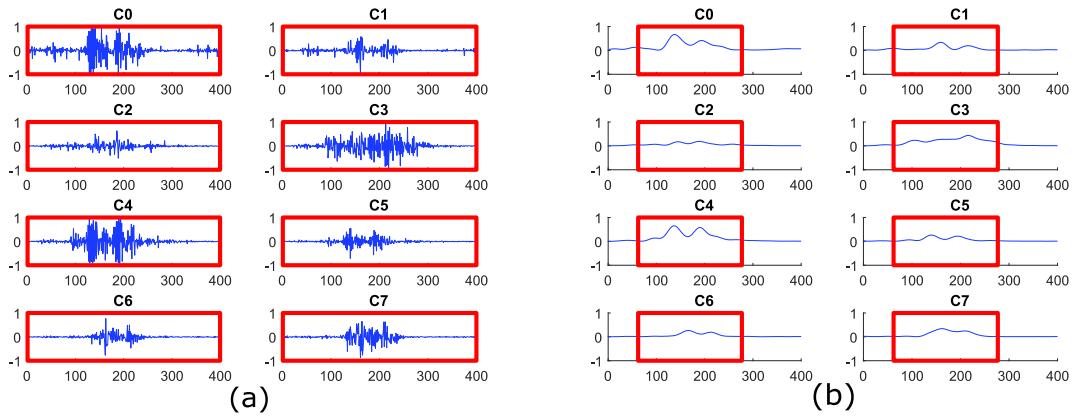


Figura 7. (a) Diagrama de las señales sEMG sin preprocesar. (b) Diagrama de las señales sEMG después de ser preprocesadas.

En cambio, para el conjunto de datos de testeo se aplica una función de verificación de actividad muscular Ω a la observación de la ventana principal. La función Ω está descrita en la Ecuación (1), donde $\mathbf{C} = \text{abs}(\mathbf{A})$ y $\tau_{preprocessing}$ es un umbral. Si $\Omega(\mathbf{C})$ es verdadero, entonces el proceso de reconocimiento continúa, caso contrario la respuesta retornada es *No Gesture* para la observación de la ventana principal MW_{test} .

$$\Omega(\mathbf{C}) = \sum_{j=1}^8 \sum_{i=1}^N \mathbf{c}_{ij} > \tau_{preprocessing}$$

Ecuación 1: Función de verificación de actividad muscular.

La función Φ se aplica solo al conjunto de entrenamiento porque retorna los límites en los cuales hay actividad muscular. Y adicionalmente, Φ permite descartar los segmentos de la señal que hacen referencia a la mano en la posición de relajación. Por otro lado, la función Ω solamente verifica si es que existe o no actividad muscular en la observación de una

ventana principal. Se probó diferentes umbrales y se encontró que $\tau_{preprocessing} = 0.39$ da los mejores resultados en el conjunto de validación preliminar.

2.2.4. Extracción de características

Para este bloque, se utilizó una sub-ventana SW para observar un segmento de la señal vista a través de la ventana principal MW (**Figura 8 (a)**). El segmento de la señal vista a través de SW es representada en una matriz \mathbf{E} de tamaño $M \times 8$; mientras, la señal observada a través de MW es representada en una matriz \mathbf{A} de tamaño $N \times 8$, donde $N > M$. Se utilizó un paso de un punto entre dos sub-ventanas consecutivas (**Figura 8 (c)**).

Las características que se tomaron para el modelo provienen de dos fuentes diferentes: los valores de la señal preprocesada y los resultados de aplicar un conjunto de funciones a los valores en bruto de la señal. Solo se utilizó funciones provenientes en el dominio del tiempo ya que el uso de funciones en el dominio de la frecuencia y tiempo-frecuencia incrementa el costo computacional de este bloque. En los siguientes pasos el índice i representa el i -ésimo instante en el cual se midió la señal sEMG vista a través de la ventana principal MW. Se aplica los siguientes pasos para extraer los vectores de características:

1. Alinear el primer punto de la sub-ventana SW con el punto $i = 1$ de la señal sEMG vista a través de la ventana principal MW.
2. Preprocesar la observación de la sub-ventana \mathbf{E} para obtener $\mathbf{F} = \psi(\text{abs}(\mathbf{E}))$. Convertir la matriz \mathbf{F} en el vector de características \mathbf{v}_i mediante la concatenación horizontal de todas las filas de \mathbf{F} .
3. Aplicar el conjunto de funciones descritas más adelante a los valores *crudos* de \mathbf{E} para obtener el vector de características \mathbf{z}_i .
4. Concatenar \mathbf{v}_i con \mathbf{z}_i de forma horizontal para generar el vector de \mathbf{x}_i .
5. Mover el primer punto de la sub-ventana SW al instante $i := i + 1$ y repetir los pasos del 2) al 5) hasta que $i = N - M + 1$.

El proceso para la extracción de características está ilustrado en la **Figura 8 (c)**. Cada \mathbf{x}_i es de longitud $|\mathbf{v}_i| + |\mathbf{z}_i|$ (donde $|\mathbf{x}|$ denota la longitud del vector \mathbf{x}) y está asociado con una etiqueta y_i correspondiente al gesto de una repetición. Empíricamente, se encontró que una sub-ventana de tamaño $M = 75$ da los mejores resultados de precisión de reconocimiento sobre el conjunto de validación preliminar. La longitud del vector \mathbf{v} es $|\mathbf{v}| = M * 8 = 75 * 8 = 600$. Esto significa que el vector \mathbf{v} tiene 600 características que son parte del vector \mathbf{x} . La aplicación del conjunto de funciones genera un vector \mathbf{z} cuyo tamaño es

$|z| = 56$, que son las características que aporta z al vector x . Por lo tanto, el vector de características x contiene $|x| = 600 + 56 = 656$ características.

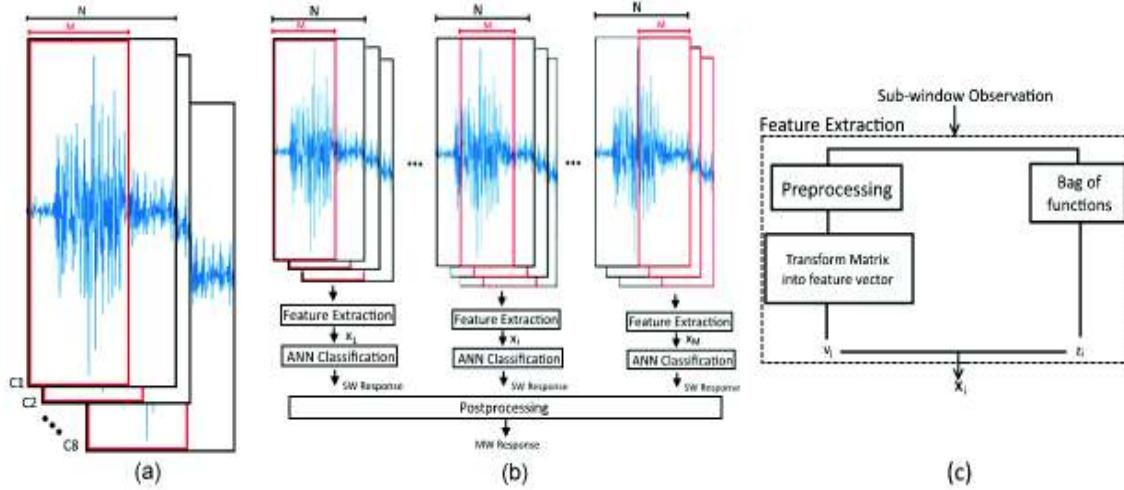


Figura 8. (a) Observación de la ventana principal MW de tamaño N y de la sub-ventana SW de tamaño M. (b) Movimiento de la sub-ventana dentro de la ventana principal para la extracción de característica, clasificación y postprocesamiento. (c) Proceso para generar un vector de características a partir de la observación de una sub-ventana.

El conjunto de funciones está conformado por: el Valor Absoluto Medio, el Cambio de Signo de Pendiente, la Longitud de forma de Onda, la Raíz Cuadrada Media y los Parámetros de *Hjorth* que son Actividad, Movilidad y Complejidad.

El Valor Absoluto Medio (MAV, por sus siglas en inglés *Mean Absolute Value*) es el valor promedio de toda la señal rectificada [22].

$$\text{MAV}(\mathbf{x}_k) = \frac{1}{M} \sum_{k=1}^M |\mathbf{x}_{i,k}|$$

Ecuación 2: Valor Absoluto Medio (MAV)

Los Cambios de Signo de Pendiente (SSC, por sus siglas en inglés *Slope Sign Changes*) mide la frecuencia de cambio del signo de la pendiente de la señal. Dadas tres muestras consecutivas $\mathbf{x}_{i-1,k}, \mathbf{x}_{i,k}, \mathbf{x}_{i+1,k}$, el valor del SSC es incrementado en uno si: $(\mathbf{x}_{i,k} - \mathbf{x}_{i-1,k}) * (\mathbf{x}_{i,k} - \mathbf{x}_{i+1,k}) \geq \epsilon$. Donde $\epsilon \geq 0$, es empleado para reducir el impacto de ruido en esta característica [22].

$$\text{SSC}(\mathbf{x}_k) = \sum_{i=1}^M f[(\mathbf{x}_{i,k} - \mathbf{x}_{i-1,k}) * (\mathbf{x}_{i,k} - \mathbf{x}_{i+1,k})]$$

$$f(v) = \begin{cases} 1, & \text{si } v \geq \epsilon \\ 0, & \text{otro caso} \end{cases}$$

Ecuación 3: Cambios de Signo de Pendiente (SSC).

La Longitud de forma de Onda (WL, por sus siglas en inglés *Waveform Length*) es una variación acumulativa que indica el grado de variaciones de las señales EMG. En la Ecuación 4 se ilustra la fórmula de cálculo de esta característica.

$$WL(\mathbf{x}_k) = \sum_{i=1}^M |\mathbf{x}_{i+1,k} - \mathbf{x}_{i,k}|$$

Ecuación 4: Longitud de forma de Onda (WL).

La Raíz Cuadrada Media (RMS, por sus siglas en inglés *Root Mean Square*) refleja el valor promedio de la señal. La RMS es calculada de la siguiente forma:

$$RMS(\mathbf{x}_k) = \sqrt{\frac{1}{M} \sum_{k=1}^M |\mathbf{x}_{i,k}|^2}$$

Ecuación 5: Raíz Cuadrada Media (RMS).

Los Parámetros de *Hjorth* son un conjunto de tres características: Actividad, Movilidad y Complejidad. Usualmente utilizados para la extracción de características de señales encefalográficas. La descripción de cada parámetro se encuentra en la **Tabla 2**.

Parámetro	Significado	Fórmula
Actividad	Representa el poder de la señal como la varianza de una función del tiempo.	$Actividad(\mathbf{x}_k) = \frac{1}{M} \sum_{k=1}^M (\mathbf{x}_{i,k} - \bar{\mathbf{x}}_k)^2$
Movilidad	Representa la frecuencia media de la serie de tiempo.	$Movilidad(\mathbf{x}_k) = \sqrt{\frac{Actividad(\mathbf{x}'_k)}{Actividad(\mathbf{x}_k)}}$
Complejidad	Indica la desviación de la pendiente y puede ser vista como una medida del cambio en la frecuencia de la señal.	$Complejidad(\mathbf{x}_k) = \frac{Movilidad(\mathbf{x}'_k)}{Movilidad(\mathbf{x}_k)}$

Tabla 2. Tabla con la descripción de los parámetros de Hjorth

Cabe mencionar que se utilizó la técnica *t-Distributed Stochastic Neighbor Embedding* (t-SNE) para visualizar cómo los vectores de entrenamiento de un usuario y de cada clase

(gesto) son agrupados en el espacio de características. Se probó diferentes tamaños de sub-ventana sobre el conjunto de entrenamiento de un usuario y se aplicó t-SNE a cada conjunto de vectores de entrenamiento. Los resultados se muestran en la **Figura 9**.

Se puede observar que cuando la longitud de la sub-ventana es menor, la sobreposición de los vectores de diferentes gestos es mayor, lo cual implica que su clasificación es más compleja. Por otro lado, cuando la longitud de la sub-ventana se incrementa, los vectores de características proyectados dentro de cada clase se acercan entre ellos y la clasificación se facilita. Sin embargo, al incrementar la longitud de la sub-ventana, el total de vectores de características de una repetición disminuye y la cantidad de características incrementa. Este efecto causa que el modelo de reconocimiento tienda a overfitting.

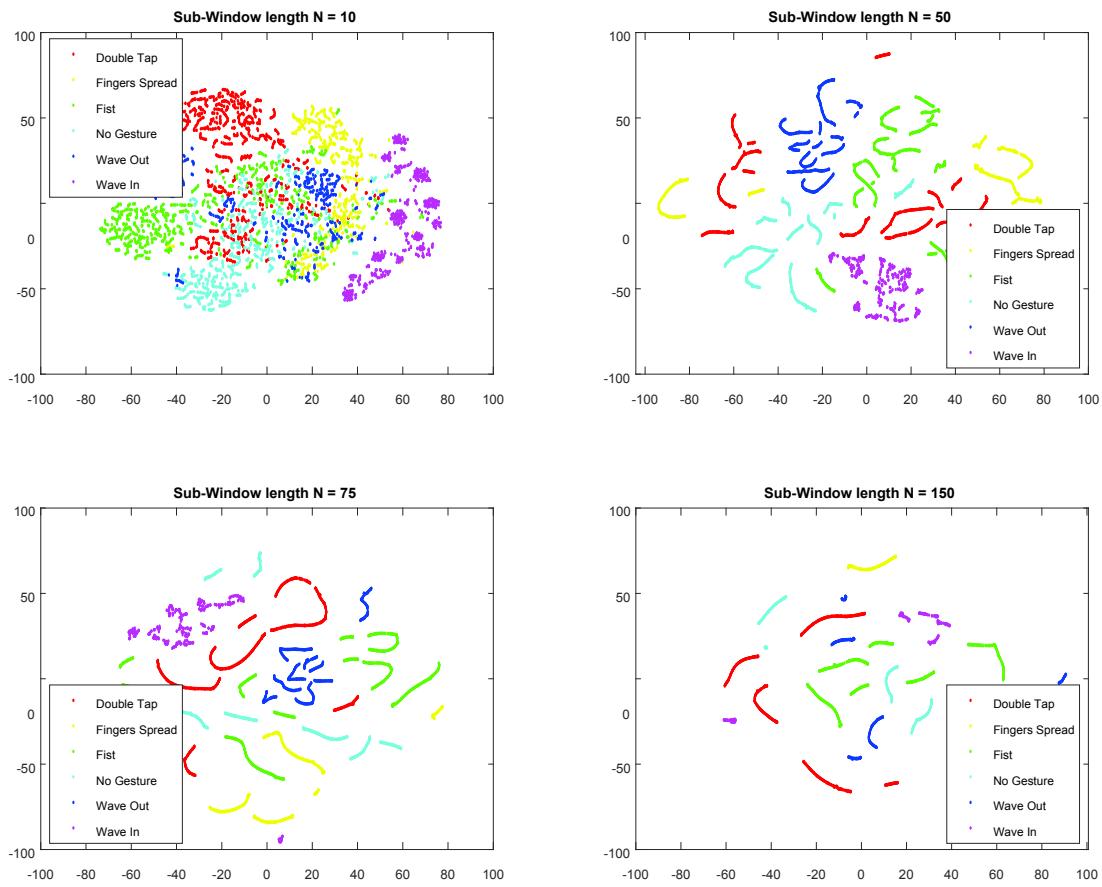


Figura 9. TSNE aplicado a diferentes vectores de características con diferentes tamaños de sub-ventana.

2.2.5. Clasificación

Para clasificación se utilizó una red artificial neuronal (ANN por sus siglas en inglés *Artificial Neuronal Network*) porque esta familia de funciones es un aproximador universal [23]: una red neuronal *feedforward* de tres capas (entrada, oculta y de salida), con funciones de

transferencia sigmoideas y un número apropiado de nodos en la capa oculta es capaz de aproximar cualquier función. Para nuestro modelo, la capa de entrada de la red tiene 656 nodos, los cuales corresponden a la longitud de los vectores de características. Para definir los nodos en la capa oculta se probó experimentalmente con un diferente número de nodos. Los mejores resultados en el conjunto de validación preliminar se obtuvieron con 328 nodos en la capa oculta, lo cual corresponde a la mitad de los nodos en la capa de entrada. La capa de salida solamente tiene 6 nodos, que corresponden a la cantidad de clases a predecir. Se probó las siguientes funciones de transferencia sigmoideas en la capa oculta: *logsig* y *tanh*. Se obtuvo los mejores resultados en el conjunto de validación con la función de transferencia *tanh* y con una función de costo con regularización. Adicionalmente, se aplicó un escalamiento de características utilizando la función: $\mathbf{x}' = (\mathbf{x}_i - \bar{\mathbf{x}}) ./ \sigma$, donde $\bar{\mathbf{x}}$ es un vector con los valores de los promedios de cada característica del vector \mathbf{x}_i , y σ es un vector con los valores de las desviaciones estándar de cada característica del vector \mathbf{x}_i , y $./$ representa la división elemento a elemento entre dos vectores.

2.2.6. Postprocesamiento

Por cada observación de la señal sEMG utilizando la ventana principal, se obtiene un vector de etiquetas donde cada etiqueta corresponde al vector de características de una observación de sub-ventana. Con la definición de un umbral $\tau_{postprocessing}$ y aplicando votación por mayoría simple se asigna una etiqueta a la observación de la ventana principal. Se asigna aquella etiqueta que tiene mayor número de ocurrencias y sobrepasa el umbral $\tau_{postprocessing}$ en el vector de etiquetas de la ventana principal. Si ninguna etiqueta cumple estas condiciones, se asigna la etiqueta *No Gesture* a la observación de la ventana principal. Se realizaron pruebas con diferentes umbrales y los mejores resultados en el conjunto de validación preliminar se obtuvieron con $\tau_{postprocessing} = 70\%$.

3. RESULTADOS Y DISCUSIÓN

3.1 Metodología de Evaluación

Para evaluar la precisión de reconocimiento del modelo propuesto, primero se entrenó un modelo por cada voluntario utilizando su correspondiente conjunto de entrenamiento. Después, se utilizó el modelo entrenado para predecir las etiquetas de cada repetición en el conjunto de testeо (30 repeticiones por clase) utilizando una ventana principal de longitud $N_{test} = 200$ con un paso de 20 puntos entre dos ventanas consecutivas. La aplicación del enfoque de ventana deslizante a cada repetición del conjunto de testeо retorna un vector con $(1000 - 20)/200 = 40$ etiquetas. Recordemos que la longitud de cada repetición en el conjunto de testeо es de alrededor de 1000 puntos. Un reconocimiento se considera

exitoso cuando todas las etiquetas en el vector de respuestas que sean diferentes de la clase *No Gesture* coinciden con la clase correspondiente de la repetición. De lo contrario, el reconocimiento en la repetición se considera erróneo y la etiqueta de respuesta para estos casos es la primera etiqueta en el vector diferente de *No Gesture* [11].

La validación preliminar se realizó sobre el conjunto de entrenamiento (5 repeticiones por cada clase) sin incluir las repeticiones de la mano en la posición de relajación. Se aplicó la técnica de validación cruzada debido a que el conjunto de entrenamiento es pequeño. En cada evaluación se utilizó una combinación de 2 repeticiones de cada clase para entrenar el modelo por usuario y 3 muestras restantes de cada clase para evaluarlo. Por lo tanto, se realizaron 10 evaluaciones por cada usuario (subconjuntos de 2 repeticiones tomadas del grupo de 5 repeticiones). El resultado de la evaluación preliminar fue el promedio de todos los resultados de cada evaluación.

Para obtener el tiempo de respuesta de los modelos evaluados, se utilizó un computador de escritorio con un procesador de 4GHz de velocidad y 16 GB de memoria RAM. El tiempo promedio reportado en este documento es la media de todos los tiempos que tardó el modelo en clasificar cada observación de la ventana principal de 200 puntos.

3.2 Resultados

En adición al modelo propuesto, se evaluó (con la misma metodología de evaluación) un modelo basado solamente en los valores de la señal preprocesada (rectificación y filtrado pasa bajos) y un modelo basado solamente en los resultados del conjunto de funciones para comparar los resultados de cada conjunto de características. También se evaluó un modelo generalizado, basado en las características del modelo propuesto, para comprobar si el modelo puede funcionar para todos los usuarios. Recordemos que los vectores de características del modelo propuesto son una combinación de los valores de la señal preprocesada y los resultados de la aplicación del conjunto de funciones.

3.2.1. Modelo propuesto

En las matrices de confusión de la **Figura 10** se muestran los resultados de la evaluación del modelo sobre el conjunto de entrenamiento para validación (**Figura 10 (a)**) y de testeo (**Figura 10 (b)**). El modelo muestra una exactitud de reconocimiento del 90.7% sobre el conjunto de testeо. La exactitud de reconocimiento, que se calcula con la fórmula de la **Ecuación 6**, es igual a la relación entre la suma de verdaderos positivos (VP) y verdaderos negativos (VN) sobre el número de casos positivos (P) más los negativos (N). La exactitud en la matriz de confusión de la **Figura 10** es igual a la suma de los valores enteros que se encuentran en la columna principal dividido entre el número total de ejemplos. En el conjunto

de testeo todos los casos son positivos y el número total de ejemplo es igual a 1500 ya que son 30 repeticiones de 5 gestos de 10 usuarios. Por lo tanto, la exactitud de reconocimiento es: $(295 + 272 + 273 + 265 + 256) / 1500 = 0.907$.

$$\text{Exactitud} = \frac{VP + VN}{P + N}$$

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

En la matriz de confusión del conjunto de datos de testeo (**Figura 10 (b)**) se observa que la clase *Fist* obtuvo la mayor sensibilidad (98.3%) y la clase *Double Tap* fue la que obtuvo la menor sensibilidad (85.3%). La sensibilidad, que se calcula con la **Ecuación 7**, es la razón entre los verdaderos positivos y la suma de verdaderos positivo con los falsos negativos. Por ejemplo, la sensibilidad de la clase *Fist* es: $295 / (295 + 5) = 0.983$.

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

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

La matriz de confusión, del conjunto de datos de testeo (**Figura 10 (b)**), muestra que la precisión de la clase *Wave Out* obtuvo los mayores resultados (99.6%) y la clase *Fist* fue aquella con el menor resultado (86.8%). La precisión, que se calcula con la **Ecuación 8**, es la razón entre los verdaderos positivos sobre la suma de verdaderos positivo más falsos positivos. Por ejemplo, la clase *Wave Out* tiene una precisión de: $273 / (273 + 1) = 0.996$.

$$\text{Precisión} = \frac{VP}{VP + FP}$$

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

Por lo tanto, el modelo propuesto es más probable que reconozca la clase *Fist* que la clase *Double Tap*. Sin embargo, el modelo es más probable de reconocer **correctamente** a la clase *Wave Out* e **incorrectamente** a la clase *Fist*. Las matrices de confusión incluyen reconocimientos de la clase *No Gesture* debido a que en algunas repeticiones no se sobrepasaron los umbrales de preprocesamiento (para verificar si existe actividad muscular) o de postprocesamiento (para procesar los resultados de la red neuronal).

Cabe mencionar, que el porcentaje que se encuentra en cada casilla es la razón entre el número de observaciones y el total de ejemplos de prueba. Por ejemplo, la casilla que indica las comparaciones de la clase *Fist* contra sí misma (casilla de color verde en la fila de la clase *Fist*) tiene 295 observaciones y el porcentaje es igual a $295 / 1500 = 0.197$.

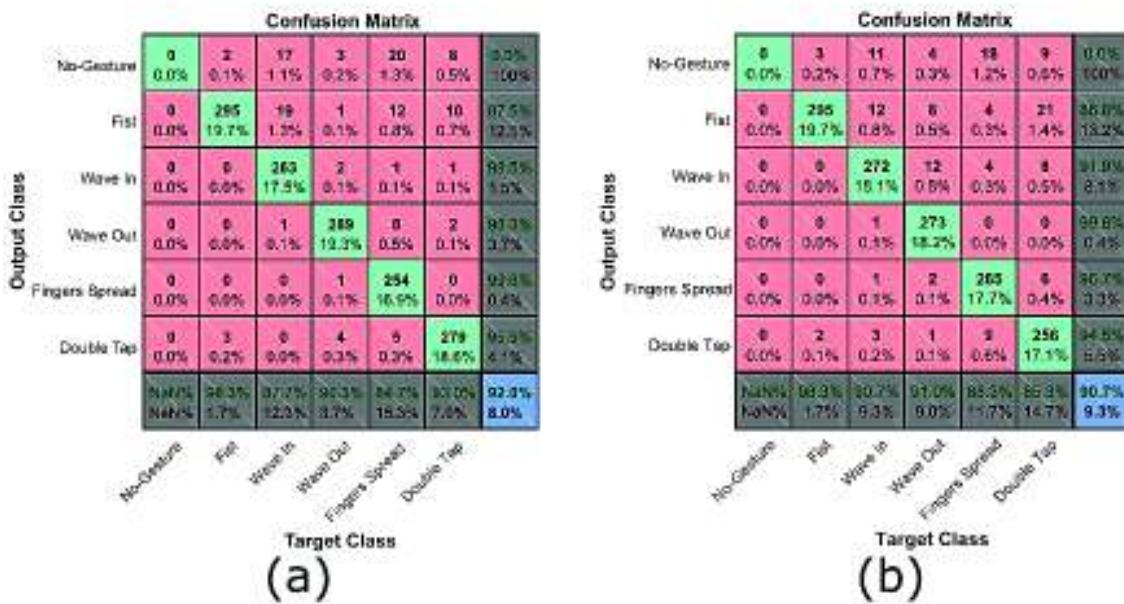


Figura 10. Matrices de confusión de la evaluación del modelo propuesto sobre en el conjunto de entrenamiento (a) y testeo (b).

3.2.2. Modelo basado solo en los valores de la señal preprocesada

Este modelo toma los valores de la señal preprocesada y los transforma en un vector de 600 características. La matriz de confusión de la evaluación del modelo basado solo en los valores de la señal preprocesada sobre el conjunto de entrenamiento **Figura 11 (a)** y testeo **Figura 11 (b)**. En la matriz de confusión de la **Figura 11 (b)** se muestra que este modelo alcanza un 88.3% de tasa de reconocimiento en el conjunto de testeo. El tiempo de respuesta de este modelo es en promedio 2.59 ms y su tiempo de entrenamiento es de alrededor de 29 segundos.

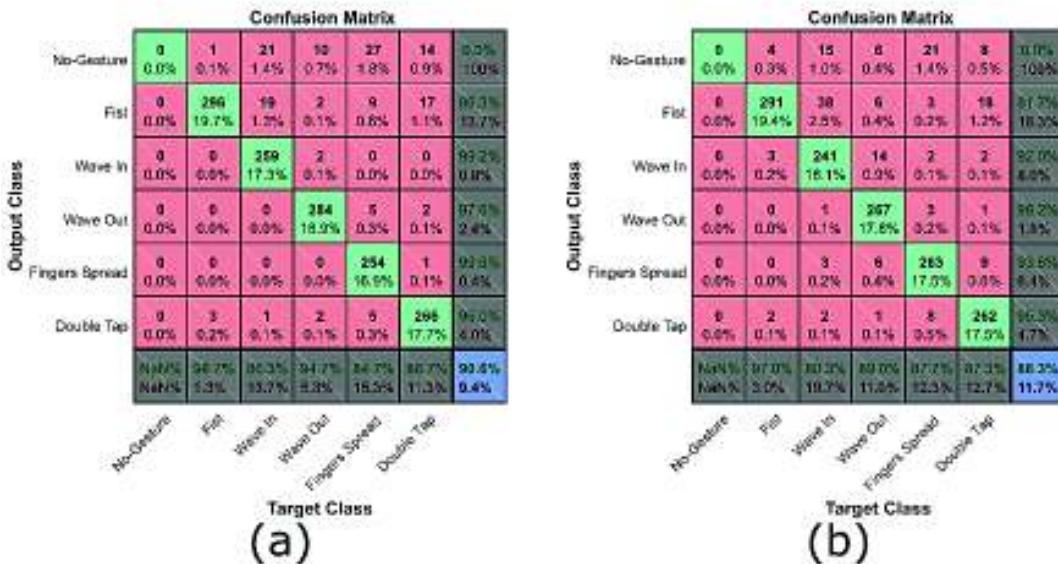


Figura 11. Matrices de confusión de la evaluación del modelo basado solo en los valores de la señal preprocesada sobre el conjunto de entrenamiento (a) y testeo (b).

3.2.3. Modelo basado solo en el resultado del conjunto de funciones

Este modelo toma los resultados de la aplicación del conjunto de funciones para formar un vector de 56 características. Las matrices de confusión de la evaluación de este modelo se encuentran en la **Figura 12**. La matriz de confusión de la **Figura 12 (b)** muestra que este modelo alcanza un 86.1% de tasa de reconocimiento en el conjunto de testeo. El tiempo de respuesta de este modelo es en promedio 26.52 ms y su tiempo de entrenamiento es de alrededor de 2 segundos.

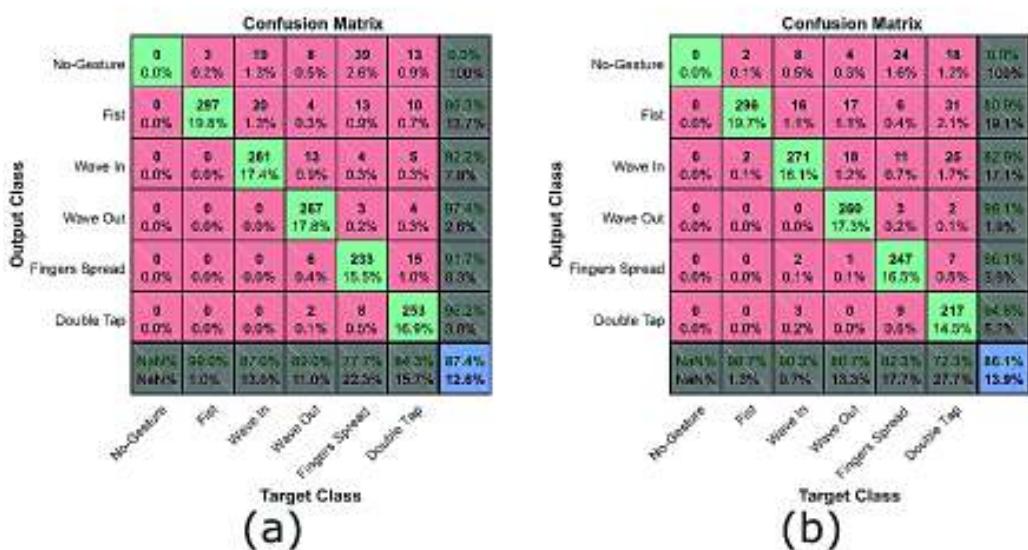


Figura 12. Matrices de confusión de la evaluación del modelo basado solo en conjunto de funciones sobre conjunto de entrenamiento (a) y testeo (b).

3.2.4. Modelo generalizado

El modelo generalizado propuesto aplica los mismos bloques de adquisición de datos, preprocesamiento, extracción de características y postprocesamiento del modelo propuesto. Se empleó el método *Stochastic Gradient Descent* para realizar un entrenamiento por lotes. Un lote consiste en los datos de testeo (30 repeticiones por clase) de cada usuario transformados en un conjunto de vectores de características. Para entrenar y validar el modelo se realizó una comparación de *uno contra nueve*. El modelo se entrenó con los datos de testeo de nueve usuarios y se evaluó con los datos de testeo del usuario restante. Se realizó la evaluación de uno contra nueve por cada usuario en el conjunto de datos. El modelo tarda alrededor de 18 minutos para ser entrenado y su tiempo de respuesta es similar al del modelo propuesto. Este modelo se validó preliminarmente sin validación cruzada y utilizando las 5 repeticiones de cada gesto en el conjunto de entrenamiento, sin las repeticiones de la mano en la posición de relajación (resultados en la **Figura 13 (a)**). La tasa de reconocimiento del modelo generalizado es de 79.9% sobre el conjunto de testeo (resultados en la **Figura 13 (b)**).

Confusion Matrix							
Output Class	No-Gesture	Fist	Wave In	Wave Out	Fingers Spread	Double Tap	
	No-Gesture	0 0.0%	5 2.0%	11 4.4%	8 3.4%	9 3.0%	1 0.4%
Fist	0 0.0%	42 16.6%	5 2.0%	0 0.0%	5 2.0%	5 2.4%	72.4% 27.6%
Wave In	0 0.0%	2 0.8%	34 13.0%	1 0.4%	0 0.0%	4 1.6%	62.3% 17.1%
Wave Out	0 0.0%	1 0.4%	0 0.0%	43 17.2%	5 2.0%	0 0.0%	67.3% 12.2%
Fingers Spread	0 0.0%	0 0.0%	0 0.0%	0 0.0%	31 12.4%	0 0.0%	100% 0.0%
Double Tap	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	39 15.8%	100% 0.0%
Total	144 54.0%	181 67.0%	186 67.1%	102 36.2%	102 36.2%	76 27.0%	75.8% 24.4%

Confusion Matrix							
Output Class	No-Gesture	Fist	Wave In	Wave Out	Fingers Spread	Double Tap	
	No-Gesture	0 0.0%	3 0.2%	42 2.8%	4 0.3%	52 3.5%	24 1.6%
Fist	0 0.0%	288 17.9%	33 2.2%	1 0.1%	39 2.6%	37 2.5%	72.6% 22.4%
Wave In	0 0.0%	26 1.7%	225 15.0%	17 1.1%	4 0.3%	23 1.5%	76.0% 23.7%
Wave Out	0 0.0%	3 0.2%	0 0.0%	276 18.4%	0 0.0%	0 0.0%	98.9% 1.1%
Fingers Spread	0 0.0%	0 0.0%	0 0.0%	0 0.0%	214 14.3%	0 0.0%	100% 0.0%
Double Tap	0 0.0%	0 0.0%	0 0.0%	2 0.1%	0 0.0%	216 14.4%	99.1% 0.9%
Total	144 54.0%	181 67.0%	186 67.1%	102 36.2%	102 36.2%	76 27.0%	75.8% 24.4%

(a) (b)

Figura 13. Matrices de confusión de la evaluación del modelo generalizado evaluado sobre el conjunto de entrenamiento (a) y testeo (b).

Los resultados de evaluar el modelo individualmente (por cada usuario) muestran una tasa de reconocimiento variable (**Figura 14 (a)**). Es decir, el modelo generalizado reconoce correctamente los gestos solo en algunos usuarios. Por otro lado, el modelo propuesto muestra una tasa de reconocimiento con poca variación al evaluar el modelo individualmente sobre los datos de testeo (**Figura 14 (b)**).

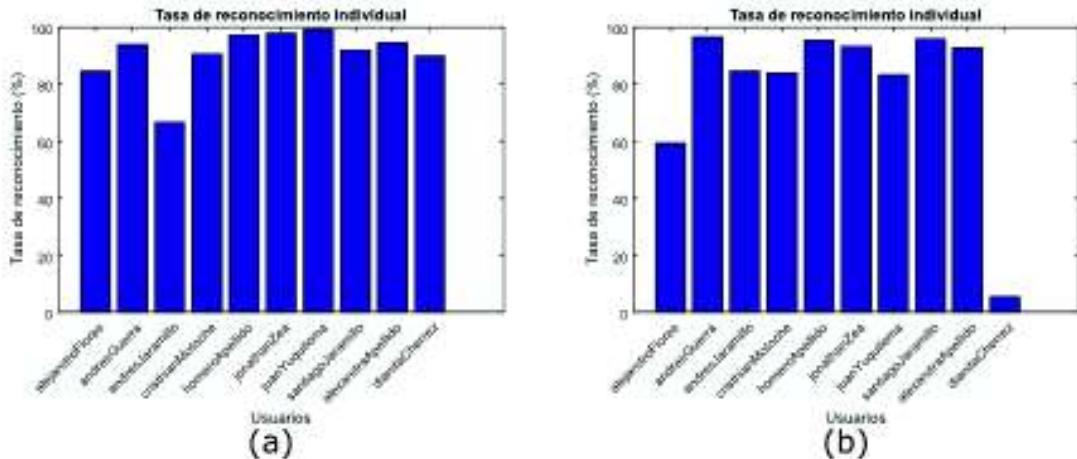


Figura 14. (a) Resultados de la evaluación individual con el modelo generalizado propuesto. (b) Resultados de la evaluación individual con el modelo propuesto.

3.3 Discusión

En el **ANEXO III** se encuentra el resumen de los resultados de los modelos que se han evaluado sobre el conjunto de entrenamiento y testeо utilizados en este y otros trabajos.

El **ANEXO III** muestra que el modelo propuesto, que utiliza ambos tipos de características (los valores de la señal preprocesada y los resultados del conjunto de funciones), obtuvo la mayor precisión de reconocimiento comparado con los demás modelos. También se puede observar que el modelo que solamente utiliza los valores de la señal preprocesada responde más rápido que los demás modelos evaluados. Adicionalmente, se puede observar que el modelo que solamente utiliza los resultados del conjunto de funciones obtuvo el menor tiempo de entrenamiento debido a que su arquitectura es menos compleja. La arquitectura de la red neuronal del modelo propuesto es más compleja y su tiempo de entrenamiento y respuesta es mayor en comparación a los otros modelos evaluados. Sin embargo, se mantiene como un modelo con respuestas en tiempo real y de alta precisión.

El **ANEXO III** también indica que el modelo propuesto responde en un promedio de 29.38 ms que es mucho menor que el valor definido para tiempo real (300 ms). Adicionalmente, el **ANEXO III** muestra que el modelo evaluado es más rápido que los modelos que utilizan el algoritmo *Dynamic Time Warping* (DTW) con el clasificador *k Nearest Neighbor* (kNN) ya que la extracción de características y la clasificación realizada por las redes neuronales artificiales es menos costosa computacionalmente.

4. CONCLUSIONES

En este documento, se ha presentado un modelo de reconocimiento de gestos de la mano humana basado en señales sEMG del antebrazo. El modelo es entrenado individualmente para cada usuario y requiere de cinco repeticiones por cada clase a reconocer. El modelo responde en un promedio de 29.38 ms, que es más bajo que el límite definido para tiempo real (300 ms), utilizando un computador con recursos limitados de memoria *RAM* y de procesamiento. Además, el modelo propuesto muestra una precisión de reconocimiento de 90.7% que es más alta que la tasa descrita en el estado del arte (85%).

Para este modelo, se aplicó un enfoque de ventana deslizante utilizando una ventana principal y una sub-ventana. La sub-ventana permite observar un segmento de la señal vista a través de la ventana principal. El modelo está compuesto de cinco bloques: adquisición de datos, preprocesamiento, extracción de características, clasificación y postprocesamiento. Para la adquisición de datos, se utilizó el dispositivo *Myo Armband* para medir la señal sEMG del antebrazo. Para el preprocesamiento, se rectificó, filtró y detectó la actividad muscular en la observación de la ventana principal. Para la extracción de características, se utilizó dos conjuntos de características: los valores de la señal preprocesados y los resultados de aplicar un conjunto de funciones. Para clasificación, se utilizó una red neuronal *feedforward* de tres capas para etiquetar cada observación de sub-ventana. Finalmente, para el postprocesamiento se aplicó una votación por mayoría simple, en base a las etiquetas de cada observación de sub-ventana, para decidir el gesto final dentro de la ventana principal.

Se encontró que la precisión de reconocimiento del modelo propuesto mejora cuando se combina los valores de la señal preprocesados con los resultados de aplicar un conjunto de funciones. Este conjunto de características da una alta tasa de reconocimiento solo cuando el modelo se adapta a cada usuario y se proporciona al menos cinco repeticiones de cada clase a reconocer. Para un modelo generalizado, estas características permiten obtener una alta precisión de reconocimiento solamente en ciertos usuarios. El trabajo futuro incluye la definición de un modelo generalizado que funcione para todos los usuarios. Los modelos que se definan para trabajo futuro deberán alcanzar una alta tasa de precisión de reconocimiento, funcionar en tiempo real y utilizar un computador con recursos limitados de memoria *RAM* y de preprocesamiento.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Konar y S. Saha, Gesture Recognition. Principles, Techniques and Applications., Cham: Springer International Publishing, 2018.
- [2] Y. Xu y Y. Dai, «Review of hand gesture recognition study and application,» *Contemporary Engineering Sciences*, vol. 10, pp. 375-384, 2017.
- [3] S. Benatti, G. Rovere, J. Bösser, F. Montagna, E. Farella, F. Glaser, P. Schonle, T. Burger, S. Fateh, Q. Huang y L. Benini, «A sub-10mW real-time implementation for EMG hand gesture recognition based on a multi-core biomedical SoC,» de *Advances in Sensors and Interfaces (IWASI), 2017 7th IEEE International Workshop on*, Vieste, Italy, 2017.
- [4] L. Chen, F. Wang, H. Deng y K. Ji, «A Survey on Hand Gesture Recognition,» *2013 International Conference on Computer Sciences and Applications*, 2013.
- [5] R. Z. Khan y N. A. Ibraheem, «Survey on Various Gesture Recognition Technologies,» *International Journal of Computer Applications*, vol. 50, nº 7, pp. 38 - 44, 2012.
- [6] R. Pradipa y S. Kavitha, «Hand Gesture Recognition – Analysis of Various Techniques, Methods and Their Algorithms,» *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, nº 3, pp. 2003 - 2010, 2014.
- [7] I. Mesa, A. Rubio, J. Diaz, J. Legarda y B. Sedano, «Reducing the Number of Channels and Signals-Features for an Accurate Classification in an EMG Pattern Recognition Task,» *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pp. 38-28, 2012.
- [8] R. Ahsan, M. Ibn y O. Khalifa , «Electromyography (EMG) Signal based Hand Gesture Recognition using Artificial Neural Network (ANN),» de *4th International Conference on Mechatronics (ICOM)*, Kuala Lumpur, Malaysia, 2011.
- [9] R. Chowdhury, M. Reaz, A. Mohd, A. Bakar, C. Kalaivani y T. Chang, «Surface Electromyography Signal Processing and Classification Techniques,» *Sensors*, vol. 13, nº 12, pp. 12431-12466, 2013.
- [10] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu y J. Li, «Gesture recognition by instantaneous surface EMG images,» *Scientific Reports*, vol. 6, nº 1, pp. 1-8, 2016.

- [11] M. Benalcázar, A. G. Jaramillo, J. A. Zea, A. Paez y V. H. Andaluz, «Hand gesture recognition using machine learning and the Myo armband,» *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017.
- [12] M. Benalcázar, C. Motoche, J. Zea, A. Jaramillo, C. Anchundia, P. Zambrano, M. Segura, F. Palacios y M. Pérez, «Real-Time Hand Gesture Recognition Using the Myo Armband and Muscle Activity Detection,» *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 2017.
- [13] Z. Xu, C. Xiang, V. Lantz , W. Kong-qiao , W. Wen-hui y Y. Ji-hai, «Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors,» *Proceedingsc of the 13th international conference on Intelligent user interfaces - IUI '09*, pp. 401-405, 2008.
- [14] Y. Sang, L. Shi y Y. Liu, «Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing,» *arXiv:1712.00216v1*, 1 12 2017.
- [15] J. Basmajian y C. De Luca, *Muscle Alive: Their Function Revealed by Electromyography.*, Baltimore: Willians & Wilkins, 1985.
- [16] K. Peter, *The ABC of EMG. A Practical Introduction to Kinesiological Electromyography.*, Arizona, USA: Noraxon U.S.A., Inc., 2006.
- [17] J. Gordon, P. Desaix, E. Johnson, J. Johnson, O. Korol, D. Kruse, B. Poe, J. Wise, M. Womble y K. Young, «Anatomy & Physiology,» de *Anatomy & Physiology*, Houston, Texas, OpenStax, 2016, pp. 405-500.
- [18] J. Bronzino, *The Biomedical Engineering Handbook: Second Edition.*, Connecticut, USA: CRC Press LLC, 1999.
- [19] J. Rodriguez-Falces, J. Navallas y A. Mal, «EMG Modeling,» *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, pp. 3 - 36, 2012.
- [20] T. L. Inc, «Myo Thalmic Labs Inc,» Thalmic Labs Inc, 2013. [En línea]. Available: <https://www.myo.com/techspecs>. [Último acceso: 26 February 2018].
- [21] T. L. Inc, «Myo Support Thalmic Labs Inc,» Thalmic Labs Inc, 2017. [En línea]. Available: <https://support.getmyo.com/hc/en-us/articles/202536726-How-do-I->

access-the-raw-EMG-data-from-the-Myo-armband-. [Último acceso: 26 February 2018].

- [22] 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.
- [23] A. Farago y G. Lugosi, «Strong Universal Consistency of Neural Network Classifiers,» de *Information Theory, 1993. Proceedings. 1993 IEEE International Symposium on*, San Antonio, 1993.

6. ANEXOS

ANEXO I

El hipervínculo redirecciona a una carpeta compartida con el conjunto de datos utilizado:

- Original: https://epnecuador-my.sharepoint.com/:f/g/personal/cristhian_motoche_epn_edu_ec/Et1vkzdyxBxDiOYcEKTjGU4BBZXsCnKSU7QqiNY-skxQyQ?e=DBTTkD
- Enlace recortado temporal (válido hasta el 30 marzo del 2019):
<https://goo.gl/vHpwqd>

ANEXO II

El hipervínculo redirecciona a un directorio compartido con el código fuente desarrollado para implementar y evaluar el modelo:

- Original: https://epnecuador-my.sharepoint.com/:f/g/personal/cristhian_motoche_epn_edu_ec/EtxyWmU7fitPgsmLMtxfWaMB9pWAYr3-pH0UeMuBpRpwgw?e=CWWfc8
- Enlace recortado temporal (válido hasta el 30 marzo del 2019):
<https://goo.gl/cTSJbw>

ANEXO III

Modelo	Características	Precisión (%)		Tiempo de respuesta (ms)	Tiempo de entrenamiento (s)
		Entren.	Testeo		
Modelo propuesto	- Señal preprocesada y transformada en un vector de características	92.0	90.7	29.38	34.78
	- Resultados del conjunto de funciones				
Modelo que utiliza solo los valores de la señal preprocesados	- Señal preprocesada y transformada en un vector de características	90.6	88.3	2.59	29.71
	- Resultados del conjunto de funciones				
Modelo que solo usa los resultados del conjunto de funciones	- Señal preprocesada y transformada en un vector de características	87.4	86.1	26.52	2.08
	- Resultados del conjunto de funciones				
Modelo generalizado propuesto	- Señal preprocesada y transformada en un vector de características	75.6	79.9	28.87	1082.39
	- Resultados del conjunto de funciones				
Modelo que usa kNN con DTW [11]	- Distancia DTW sobre 5 señales de cada clase	-	86.1	245.5	-
	- Distancia DTW sobre 5 señales de cada clase				
Modelo que usa kNN con DTW y aplica detección de actividad muscular [12]	- Se desconoce las características que utiliza este modelo.	-	89.5	193.1	-
	- Se desconoce las características que utiliza este modelo.				
Sistema propietario del Myo Armband [12]	- Se desconoce las características que utiliza este modelo.	-	83.1	-	-

Nota: Estos modelos reconocen los mismos cinco gestos definidos y fueron evaluados con la misma metodología de evaluación sobre el mismo conjunto de datos.