

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

**DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL QUE
PERMITA LA CLASIFICACIÓN ENTRE PERSONAS QUE
PRESENTEN O NO UNA PATOLOGÍA A PARTIR DE SEÑALES
ELECTROMIOGRÁFICAS DE LA EXTREMIDAD INFERIOR
USANDO MÁQUINAS DE SOPORTE DE VECTORES**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERA EN ELECTRÓNICA Y CONTROL**

EMILIA ABIGAIL MEZA PAREDES

DIRECTOR: Dr. JORGE ANDRÉS ROSALES ACOSTA

CODIRECTOR: MSc. MARÍA FERNANDA TRUJILLO GUERRERO

Quito, noviembre 2021

AVAL

Certificamos que el presente trabajo fue desarrollado por Emilia Abigail Meza Paredes, bajo nuestra supervisión.



Firmado electrónicamente por:
**JORGE ANDRES
ROSALES ACOSTA**

Dr. JORGE ANDRÉS ROSALES ACOSTA
DIRECTOR DEL TRABAJO DE TITULACIÓN



Firmado electrónicamente por:
**MARIA FERNANDA
TRUJILLO
GUERRERO**

MSc. MARÍA FERNANDA TRUJILLO GUERRERO
CODIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo Emilia Abigail Meza Paredes, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado 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 dejo constancia de que la Escuela Politécnica Nacional podrá hacer uso del presente trabajo según los términos estipulados en la Ley, Reglamentos y Normas vigentes.



Emilia Abigail Meza Paredes

DEDICATORIA

Este proyecto de titulación, se lo quiero dedicar a Jehová Dios: mi padre, mi amigo y mi confidente, por brindarme la fuerza necesaria en todo momento, para no desistir y alcanzar esta meta tan anhelada. Todo lo que se haga que sea con el corazón, como si siguiéramos a Dios. Colosenses 3:23-24.

AGRADECIMIENTO

Quiero agradecer a Dios por brindarme salud y por sus infinitas bendiciones, porque nunca me desamparó en tiempos oscuros y temerosos.

A mis padres Jacqueline y Jorge por el regalo de la vida, en especial a mi madre por su esfuerzo y sacrificio diario, quiero que sepas que admiro tu entereza.

A mi querida Escuela Politécnica Nacional, la cual me acogió, me permitió formarme para ser una profesional y por darme la oportunidad de cumplir mi mayor sueño, ser una Ingeniera. En especial a la Facultad de Ingeniería Eléctrica y Electrónica por brindarme la mejor educación de la mano de docentes de calidad.

A mi director el Dr. Andrés Rosales, por su apoyo y orientación durante el planteamiento y ejecución del presente trabajo, agradezco su gran disposición y su interés en los estudiantes.

A mi codirectora a quien admiro, la MSc. María Fernanda Trujillo, que además de ser una gran maestra y de quien tuve la fortuna de tener su guía, sus consejos fueron primordiales para terminar la tesis, agradezco su paciencia y sus conocimientos que me ayudaron en cada paso que di durante este arduo camino.

A mi difunta abuelita, mi Luz Clarita, fue mi segunda madre y a quien extraño profundamente, le agradezco por sus consejos, su preocupación, su generosidad, te dedico mi éxito, este logro no lo hubiera conseguido sin tus bendiciones, desde el fondo de mi corazón espero donde quiera que estés te sientas orgullosa de tu nieta.

A mis tíos Fernando y Gladys, gracias por creer siempre en mí, por darme fortaleza cada que quise tirar la toalla, por decirme cada día que, si lo iba a lograr, sus palabras fueron el impulso que necesité para salir adelante, este logro también es de ustedes, gracias por su apoyo moral, económico y en todos los ámbitos de mi vida.

A mi familia, a mis tíos, mis primos y mis hermanas, especialmente a mi abuelita Rosita, gracias por ser tan unidos como familia a pesar de las adversidades, de quienes he recibido todo el cariño, durante toda la carrera.

A mi mejor amiga Maye, a mis compañeros de facultad, a mis amigos los cuales he ido cultivando desde el primer día de la universidad en prepo, algunos que conocí en los últimos semestres, a pesar de la distancia, de la pandemia o de estar en otras carreras

fueron mi bastón en los buenos y malos momentos. Con quienes pasamos noches completas estudiando, espero en un futuro nos reunamos como colegas y recordemos la maravillosa vida universitaria. Finalmente, a una persona muy especial para mi A Fernando C., por los bellos momentos compartidos.

ÍNDICE DE CONTENIDO

AVAL.....	I
DECLARACIÓN DE AUTORÍA	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO.....	VI
RESUMEN.....	VIII
ABSTRACT	IX
1 INTRODUCCIÓN.....	1
1.1 OBJETIVOS	2
1.2 ALCANCE	2
1.3 MARCO TEÓRICO.....	3
1.3.1 MIEMBRO INFERIOR.....	3
1.3.2 SEÑALES ELECTROMIOGRÁFICAS EMG	7
1.3.2.1 Tipos de señales EMG	7
1.3.2.2 Características y estados de una señal EMG.....	7
1.3.2.3 Patrones neuromusculares de una EMG patológica.	8
1.3.3 APRENDIZAJE AUTOMÁTICO.....	9
1.3.3.1 Aprendizaje supervisado	9
1.3.3.2 Técnicas de preprocesamiento	10
1.3.3.3 Extracción de características.....	11
1.3.3.4 Partición de datos.....	16
1.3.3.5 Métodos de validación	17
1.3.3.6 Máquina de soporte vectorial SVM.....	18
1.3.3.7 Métodos de Kernel	20
2 METODOLOGÍA.....	21
2.1 BASE DE DATOS.....	21
2.1.1 SENSOR COMERCIAL DATALOG MWX8 [27]	21
2.1.2 DISPOSICIÓN DE LA BASE DE DATOS.....	22
2.1.2.1 Formación de los archivos	23
2.2 PRE-PROCESAMIENTO	24
2.2.1 SEGMENTACIÓN	24
2.2.2 RECTIFICACIÓN Y FILTRADO	28
2.2.3 EXTRACCIÓN DE CARACTERÍSTICAS	28
2.3 CLASIFICACIÓN.....	30

3	RESULTADOS Y DISCUSIÓN.....	32
3.1	MÉTODOS DE EVALUACIÓN	32
3.2	RESULTADOS.....	33
3.2.1	RESULTADOS USANDO CARACTERÍSTICAS TRADICIONALES.....	33
3.2.1.1	Validación simple por exclusión	34
3.2.1.2	Validación Cruzada K-fold.....	35
3.2.2	RESULTADOS USANDO CARACTERÍSTICAS MEJORADAS	37
3.2.2.1	Validación simple por exclusión	38
3.2.2.2	Validación Cruzada K-fold.....	39
3.3	DISCUSIÓN.....	41
4	CONCLUSIONES Y RECOMENDACIONES.....	43
4.1	CONCLUSIONES.....	43
4.2	RECOMENDACIONES.....	44
5	REFERENCIAS BIBLIOGRÁFICAS	45
	ANEXOS	48

RESUMEN

Esta tesis propone una herramienta para estudiar la relación de la electromiografía superficial (sEMG) entre personas sanas y alguien con una patología en el miembro inferior. La Máquina de Soporte de Vectores (SVM) se utiliza para clasificar señales electromiográficas debido a que, los modelos que se obtienen son resistentes al sobreajuste. Para el análisis se ha tomado la base de Datos EMG Dataset de UCI Machine Learning, esta contiene señales de 11 sujetos con anomalía de rodilla y 11 normales, previamente diagnosticados por un profesional. Se someten a tres movimientos para analizar el comportamiento asociado con la extremidad inferior: la marcha, la extensión de la pierna desde una posición sentada y la flexión de la pierna hacia arriba. El análisis tiene etapas como preprocesamiento, extracción de características, entrenamiento y validación. Varias características de electromiografía convencionales se utilizan para realizar la comparación con un conjunto de características mejoradas. El proceso de entrenamiento se logró con el 60% de los datos y la validación del algoritmo con el 40% de ellos. Finalmente, a partir de la matriz de confusión, evaluada de los modelos finales, se calcularon métricas como Exactitud, Sensibilidad, Especificidad y Precisión para seleccionar el modelo con desempeño óptimo en la vida real.

PALABRAS CLAVE: aprendizaje de máquina, clasificación, electromiografía, extracción de características, preprocesamiento, máquina de soporte de vectores, electromiografía de superficie.

ABSTRACT

This thesis proposes a tool for studying the relationship of surface electromyography between healthy people and someone with some pathology in the lower limb. Support Vector Machine (SVM) is used to classify electromyographic signals because models are robust to overfitting. For the lower limb, analysis has been taken the EMG Dataset from UCI Machine Learning. This database contains signals from 11 subjects with knee abnormality and 11 normally, previously diagnosed by a professional. They undergo three movements to analyze the behavior associated with the lower limb, gait, leg extension from a sitting position, and flexion of the leg up. Analysis has stages as preprocessing, features extraction, training, and validation. Several conventional electromyography features are used in performance comparison with it combined with Enhanced features. The training process was accomplished with 60% of the data and the validation of the algorithm with 40% of them. Finally, from the confusion matrix, evaluated from the final models, metrics such as Accuracy, Sensitivity, Specificity and Precision were calculated to select the model with optimal performance in real life.

KEYWORDS: classification, electromyography EMG, feature extraction, machine learning, preprocessing, signal, support vector machine SVM, surface electromyography

1 INTRODUCCIÓN

Este proyecto técnico busca crear una herramienta que determine si una persona presenta o no una patología en su extremidad inferior mediante la clasificación de señales electromiográficas EMG. Se empleará la base de datos EMG Dataset in Lower Limb obtenida de UCI Machine Learning [1], que tiene dos grupos de personas: 11 normales y 11 con anormalidad en la rodilla que realizan tres movimientos, previamente, diagnosticado por un profesional.

El contenido desarrollado en el presente trabajo técnico es investigativo, no se recomienda emplearlo para diagnóstico de enfermedades sin supervisión de un profesional.

La electromiografía permite determinar patrones neuromusculares que en condiciones de lesión muestran alteraciones en las activaciones motoras que son diferentes a los patrones de una persona sana [2]. Los análisis clínicos actuales son empíricos y dependen del médico especialista, estos diagnósticos son subjetivos y muchas veces limitan los criterios a resultados visuales, pero se pretende desarrollar un algoritmo que pueda ser empleado como soporte técnico. El algoritmo basado en electromiografía buscará integrar soluciones que reconozcan si un sujeto presenta alguna patología para que esta disciplina no quede únicamente en la solución de problemas de ingeniería, sino que, de alguna forma tenga inclusión en la práctica clínica.

El estudio e interpretación de las señales EMG se empleará para detectar si un sujeto presenta una patología, lo cual, permitirá a los médicos diagnosticar, clasificar y en varios casos prevenir enfermedades, que pueden desarrollarse en los músculos de las extremidades inferiores. Se busca construir un modelo que permita predecir o determinar la respuesta normal o anormal de los músculos involucrados, mediante el preprocesamiento y la digitalización gráfica de las señales biológicas; además, se incorporará la tecnología, usada en la adquisición de señales EMG, para la detección de problemas en el ser humano, de tal manera que, en el futuro se constituya en una herramienta que brinden soporte técnico médico.

La creación de sistemas de soporte para la detección de patologías en la teoría de aprendizaje de máquina conocido como Machine Learning, permite que el aprendizaje supervisado desarrolle algoritmos para la clasificación de señales EMG utilizando varias técnicas, de modo que, una vez analizadas varias de estas técnicas se empleará máquina de vectores de soporte (SVM), porque es un método eficiente que presenta alta precisión y buenos resultados según varios estudios [3] [4] [5], para determinar el

mejor modelo de entrenamiento se realizarán diversas pruebas de validación y varias combinaciones de hiperparámetros.

Por los motivos descritos, el presente trabajo pretende desarrollar una herramienta de software que permita la clasificación de señales normales y anormales con mayor exactitud, para elaborar el algoritmo propuesto se considerarán procesos como: preprocesamiento de la señal, extracción de características y clasificación mediante SVM.

1.1 OBJETIVOS

El objetivo general de este Proyecto Técnico es:

- Desarrollar una herramienta computacional que permita la clasificación entre personas que presenten o no una patología a partir de señales electromiográficas de la extremidad inferior usando máquinas de soporte de vectores.

Los objetivos específicos de este Proyecto Técnico son:

- Estudiar los métodos de SVM enfocados a la clasificación de señales, mediante, técnicas de preprocesamiento y procesamiento de las mismas.
- Diseñar un algoritmo que permita la clasificación de señales electromiográficas a partir de una base de datos para determinar si una persona tiene o no una patología en su extremidad inferior.
- Implementar el algoritmo de clasificación de las señales electromiográficas.
- Comprobar la exactitud del modelo planteado como clasificador de personas con alguna patología, mediante una evaluación de los resultados obtenidos.

1.2 ALCANCE

En el presente proyecto técnico se estudiará al menos dos métodos de Kernel asociados al SVM que serán utilizados para la clasificación de señales EMG, al igual que las técnicas aplicadas a señales electromiográficas (EMG) para su preprocesamiento y procesamiento.

Se analizará al menos dos técnicas del preprocesamiento y procesamiento de señales que permitan realizar un análisis matemático para extraer las características de varias señales electromiográficas en estado estacionario.

Se obtendrá la base de datos EMG Dataset in Lower Limb del repositorio de UCI Machine Learning que contenga las señales de la extremidad inferior para analizar el comportamiento muscular asociado a la rodilla, mediante la extracción de señales de cada electrodo.

Se desarrollará un algoritmo utilizando el software más apropiado que permita la clasificación de señales EMG, a partir de una base de datos con al menos 22 sujetos, estos realizan aproximadamente cinco repeticiones por movimiento para determinar si una persona tiene alguna patología en su extremidad inferior.

Se diseñará e implementará un clasificador con máquina de soporte de vectores para que la base de datos seleccionada presente mayor exactitud.

Se entrenará el algoritmo empleando el 60% de las señales electromiográficas para clasificar personas que presenten o no una patología.

Se validará el algoritmo desarrollado, implementando el 40% de las señales electromiográficas, a partir de la base de datos seleccionada.

1.3 MARCO TEÓRICO

La anatomía topográfica es la ciencia que estudia las semejanzas que guardan las diferentes estructuras del organismo, sirve como base para el estudio de patologías internas, externas y medicina operatoria, esta rama de la anatomía es el soporte diario de los médicos en los diagnósticos [6].

1.3.1 MIEMBRO INFERIOR

En la presente sección, se realizará un análisis de la extremidad inferior que incluye músculos y articulaciones, es de gran importancia, puesto que, la base de datos en la que se va a trabajar considera anomalías presentes en este miembro cuya estructura se detalla en la Figura 1.1 que abarca desde la cadera hasta el pie.

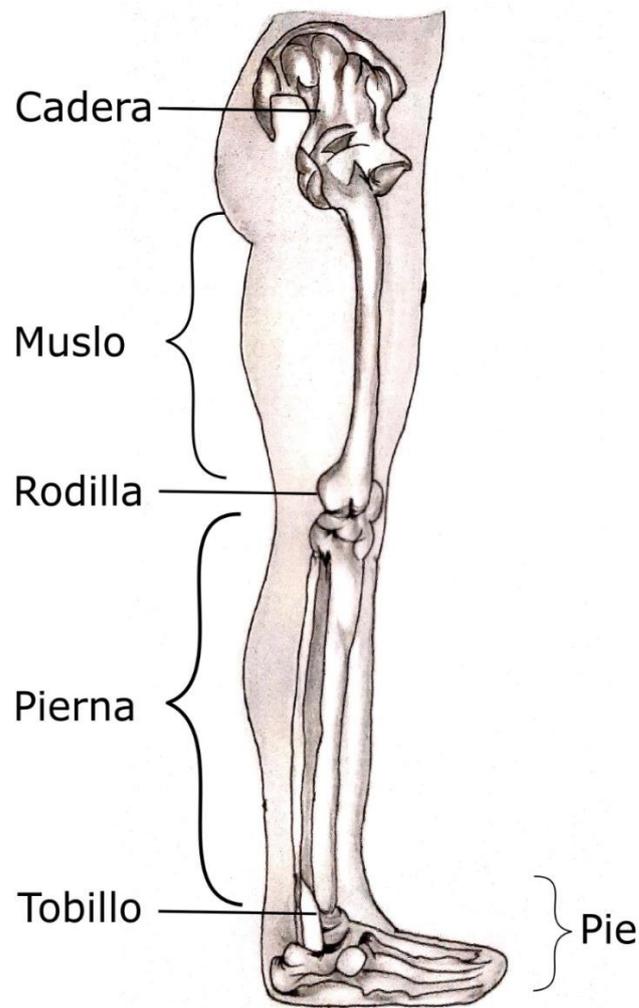


Figura 1.1 Anatomía estructural

- La cadera une el miembro inferior a la pelvis y comprende tres partes: región inguinofemoral, región obturatriz y región glútea.
- El muslo es la sección comprendida entre la cadera y la rodilla, esta sección envuelve el hueso fémur y divide al muslo en dos partes: femoral anterior y posterior.
- La rodilla se encuentra soportada por la rótula entre los extremos inferior del fémur y superior de la tibia.
- La pierna se encuentra entre la rodilla y el tobillo; está conformado por: los huesos tibia y peroné.
- La sección talocrural corresponde a la articulación del tobillo y finalmente, el pie que se extiende hasta los dedos [6].

Músculos de la extremidad inferior

Comprender la funcionalidad de algunos músculos, es indispensable para determinar el comportamiento muscular asociado a la extremidad inferior, es necesario establecer en donde se realizó la ubicación de los electrodos durante el estudio. En la

Tabla 1.1 se muestra los grupos musculares involucrados en los movimientos realizados para la adquisición de la base de datos.

Tabla 1.1 Músculos sujetos a medición [7]

Músculo	Función
Recto Femoral (RF)	Permite flexionar el muslo hacia la cadera y trabaja la rodilla como extensor de pierna. La localización se presenta en la Figura 1.2a.
Vasto Medial (VM)	Es de gran importancia para la rodilla porque ayuda en la firmeza anatómica de la rótula y evita que tienda a desviarse lateralmente. La localización se presenta en la Figura 1.2b.
Semitendinoso (ST)	Permite flexionar la rodilla y trabaja la cadera como extensor. La localización se presenta en la Figura 1.2c.
Bíceps Femoral (BF)	Permite flexionar la rodilla y trabaja el muslo como extensor hacia el tronco. La localización se presenta en la Figura 1.2d.

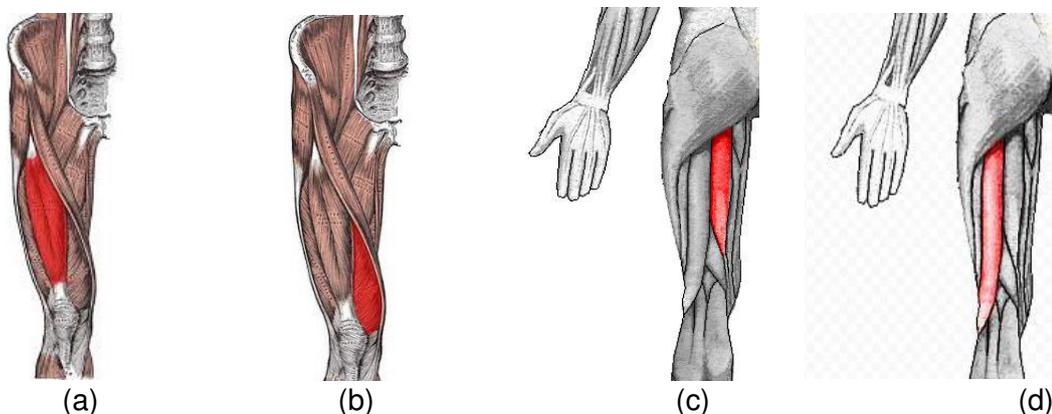


Figura 1.2 (a) (b) Músculos en la cara anterior, (c) (d) Músculos en la cara posterior [7]

Articulaciones de la extremidad inferior

Las articulaciones son la unión entre los huesos, su función es permitir la movilidad de las extremidades y aportar flexibilidad al cuerpo humano. Las principales articulaciones de la extremidad inferior a tratar en el presente proyecto son: cadera, rodilla y tobillo, puesto que, intervienen en los movimientos realizados por la persona sujeta a evaluación.

La cadera proporciona estabilidad a todo el cuerpo, soporta el peso de la parte superior y se encuentra ubicada entre el fémur y la parte lateral inferior de la pelvis.

La rodilla es la articulación más importante porque brinda un punto de apoyo, además de ser la más grande relaciona tres huesos: la rótula soportada por el fémur y la tibia que permiten trabajar a los músculos extensores y flexores en el plano sagital. Es fundamental mencionar que el rango de flexión de la rodilla depende de la cadera. Si la rodilla está completamente extendida se traduce en 0° y si está completamente flexionada tocando el glúteo con el talón (movimiento pasivo) alcanza los 160°, esto depende de cada sujeto, al estar la cadera flexionada, se limita a 140° el ángulo de flexión en la rodilla y si la articulación de la cadera está extendida este ángulo se limita a 120° [7].

El tobillo es una articulación ligeramente débil, reforzada con el revestimiento de ligamentos que conectan la pierna con el pie, creada para el soporte de un peso considerablemente fuerte.

Movimiento de la extremidad inferior

Una vez se ha realizado el análisis estructural de músculos y articulaciones, se puede deducir que estudiando la rodilla sin excluir los elementos que se interrelacionan, al verse afectados pueden emplearse para diagnosticar anomalías presentes durante una patología que afecta a la extremidad inferior. Los principales movimientos pertinentes al estudio que una persona puede realizar son.

- Marcha.
- Extensión de pierna sentado.
- Flexión de pierna de pie.

Patologías en la extremidad inferior

La extremidad inferior tiene una estructura compleja y al estar cada día sometida a un esfuerzo significativo, en atletas o en personas con factores de predisposición como sobrepeso, edad o anomalías congénitas es muy común que se presenten inflamaciones, torceduras, desgarros y fracturas, todas estas lesiones comúnmente denominadas patologías, son causadas por accidentes y tienen su origen en el sistema músculo esquelético; sin embargo, existen patologías que tienen origen neuromuscular y son más difíciles de diagnosticar.

En la rutina clínica un especialista clasifica una patología, mediante la interpretación de síntomas, diagnóstico y evaluación de exámenes, lo que resulta insuficiente para patologías tempranas cuyos signos son poco evidentes entre estas se pueden

mencionar: Según [2], estados tempranos de Parkinson, marcha de Trendelenburg o marchas patológicas de origen psiquiátrico. Según [7], una fractura de tobillo puede no presentar signos evidentes en las radiografías, ya que el daño se presenta en los ligamentos. El síndrome del dolor patelofemoral se produce cuando el sujeto muestra dolor en la parte anterior de la rodilla por una lesión en el músculo vasto medial, que provoca debilitación causando dificultad en el movimiento de la rótula. Las señales EMG se usan en aplicaciones de diagnóstico y rehabilitación como pruebas complementarias.

1.3.2 SEÑALES ELECTROMIOGRÁFICAS EMG

La electromiografía es la ciencia que estudia los biopotenciales generados por la actividad de las unidades motoras controladas por el sistema nervioso [8], implica la interpretación y el análisis de gráficos derivados de impulsos eléctricos provenientes de los músculos.

1.3.2.1 Tipos de señales EMG

Las señales electromiográficas pueden clasificarse en dos: las señales EMG de superficie o no invasivas, que en el ámbito clínico se usan para la rehabilitación y las señales EMG de aguja o invasivas se emplean en el estudio de casos miógenos según [9]. En la actualidad, han ido ganando acogida las señales EMG que sirven para monitorear y analizar el trabajo realizado por las fibras musculares. Las diferencias entre estas señales se describen en la Tabla 1.2.

Tabla 1.2 Diferencias de una EMG superficial y una EMG de aguja [2] [8]

Superficiales	Invasivas
Es una señal adquirida de forma no invasiva con electrodos colocados en la piel.	La señal se adquiere a nivel intramuscular.
Susceptible a ruido externo.	Presentan menor ruido.
Se obtiene una superposición de varias unidades motoras.	Se pueden descomponer en unidades motoras individuales.
Influyen características bioquímicas y externas como la humedad de la piel.	Requiere vigilancia médica.

1.3.2.2 Características y estados de una señal EMG.

Para iniciar con el estudio y tratamiento de señales EMG es necesario mencionar varias características importantes como indican diversos estudios [2] [8] [9] [10].

- Amplitudes entre 0.05 mV – 5 mV pico
- Frecuencia entre 0 y 500 Hz

- Frecuencia dominante entre 50 y 150 Hz
- La adquisición de señales se ve afectada por el ruido en la línea de alimentación de 60 Hz.

Los estados de una señal EMG dependen de lo que registre un electrodo, se considera estado estacionario o EMG estable, si la contracción muscular dura un periodo de tiempo largo, caso contrario, es una señal EMG de estado transitorio, si la contracción muscular se da en un corto periodo de tiempo, es repentina y se retorna pronto al estado de reposo [11]. Entiéndase por contracción muscular el estado activo y reposo como estado inactivo, como se muestra en la Figura 1.3.

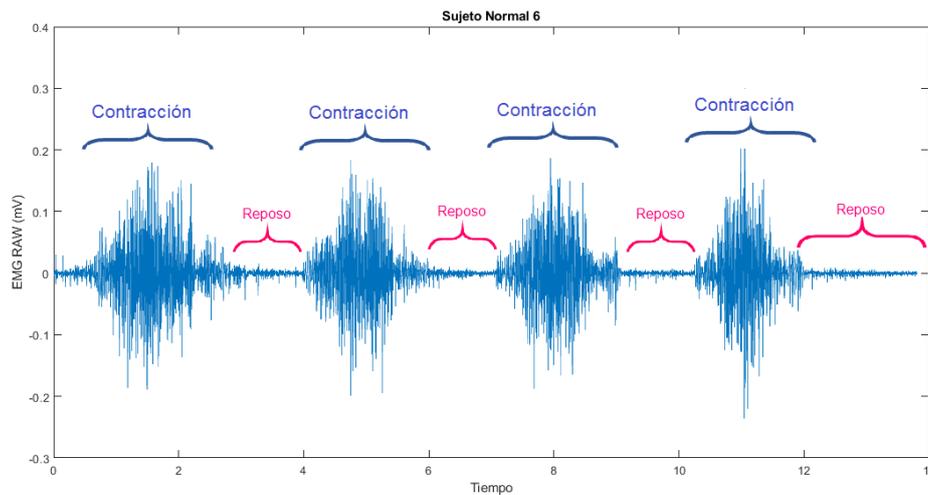


Figura 1.3 Estados de una señal EMG [12]

La Figura 1.3 presenta la señal EMG en bruto para el músculo semitendinoso de la persona normal número 6 que realiza el movimiento Flexión de pierna de pie, se puede distinguir las dos etapas contracción muscular y reposo.

1.3.2.3 Patrones neuromusculares de una EMG patológica.

Como se ha mencionado, la activación de una fibra muscular provoca que varias unidades motoras se involucren en la generación de un biopotencial, si el músculo se encuentra deteriorado o fatigado debido a una patología las características de la señal EMG son diferentes, por ejemplo, la onda generada puede ser más pequeña; caso contrario, si la patología no es a nivel muscular, si no, a nivel nervioso las unidades motoras realizan trabajo extra para compensar esta afectación [12].

En ámbitos de investigación médica, como análisis de movimiento o rehabilitación, emplear las señales EMG en la actualidad, desempeñan un papel fundamental para comprender los patrones neuromusculares que se presentan en los sujetos con alteraciones patológicas.

Una vez conocida la composición anatómica y funcional de la extremidad inferior podemos establecer consideraciones dentro del preprocesamiento, análisis y respuestas de la herramienta de software que se desarrollara mediante técnicas de aprendizaje automático.

1.3.3 APRENDIZAJE AUTOMÁTICO

Conocido como Machine Learning, es una parte de la Inteligencia Artificial que emplea algoritmos computacionales con el fin de aprender; mediante, el reconocimiento de patrones y extracción de características busca la solución de un problema independientemente de una programación explícita o de una ecuación predeterminada. Emplea los datos, de forma que realice predicciones por sí mismo. Existen dos tipos de Aprendizaje Automático (AA): supervisado y no supervisado [13].

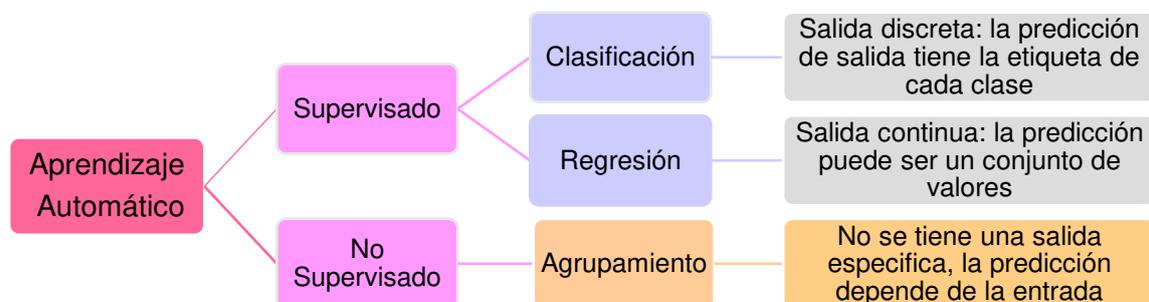


Figura 1.4 Tipos de AA [14]

En la Figura 1.4 se presenta un resumen de los tipos de Aprendizaje Automático. Entre los principales algoritmos de clasificación se encuentran: Máquina de vectores de soporte (SVM), Análisis Discriminante, Redes bayesianas (Naive Bayes) y K-vecino más cercano (K-NN). Entre los algoritmos de regresión más comunes se tienen: Regresión Lineal (GLM), Regresión de vectores de soporte (SVR), Árboles de regresión y Redes Neuronales. Para aprendizaje no supervisado existen Algoritmos como Fuzzy C-means, Modelos Ocultos de Markov, K-Means, entre otros [14].

1.3.3.1 Aprendizaje supervisado

En el aprendizaje supervisado el conjunto de observaciones para entrenamiento incluyen valores de entrada y salida [14]. El algoritmo en el presente proyecto va a predecir si una persona es sana o tiene alguna patología, a partir del análisis de una señal EMG estimará si es normal o anormal, siendo estas etiquetas las salidas o clases de modelo. Se entrenará un clasificador con Máquina de soporte de vectores (SVM),

empleando un conjunto de señales obtenidas de una base de datos en la extremidad inferior.

El algoritmo desarrollado tendrá un conjunto de entrada representado por las características extraídas de los datos. Estos datos serán las señales tomadas de cada electrodo sometidas a una etapa de preprocesamiento, este conjunto se particionará en un conjunto de entrenamiento y un conjunto de prueba destinados para el proceso de aprendizaje, culminada la etapa de entrenamiento, se emplearán los patrones aprendidos para verificar el funcionamiento del modelo entrenado.

Para desarrollar un modelo predictivo robusto será necesario ejecutar el diagrama de flujo mostrado en la Figura 1.5, que permite el entrenamiento, comparación de múltiples modelos, la optimización de parámetros del modelo y la validación del rendimiento del modelo [3].

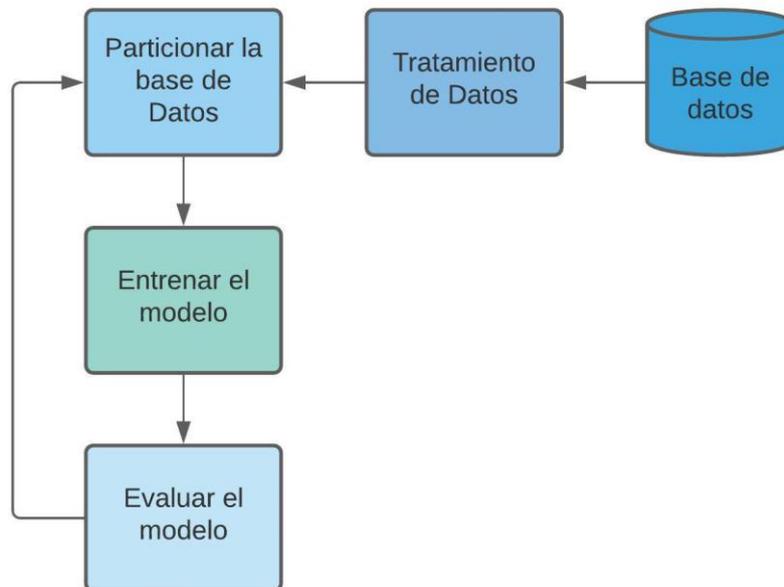


Figura 1.5 Procedimiento de Aprendizaje Supervisado

1.3.3.2 Técnicas de preprocesamiento

Realizar la clasificación de señales requiere un tratamiento de datos, al adquirir los datos de sensores o bases de datos se puede incluir valores atípicos y ruido; que son propios de cualquier tipo de señal como se observa en la Figura 1.6, lo cual, dificulta la distinción entre las clases sano o con patología, la finalidad de hacer un preprocesamiento es realizar una limpieza, que busca convertir los datos en señales con un formato más fácil de interpretar. Cambiar la perspectiva de una señal EMG en bruto o EMG RAW permite

la extracción de características que ayudan al aprendizaje automático a distinguir fácilmente cual es una señal normal o anormal [15].

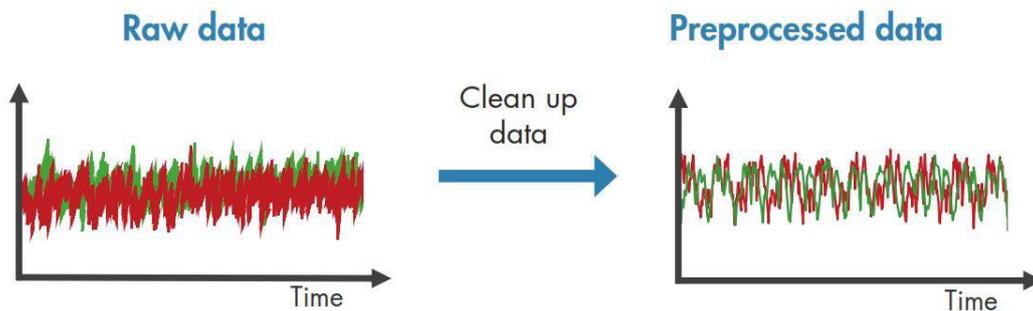


Figura 1.6 Diferencia entre señales en bruto y pre procesadas [16]

Las señales EMG que se adquieren mediante electrodos ubicados en la piel contienen información muy importante, por lo que se aplican varias técnicas de preprocesamiento para analizarlas. La rectificación de una señal consiste en obtener el valor absoluto de la señal en bruto EMG RAW, que es útil para detectar frecuencias bajas. Existen diversas técnicas de suavizado que permitirán la reducción de picos indeseados manteniendo la tendencia de la señal EMG según [8], entre estas se encuentran la segmentación, el valor absoluto o la media (MA) de un segmento de señal, el valor RMS y los filtros de paso bajo para reducir la diafonía en la alimentación, no es recomendable el uso de un filtro Notch, porque se superpone a la señal EMG, recayendo en una gran pérdida de información [2] [10].

1.3.3.3 Extracción de características

Las características o denominados indicadores de señal se usan para construir los patrones que un clasificador utiliza durante el proceso de aprendizaje. Entre los indicadores más comunes se tienen: tiempo, frecuencia y tiempo-frecuencia [16]. Existen varios métodos basados en señales. Para la presente investigación se determinarán indicadores en el dominio del tiempo.

Al no existir una característica o conjunto de características que estén generalizadas por la comunidad científica, según [17] las que se van a usar brindan mejores resultados. La extracción de características es un proceso que permite comprender la información que es transmitida por las señales EMG RAW, con la finalidad de mejorar el proceso de clasificación, buscando crear un vector de características relevantes que brinden un buen resultado a las predicciones con una menor demanda computacional.

Según [18] las características que se van a detallar a continuación fueron seleccionadas por sus excelentes resultados en trabajos anteriores. En formación de los vectores se emplearán las características más comunes [19], que además, permitirán capturar la información de calidad en una señal EMG, estas son:

Valor Medio Absoluto (MVA)

En la Ecuación (1.1), se presenta el MVA que es el promedio del sumatorio en valor absoluto de las muestras de una señal EMG.

$$MVA = \frac{1}{L} \left(\sum_{i=1}^L |F_i| \right) \quad (1.1)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Además, se empleará una función matemática ventana de ponderación para modificar esta característica como MVA1, lo que se tiene en la Ecuación (1.2) y puede aventajar el presente análisis según [4] [18] [20].

$$MVA1 = \frac{1}{L} \left(\sum_{i=1}^L e_i |F_i| \right) \quad e_i = \begin{cases} 1, & \frac{1}{4}L \leq i \leq \frac{3}{4}L \\ 0.5, & \text{En otro caso} \end{cases} \quad (1.2)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Empleando el mismo criterio anterior MVA2 es otra característica observada en la Ecuación (1.3).

$$MVA2 = \frac{1}{L} \left(\sum_{i=1}^L e_i |F_i| \right) \quad e_i = \begin{cases} 1, & \frac{1}{4}L \leq i \leq \frac{3}{4}L \\ \frac{4i}{L}, & i < \frac{1}{4}L \\ \frac{4(i-L)}{L}, & \text{En otro caso} \end{cases} \quad (1.3)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Longitud de forma de onda (WL)

Es la longitud acumulada en un segmento de señal, permite analizar la duración de la señal y su representación matemática se tiene en la Ecuación (1.4).

$$WL = \sum_{i=2}^L |F_i - F_{i-1}| \quad (1.4)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Cruces por cero (ZC)

Permite realizar una estimación de frecuencia, la representación matemática se tiene en la Ecuación (1.5).

$$ZC = \sum_{i=1}^{L-1} f(F_i) \quad (1.5)$$

$$f(F_i) = \begin{cases} 1, Si \{(F_i > 0 \& F_{i+1} < 0) | (F_i < 0 \& F_{i+1} > 0)\} \& |F_i - F_{i+1}| \geq T \\ 0, & En\ otro\ caso \end{cases}$$

Donde, L es la longitud de la señal, Fi es la señal EMG pre procesada y T es un valor de umbral.

Cambios en la pendiente de la señal (SSC)

Permite calcular el número de veces que la señal cambia de pendiente, la representación matemática se tiene en la Ecuación (1.6).

$$SSC = \sum_{i=2}^{L-1} f(F_i) \quad (1.6)$$

$$f(F_i) = \begin{cases} 1, Si \{(F_i > F_{i-1} \& F_i > F_{i+1}) | (F_i < F_{i-1} \& F_i < F_{i+1})\} \\ \& \{(|F_i - F_{i+1}| \geq T) | (|F_i - F_{i+1}| \geq T)\} \\ 0, & En\ otro\ caso \end{cases}$$

Donde, L es la longitud de la señal, Fi es la señal EMG pre procesada y T es un valor de umbral.

Valor medio Cuadrático (RMS)

En condiciones de no fatiga esta característica sirve para determinar el comportamiento muscular, la representación matemática se tiene en la Ecuación (1.7).

$$RMS = \frac{1}{L} \left(\sum_{i=1}^L (F_i)^2 \right) \quad (1.7)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Cambio de amplitud promedio (AAC)

Es la relación entre los valores instantáneos de cada ciclo, la representación matemática se tiene en la Ecuación (1.8).

$$AAC = \frac{1}{L} \sum_{i=1}^{L-1} |F_{i+1} - F_i| \quad (1.8)$$

Donde, L es la longitud de la señal y Fi es la señal EMG pre procesada.

Diferencia de desviación estándar absoluta (DASDV)

Es una forma de medir la dispersión de los datos, la representación matemática se tiene en la Ecuación (1.9).

$$DASDV = \sqrt{\frac{\sum_{i=1}^{L-1} (F_{i+1} - F_i)^2}{L - 1}} \quad (1.9)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Porcentaje de miopulso (MYOP)

Se define como porcentaje de salida Myopulse en donde el valor absoluto de la señal EMG supera el valor umbral predefinido, la representación matemática se tiene en la Ecuación (1.10).

$$MYOP = \frac{1}{L} \sum_{i=1}^L f(F_i) \quad f(F_i) = \begin{cases} 1, & F_i \geq T \\ 0, & \text{En otro caso} \end{cases} \quad (1.10)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Amplitud Willison (WA)

Determina el número de veces que las unidades motoras superan el umbral de los potenciales de activación, la representación matemática se tiene en la Ecuación (1.11).

$$WA = \sum_{i=1}^{L-1} f(F_i) \quad f(F_i) = \begin{cases} 1, & |F_i - F_{i+1}| \geq T \\ 0, & \text{En otro caso} \end{cases} \quad (1.11)$$

Donde, L es la longitud de la señal, F_i es la señal EMG pre procesada y T es un valor de umbral.

Integral cuadrada simple (SSI)

Es la suma de los valores cuadráticos de una señal EMG, la representación matemática se tiene en la Ecuación (1.12).

$$SSI = \sum_{i=0}^L (F_i)^2 \quad (1.12)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Varianza (VAR)

Esta característica permite determinar la potencia de una señal, la representación matemática se tiene en la Ecuación (1.13).

$$VAR = \frac{1}{L-1} \sum_{i=1}^L (F_i)^2 \quad (1.13)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Detector de registros (LD)

Permite determinar la fuerza ejecutada sobre los músculos, la representación matemática se tiene en la Ecuación (1.14).

$$LD = \exp\left(\frac{1}{L} \sum_{i=0}^L \log(|F_i|)\right) \quad (1.14)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Longitud Máxima Fractal (MFL)

Es una característica común, la representación matemática se tiene en la Ecuación (1.15).

$$MFL = \log_{10}\left(\sqrt{\sum_{i=1}^{L-1} F_{i+1} - F_i^2}\right) \quad (1.15)$$

Donde, L es la longitud de la señal y F_i es la señal EMG pre procesada.

Además, se procederá a introducir dos características que se proponen en [18] para mejorar la exactitud del clasificador, ya que, son herramientas valiosas en aplicaciones clínicas y de rehabilitación, también, se realizará una comparación con los resultados tradicionales.

Valor medio absoluto mejorado (EVAM)

Es una modificación del valor absoluto medio revisado en la Ecuación (1.1), la representación matemática se tiene en la Ecuación (1.16).

$$EMAV = \frac{1}{L} \sum_{i=1}^L |(F_i)^w| \quad (1.16)$$

$$w = \begin{cases} 0.75 & i \geq 0.2L \ \& \ i \leq 0.8L \\ 0.5 & \text{En otro caso} \end{cases}$$

Donde, L es la longitud de la señal, F_i es la señal EMG pre procesada y w es un parámetro entre el 20% y el 80% de las muestras en un segmento de la señal EMG, para distinguir la región media, que es donde se concentra la mayor información durante la contracción muscular.

Longitud de forma de onda mejorada (EWL)

Es una modificación mínima de la longitud de forma de onda revisada en la Ecuación (1.4), la representación matemática se tiene en la Ecuación (1.17).

$$EWL = \sum_{i=2}^L |(F_i - F_{i-1})^w| \quad (1.17)$$
$$w = \begin{cases} 0.75 & i \geq 0.2L \text{ \& } i \leq 0.8L \\ 0.5 & \text{En otro caso} \end{cases}$$

Donde, L es la longitud de la señal, F_i es la señal EMG pre procesada y w es un parámetro entre el 20% y el 80% de las muestras en un segmento de la señal EMG, para distinguir la región media, que es donde se concentra la mayor información durante la contracción muscular.

1.3.3.4 Partición de datos

Durante el proceso de aprendizaje supervisado Figura 1.5, después de realizar el respectivo tratamiento de datos conformado por: preprocesamiento y extracción de características, a continuación, se efectuará la partición de datos, es de suma importancia aclarar algunos términos que generalmente son mal interpretados durante el desarrollo de un algoritmo para clasificación. La implementación de un algoritmo requiere normalmente de 2 conjuntos de datos: conjunto de datos de prueba (Testing set) y conjunto de datos de entrenamiento (Training set), a menudo, el conjunto de datos de entrenamiento se fracciona para obtener un conjunto de datos de validación (Validation set) como se observa en la Figura 1.7.



Figura 1.7 Formas de particionar un conjunto de Datos

Se describirá brevemente cada partición de datos y su propósito general.

- **Conjunto de datos de entrenamiento:** sirve para la etapa de entrenamiento del modelo, de este surgen los patrones que permiten al algoritmo realizar las predicciones.
- **Conjunto de datos de validación:** sirve para medir y comparar el rendimiento del modelo mientras se ajustan los hiperparámetros hasta conseguir un modelo robusto y preciso. Este conjunto influye directamente en el proceso de aprendizaje del modelo.
- **Conjunto de datos de prueba:** sirve para brindar una evaluación final del modelo y de su desempeño con los datos u observaciones no vistas, es decir, se emplea cuando el clasificador puede dar una valoración imparcial a los nuevos datos.

No todos los algoritmos necesitan los 3 conjuntos de datos, esto depende de la naturaleza de cada algoritmo y del método de validación que se emplea durante el entrenamiento [17]. Según [21], las particiones más comunes son 60%-40%, 70%-30%, 80%-20% y 50%-50%.

La división de los conjuntos de datos depende de la robustez de la base de datos, en el presente proyecto se va a realizar una partición de datos del 60% para el entrenamiento y 40% para la prueba, debido a que el conjunto de observaciones es mayor de 100 muestras, pero menor que 1'000 000 [22].

1.3.3.5 Métodos de validación

Existen varios métodos de validación, para el presente proyecto se va a emplear el método de validación por exclusión. Se divide de forma aleatoria en la proporción 60%-40% al conjunto de datos: conjunto de entrenamiento y conjunto de prueba respectivamente ver Figura 1.7. La ventaja de este método es disminuir el tiempo de procesamiento porque se entrena una vez. Según [23] se sugiere emplear este método de validación en bases de datos grandes, en vista de que utiliza una parte del conjunto de datos de entrenamiento para la validación.

El conjunto de datos de prueba son observaciones nuevas para el modelo, a partir de las cuales, se conseguirá medir la precisión del algoritmo, se escoge un porcentaje de prueba alto para evitar el sobreajuste al tener demasiadas características en el conjunto de entrenamiento, se puede sobreentrenar o se puede subajustar al no emplear suficientes observaciones el modelo no podría generalizar parámetros que determinen si es una señal EMG normal o anormal.

Además, se pretende entrenar el modelo bajo el método de validación cruzada para realizar una comparación con [19]. La validación cruzada es un método en el cual, un conjunto de observaciones se subdivide en k partes distintas de forma aleatoria y del mismo tamaño. De las k partes se emplea una para validación y las partes restantes se utilizan para entrenar el modelo. Este proceso reiterado se ejecuta k veces de modo que los k resultados son promediados para medir la precisión del modelo. La ventaja de este método es que se garantiza que los datos de entrenamiento y prueba son independientes y se emplean todas las observaciones en el proceso de aprendizaje; mientras que, la desventaja es que requiere una alta demanda computacional para su entrenamiento y puede tomar un largo periodo de tiempo por lo que es recomendable para bases de datos pequeñas [24].

Partición 1	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Exactitud 1
Partición 2	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Exactitud 2
Partición 3	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Exactitud 3
Partición 4	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Exactitud 4
Partición 5	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Exactitud 5

Datos para validación	Datos para entrenamiento
-----------------------	--------------------------

Figura 1.8 Particiones de un conjunto de datos para validación cruzada [24]

En la Figura 1.8, se puede observar un conjunto de datos fraccionado en 5 particiones, en cada iteración para el entrenamiento se emplea una partición diferente para la validación y el resto para el entrenamiento; finalmente se tienen 5 mediciones de exactitud, las cuales se promedian para determinar la precisión más aproximada a la realidad, en este tipo de validación cada pasada emplea todas las observaciones.

1.3.3.6 Máquina de soporte vectorial SVM

Como se ha mencionado anteriormente, la SVM es un algoritmo de predicción robusto porque es capaz de detectar patrones complejos a partir de la construcción de un hiperplano o conjunto de hiperplanos en un espacio multidimensional. Los hiperplanos permiten separar los predictores en dos o más clases, por eso, es conocido como un clasificador binario y multiclase [3]. Se pueden aplicar métodos de Kernel para problemas con datos que no son linealmente separables y así transformarlos en un clasificador lineal. Una máquina de soporte vectorial puede realizar predicciones eficientes si el problema de clasificación tiene un gran número de características,

incluso, si el conjunto de observaciones para entrenamiento es pequeño [25]. Debido a su alta precisión y a sus buenos resultados se ha ido posicionando como el mejor algoritmo si se requiere una alta exactitud, en la Tabla 1.3 se resumen algunas ventajas y desventajas.

Los vectores de soporte son un pequeño subconjunto de predictores, gráficamente son los datos más cercanos a un hiperplano que sirven para el ubicar de forma óptima la superficie de decisión entre las clases [25].

Tabla 1.3 Características de una SVM [3]

Ventajas	Desventajas
Es un algoritmo poco propenso a tener sobreajuste.	Puede ser lento de procesar a nivel computacional si se cuenta con varias características para entrenamiento.
Más fácil de implementar que una red neuronal.	Encontrar el modelo más eficiente implica varios entrenamientos con diferentes tipos de Kernel.
Mayor rendimiento si se tienen únicamente dos clases para clasificación o predicción.	La interpretación de resultados resulta bastante compleja.

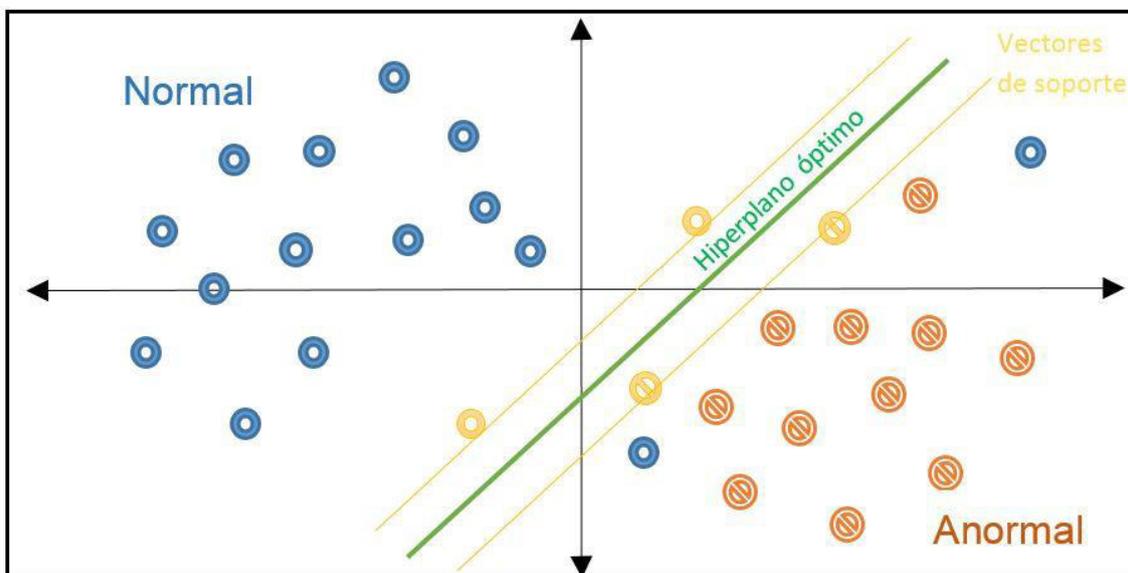


Figura 1.9 Vectores de soporte [26]

La Figura 1.9, presenta como se vería la separación entre las clases normal y anormal de las señales EMG en estudio, además, se visualizan los vectores de soporte próximos al hiperplano óptimo de separación entre las clases, cabe recalcar, que en el semiplano anormal hay dos observaciones normales, se los puede reconocer como valores

perdidos o simplemente aumentar a un espacio multidimensional, para conseguir esto se aplican los métodos de Kernel.

1.3.3.7 Métodos de Kernel

Existen varios métodos de Kernel que son aplicados dependiendo del problema de clasificación, la solución más sencilla al problema del presente proyecto, es separar las clases mediante una línea recta como se observa en la Figura 1.9, pero esta solución no necesariamente puede ser la más eficiente, es por esto que se realizará una comparación entre los diferentes resultados ajustando los hiperparámetros del modelo, de forma manual, esto quiere decir que se procederá a entrenar varios modelos con diferente tipo de Kernel y se medirá cual es más preciso y eficiente. Se hará una breve revisión de los tipos de Kernel [26].

Lineal

Ideal para realizar separación binaria o de dos clases, la representación matemática se encuentra en la Ecuación (1.18).

$$K(x_i, x_j) = x_i^T x_j \quad (1.18)$$

Donde, K es el espacio nuevo de mapeo y X_i y X_j son las clases.

Polinomial

Permite realizar la separación de clases con funciones polinomiales, se recomienda de segundo y tercer orden, la representación matemática se encuentra en la Ecuación (1.19).

$$K(x_i, x_j) = (x_i^T x_j + 1)^p \quad (1.19)$$

Donde, K es el espacio nuevo de mapeo, X_i y X_j son las clases y p es el orden del polinomio.

Función de base Radial RBF o Gaussiana

Permite separar las clases en un mapeo multidimensional empleando una campana de Gauss, puede emplearse para múltiples clases, la representación matemática se encuentra en la Ecuación (1.20).

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1.20)$$

Donde, K es el espacio nuevo de mapeo, X_i y X_j son las clases y σ es el ancho.

2 METODOLOGÍA

En el presente capítulo se emplea las técnicas investigadas en la sección anterior con el fin de obtener los conjuntos de datos de entrenamiento y prueba, para validar los algoritmos empleados. Partiendo por la depuración de la base de datos seleccionada, mediante las herramientas del software Matlab ®, además, se detallará el diseño del algoritmo de tratamiento de las señales EMG obteniendo un rendimiento óptimo.

2.1 BASE DE DATOS

La base de datos empleada “EMG Dataset in Lower Limb” [1] contiene señales EMG superficiales para el análisis de la extremidad inferior, que fueron adquiridas mediante un sensor comercial DataLOG MWX8 de Biometrics Ltd, colocado sobre la región contigua a la articulación de la rodilla.

2.1.1 SENSOR COMERCIAL DATALOG MWX8 [27]

El sensor DataLOG MWX8 se puede emplear como un dispositivo médico de investigación con la finalidad de obtener datos de forma portátil, puede usarse en rehabilitación y sirve para monitorear el rendimiento durante una actividad humana. Se usa en entornos educativos como laboratorios de marcha y en la ergonomía laboral para reducir lesiones ocupacionales, y por consiguiente, mejorar la calidad de vida, confort y seguridad de una persona en un trabajo industrial [28].

El sensor es de fácil adaptación porque puede ser ubicado en la pierna o en el brazo como se muestra en la Figura 2.1, incorpora: un cinturón, una pantalla gráfica LCD, una tarjeta microSD para almacenamiento y se conecta a la computadora de forma inalámbrica mediante Bluetooth, que es ideal para adquisición de datos en tiempo real.



Figura 2.1 Sensor DataLOG MWX8 [27]

ESPECIFICACIONES TECNICAS

El dispositivo tiene una capacidad de 8 canales analógicos y 4 digitales, el usuario puede seleccionar uno o varios canales analógicos para goniometría, cada sensor mide el biopotencial muscular a un rango de frecuencia de 1-20 KHz y una resolución de 14 bits con datos entre -8191 a 8192. Este dispositivo requiere de dos baterías de litio AA de 1.5 V, su duración es de 5 a 20 horas y depende del tipo de sensor. Tiene una tarjeta de memoria interna de 2GB y un adaptador compatible de Bluetooth Microsoft. Su ancho de banda para una entrada del sensor entre 1mV y 3mV DC es de 1KHz (+0dB/-3dB), tiene un filtro Anti-Aliasing automático aproximadamente de -3dB (0.45 x frecuencia de muestreo) y máximo de -13dB (0.8 x frecuencia de muestreo). Cada canal tiene una corriente menor a 20 mA y un voltaje regulable de 0 a 4.6 VDC [27].

Una vez conocidas las especificaciones generales del sensor, se detalla el protocolo de adquisición empleado para la base de datos del presente trabajo técnico. Se utilizaron 5 canales analógicos: 4 canales para muestreo de las sEMG y un canal para goniometría. Los 4 canales son para los electrodos colocados en los siguientes músculos de la extremidad inferior: Recto Femoral (RF), Vasto Medial (VM), Semitendinoso (ST), Bíceps Femoral (BF) sus funciones se detallaron en la Tabla 1.1 y sus ubicaciones fueron presentadas en la Figura 1.2. Los datos obtenidos de cada músculo se adquirieron en la tarjeta microSD de almacenamiento interno y fueron trasladados en tiempo real al software del sensor mediante un adaptador bluetooth, con la resolución mencionada y a una frecuencia de muestreo de 1 KHz [1].

2.1.2 DISPOSICIÓN DE LA BASE DE DATOS

La base de datos mencionada contiene 22 muestras de sujetos del sexo masculino, 11 personas normales y 11 personas con alguna anomalía en la rodilla, cada uno fue diagnosticado por un profesional y realizó tres ejercicios que permitieron analizar el comportamiento muscular, estos movimientos que han sido mencionados anteriormente e involucran a toda la extremidad inferior, se describen así:

1. El sujeto realiza el ciclo de marcha.
2. Desde una posición de pie el sujeto flexiona la rodilla.
3. Sentado, el sujeto extiende la pierna.

Cada sujeto tiene 3 tomas diferentes almacenadas en 3 distintos archivos, cada archivo tiene el nombre en el siguiente formato: número de sujeto del 1 al 11, N si es normal o A si es anormal y la contracción del nombre del movimiento que realizó "mar, pie, sen". Por ejemplo: 1Amar se refiere al sujeto anormal número 1 que realiza el ciclo de marcha, 11Npie hace referencia al sujeto normal número 11 que realiza la flexión de la rodilla.

Las series se encuentran en el dominio del tiempo y contienen entre 2 a 5 repeticiones por movimiento para cada sujeto con diferente duración. La base de datos contiene 4 carpetas: A_LOG, A_TXT, N_LOG y N_TXT. Las carpetas _LOG contienen las series en formato .log, que sirven para analizarse mediante el software de la empresa Biometrics Ltd. Fabricante del sensor con el que se adquirieron las señales. Las carpetas _TXT, contiene las series en un archivo de texto.

2.1.2.1 Formación de los archivos

Las señales EMG superficiales para el presente proyecto fueron tomadas de las carpetas con los archivos de texto organizadas de la siguiente forma: cada archivo contiene 7 líneas de encabezado para descripción y está compuesto por 5 columnas como se observa en la Tabla 2.1.

Tabla 2.1 Atributos de información de los archivos de texto

Canal	Ch1	Ch2	Ch3	Ch4	Ch5
Músculo	RF	BF	VM	ST	FX
Columna	0	1	2	3	4

Canal: corresponde al electrodo fijado a un músculo.

Músculo: corresponde al músculo sujeto a medición.

RF: Recto Femoral.

BF: Biceps Femoral.

VM: Vasto Medial.

ST: Semitendinoso.

FX: Ángulo de Flexión en la rodilla.

Cada columna tiene el mismo número de muestras, en caso de no ser así, se procedió a depurar un par de archivos que tenían datos extra en una o dos columnas, de tal manera que, en cada electrodo quede la misma cantidad de muestras. En la fase de entrenamiento del presente proyecto, se tomará en consideración las primeras cuatro columnas que contienen información relevante respecto a la naturaleza de la señal normal o anormal, el ángulo de flexión daría relevancia a estudios de clasificación de actividades o detección de movimientos.

En la primera fase se analizó y se estudió la base de datos seleccionada, realizando las transformaciones necesarias en los archivos que contienen las señales sEMG para

determinar la mejor representación posible y facilitar la ejecución de algoritmos de Machine Learning en el software escogido. Debido a que el análisis no se realizará en tiempo real y el objetivo del trabajo de titulación es desarrollar una herramienta de detección ideal que permita averiguar si un individuo está presentando una anomalía en la rodilla o está en riesgo para que un profesional pueda emplear la herramienta como soporte y someter a la persona a más pruebas o un tratamiento adecuado, el entrenamiento del algoritmo se va a desarrollar implementado el software Matlab ®. Matlab ® es un software que permite interactuar con diversos tipos de datos: texto, señales o imágenes de forma directa, cuenta con librerías de alta calidad a través de algoritmos y funciones estándar de la industria, las cuales, proporcionan herramientas para evaluar y seleccionar el conjunto correcto de características, a partir de un conjunto determinado por el usuario, con el fin, de desarrollar un algoritmo de aprendizaje automático e iniciar la creación aplicaciones, no es necesario dominar una programación complicada porque Matlab ® permite exportar los modelos creados en este entorno de programación flexible a un código en lenguaje “C”, que sirve en la implementación de sistemas embebidos.

2.2 PRE-PROCESAMIENTO

La mayor concentración de información se encuentra en la región de contracción de una señal sEMG ver Figura 1.3, siendo este un sistema estocástico que puede ser sometido a un análisis estadístico, por lo que, se emplearon los métodos antes mencionados segmentación, rectificación y filtrado para conseguir reducir cambios de varianza propios de las señales no estacionarias [11].

2.2.1 SEGMENTACIÓN

La segmentación consiste en dividir a una señal EMG en ventanas de tiempo. Esta fase es una de las más importantes durante el proceso de clasificación, dado que, al no existir una ventana con un tamaño estándar o tiempo fijo la mayoría de investigadores emplean técnicas basadas en trabajos anteriores, sin estudios concretos o resultados que los respalden [29].

Existen tres métodos de segmentación: ventanas definidas por eventos realizados, ventanas deslizantes y ventanas definidas por actividad.

Ventanas definidas por evento

Los eventos son las actividades realizadas con una secuencia de dos o más movimientos consecutivos, esta técnica se emplea generalmente para reconocimiento

de gestos específicos realizados esporádicamente y que contienen una o más actividades intercaladas unas con otras.

Ventanas deslizantes

Es el método de segmentación más utilizado en varios estudios según [29], su fácil implementación ha permitido que vaya ganando acogida para aplicaciones que demandan reconocimiento de gestos y actividades en tiempo real. Según [30], existen dos técnicas de ventanas deslizantes: ventaneado con solapamiento o denominado “overlapping window” y ventaneado disjunto o denominado “non-overlapping window”, las ventanas solapadas son ventanas adyacentes con un segmento de señal superpuesto, las ventanas no solapadas son ventanas adyacentes sin superposición de ningún segmento de señal como se observa en la Figura 2.2. No siempre es correcto emplear ventanas superpuestas, pues a pesar de su bajo costo computacional para su tratamiento, en las señales electromiográficas se puede perder información valiosa, especialmente de los bordes o límites de las ventanas porque cada segmento en la etapa de contracción muscular representa un conjunto particular de activaciones de las unidades motoras. Otra desventaja de usar esta técnica para pre procesar datos de entrenamiento, es que cada segmento tiene un tamaño de tiempo fijo, lo que en varios casos incluye el estado de reposo de la señal.

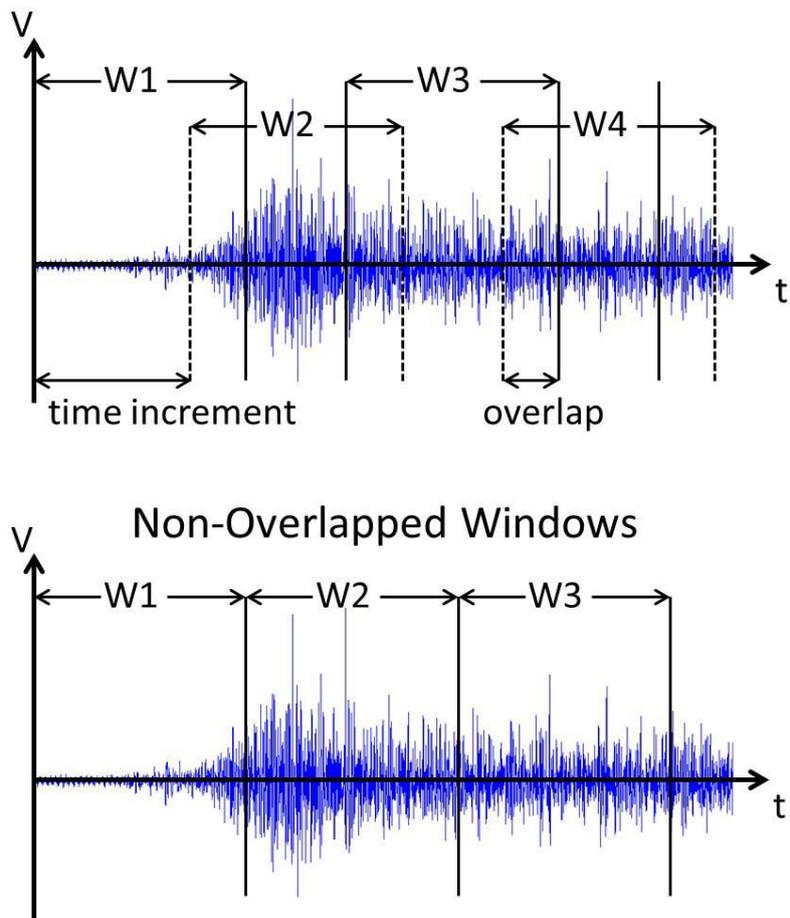


Figura 2.2 Tipos de ventanas deslizantes [17]

Ventanas definidas por actividad

En esta técnica se emplea una función capaz de segmentar la actividad muscular, eliminando las partes del gesto de relajación dentro de cada señal [11]. Los puntos inicial y final de cada contracción muscular se pueden determinar de diferentes maneras, pero algo es seguro, este es un proceso empírico [29]. Normalmente esta técnica se emplea para algoritmos de clasificación de actividades específicas, por lo que, se recomienda encontrar una segmentación particular dependiendo de los requerimientos de la aplicación. Este método se usó en el desarrollo de la herramienta del presente proyecto.

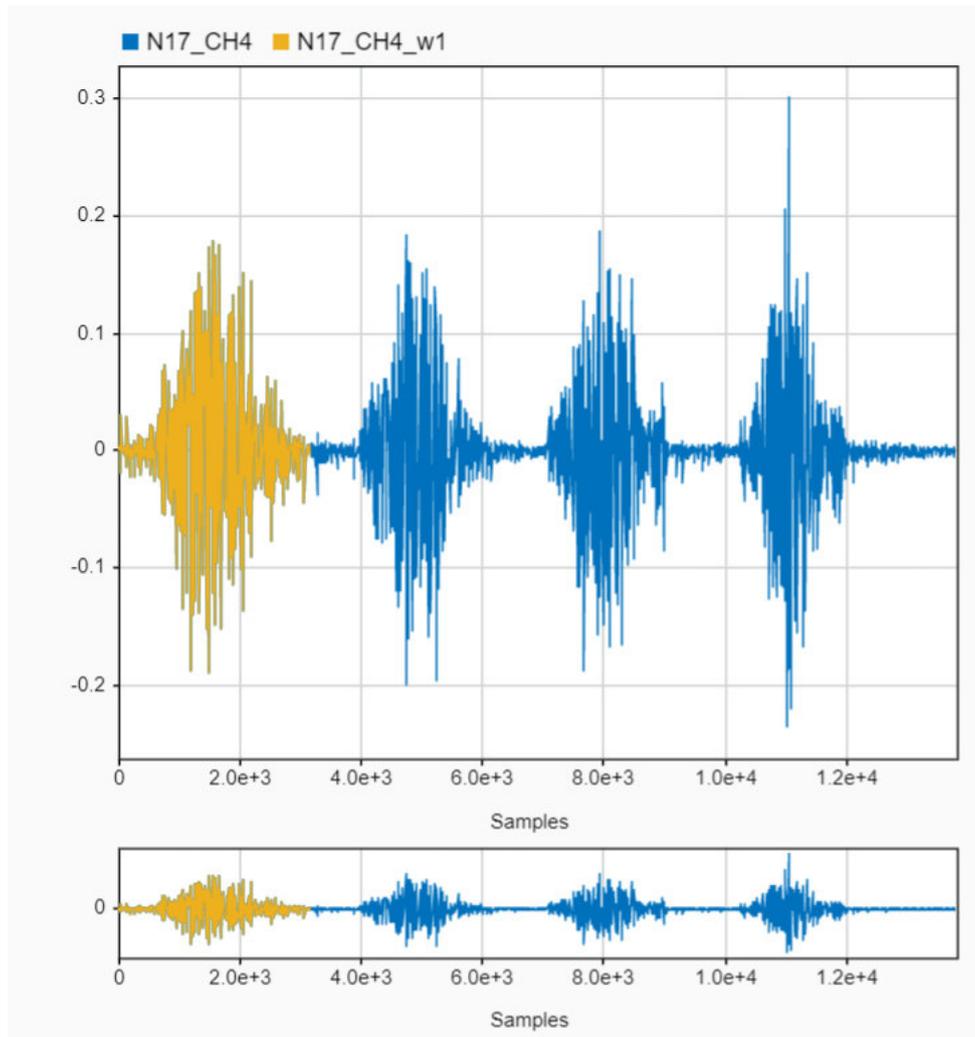


Figura 2.3 Segmentación con herramienta Panner

Empleando la herramienta Panner en Signal Analyzer App de Matlab ® se procedió a determinar de forma visual el tamaño óptimo de cada segmento o ventana, este tiempo resultó ser diferente para cada señal. El segmento puede tener cualquier tamaño, la selección de la duración de la ventana se realizó buscando la presencia de suficientes muestras. El ciclo de activación muscular es diferente para cada señal y en vista que las señales de la base de datos tienen diferente número de repeticiones. Se determinó los segmentos de forma particular en cada ciclo; sin embargo, se obtuvo el mismo número de ventanas para cada electrodo. Consiguiendo, 152 ventanas para las señales anormales y 150 ventanas para las señales normales, dando un total de 302 ventanas en cada electrodo. Finalmente, el número de ventanas pre procesadas fue de 1208.

En la Figura 2.3, se observa el gráfico de una de las señales EMG RAW, el proceso de segmentación se realizó con la respectiva extracción de la ventana, con la herramienta Panner se seleccionó el tamaño adecuado de la ventana que fue cortada en los

extremos de la contracción muscular, separando la etapa de reposo. Para la señal expuesta se obtuvo 4 ventanas.

2.2.2 RECTIFICACIÓN Y FILTRADO

Existen otras técnicas de preprocesamiento que se emplearon para suavizar las señales EMG registradas durante la adquisición de datos, estas pueden ser perturbadas por ruido electrónico u otros tipos de artefactos. Las perturbaciones se suelen eliminar mediante la aplicación de un filtro.

En el presente proyecto se rectificó cada señal EMG RAW con la función “abs” aplicando valor absoluto a cada elemento de las ventanas o segmento de muestras obtenido previamente. Posteriormente, se empleó un filtro pasa-bajos Butterworth de primer orden con una frecuencia de corte de 60 Hz (i.e., frecuencia normalizada de 0.12π rad/sample) para limitar el ruido de línea. El preprocesamiento de un segmento se presenta en la Figura 2.4.

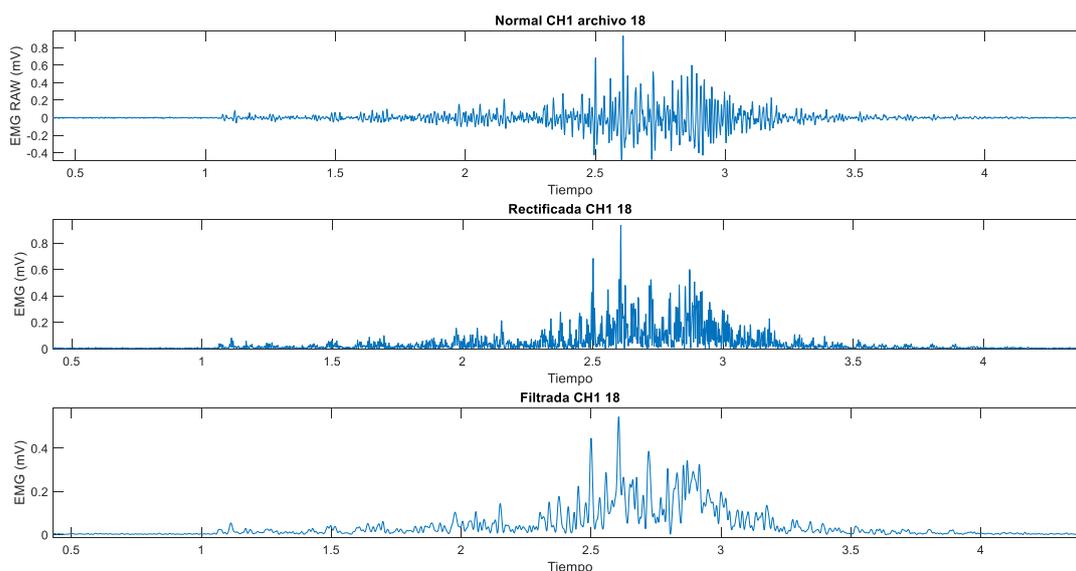


Figura 2.4 Preprocesamiento de una ventana

2.2.3 EXTRACCIÓN DE CARACTERÍSTICAS

En el apartado teórico se mencionó los tipos de indicadores de una señal y se realizó el análisis matemático necesario para determinar las características más comunes en el dominio del tiempo que se emplearon durante el desarrollo del algoritmo. La extracción de características fue aplicada para todas las ventanas preprocesadas, se realizó sobre los datos de entrenamiento y de prueba, con la finalidad de desarrollar un conjunto de patrones útil, permitiendo al algoritmo de clasificación alcanzar un buen rendimiento.

Se partió de buenos resultados obtenidos en el algoritmo de clasificación en [18], empleando el mismo conjunto de características, es decir, un conjunto de indicadores convencionales y un conjunto combinado con características mejoradas, para determinar el set de vectores de características más adecuado. En la construcción de los vectores de características se empleó el método más común usado en los algoritmos de reconocimiento de movimientos. Los vectores de características se forman como se observa en la Figura 2.5.

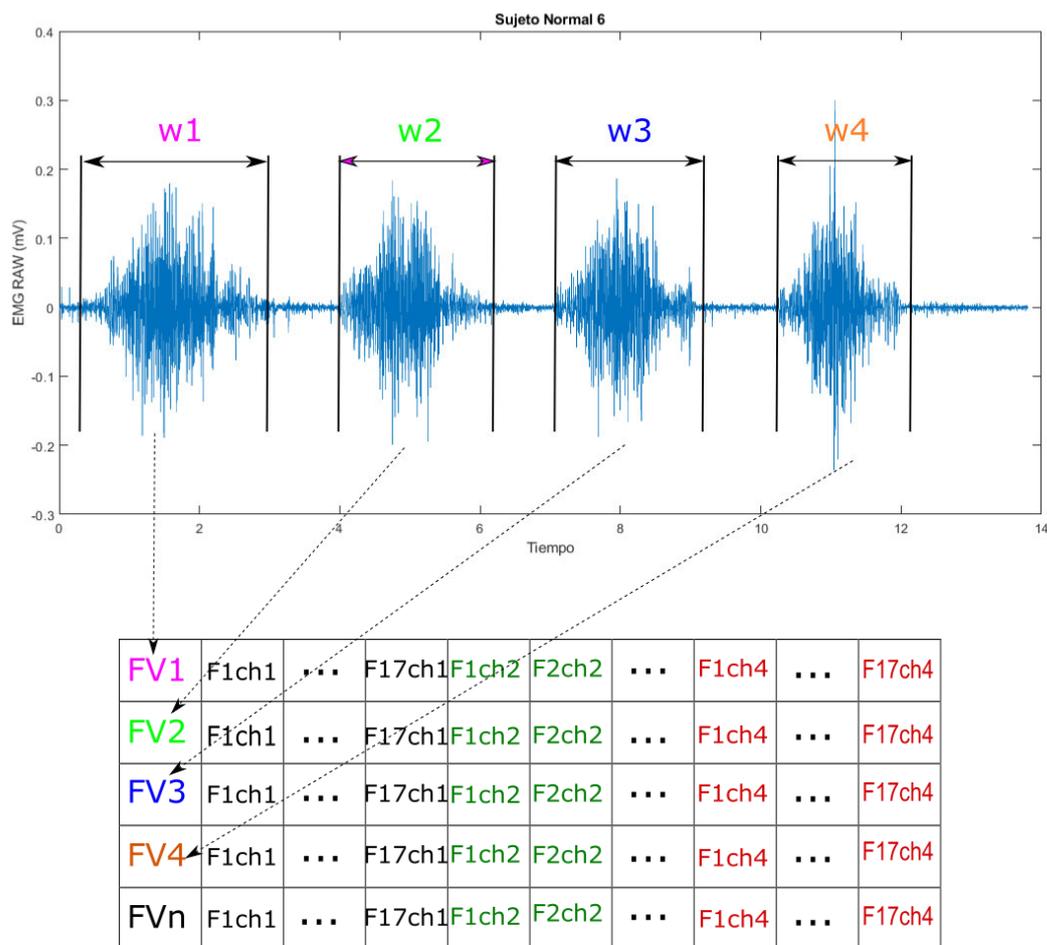


Figura 2.5 Construcción de los vectores de características[17]

Se extraen las características de cada electrodo, para ventanas preprocesadas como se observa en la Ecuación (2.1):

$$FV_n = [Feat_1, Feat_2, Feat_3, \dots, Feat_j] * [Ch_1, Ch_2, \dots, Ch_k] \quad (2.1)$$

Se estableció el número de canales $k=4$, que representa al número de electrodos colocados en cada músculo, $j=17$ son las características extraídas en el dominio del tiempo para cada ventana, existen un total de $n=302$ segmentos normales y anormales.

Cada FV tiene 68 características. Este procedimiento permite construir el conjunto de datos para entrenar al modelo.

2.3 CLASIFICACIÓN

En el presente proyecto se usó la Máquina de Soporte de Vectores, debido a las ventajas descritas previamente, en esta etapa se usó SVM multiclase con enfoque uno contra uno, porque es el método de clasificación más adecuado para dos clases. Este enfoque transforma un problema de k clases en $\frac{k(k-1)}{2}$ problemas binarios, el proyecto se centra en dos clases normal y anormal, dicho enfoque, aumenta la exactitud de clasificación, más que un enfoque de uno contra todos, además lo hace, sin sobre ajustar los datos de entrenamiento [11]. Este clasificador mapea el conjunto de observaciones sobre un plano en el espacio mientras busca el hiperplano de separación óptimo entre las clases, tratando de agrupar estas con un margen de distancia máximo, siempre y cuando los vectores de soporte sean los más cercanos a cada clase. En caso de que una SVM lineal no satisfaga los requerimientos de la aplicación se mapea las observaciones de entrenamiento en un espacio de alta dimensión y se construye el hiperplano de separación óptimo en este espacio, a continuación, la redimensión se deshace transformando el hiperplano encontrado al espacio de origen. Para entrenar los modelos de clasificación se empleó la herramienta "Classification Learner App" de Matlab®, que sirve para entrenar modelos de clasificación de datos usando aprendizaje automático supervisado. En la aplicación se usó la función Kernel para mapear las muestras en un plano N-dimensional. Se procedió a variar los tipos de Kernel manualmente entre lineal y polinomial de grado dos y tres, pues el conjunto de muestras obtenido para el entrenamiento como se observa en la Figura 2.6, no se puede separar linealmente, a menos que, la SVM aplique un margen suave entre las clases.

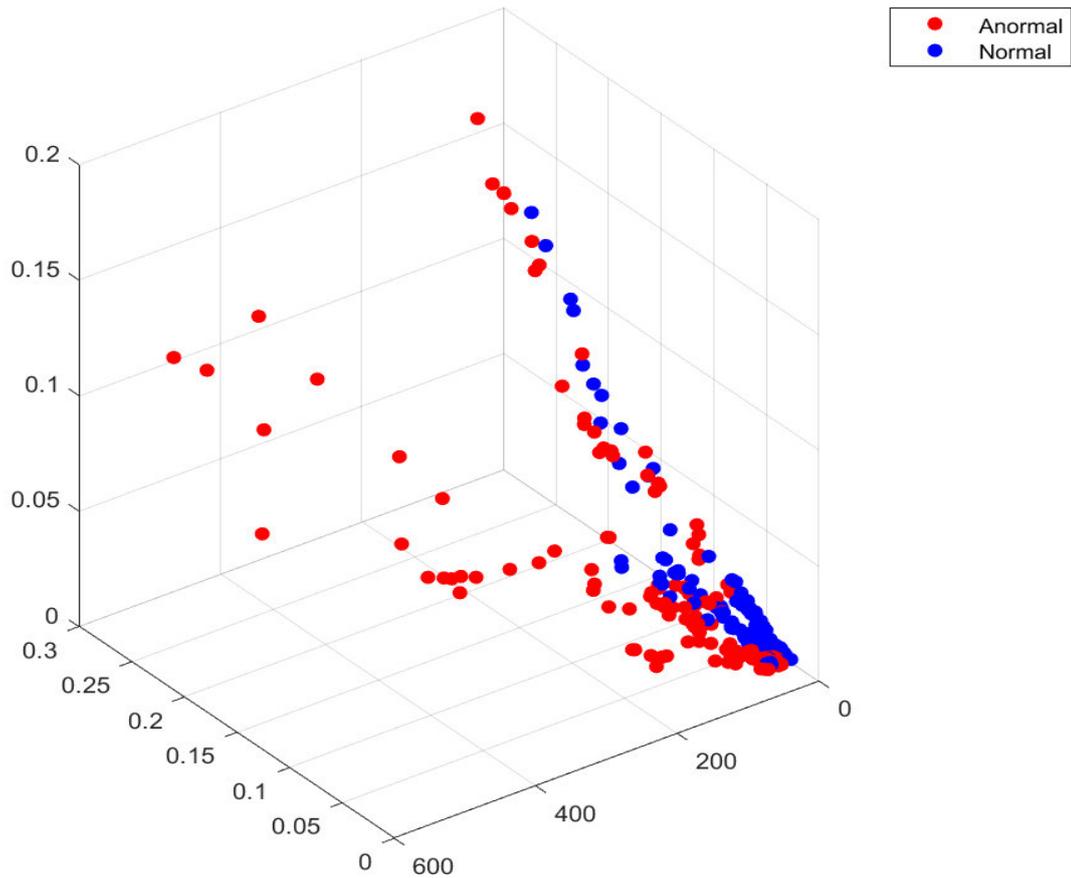


Figura 2.6 Distribución de los indicadores normales y anormales sobre un plano 3D

Entrenamiento y Validación

El conjunto de datos de entrenamiento o conjunto de características que contienen los indicadores en el dominio del tiempo (FVs), se procedió a realizar el respectivo etiquetado indicando al clasificador cuales son los parámetros de entrenamiento usados como predictores y las clases o variables de respuesta, en este caso se identificó si una señal EMG es normal o anormal. La validación se realizó con los métodos de validación cruzada con 5-particiones y validación por exclusión con el 40% de observaciones o predictores no vistos por el modelo durante la etapa de entrenamiento.

3 RESULTADOS Y DISCUSIÓN

En el presente capítulo se muestran los resultados obtenidos para los diferentes modelos entrenados. Se detallan las pruebas realizadas con un conjunto de características tradicionales, a continuación, con la misma partición de datos, se presentan las pruebas realizadas para el conjunto combinado con características mejoradas, en las cuales, se varían los tipos de Kernel, de manera que, se podrá realizar una evaluación comparativa de los resultados de ambas combinaciones con diferentes tipos de Kernel. La verificación del funcionamiento del algoritmo se realizará mediante la evaluación de los modelos usando diferentes métodos de validación.

3.1 MÉTODOS DE EVALUACIÓN

Los modelos entrenados emplearon los indicadores en el dominio del tiempo para desarrollar un conjunto de reglas que permiten mapear las correspondencias entre las características y las clases. Este conjunto de reglas fue sometido a varios criterios de evaluación cualitativos y cuantitativos.

Entre los criterios cualitativos para el desarrollo de un algoritmo de clasificación se tiene que: escoger el más simple, que sea de fácil interpretación, cuyo consumo de energía y recursos de capacidad sean reducidos, además, que permita explorar, describir y visualizar los datos a través, de métodos gráficos o mediante un análisis exploratorio de datos.

Entre los métodos cuantitativos se puede medir la exactitud, sensibilidad, especificidad y precisión a partir de la matriz de confusión, esto es indispensable para escoger los modelos que pueden tener mejor desempeño en la vida real y descartar los modelos subóptimos.

Una matriz de confusión permite determinar el desempeño de los algoritmos de aprendizaje supervisado entrenados, su tamaño es de $m \times n$, donde m , es el número de filas o clases ideales y n es el número de columnas que representan las predicciones del modelo, exponen de forma explícita cuando una clase es predicha de manera correcta o errónea. [31]

Exactitud

Determina el porcentaje de datos clasificados correctamente y es igual a la relación entre la suma de los verdaderos positivos (VP) y verdaderos negativos (VN) sobre la suma del número de casos positivos totales (P) y los negativos totales (N). Los casos positivos incluyen los verdaderos positivos (VP) y los falsos negativos (FN), mientras

que, los casos negativos incluyen los verdaderos negativos (VN) y los falsos positivos (FP), cuya representación matemática se encuentra en la Ecuación (3.1).

$$Exactitud = \frac{VP + VN}{P + N} = \frac{VP + VN}{(VP + FN) + (VN + FP)} \quad (3.1)$$

Sensibilidad y Especificidad

Es la razón de los datos clasificados correctamente sobre el número de casos. La sensibilidad es la tasa de los verdaderos positivos, la representación matemática se encuentra en la Ecuación (3.2). La especificidad es la tasa de los verdaderos negativos, la representación matemática se encuentra en la Ecuación (3.3).

$$Sensibilidad = \frac{VP}{VP + FN} \quad (3.2)$$

$$Especificidad = \frac{VN}{VN + FP} \quad (3.3)$$

Precisión

Determina el porcentaje de datos predichos por el modelo correctamente. El valor predictivo positivo es igual a la relación entre los verdaderos positivos (VP) y las predicciones positivas incluyen los verdaderos positivos (VP) y los falsos positivos (FP), la representación matemática se encuentra en la Ecuación (3.4). El valor predictivo negativo es igual a la relación entre los verdaderos negativos (VN) y las predicciones negativas, incluyen los verdaderos negativos (VN) y los falsos negativos (FN), la representación matemática se encuentra en la Ecuación (3.5).

$$VPP = \frac{VP}{VP + FP} \quad (3.4)$$

$$VPN = \frac{VN}{VN + FN} \quad (3.5)$$

3.2 RESULTADOS

3.2.1 RESULTADOS USANDO CARACTERÍSTICAS TRADICIONALES

Los resultados usando características tradicionales toman en cuenta las características presentadas desde la Ecuación (1.1) hasta la Ecuación (1.15).

3.2.1.1 Validación simple por exclusión

El modelo 1, consideró Kernel lineal, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.1.

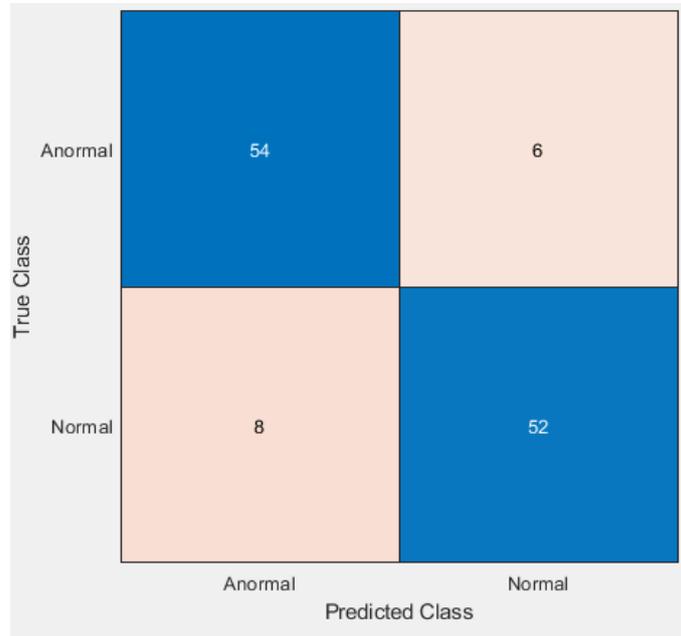


Figura 3.1 Matriz de confusión modelo 1

El modelo 2, consideró Kernel cuadrático, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.2.

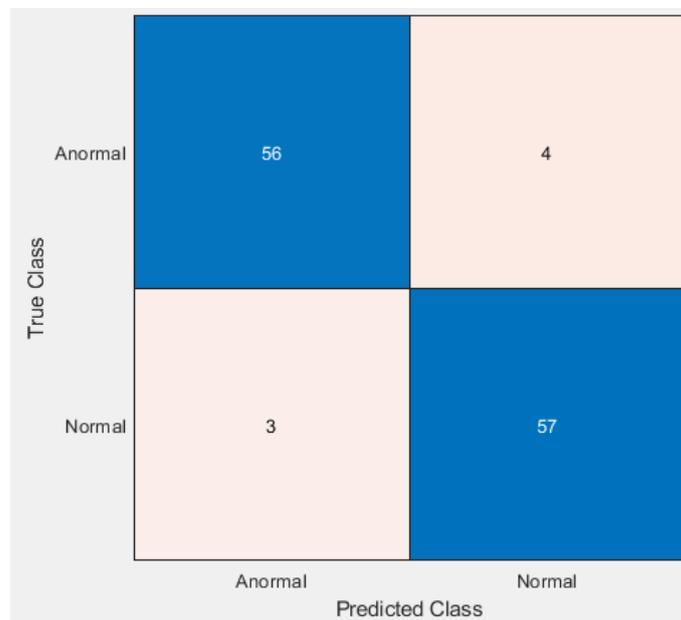


Figura 3.2 Matriz de confusión modelo 2

El modelo 3, consideró Kernel cúbico, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.3.

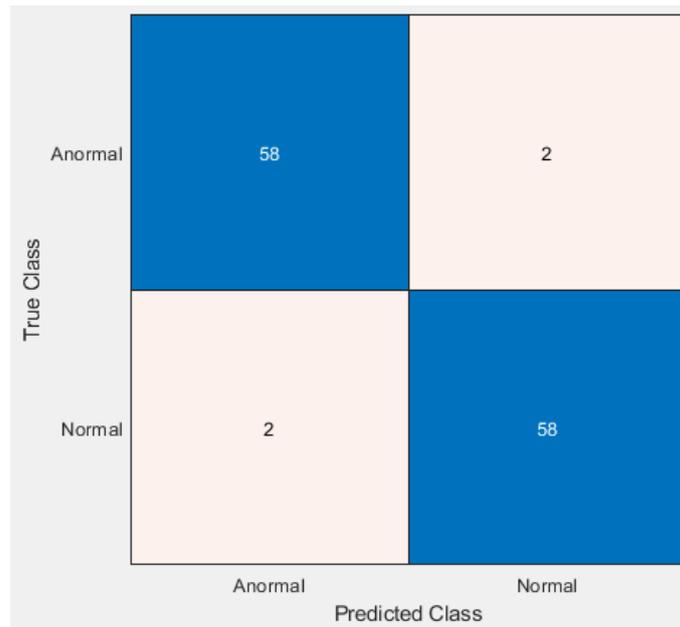


Figura 3.3 Matriz de confusión modelo 3

La Tabla 3.1 muestra los resultados de los tres modelos entrenados, en el primer apartado de resultados se consideraron 15 características tradicionales obtenidas previamente, se varió el Kernel entre lineal, polinomial de grado dos y tres, y se evaluó en cada modelo exactitud, sensibilidad, especificidad y precisión.

Tabla 3.1 Métricas de evaluación

Métrica / Kernel	Exactitud	Sensibilidad	Especificidad	Precisión (VPP)	Precisión (VPN)
Lineal	0.883	0.9	0.867	0.871	0.897
Cuadrático	0.942	0.933	0.95	0.949	0.934
Cúbico	0.967	0.967	0.967	0.967	0.967

3.2.1.2 Validación Cruzada K-fold

El modelo 1, evaluado con validación cruzada, consideró Kernel lineal, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.4.

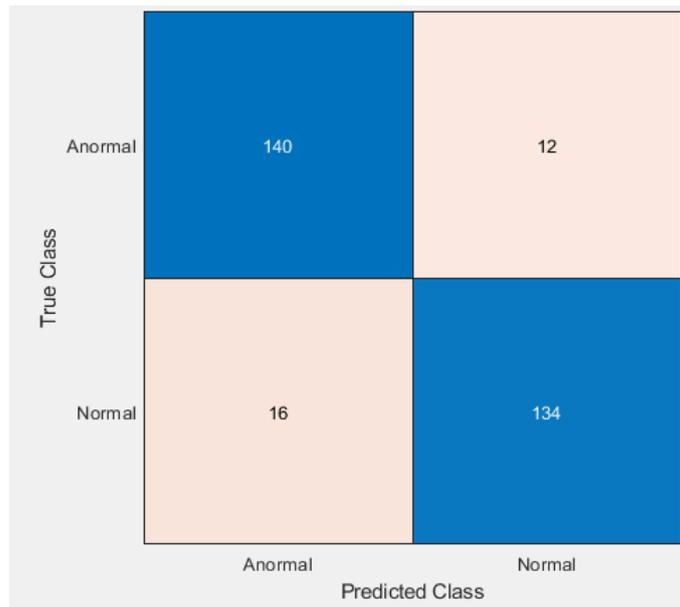


Figura 3.4 Matriz de confusión modelo 1 con Cross-Validation

El modelo 2, evaluado con validación cruzada, consideró Kernel cuadrático, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.5.

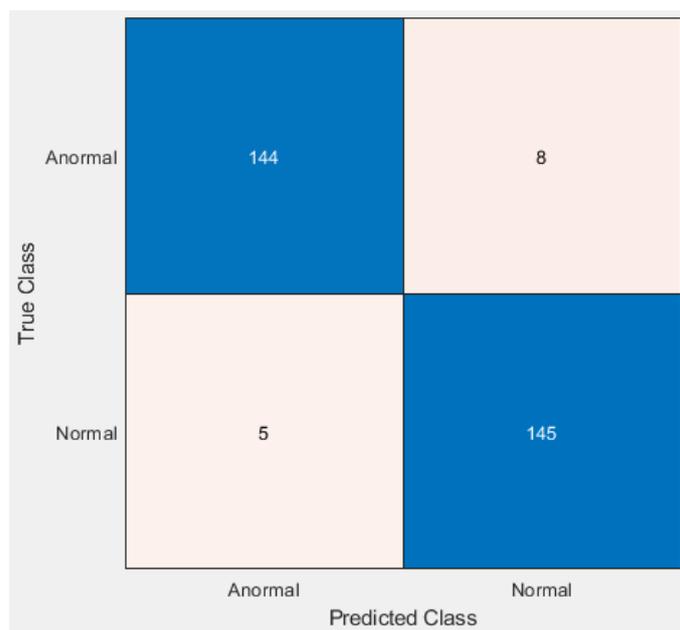


Figura 3.5 Matriz de confusión modelo 2 con Cross-Validation

El modelo 3, evaluado con validación cruzada, consideró Kernel cúbico, usando características tradicionales y se obtuvo la matriz de confusión presentada en la Figura 3.6.

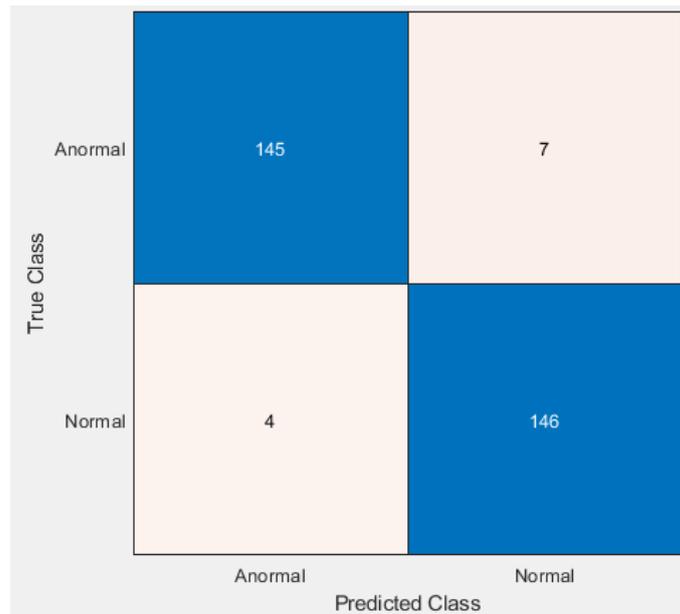


Figura 3.6 Matriz de confusión modelo 3 con Cross-Validation

La Tabla 3.2 muestra los resultados de los tres modelos evaluados con validación cruzada, en la segunda sección de resultados se consideraron 15 características tradicionales obtenidas previamente, se varió el Kernel entre lineal, polinomial de grado dos y tres, y se evaluó en cada modelo exactitud, sensibilidad, especificidad y precisión.

Tabla 3.2 Métricas de evaluación

Métrica / Kernel	Exactitud	Sensibilidad	Especificidad	Precisión (VPP)	Precisión (VPN)
Lineal	0.907	0.921	0.893	0.897	0.918
Cuadrático	0.957	0.947	0.967	0.966	0.948
Cúbico	0.964	0.954	0.973	0.973	0.954

3.2.2 RESULTADOS USANDO CARACTERÍSTICAS MEJORADAS

En los resultados usando características mejoradas se procedió a cambiar las características MAV (1.1) y WL (1.4) por EMVA (1.16) y EWL (1.17), del mismo modo, se toman en cuenta las características presentadas en las Ecuaciones: (1.2), (1.3), y desde la Ecuación (1.5) hasta la Ecuación (1.17). Se varió el set de características de entrenamiento con la finalidad de mejorar el rendimiento del algoritmo. Las características mejoradas son una extensión de las características matemáticas iniciales con una pequeña variación, lo que permite, examinar como contribuye el muestreo en la señal y reducir el tiempo de procesamiento durante la evaluación [18].

3.2.2.1 Validación simple por exclusión

El modelo 4e, se consideró Kernel lineal, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.7.

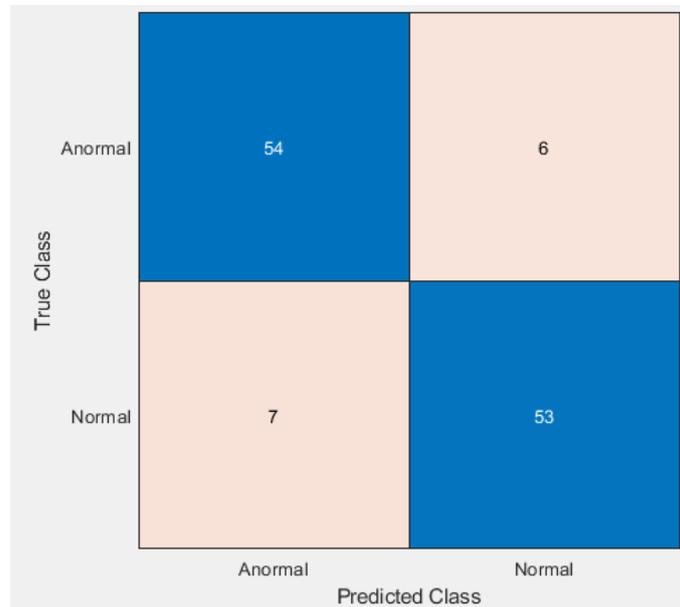


Figura 3.7 Matriz de confusión modelo 4e, usando características mejoradas

El modelo 5e, consideró Kernel cuadrático, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.8.

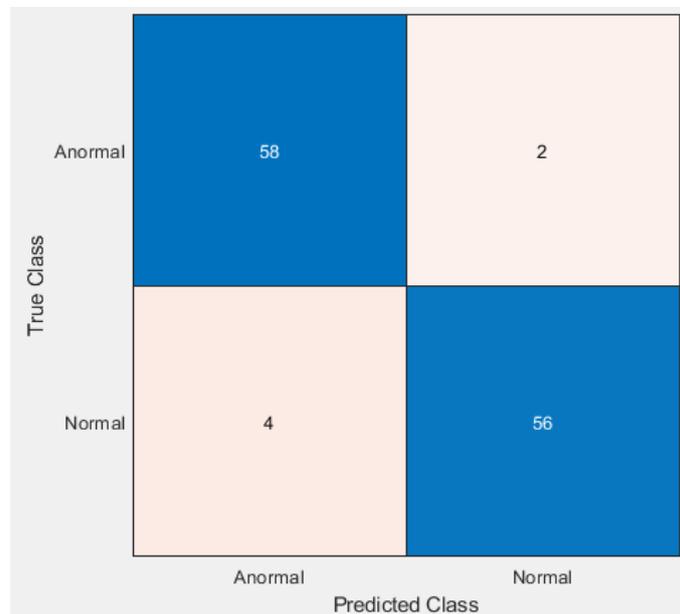


Figura 3.8 Matriz de confusión modelo 5e, usando características mejoradas

El modelo 6e, consideró Kernel cúbico, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.9.

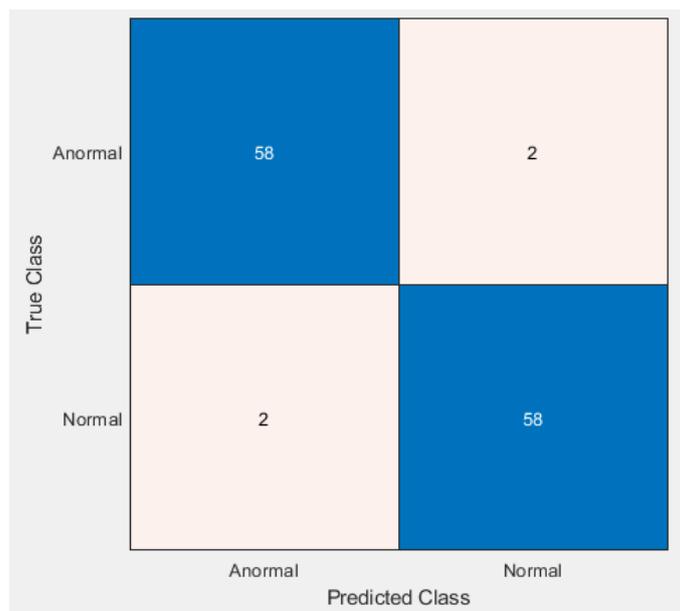


Figura 3.9 Matriz de confusión modelo 6e, usando características mejoradas

La Tabla 3.3 muestra los resultados de los tres modelos entrenados, en el tercer apartado de resultados se consideraron 13 características tradicionales obtenidas previamente y dos variaciones que se designaron como características mejoradas, se varió el Kernel entre lineal, polinomial de grado dos y tres, y se evaluó en cada modelo exactitud, sensibilidad, especificidad y precisión.

Tabla 3.3 Métricas de evaluación para los modelos con características mejoradas

Métrica / Kernel	Exactitud	Sensibilidad	Especificidad	Precisión (VPP)	Precisión (VPN)
Lineal	0.892	0.900	0.883	0.885	0.898
Cuadrático	0.95	0.967	0.933	0.935	0.966
Cúbico	0.967	0.967	0.967	0.967	0.967

3.2.2.2 Validación Cruzada K-fold

El modelo 4e, evaluado con validación cruzada, consideró Kernel lineal, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.10.

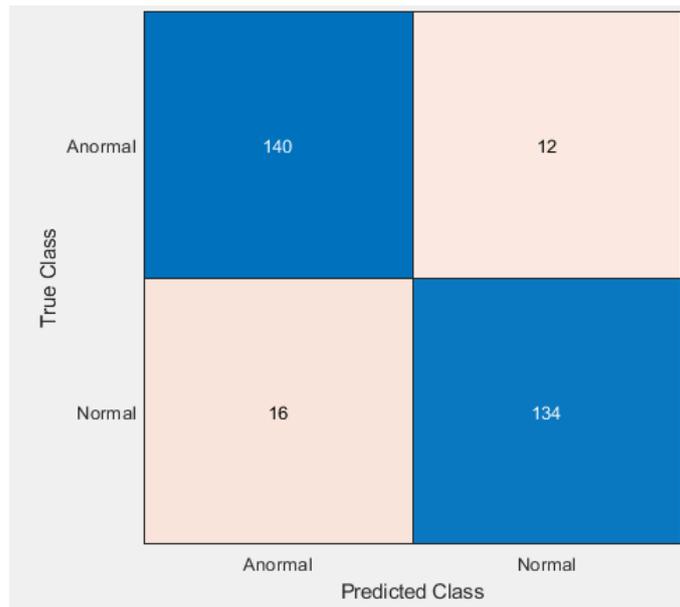


Figura 3.10 Matriz de confusión modelo 4e con Cross-Validation

El modelo 5e, evaluado con validación cruzada, consideró Kernel cuadrático, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.11.

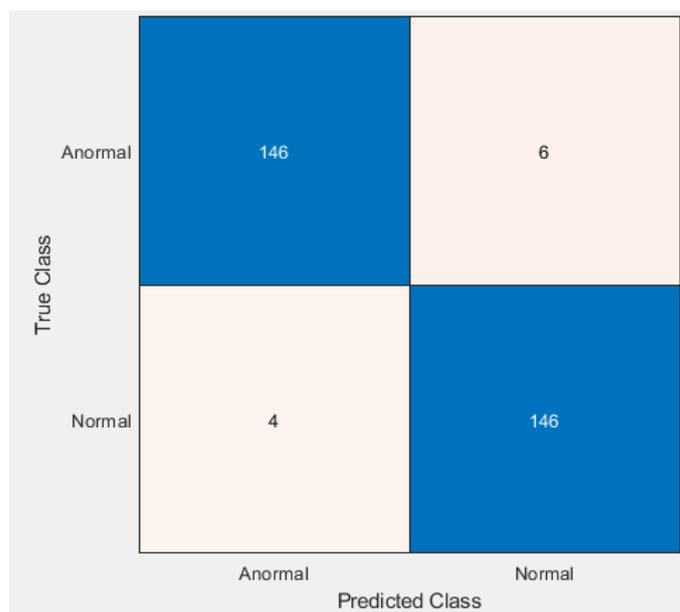


Figura 3.11 Matriz de confusión modelo 5e con Cross-Validation

El modelo 6e, evaluado con validación cruzada consideró Kernel cúbico, usando características mejoradas y se obtuvo la matriz de confusión presentada en la Figura 3.12 .

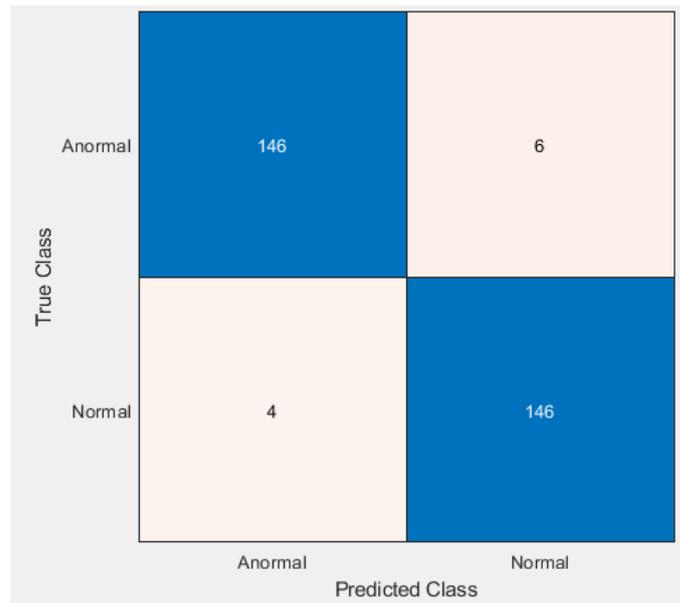


Figura 3.12 Matriz de confusión modelo 6e con Cross-Validation

La Tabla 3.4 Tabla 3.3 Métricas de evaluación muestra los resultados de los tres modelos entrenados con validación cruzada, el cuarto apartado de resultados se consideró 13 características tradicionales obtenidas previamente y dos variaciones que se designaron como características mejoradas, se varió el Kernel entre lineal, polinomial de grado dos y tres, y se evaluó en cada modelo exactitud, sensibilidad, especificidad y precisión.

Tabla 3.4 Métricas de evaluación de los modelos con características mejoradas

Métrica / Kernel	Exactitud	Sensibilidad	Especificidad	Precisión (VPP)	Precisión (VPN)
Lineal	0.907	0.921	0.893	0.897	0.918
Cuadrático	0.957	0.947	0.967	0.966	0.948
Cúbico	0.967	0.961	0.973	0.971	0.961

3.3 DISCUSIÓN

La Tabla 3.1 y Tabla 3.2, muestran los resultados de los modelos entrenados con 15 características tradicionales, los modelos óptimos en base a las métricas obtenidas se consiguieron con el Kernel polinomial de grado 3, la exactitud y precisión del mejor modelo que empleo Kernel cúbico y validación por simple exclusión fue 96.7% y 96.7%, respectivamente; la exactitud y precisión del modelo que empleó validación cruzada alcanzó 96.4% y 97.3%, respectivamente. El modelo de validación por simple exclusión da un resultado excelente; además, es el más cercano a la realidad porque utiliza observaciones no vistas por el modelo durante el entrenamiento para medir su eficacia. Sin embargo, los algoritmos entrenados con el método de validación cruzada tienen

mayor exactitud y precisión usando Kernel polinomial de grado 2 y lineal porque se promedian los resultados de las métricas alcanzadas en cada partición.

La Tabla 3.3 y Tabla 3.4 muestran los resultados de los modelos entrenados que usaron características mejoradas. El método de validación por simple exclusión como se esperaba aumenta su eficacia en cada modelo, exceptuando la configuración de Kernel cúbico donde solo mejora su eficiencia, sin embargo, con el método de validación cruzada mejora su exactitud únicamente el modelo que se entrena con Kernel cúbico, mientras los modelos configurados con Kernel lineal y cuadrático no aumentan su exactitud solo su eficiencia.

Al comparar los resultados de los datos de evaluación final, es notable que los modelos propuestos, a pesar de ser sometidos a diferentes métodos de validación para clasificar señales EMG normales y anormales de la base de datos obtenida, alcanzan un eficacia y eficiencia considerable sobre todo empelando el método Kernel cúbico. Esto puede deberse a que a simple vista parece un problema de clasificación lineal pero el preprocesamiento y extracción de características le permiten al modelo que emplea Kernel polinomial de grado 3, obtener una mejor caracterización de patrones. Además, se tiene que un adecuado tratamiento de datos de entrenamiento, permite mejorar los resultados de los modelos entrenados, pues la máxima exactitud alcanzada fue de 96.7%, mientras que, en el estudio [19], usan la misma base de datos con diferente preprocesamiento y un método de validación cruzada de 10 K-fold, logrando alcanzar una exactitud de 70.1% con el mismo método de clasificación SVM.

Se elaboró una interfaz de usuario que permite ingresar nuevos datos, adquiridos con el mismo sensor u otro diferente, siempre y cuando se mantenga el formato de archivo. Además, se observa de forma interactiva los resultados de la clasificación y el funcionamiento de los modelos entrenados; facilitando al usuario la comparación de las señales en crudo y preprocesadas digitalizadas gráficamente su funcionalidad se presenta en el Anexo A.

Una vez presentados los resultados, es necesario comentar que, a pesar limitada cantidad de datos, se obtuvieron buenos resultados sin incrementar señales en la base de datos, finalmente, se presenta las características del ambiente en donde fueron entrenados los algoritmos, se empleó una computadora HP ® Pavilion ® g4 Notebook con un procesador AMD CPU E2-1800 APU, Gráficos AMD Radeon™ HD 7340 a 1.70 GHz, y con 8 GB de RAM.

4 CONCLUSIONES Y RECOMENDACIONES

En el presente capítulo se presentan las conclusiones y recomendaciones del trabajo de titulación, debido a que, finalizadas las etapas se cumple satisfactoriamente con todos los objetivos y alcances planteados.

4.1 CONCLUSIONES

- Se emplearon tres técnicas de Kernel: lineal, cuadrático y cúbico para entrenar doce modelos de clasificación de señales EMG, beneficiando el mapeo del conjunto de observaciones no linealmente separable en un espacio de alta dimensión y permitiendo transformarlas en linealmente separable.
- Se diseñó un algoritmo de clasificación que incluyó etapas de preprocesamiento (segmentación, rectificación y filtrado) que contribuyeron para la extracción de características en el dominio del tiempo y al entrenamiento de varios modelos mediante aprendizaje supervisado.
- Se analizó el comportamiento muscular asociado a la extremidad inferior mediante las señales de los electrodos ubicados en los músculos Recto Femoral, Biceps Femoral, Vastomedial y Semitendinoso que permitieron el estudio de la actividad de las neuromotoras durante la ejecución del movimiento en la rodilla, mediante digitalización y preprocesamiento de bioseñales.
- Se desarrolló un algoritmo, que permite detectar si una persona presenta una patología en la rodilla mediante la comparación de distintos conjuntos de características comunes y mejoradas, obteniendo que la SVM tiene mejor desempeño al trabajar con un conjunto adecuado de características mejoradas, lo cual es útil en aplicaciones donde se tienen pocas observaciones de entrenamiento.
- Se obtuvieron buenas métricas de exactitud en los modelos propuestos gracias a varias particiones de las observaciones de entrenamiento y validación generados de forma aleatoria para la base de datos seleccionada hasta conseguir modelos óptimos de clasificación. Además, durante estas pruebas se utilizó una SVM multiclase y un enfoque uno contra uno que permitió no sobreajustar los datos de entrenamiento.
- Se realizó la evaluación de cada modelo propuesto a través de dos métodos: validación por exclusión y validación cruzada, los cuales permitieron ver los

modelos que cumplen con los criterios cualitativos y cuantitativos más aptos para ser probados en la vida real.

- Se concluye que el algoritmo basado en electromiografía es el primer paso para integrar una herramienta valiosa en el diagnóstico clínico, que pretende servir como soporte de diagnóstico al profesional médico/fisioterapeuta, a fin de, determinar que individuos están en riesgo para ser sometidos a exámenes más exhaustivos. Como las EMG son señales complejas, realizar un diagnóstico fiable requiere de un médico especialista.

4.2 RECOMENDACIONES

- Se recomienda el uso de una mayor cantidad de datos o señales durante el entrenamiento para mejorar la efectividad del algoritmo, ya que, la base de datos que se escogió cumplió con los requerimientos de estudio, pero mientras más señales se tenga, el algoritmo caracterizará los patrones de clasificación de mejor manera.
- Para adquirir nuevas señales durante el uso de la herramienta, los electrodos se deben colocar en los mismos músculos de estudio: Recto Femoral, Biceps Femoral, Vasto Medial y Semitendinoso, para garantizar el correcto funcionamiento del software desarrollado; se sugiere una adecuada adquisición de datos con el sensor, esto incluye: limpiar la piel de sudor u otras impurezas que puedan afectar al electrodo y evitar el ruido excesivo externo sea electrónico u otro tipo.
- Se sugiere que el archivo de carga tenga formato de texto (.txt), o de ser necesario se realice una conversión previa empleando otro Software apropiado para que el archivo tenga el formato adecuado.
- A pesar de que la herramienta clasifica a los sujetos normales y los sujetos con una anormalidad en su extremidad inferior, se propone en futuros trabajos adquirir una base de datos más grande, que permita a la herramienta diferenciar el tipo de patología que el sujeto tiene, además se recomienda emplear un clasificador multiclase para esto.

5 REFERENCIAS BIBLIOGRÁFICAS

- [1] O. Aviles, J. L. Rodriguez, M. Herrera, and G. Martínez, “EMG Data set in Lower Limb”, 2013. [Online]. Disponible: <http://archive.ics.uci.edu/ml/datasets/emg+dataset+in+lower+limb>.
- [2] C. Cifuentes, F. Martínez, and E. Romero, “Análisis teórico y computacional de la marcha normal y patológica: una revisión,” *Revista Med*, vol. 18, no. 2, pp. 182–196, 2010, doi: <https://doi.org/10.18359/rmed.1311>.
- [3] R. Gago, “Uso de algoritmos de aprendizaje automático aplicados a bases de datos genéticos”, MSc. Tesis en Bioinformática y Bioestadística, Universitat Oberta de Catalunya, 2017.
- [4] M. Oskoei and H. Hu, “Support vector machine-based classification scheme for myoelectric control applied to upper limb,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, 2008, doi: [10.1109/TBME.2008.919734](https://doi.org/10.1109/TBME.2008.919734).
- [5] M. Atzori *et al.*, “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” *Sci. Data*, vol. 1, pp. 1–13, 2014, doi: [10.1038/sdata.2014.53](https://doi.org/10.1038/sdata.2014.53).
- [6] M. Noguera, “Bases de Anatomía Topográfica,” in *Generalidades de Anatomía Topográfica*, San Miguel de Tucumán, 2015, pp. 21–24.
- [7] C. Beltre, “Anatomía Topográfica,” 2021.[Online]. Disponible: <https://anatomiatopografica.com/>.
- [8] Á. Moreno Sanz, “Procesado Avanzado De Señal Emg”, Tesis de Electrónica Industrial y Automática, Universidad Carlos III de Madrid, Escuela Politécnica Superior. 2017.
- [9] C. Rios and V. Quintero, “Control de una mano virtual usando señales electromiográficas”, Tesis de Mecatrónica, Universidad Militar de Nueva Granada, Bogotá, 2010.
- [10] E. Avila and E. Montes, “Obtención y análisis de señales electromiográficas de las articulaciones tibiofemoral y femororotuliana aplicada a la detección y rehabilitación de problemas musculares en el movimiento de la rodilla”, Tesis de Electrónica, Universidad Politécnica Salesiana Sede Cuenca, 2012.
- [11] L. Unapanta, “Reconocimiento de gestos de la mano en tiempo real basado en señales electromiográficas utilizando myo armband con wavelets y máquinas de

vectores de soporte”, Tesis Facultad de Sistemas, *Escuela Politécnica Nacional*, Quito, Ecuador, 2019, pp. 4-18.

- [12] E. Guzmán and G. Méndez, “Electromyography in the rehabilitation sciences,” *Salud Uninorte*, vol. 34, no. 3, pp. 753–765, 2018, doi: 10.14482/sun.34.3.616.74.
- [13] J. Bagnato, “¿Qué es Machine Learning? Aprende Machine Learning,” 2017, Agosto 27. [Online]. Disponible: <https://www.aprendemachinelearning.com/que-es-machine-learning/>.
- [14] P. Tárrega, “Herramientas Imprescindibles para Machine Learning Video-MATLAB”, 2016. [Online] Disponible: <https://www.mathworks.com/videos/essential-tools-for-machine-learning-1488913197554.html>.
- [15] “Machine Learning with MATLAB,” MathWorks, 2020. [Online]. Disponible: <https://uk.mathworks.com/solutions/machine-learning.html>.
- [16] “MATLAB for Predictive Maintenance - MATLAB & Simulink”, Mathworks, [Online]. Disponible: <https://www.mathworks.com/solutions/predictive-maintenance.html>.
- [17] M. Ortiz Catalan, R. Branemark, and B. Hakansson, “BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms,” *Source Code Biol. Med.*, vol. 8, pp. 1–18, 2013, doi: 10.1186/1751-0473-8-11.
- [18] J. Too, A. R. Abdullah, and N. M. Saad, “Classification of Hand movements based on discrete wavelet transform and enhanced feature extraction,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 6, pp. 83–89, 2019, doi: 10.14569/ijacsa.2019.0100612.
- [19] A. Vijayvargiya, R. Kumar, N. Dey et al., “Comparative Analysis of Machine Learning Techniques for the Classification of Knee Abnormality,” *IEEE Transactions on Biomedical Engineering International Conference on Computing Communication and Automation. ICCCA*, pp. 1–6, 2020, doi: 10.1109/ICCCA49541.2020.9250799.
- [20] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Feature reduction and selection for EMG signal classification”, *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012, doi: 10.1016/j.eswa.2012.01.102.
- [21] D. Prashanth, R. Mehta, and N. Sharma, “Classification of Handwritten Devanagari Number An analysis of Pattern Recognition Tool using Neural

- Network and CNN,” in *Procedia Computer Science*, Enero, 2020, vol. 167, pp. 2445–2457, doi: 10.1016/j.procs.2020.03.297.
- [22] S. Kumar, “Data splitting technique to fit any Machine Learning Model,” 2020. [Online]. Disponible: <https://towardsdatascience.com/data-splitting-technique-to-fit-any-machine-learning-model-c0d7f3f1c790>.
- [23] Mathworks, “Select Data and Validation for Classification Problem | Matlab & Simulink”, 2021. [Online] Disponible: <https://www.mathworks.com/help/stats/select-data-and-validation-for-classification-problem.html>.
- [24] E. Allibhai, “Hold-out vs. Cross-validation in Machine Learning,” 2018. [Online]. Disponible: <https://medium.com/@ejjaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f>.
- [25] A. Godoy, “Machine learning techniques used for text mining”, *Investigación bibliotecológica*, vol. 31, no. 71, pp. 103–126, 2017, doi: 10.22201/iibi.0187358xp.2017.71.57812.
- [26] “Support Vector Machine (SVM) - MATLAB & Simulink.” [Online]. Disponible: <https://www.mathworks.com/discovery/support-vector-machine.html>.
- [27] Biometrics Ltd, “DataLOG para adquisición y supervisión de datos portátiles,” 2021. [Online]. Disponible: <https://www.biometricsltd.com/datalog.htm> (accessed Jun. 25, 2021).
- [28] A. García, “12 consejos de ergonomía en el trabajo industrial.” [Online]. Disponible: <https://www.ofiprix.com/blog/12-consejos-de-ergonomia-en-el-trabajo-industrial/> (accessed Jun. 25, 2021).
- [29] O. Banos, J. Galvez, M. Damas, H. Pomares, and I. Rojas, “Window size impact in human activity recognition,” *Sensors (Switzerland)*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014, doi: 10.3390/s140406474.
- [30] A. Dehghani, O. Sarbishei, T. Glatard, and E. Shihab, “A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors,” *Sensors (Switzerland)*, vol. 19, no. 22, pp. 10–12, 2019, doi: 10.3390/s19225026.
- [31] P. Recuero de los Santos, “Machine Learning a tu alcance: La matriz de confusión,” *Blog think big- Telefónica Tech*, 2018. [Online]. Disponible: <https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/>.

ANEXOS

Anexo A. Manual de Usuario

ANEXO A

MANUAL DE USUARIO

CONTENIDO

A.1 Descripción de la herramienta

A.2 Ejecución de la interfaz

A.3 Pantallas de la herramienta

A.4 Formato de Archivos para cargar

A.1 DESCRIPCIÓN DE LA HERRAMIENTA

La interfaz de usuario se desarrolló empleando el software de MATLAB ® versión 2021A, con el objetivo de conseguir una visualización dinámica con el resultado de los modelos entrenados. Se utilizó la herramienta App Designer de Matlab ® para integrar nuevos datos, esto incluye, gráficas de señales en crudo y señales preprocesadas, además, la clasificación con los modelos entrenados y las predicciones resultantes.

A.2 EJECUCIÓN DE LA INTERFAZ

Ejecutar la interfaz en una versión de Matlab 2021A o superior y en la Carpeta Actual cargar “1.6 Interfaz” que contiene los archivos de la herramienta. A continuación, correr el archivo “Inicio.mlapp” como se muestra en la Figura A.1.

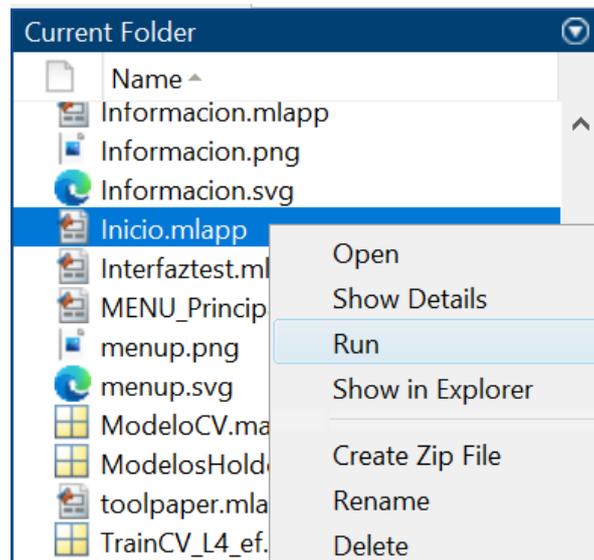


Figura A.1 Correr la ventana Inicial de la interfaz de usuario

A.3 PANTALLAS DE LA HERRAMIENTA

Dentro de la interfaz se tienen 4 ventanas: inicio, menú principal, información, clasificador.

La Figura A.2 muestra la ventana inicial de la interfaz y cuenta con un botón que lleva al menú principal.



Figura A.2 Ventada de inicio

La Figura A.3 despliega la ventana menú que permite navegar por las distintas opciones de la herramienta.

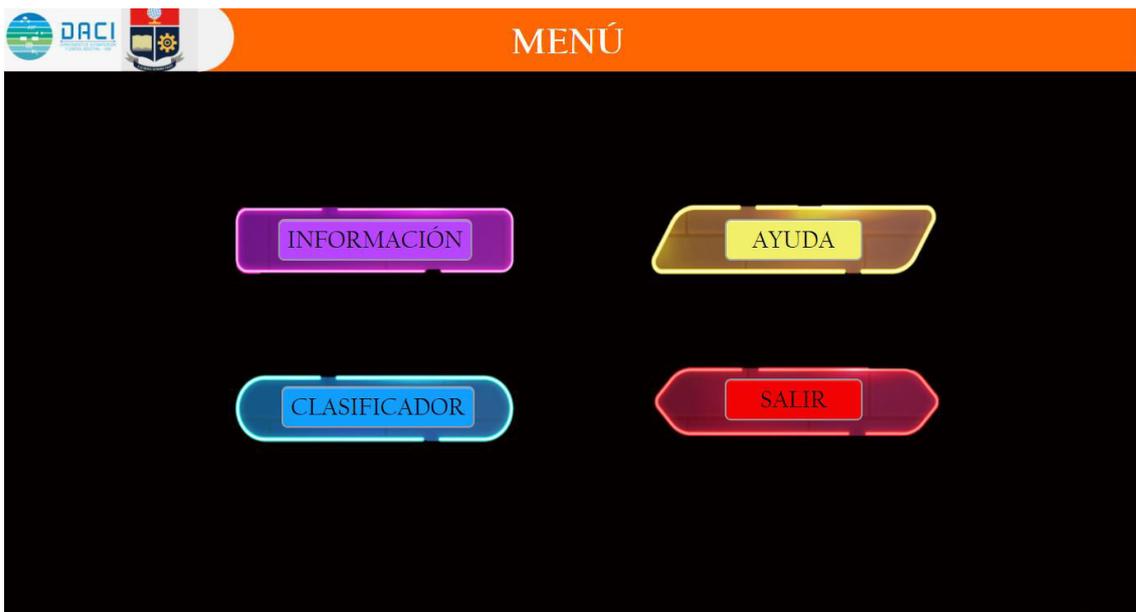


Figura A.3 Menú principal

La Figura A.4 despliega una breve descripción que le permite al usuario familiarizarse con la herramienta y su utilidad.



Figura A.4 Ventana Información

La Figura A.5 muestra las cuatro etapas principales del clasificador de señales: lectura de nuevos datos, gráficas de las señales, modelos entrenados y finalmente resultados.



Figura A.5 Herramienta de clasificación

Sección A. Permite al usuario cargar nuevos archivos de datos, siempre que estos tengan formato válido, ver apartado A4.

Sección B. Presenta los modelos entrenados con su respectiva exactitud.

Sección C. El usuario selecciona entre la visualización en el dominio del tiempo de las señales en crudo o preprocesadas de los electrodos colocados en los músculos: Recto Femoral, Biceps Femoral, Vasto Medial, Semitendinoso.

Sección D. Muestra el resultado de la herramienta para modelo escogido, sea Normal si el sujeto no presenta anomalías o Anormal en caso de presencias patológicas.

A.4 FORMATO DE ARCHIVOS PARA CARGAR

El archivo debe tener formato de texto (i.e., .txt), con 7 líneas de encabezado para descripción y estar compuesto por 5 columnas como se observa en la Tabla A.1.

Tabla A.1 Atributos de información de los archivos de texto

Canal	Ch1	Ch2	Ch3	Ch4	Ch5
Músculo	RF	BF	VM	ST	FX
Columna	0	1	2	3	4

Canal: corresponde al electrodo fijado a un músculo.

Músculo: corresponde al músculo sujeto a medición.

RF: Recto Femoral.

BF: Biceps Femoral.

VM: Vasto Medial.

ST: Semitendinoso.

FX: Ángulo de Flexión en la rodilla.

Cada columna de cada electrodo debe tener el mismo número de muestras.

Finalmente, el botón de ayuda despliega el manual de usuario.

Si se desea Salir de la Herramienta se puede seleccionar el botón Salir del Menú, o a su vez, el botón de cierre en la esquina superior derecha de cada ventana, que despliega una ventana de confirmación

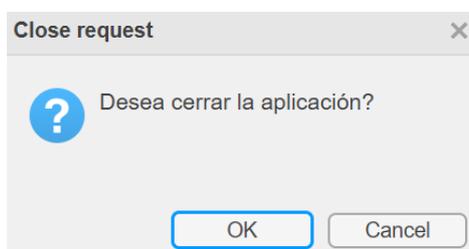


Figura A.6 Mensaje de confirmación de cierre

Existen dos tipos de errores, el primero es cargar un archivo con un formato inválido o en caso de no existir ningún archivo de datos cargado se despliega un mensaje de error, se debe cargar un archivo con datos en formato de texto.

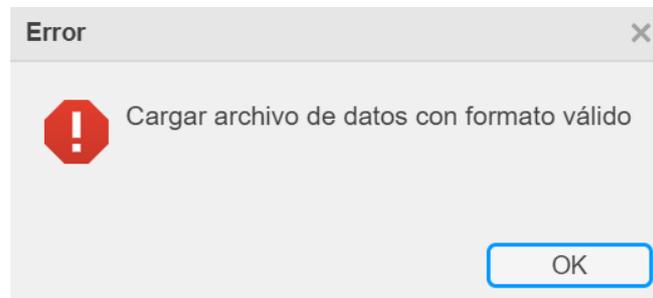


Figura A.7 Mensaje de error, archivo inválido.

Para realizar una predicción con cualquier modelo se debe seleccionar uno, caso contrario la herramienta despliega un mensaje de error.

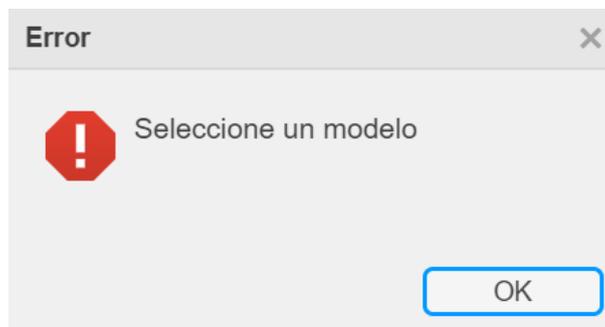


Figura A.8 Mensaje de error, resultado inválido