

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California**



**Maestría en Ciencias
en Ciencias de la Computación**

**Detección de gestos por medio de relojes inteligentes
para el etiquetado de datos en el sensado participativo para
"reconocimiento de actividad"**

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Luis Antonio González Jasso

Ensenada, Baja California, México

2017

Tesis defendida por
Luis Antonio González Jasso

y aprobada por el siguiente Comité

Dr. Jesús Favela Vara
Director de tesis

Dra. María del Carmen Maya Sánchez

Dr. José Antonio García Macías



Dr. Jesús Favela Vara
Coordinador del Posgrado en Ciencias de la Computación

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Luis Antonio González Jasso © 2017

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis.

Resumen de la tesis que presenta Luis Antonio González Jasso como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación.

Detección de gestos por medio de relojes inteligentes para el etiquetado de datos en el sensado participativo para "reconocimiento de actividad"

Resumen aprobado por:

Dr. Jesús Favela Vara
Director de tesis

Para hacer uso de los algoritmos supervisados de reconocimiento de actividades, es necesario contar con datos etiquetados para entrenar modelos de clasificación. El etiquetado de la actividad se puede realizar mediante la **observación**, en **condiciones controladas** y por medio de **auto-etiquetado**. Las dos primeras formas de etiquetado se realizan de forma intrusiva, lo que la hace una tarea tediosa, tanto para quien realiza la actividad a etiquetar como para quien registra y etiqueta la información generada. En esta tesis se propone una técnica para el etiquetado de datos de la actividad de personas, por medio de gestos sutiles y fáciles de realizar, que son sensados y reconocidos mediante relojes inteligentes. Las señales obtenidas del sensor inercial del reloj inteligente son usadas para entrenar algoritmos de clasificación para poder detectar estos gestos. Se obtuvieron datos de 15 participantes que realizaron los seis gestos propuestos en tres posiciones diferentes. Con 208 características obtenidas a partir de las señales del acelerómetro y giroscopio se entrenaron dos algoritmos de clasificación (Maquinas de Soporte Vectorial y Optimización Mínima Secuencial), para detectar los seis gestos propuestos. El resultado obtenido es de una precisión del 81% para el uso de los seis gestos sutiles y un 91% en el uso de los primeros tres gestos sutiles.

Palabras clave: **Reconocimiento de gestos, sensado participativo, interfaces vestibles, etiquetado de datos, procesamiento de señales.**

Abstract of the thesis presented by Luis Antonio González Jasso as a partial requirement to obtain the Master of Science degree in in Computer Science

**Detection of gestures using smart watches for data labeling
in participatory sensing for "activity recognition"**

Abstract approved by:

Dr. Jesús Favela Vara
Director de tesis

The use of supervised activity recognition algorithm requires labeled data in order to train classification models. Labeling an activity can be performed through **observation**, in **controlled conditions**, or by **self-labeling**. The two first approaches are intrusive, which makes the task tedious for the person performing the activity, as well as for the one tagging the activity. In this work I propose a technique for activity labeling using subtle gestures that are simple to execute, and that can be sensed and recognized using smartwatches. The signals obtained by the inertial sensor in a smartwatch are used to train classification algorithms to identify the gesture. We obtained data from 15 participants who executed 6 proposed gestures in 3 different positions. 208 characteristics were computed from the accelerometer and gyroscope signals and were used to train to classification algorithms (Support Vector Machine and Sequential Minimal Optimization) to detect the six proposed gestures. The results obtained achieve a precision of 81% for the 6 subtle gestures, and 91% when using only the first 3 gestures.

Keywords: **Gesture recognition, participatory sensing, wearable interfaces, data labeling, signal processing.**

Dedicatorias

A mi familia, por acompañarme
en esta increíble aventura.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología por tener estos excelentes programas de becas.

Al Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), por darme abrigo durante el proceso de formación.

Al Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP), por permitirme continuar preparándome.

Al Dr. Jesús Favela Vara por su sencillez, sapiencia y por su apoyo.

A mi comité de tesis, al Dr. José Antonio García Macías y a la Dra. María del Carmen Maya Sánchez por sus aportaciones a este trabajo.

A todos los doctores del posgrado de Ciencias de la Computación.

A todas aquellas personas que participaron de alguna manera en la elaboración de este estudio y que me es imposible mencionarlas a todas.

A la ciudad de Ensenada por permitirme a mí y a mi familia enamorarme de ella.

Tabla de contenidos

Resumen en español	ii
Resumen en inglés	iii
Dedicatorias	iv
Agradecimientos	v
Lista de figuras	ix
Lista de tablas.....	xii
Capítulo 1. Introducción.....	1
1.1. Planteamiento del problema	1
1.2. Objetivos	2
1.2.1. Objetivo general	2
1.2.2. Objetivos específicos	2
1.3. Preguntas de investigación	2
1.4. Propuesta de solución.....	2
1.5. Metodología.....	3
1.6. Contribución.....	4
1.7. Organización de la tesis	4
Capítulo 2. Fundamentos teóricos.....	6
2.1. Introducción.....	6
2.2. Etiquetado de datos.....	6
2.2.1. Etiquetado de datos por medio de la observación.....	8
2.2.2. Etiquetado de datos en condiciones controladas de sensado	9
2.2.3. Auto-etiquetado de la actividad	10
2.3. Sensado participativo y cómputo vestibular.....	11
2.4. Uso del cómputo vestibular para la detección de los gestos	12
2.5. Estado del arte en la detección de gestos	13
2.6. Eliminación de la gravedad	16
2.7. Conclusión.....	18
Capítulo 3. Propuesta de reconocimiento de gestos sutiles realizados con la mano	19

3.1. Posición y orientación de los sensores	19
3.2. Detectar el inicio y el fin de un gesto sutil	20
3.3. Distorsión Dinámica Temporal para determinar si en la señal se encuentra un gesto	20
3.4. Segmentación de la secuencia de datos	30
3.4.1. Establecer el inicio y el fin de un gesto sutil	33
3.5. Extracción de características.....	36
3.6. Uso de aprendizaje máquina para establecer dos modelos en la detección de los gestos sutiles	39
3.6.1. Máquinas de Soporte Vectorial	40
3.6.2. Optimización Mínima Secuencial.....	43
3.6.3. Escalamiento de datos.....	44
Capítulo 4. Evaluación de estrategia propuesta para el reconocimiento de gestos sutiles	45
4.1. Selección del reloj inteligente.....	45
4.2. Selección de gestos sutiles y posturas para realizar estos gestos	46
4.3. Selección de sujetos para entrenamiento y pruebas de los algoritmos.....	48
4.4. Adquisición de datos.....	51
4.5. Validación cruzada para determinación de parámetros.....	53
4.6. Uso de herramienta de minería de datos WEKA	55
4.7. Filtrado de datos	57
4.8. Implementación de modelos	57
4.9. Evaluación de la detección de los gestos sutiles.....	59
4.10. Conclusión.....	60
Capítulo 5. Resultados	63
5.1. Desarrollo de los gestos sutiles.....	63
5.2. Uso de la segmentación de la secuencia de datos.....	63
5.3. Detección de posibles gestos por medio del algoritmo DTW	65
5.4. Selección de parámetros para SVM	65
5.5. Efectividad de cada uno de los gestos en SVM.....	66
5.6. Selección de parámetros para SMO.....	68
5.7. Efectividad de cada uno de los gestos en SMO	68
5.8. Clasificación de los tres primeros gestos	69

5.9. Clasificación de los últimos tres gestos.....	70
5.10. Clasificación de un conjunto de datos suministrado	70
5.11. Discusión	71
Capítulo 6. Conclusiones y trabajo futuro	73
6.1. Conclusiones	73
6.2. Contribución.....	73
6.3. Limitaciones	74
6.4. Trabajo futuro	74
Literatura citada	76

Lista de figuras

Figura	Página
1. Figura 1: Proceso para crear un modelo para el reconocimiento de patrones a usarse en alguna aplicación. [6]	8
2. Figura 2: Representación de la observación directa al dormir, sin intervenir en las actividades del sujeto a observar.....	9
3. Figura 3: Representación del etiquetado en condiciones controladas de la actividad de dormir.	10
4. Figura 4: Representación del auto-etiquetado de la actividad de dormir. El usuario indica el inicio y el fin del sueño.....	11
5. Figura 5: Efecto de la aceleración de la gravedad sobre cada eje dependiendo de la posición de la mano. a) posición de la mano en la que afecta al eje X y b) posición de la mano en la que afecta al eje Z.	18
6. Figura 6: Orientación de los ejes de los sensores con respecto a la mano.....	19
7. Figura 7: Matriz de costos de dos secuencias X (eje vertical) y Y (eje horizontal) generadas por el método DTW. En este caso, ambas secuencias son de distinto tamaño.	22
8. Figura 8: Ruta de distorsión con costo mínimo entre la secuencia X y Y, llamada ruta de distorsión.....	23
9. Figura 9: Ruta de distorsión que no cumple con la monotonicidad. En el rectángulo rojo aparece un elemento que no cumple con la regla $n_k - n_{k-1} \geq 0$, ya que $2 - 3 < 0$	23
10. Figura 10: Movimiento para el cálculo de los costos de la matriz de costos, realizado en un tiempo $O(nm)$, siempre cumpliendo con las 3 condiciones.....	25
11. Figura 11: Ventana de distorsión de tamaño r . La ruta de distorsión debe mantenerse dentro de los límites establecidos por esta ventana.	27
12. Figura 12: Restricción de pendiente donde la relación proporcional entre p y q no debe ser mayor al 20%, es decir, de cada 5 celdas que se acumulen en q debe de haber al menos 4 celdas previas en p y viceversa.	28
13. Figura 13: Ejemplo de (a) señal de referencia, (b) posible señal de un gesto y (c) señal del movimiento de la mano formando círculos manteniendo el antebrazo estático.	29
14. Figura 14: Ruta de distorsión generada a partir de la comparación de la señal de referencia y un posible gesto.	29
15. Figura 15: Matriz de distorsión generada a partir de la comparación de la señal de referencia y la señal con ruido que no pertenece a ninguno de los gestos sutiles.	30

16. Figura 16: A la izquierda está la segmentación en secuencia de seis datos (salvo el final) provenientes del giroscopio para encontrar el segmento donde inicia y termina el gesto. En cada segmento se obtendrá el promedio (marcado por la línea verde punteada) para poder simplificar la señal y obtener una señal simplificada, como se muestra en la figura de la derecha, que permita eliminar datos con ruido.31
17. Figura 17: Buscando el inicio y el fin preliminares en la serie de tiempo proveniente de la segmentación.32
18. Figura 18: Secuencia de datos provenientes de la suma de las magnitudes de los ejes del giroscopio acotados según los umbrales aplicados a la segmentación.35
19. Figura 19: Búsqueda del inicio y fin del gesto mediante umbrales tanto de inicio, en color rojo, y de fin, en color verde, fijados en 0.19 y 0.17 respectivamente.36
20. Figura 20: Recorrido hacia la izquierda por no cumplir la regla de estar a una distancia de 220 ms de un valor mínimo de 0.4. El inicio y final quedarían en la posición 13 y 33 respectivamente.36
21. Figura 21: Series de tiempo de los tres ejes del acelerómetro (izq.) y giroscopio (der.) acotados por el inicio y fin preestablecido.37
22. Figura 22: Eje X Promedio del acelerómetro vs Eje X promedio del giroscopio, generado a partir de la Tabla 4.41
23. Figura 23: Línea que separa ambos gestos para lograr de esta forma su clasificación. Todo lo que esté a la izquierda de la línea será considerado como pellizco con el índice, de otra forma será pellizco con el dedo medio.42
24. Figura 24: El número de hiperplanos que se pueden generar para separar ambos gestos llega a ser infinito.42
25. Figura 25: Hiperplano en forma de círculo debido a un valor de gama que así lo permite.43
26. Figura 26: Durante el desarrollo del experimento en esta tesis se utilizaron dos relojes inteligentes: a) LG G100; b) Asus ZenWatch.46
27. Figura 27: Gestos seleccionados para ser usados. a) es un pellizco con el dedo índice, b) pellizco con el dedo medio y c) pellizco con el dedo anular. Los restantes tres gestos son idénticos a los primeros, pero repitiendo dos veces el gesto.47
28. Figura 28: Diferentes posturas en los que se realiza cada uno de los gestos; a) el usuario sentado, con el brazo soportado por un libro a caja estando encima de una mesa; b) el usuario parado como si estuviera observando la hora en el reloj; y c) el usuario parado en descanso con los brazos a los costados.48
29. Figura 29: Gesto de pellizco de a) un niño de 8 años, b) un adulto de 25 años y c) un adulto mayor de más de 70 años. Se puede observar como las condiciones degenerativas naturales puede afectar el nivel de precisión en los adultos mayores (figura c).50

30. Figura 30: Comunicación entre teléfono-reloj para coordinar la toma de los datos de los sensores en cada gesto. Al finalizar la toma de datos, todos los datos son comprimidos para ser recuperados posteriormente.	52
31. Figura 31: Captura de pantalla de la app en el teléfono que controla la captura de gestos en el reloj.	53
32. Figura 32: Validación cruzada con 5-iteraciones. Las instancias encerradas en el recuadro negro son consideradas como entrenamiento y las que están encerradas en recuadro amarillo son de prueba.	54
33. Figura 33: Ventana de Explore, con las pestañas de Preprocess y Classify, que son las opciones que más serán utilizadas.	55
34. Figura 34: Pestaña de clasificación donde se puede realizar la validación cruzada para encontrar los mejores parámetros y evaluar el algoritmo.	57
35. Figura 35: Guardando el archivo binario que contendrá el modelo para la detección de los gestos sutiles.	59
36. Figura 36: Diagrama de flujo que representa de forma general el trabajo realizado en esta tesis.	62
37. Figura 37: Gesto anular con doble repetición realizado por uno de los participantes donde existe ruido en la señal, mismo que es discriminado al seleccionar, mediante el ventaneo, solo la señal del gesto mostrada en b). La duración total del gesto es de aproximadamente 720 milisegundos.	64

Lista de tablas

Tabla	Página
1. Tabla 1: Información relevante de otras investigaciones revisadas sobre la detección de gestos y/o uso de reconocimiento de actividad, incluido el de esta investigación.	14
2. Tabla 2: Combinación de los ejes de cada sensor que permite crear nuevas características.	37
3. Tabla 3: Representación simplificada de la extracción de 208 características a partir de los tres ejes tanto del acelerómetro como del giroscopio de los gestos sutiles.	39
4. Tabla 4: Datos promedio del acelerómetro (columna 1) y del giroscopio (columna 2) en su eje X, así como la etiqueta del gesto sutil al que pertenece. Estos datos son considerados de entrenamiento para los algoritmos de clasificación.	40
5. Tabla 5: Escalamientos propuestos que permitirán que cada vector de características esté dentro del mismo rango de valores.	44
6. Tabla 6: Criterios de inclusión y exclusión para la selección de participantes en el estudio.	49
7. Tabla 7: Estructura de los datos del estudio según el formato arff.	56
8. Tabla 8: Porcentaje de gestos rechazados por no cumplir lo predispuesto en la sección 3.3.	64
9. Tabla 9: Porcentaje de reconocimiento de un posible gesto por parte del algoritmo DTW.	65
10. Tabla 10: Valores del costo y gamma para el algoritmo SVM según lo estipulado en la sección 4.8.	66
11. Tabla 11: Matriz de confusión obtenida a partir de la validación cruzada con 10 repeticiones de la clasificación de los seis gestos. Los gestos que son más frecuentemente confundidos es el índice con repetición con el medio con repetición y el medio con repetición con el anular con repetición.	67
12. Tabla 12: Porcentaje de gestos correctamente clasificados (también llamado recall) del total de cada uno de los gestos.	68
13. Tabla 13: Evaluación de los tres primeros gestos como una opción a usarse como respuestas a preguntas ternarias como Si, No, No lo sé.	69
14. Tabla 14: Matriz de confusión de los tres primeros gestos sutiles.	69
15. Tabla 15: Resultados obtenidos de WEKA con referencia a los tres últimos gestos sutiles; Matriz de confusión (izq.) y la precisión (der.)	70
16. Tabla 16: Matriz de confusión (izq.) de las muestras de los seis gestos sutiles y la precisión (der.) de cada uno de ellos, haciendo uso del clasificador SVM.	71

Capítulo 1. Introducción

Con el uso de algoritmos de reconocimiento de patrones, se han logrado detectar algunas de las actividades que un ser humano realiza a partir de información capturada por sensores que registran señales de video, audio, aceleración, entre otros. Detectando la actividad es posible asistir en distintas tareas que realizan las personas. Por ejemplo, se puede monitorear la reactivación física de una persona que está en rehabilitación; también existen aplicaciones en los teléfonos inteligentes que, mediante el reconocimiento de la actividad, al detectar que la persona va en bicicleta inicia una aplicación de música o cuando va manejando abre la aplicación de navegación con GPS.

Una de las dificultades que tienen los algoritmos de reconocimiento de patrones es la compensación que se debe realizar, en algunos casos, a la señal registrada para permitir que pueda ser clasificada por los algoritmos, debido a la variabilidad que puede presentarse en la señal al ser realizada en diferentes escenarios (Spanias et al. 2014). Un ejemplo de lo anterior es el reconocimiento de la actividad de una persona, según su color de piel, mediante video, en donde según la iluminación cambia ligeramente el color que percibe el dispositivo que está grabando.

Los algoritmos de reconocimiento de patrones supervisados pueden llegar a resolver el problema de la compensación, pero requieren de información de entrenamiento correctamente etiquetada (Spanias et al. 2014). El poder contar con la señal etiquetada de una actividad en particular, permite a los algoritmos de reconocimiento de patrones determinar con mayor precisión la actividad que se realiza.

1.1. Planteamiento del problema

Una de las formas más comunes de realizar el etiquetado de la actividad es mediante la observación directa del investigador, quien registra en qué momento inicia y termina la actividad, para posteriormente extraer la señal correspondiente a ese periodo para finalmente ser etiquetada. Esta forma de etiquetado se realiza normalmente de forma intrusiva, lo que puede provocar estrés, incomodidad o nerviosismo en el usuario, haciendo que la actividad la realice de forma distinta a como generalmente la hace.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo general de esta tesis consiste en proponer y evaluar mecanismos de retroalimentación para el etiquetado de datos de actividad y comportamiento de las personas, por medio de gestos mediante el uso de relojes inteligentes.

1.2.2. Objetivos específicos

- Seleccionar los gestos naturales que se pueden emplear para el etiquetado de actividades.
- Identificar los patrones de los gestos seleccionados y verificar su pertinencia para que el usuario no sea distraído de su actividad principal.
- Proponer técnicas para el reconocimiento de los gestos propuestos
- Detectar la eficiencia de los algoritmos propuestos para reconocer los gestos seleccionados.

1.3. Preguntas de investigación

- ¿Qué gestos son los más apropiados para que el usuario etiquete sus datos o que puedan darle retroalimentación?
- ¿Qué técnicas de procesamiento de señales y reconocimiento de patrones son adecuadas para detectar gestos usando relojes inteligentes?
- ¿Pueden detectarse los gestos mediante el uso de relojes inteligentes con suficiente precisión para ser usado en aplicaciones prácticas?

1.4. Propuesta de solución

En este estudio se propone el diseño y la evaluación de un experimento que permita recabar datos de las señales del giroscopio y acelerómetro provenientes de un reloj inteligente, que permitan identificar gestos no intrusivos para etiquetar la actividad que realiza, como parte del sensado participativo. Determinando para esto, los gestos que se consideren no intrusivos, cómodos y de fácil ejecución.

1.5. Metodología

La metodología aplicada en el trabajo de investigación consiste en **seis etapas** que son: Análisis de la literatura, identificación de gestos, tratamiento de la señal y extracción de características, clasificación de los gestos y evaluación del método propuesto.

- **Análisis de la literatura:** Se revisó la literatura para detectar los diferentes tipos de gestos usados para diversos objetivos, con la finalidad de encontrar algunos que se puedan reusar en este estudio o que puedan ser mejorados para su uso, y que cumplan con el objetivo de ser apropiados, es decir, que no sean intrusivos, que sean cómodos y de fácil ejecución. Además de los gestos, se buscó el algoritmo de reconocimiento de patrones que mejores resultados presentara en la literatura y que además estuviera presente dentro del software WEKA.
- **Reestructura de preguntas de investigación y objetivos:** Al verificar qué tan pertinentes o suficientes son las preguntas de investigación y los objetivos planteados, se reestructuraron las preguntas de investigación y/o los objetivos.
- **Identificar y establecer de 4 a 8 gestos:** Buscar aquellos gestos sutiles que no sean intrusivos y además que sean de fácil ejecución. Siempre tomando en cuenta que la memoria de corto plazo nos permite recordar 7 ± 1 números (Miller 1956).
- **Tratamiento de la señal y extracción de características:** En esta etapa se buscará extraer la señal que contiene los gestos sutiles seleccionados, tratando de quitar posibles ruidos que pudiera tener la señal. En base a esta señal resultante, se determinarán las características que se consideren relevantes para definir a cada uno de los gestos sutiles.
- **Clasificación de los gestos:** Se establecieron los parámetros de dos modelos que, a partir de la extracción de características, clasificaban los datos en alguno de los gestos sutiles.
- **Evaluación del método propuesto:** Mediante la validación cruzada y datos suministrados se evaluó la eficiencia de ambos modelos.

1.6. Contribución

Las contribuciones esperadas de este estudio son:

- **Gestos sutiles con poco esfuerzo:** Se espera encontrar si es factible detectar, mediante un dispositivo vestible, gestos sutiles que no sean intrusivos y que se realicen con poco esfuerzo.
- **Modelo para filtrado:** Se usará un modelo para el filtrado de la señal que permita el preprocesamiento de la misma, con la certeza de que se trata de la señal de un gesto o no.
- **Selección de parámetros:** Se establecerá un procedimiento para poder determinar de forma rápida y los mejores parámetros que un clasificador requiere, haciendo uso de la validación cruzada y de aproximaciones.

1.7. Organización de la tesis

El resto del documento se organiza de la siguiente manera:

En el **Capítulo 2** (*Fundamentos teóricos*) En esta sección se presenta la importancia de realizar el etiquetado de la actividad y algunas formas de llevarla a cabo, así como su uso en el sensado participativo por medio del cómputo vestible. Finalmente se presenta el estado del arte en cuanto a detección de gestos se refiere.

En el **Capítulo 3** (*Propuesta de reconocimiento de gestos sutiles con la mano*) En esta sección se desarrolla el análisis de la señal proveniente del acelerómetro y giroscopio para poder determinar el inicio y fin del gesto sutil dentro de esta señal, para que por medio de los algoritmos de aprendizaje supervisado se logre el reconocimiento del gesto sutil.

En el **Capítulo 4** (*Evaluación de estrategia propuesta*) En esta sección se definen las características del reloj inteligente que se usara en los gestos sutiles establecidos, así como las posturas de los participantes para desarrollar dichos gestos por medio de la aplicación generada para este fin.

En el **Capítulo 5** (*Resultados*) En esta sección se presentan los resultados obtenidos del ventaneo de la señal para determinar el inicio y el fin de los gestos, así como la eficiencia encontrada en cada uno de los algoritmos utilizados. Además, se presenta algunas observaciones sobre los resultados obtenidos.

En el **Capítulo 6** (*Conclusión y trabajo futuro*) En esta sección se resumen las principales contribuciones del estudio, así como las limitaciones que tienen para finalmente establecer alternativas de trabajo futuro.

Capítulo 2. Fundamentos teóricos

2.1. Introducción

El reconocimiento de la actividad es uno de los problemas de investigación más activos en áreas como Computación Ubicua y visión por computadora (Avgerinakis et al. 2013). Para lograr detectar la actividad que realiza un individuo, se han desarrollado algoritmos que utilizan datos obtenidos por una variedad de sensores. La mayor parte de estos algoritmos son de aprendizaje supervisado, es decir, a partir de datos de entrenamiento previamente etiquetados se crea un modelo que infiere una función en específico. Por ejemplo, para entrenar un modelo que detecte si una persona está dormida, se requiere conocer en qué momento la persona se duerme y en qué momento se despierta. Algunas soluciones al reconocimiento de actividad hacen uso de video, de sensores lógicos y ambientales relativamente simples como el acelerómetro, giroscopio o sensores de presión. Por ejemplo, para detectar si una persona está dormida se suele usar el acelerómetro y/o un sensor de presión.

Algunas de las formas en que se puede etiquetar la actividad, es por medio de observaciones por parte de terceros y/o por medio de grabaciones de video. Algunos de los problemas al utilizar estas formas de etiquetado de datos para la detección de la actividad, es que las personas se sienten vigiladas y esto puede afectar sus actividades y/o comportamientos, y que este tipo de sistemas tiene movilidad limitada (Lane et al. 2011).

2.2. Etiquetado de datos

El reconocimiento de patrones estudia cómo las máquinas observan el ambiente y aprenden a distinguir los patrones de interés de sus antecedentes (Jain et al. 2000). Mediante el uso del reconocimiento de patrones se puede realizar la detección de la actividad, haciendo uso de una serie de ejemplos para realizar esta tarea. Este tipo de ejemplos no son otra cosa que información de la actividad previamente etiquetada como tal. La Figura 1 muestra el proceso general para resolver un problema de reconocimiento de patrones. El proceso del reconocimiento es iterativo para mejorar la eficiencia del clasificador por medio de, por ejemplo, la selección de atributos o clasificadores apropiados.

La tarea de etiquetar datos se considera tradicionalmente como tediosa e inevitable para la elaboración de un clasificador supervisado (Sun et al. 2017). Generalmente, el etiquetado de datos se realiza por personas familiarizadas con el funcionamiento de los clasificadores (Sun et al. 2017), y son ellos quienes identifican la información que corresponde a la actividad asociada para poder ser etiquetada como tal. Esta forma de realizar el etiquetado de la información podría no resultar ser una tarea trivial para personas sin experiencia, a fin de que puedan implementar con facilidad un clasificador determinado.

Una limitante en la elaboración del etiquetado para la detección de actividad bajo ambientes controlados es que, las personas que realizaran la actividad deben acudir directamente al lugar y hora indicada para que el investigador tome las muestras de datos y las etiquete. Con esta técnica de etiquetado de actividad se pierde la posibilidad de realizar sensado participativo. Además, para poder realizar sensado participativo los dispositivos deben tener la suficiente movilidad (Korpela et al. 2015) para poder ejecutar la actividad en lugares donde no se encuentre presente el investigador.

Etiquetar los datos permite generar mejores clasificadores que si no se realizara el etiquetado, incluso si se realiza un etiquetado parcial (Hayashi & Takai 2015).

El etiquetado de la actividad se puede realizar mediante la **observación**, en **condiciones controladas** y por medio de **auto-etiquetado**.

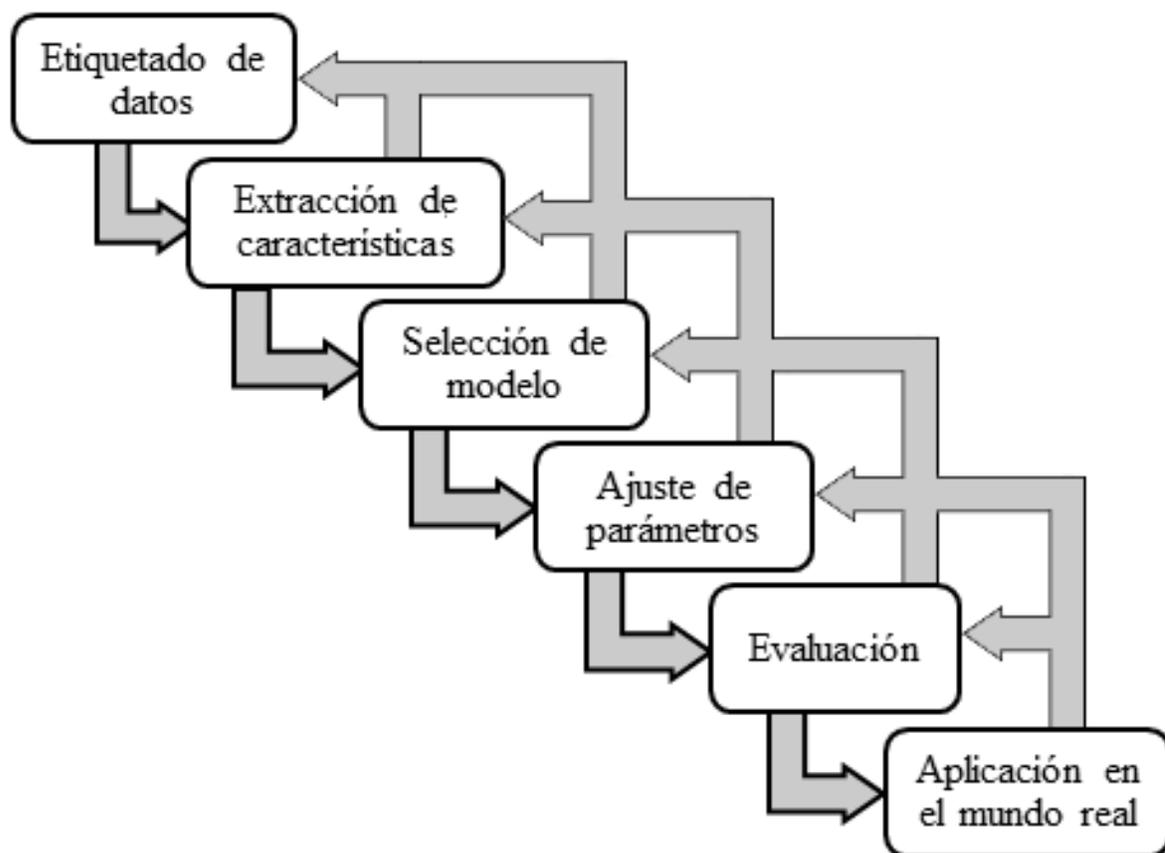


Figura 1: Proceso para crear un modelo para el reconocimiento de patrones a usarse en alguna aplicación. [6]

2.2.1. Etiquetado de datos por medio de la observación

Esta forma de etiquetado consiste en observar directamente las acciones que está realizando el individuo que está siendo sentido. En la observación directa el investigador está presente al momento de realizar las actividades a etiquetar, mientras los dispositivos encargados de registrar las señales de la actividad se encuentran en funcionamiento, para posteriormente hacer un cruce de datos en los tiempos registrados por los sensores y lo observado para así etiquetar finalmente la actividad. Durante todo el proceso de etiquetado el investigador sigue al sujeto a donde quiera que este vaya, pero sin interactuar con él, actuando como si se tratara de su sombra, como se ejemplifica en la Figura 2. Una de las ventajas de esta forma de etiquetado es el poder obtener información adicional, independientemente de que no exista el deseo de proporcionarse. Por el contrario, se puede presentar la **expectativa del observador**, en la cual la presencia de un observador influye en el actuar del participante (Ko 2017).



Figura 2: Representación de la observación directa al dormir, sin intervenir en las actividades del sujeto a observar.

Una variante de esta forma de etiquetado es utilizar video que registre la actividad, para posteriormente ser analizado y así poder etiquetar los datos. El inconveniente de esta técnica es que si la actividad involucra movilidad, resulta difícil seguir al individuo con una cámara de video.

2.2.2. Etiquetado de datos en condiciones controladas de sensado

Esta forma de etiquetado es una de las más usadas debido a que es más fácil lograr disminuir los factores de ruido que se pueden presentar en la señal de los sensores (Yang et al. 2015), debido al control de las condiciones en las que se realiza la actividad. Permite además la observación minuciosa y detallada de la actividad en un sitio especialmente previsto para realizar el etiquetado de la información. La Figura 3 muestra un ejemplo de etiquetado de datos en condiciones controladas de la actividad de dormir.



Figura 3: Representación del etiquetado en condiciones controladas de la actividad de dormir.

Una de las complicaciones del etiquetado de la actividad mediante este método, es que los clasificadores obtenidos podrían no funcionar correctamente en ambientes realistas.

2.2.3. Auto-etiquetado de la actividad

Esta forma de etiquetar la actividad consiste en dejar que sea el mismo participante quien realice el etiquetado. Para realizar el etiquetado se provee al participante de las instrucciones necesarias (por parte del investigador) para realizar correctamente el etiquetado. Con esta manera de etiquetar la actividad, se logran romper las barreras del etiquetado a través del tiempo, ya que en los dos casos anteriores el etiquetado depende del tiempo disponible del investigador y con esta técnica puede realizarse casi en cualquier momento. La Figura 4 muestra la representación del auto-etiquetado de la información mediante el uso de un dispositivo móvil.



Figura 4: Representación del auto-etiquetado de la actividad de dormir. El usuario indica el inicio y el fin del sueño.

Una desventaja que presenta esta técnica es que se deja al usuario el control total del etiquetado de la actividad, confiándole la integridad de la información recabada.

2.3. Sensado participativo y cómputo vestible

Las capacidades de los dispositivos como teléfonos inteligentes y en general del cómputo vestible, han incrementado la capacidad de mejorar, transformar y mezclar los datos procedentes de cualquiera de sus sensores (Salomoni et al. 2012). El cómputo vestible es aquel que nos permite llevar puesto dispositivos con sensores como si se tratara de ropa o accesorios vestibles que usamos cotidianamente y que gracias a la cercanía que guardan con el usuario, permiten el monitoreo constante y más preciso de lo que se pudieran hacer con otro tipo de sistema (Bojórquez 2015).

En la actualidad, se puede adquirir una gran variedad de dispositivos electrónicos vestibles, que permiten sensar una variedad de información por medio de sus múltiples sensores como el acelerómetro, giroscopio, GPS, ritmo cardiaco, entre otros. Además, gracias al auge del uso de redes telefónicas 3g (actualmente 4g) y al Wi-Fi, los usuarios pueden compartir la información generada en sus dispositivos a cualquier parte del mundo en cualquier momento. El sensado participativo, aprovecha estas dos grandes

ventajas (datos generados y comunicación permanente entre usuarios) para que los usuarios puedan recopilar y compartir datos generados en el entorno que se encuentren.

Por medio del uso del cómputo vestible y del sensado participativo, se plantea la posibilidad de realizar el etiquetado de la actividad mediante el uso de gestos no intrusivos y de fácil y cómoda ejecución, que permita a cualquier usuario etiquetar la actividad que esté realizando sin interrumpir las tareas que se encuentre realizando en ese momento.

2.4. Uso del cómputo vestible para la detección de los gestos

Uno de los dispositivos más importantes y de mayor aceptación en el cómputo vestible es el teléfono inteligente[4,13]. Adicionalmente, los teléfonos inteligentes están siendo usados junto con otros dispositivos vestibles, como los relojes inteligentes, para hacer un seguimiento a una amplia gama de actividades humanas [4,13]. Este tipo de seguimiento está unido a una serie de desafíos técnicos comunes asociados con los sistemas de sensado móvil, incluyendo: i) la detección de las personas y su entorno; ii) el uso eficiente de la energía de los recursos de dispositivos móviles; iii) la minería y la interpretación de los datos del sensor móvil; y iv) la interacción con los usuarios, en gran parte para proporcionar retroalimentación e información a los usuarios (Lane et al. 2011).

Con el uso de dispositivos vestibles se puede lograr etiquetar la actividad que esté realizando una persona. El etiquetado puede realizarse por medio de movimientos finos y discretos realizados con la mano, para establecer de forma rápida y poco intrusiva el momento en que inicia y termina la actividad, así como el tipo de actividad que esté realizando. El tipo de gestos discretos deben poder ser recordados de forma fácil para lograr etiquetar distintos tipos de actividad y/o comportamiento como la ansiedad, cansancio, estrés, caminar, comer, correr, dormir, entre otros. Además, de etiquetar la actividad, se pueden usar estos mismos gestos para establecer la percepción de los niveles de intensidad de la actividad, como muy cansado, poco cansado o descansado. El tipo de gestos seleccionados deben ser adecuados para etiquetar la actividad con algún dispositivo vestible.

2.5. Estado del arte en la detección de gestos

La presencia generalizada de sensores de movimiento en los dispositivos móviles de los usuarios, ha generado un creciente interés en la investigación en reconocimiento de la actividad humana (Stisen et al. 2015). Para el reconocimiento de gestos que se realizan con las manos se ha desarrollado amplio trabajo. En (Lu et al. 2014) Lu Z., Chen X., et. al. usan un dispositivo de detección de gestos portátil con el uso de un acelerómetro y cuatro sensores de electromiografía de superficie (SEMG), para detectar 19 gestos mediante el uso del antebrazo con una precisión del 95% con gestos dependientes del usuario y 89.6% en pruebas independientes del usuario.

En los trabajos [16,17 y 18] los autores realizan el reconocimiento de gestos utilizando las manos, la cadera o los pies, con precisión de más del 90% en todos los casos. En estos estudios se usaron una combinación de sensores como el acelerómetro, giroscopio y sensor de presión (Song et al. 2015); acelerómetro, giroscopio y magnetómetro (Alavi et al. 2016); acelerómetro y video (Porzi et al. 2013) con muy buenos resultados, sobre todo aquellos que usaron el acelerómetro con algún sensor adicional. En la investigación hecha en (Bobeth et al. 2012), los autores reconocen gestos realizados por adultos mayores mediante el uso de Kinect para cambiar de canal en un televisor, logrando buenos resultados y buen nivel de aceptación por parte de los usuarios.

En el trabajo reportado en (Takahashi & Kishino 1991), Takahashi T. y Kishino F. usaron un guante de datos VPL. Este guante es un dispositivo de interfaz de gestos con la mano que ofrece información en tiempo real sobre el movimiento de esta. Esta interfaz usa gestos estáticos de la mano para reconocer los 46 gestos del kana Japones (alfabeto Japones), de los cuales logran reconocer 34 gestos de los 46 de forma correcta.

En (Kellogg et al. 2014), Kellogg et. al., argumentan que con su sistema AllSee, puede reconocer gestos usando tarjetas RFID. Logrando una precisión del 97% al realizar ocho gestos diferentes.

En (Shoaib et al. 2016), Shoaib M, et. al., se reporta el uso de un dispositivo en la muñeca y un teléfono inteligente en el bolsillo del pantalón para detectar la actividad del individuo, mediante la lectura de los sensores acelerómetro y giroscopio en ambos dispositivos (para complementarse uno al otro).

Tabla 1: Información relevante de otras investigaciones revisadas sobre la detección de gestos y/o uso de reconocimiento de actividad, incluido el de esta investigación.

Año	Título	Autores	Características	Gestos	Sensores usados
2009 (Neto et al. 2009)	Accelerometer-Based Control of an Industrial Robotic Arm	Pedro Neto, J. Norberto Pires, A. Paulo Moreira	Sistema basado en dos acelerómetros para controlar un robot industrial por medio de gestos con los brazos y manos.	5 gestos	2 acelerómetros, 1 por cada mano
2014 (Lu et al. 2014)	A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices	Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, Ping Zhou	Con el acelerómetro y la electromiografía de superficie (SEMG) se mide la actividad de los músculos para poder capturar movimientos finos con la muñeca y los dedos.	Gestos de escala corta y larga	Acelerómetro y SEMG
2013 (Porzi et al. 2013)	A SmartWatch-based Gesture Recognition System for Assisting People with Visual Impairments	Porzi Lorenzo, Messelodi Stefano, Modena Carla Maria, Ricci Elisa.	Mediante el uso de un teléfono y un reloj inteligente diseñan un sistema para ayudar a los ciegos y débiles visuales a guiarse mediante el uso del acelerómetro.	8 gestos diferentes	Acelerómetro
2015 (Zhao et al. 2015)	Demo: Finger and Hand Gesture Recognition using Smartwatch	Zhao Yixin, Pathak Parth, Xu Chao, Mohapatra Prasant	Afirman que los diferentes tendones crean una firma única del movimiento de la muñeca mientras se realizan diferentes gestos.	27 gestos	Acelerómetro y giroscopio
2009 (Wenhui et al. 2009)	Dynamic Gesture Recognition based on Multiple Sensors Fusion Technology	Wenhui Wang, Xiang Chen, Kongqiao Wang, Xu Zhang, Jihai Yang	Estudian la eficiencia de usar sensores combinados para el reconocimiento de gestos. Evalúan primero los gestos con un solo sensor y después combinados.	20 gestos	Acelerómetro, SEMG y una cámara web
2016 (Wu et al. 2016)	Natural Gesture Modeling and Recognition Approach Based on Joint Movements	Wu Yuanhao, Chen Ken, Fu Chenglong	Presenta un modelo de reconocimiento de gestos que toma en cuenta los movimientos, ángulos y orientación del brazo. Toma en cuenta la forma expresiva, natural, flexible y eficiente.	12 gestos	3 IMUs

2014 (Costante et al. 2014)	PERSONALIZING A SMARTWATCH-BASED GESTURE INTERFACE WITH TRANSFER LEARNING	Costante Gabriele, Porzi Lorenzo, Lanz Oswald, Valigi Paolo, Ricci Elisa	Dan un enfoque más semántico a los gestos además de una métrica de transferencia del algoritmo de aprendizaje, donde se puede reconfigurar el sistema para reconocer nuevos gestos.	19 gestos	Acelerómetro
2016 (Aoyama et al. 2016)	ThumbSlide: An Interaction Technique for Smartwatches using a Thumb Slide Movement	Aoyama Shuhei, Shizuki Buntarou, Tanaka Jiro	En este artículo detectan el movimiento del pulgar (deslizamiento a la derecha o la izquierda sobre el dedo índice) causado por los músculos y tendones conectados a este.	1 gesto	16 foto reflectores
2010 (Rico & Brewster 2010)	Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability	Rico Julie, Brewster Stephen	Establecen que la adopción de gestos requiere de comportamientos que no sean vergonzosos o perjudiciales y es el usuario quien debe decidir si su motivación para usar esa tecnología supera el riesgo a un error social.	Aplauso, rotación de la muñeca, midiendo la distancia entre dos dedos	16 foto reflectores
2014 (Dementyev & Paradiso 2014)	WristFlex: Low-Power Gesture Input with Wrist-Worn Pressure Sensors	Dementyev Artem, Paradiso Joseph	Proponen una interfaz que utiliza la distribución de presión alrededor de la muñeca para discernir los movimientos del dedo, ya que los músculos que mueven los dedos y la mano se localizan principalmente en el antebrazo.	5 gestos	
2014 (Liu et al. 2014)	Visual Gesture Recognition for Human Robot Interaction Using Dynamic Movement Primitives	Liu Zhan, Hu Fan, Luo Dingsheng, Wu Xihong	Mediante una cámara monocular detectan el movimiento de las manos desnudas, por lo que siguen su movimiento según el color de la piel.		Cámara monocular montada en un robot
2015 (Wilhelm et al. 2015)	eRing: Multiple Finger Gesture Recognition with one Ring Using an Electric Field	Wilhelm Mathias, Krakowczyk Daniel, Trollmann Frank, Albayrak Sahin	Con el uso de un anillo equipado con un campo eléctrico, detectan gestos por medio de variaciones producidos en este campo por el movimiento de la mano y los dedos.	6 posturas y 6 gestos	Anillo con campo eléctrico

2017	Detección de gestos por medio de relojes inteligentes en el etiquetado de datos para el sensado participativo para “reconocimiento de actividad”.	Gonzalez Luis	Detección de seis gestos sutiles no intrusivos y de fácil ejecución para el etiquetado de la actividad.	6 gestos	Acelerómetro y giroscopio
------	---	---------------	---	----------	---------------------------

En (Kalantarian et al. 2015), Kalantarian H., et. al., buscan predecir la ingesta de medicamentos por parte de los usuarios, al reconocer esta actividad mediante la lectura del acelerómetro y el giroscopio al momento de abrir el frasco del medicamento o al poner la pastilla en la palma de mano. Para la detección de esta actividad, mediante su gesto correspondiente, se requiere que la píldora de la botella o la cartera sean sacadas de una forma determinada para poder reconocerlo correctamente. En la Tabla 1 se resume la información de estas y otras investigaciones trascendentales en la detección de gestos y uso de reconocimiento de patrones, incluido el de esta investigación.

En este trabajo se usará el acelerómetro y giroscopio para detectar los gestos sutiles, ya que en todos los trabajos previos se obtuvieron muy buenos resultados, superiores al 80% de precisión. Además, estos dos sensores son económicos y muy probablemente se pueden encontrar en algún dispositivo móvil que sea vestible. Los trabajos que hacen uso de sensores que no son movibles con facilidad como Kinect, RFID (por los sensores lectores fijos) o cámaras, no son contemplados porque tienen poca movilidad y se requiere mayor inversión en infraestructura para colocarlos.

2.6. Eliminación de la gravedad

El acelerómetro mide la aceleración con la que se mueve el sensor, así como la aceleración ejercida hacia el centro de la tierra por medio de la gravedad. Dependiendo de la posición de la mano la fuerza de la gravedad afectará a uno varios ejes del sensor, como se muestra en la Figura 5. Para cumplir con el objetivo de detectar el gesto independientemente de la posición de la mano, la aceleración ejercida por

la gravedad debe ser eliminada. Para poder quitar los efectos de la gravedad, se aplicará un filtro pasa bajas sobre las fuerzas que se le aplican al dispositivo (A_d) como muestra la ecuación (1)¹.

$$A_d = -g - \sum F/mass \quad (1)$$

Donde A_d es la aceleración que se aplica al dispositivo, g es la constante de gravedad con valor de 9.81 m/s^2 , y $\sum F$ son las fuerzas aplicadas al dispositivo que no sean la gravedad.

Este procedimiento implementa un factor de filtrado a la gravedad de 0.2, lo que significa que al 20% de su valor actual se le agregará el 80% del valor medio anterior, como se muestra en la ecuación (2), para así finalmente ser eliminada de la aceleración total aplicada al dispositivo como se muestra en la ecuación (3).

$$g^* = 0.8 * g^* + 0.2 * A_d \quad (2)$$

$$A_d = A_d - g^* \quad (3)$$

Donde g^* es el valor de la gravedad filtrado hasta el momento (al iniciar la iteración el valor de g es 0, dado que no se ha filtrado nada). Esta iteración se repite hasta que se han cubierto la totalidad de los datos.

Debido a que una parte de la señal del acelerómetro contendrá la señal de la fuerza de gravedad atenuándose, los sensores empezarán a tomar lecturas 100 milisegundos antes de que sea lanzada la petición de cualquier gesto, para que esta atenuación no sea tomada en cuenta.

¹ <https://developer.android.com/reference/android/hardware/SensorEvent.html#values>

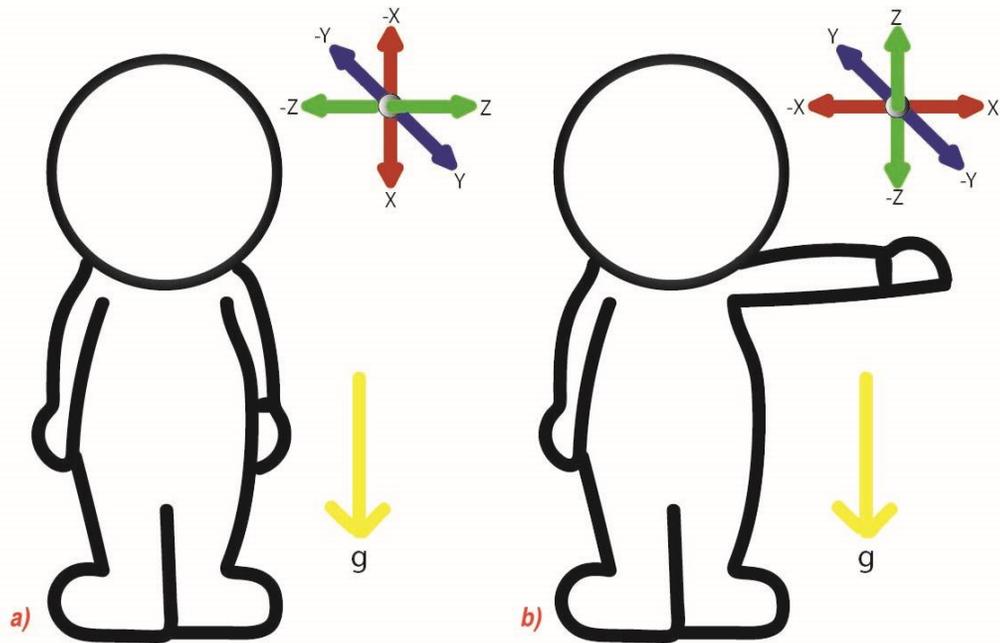


Figura 5: Efecto de la aceleración de la gravedad sobre cada eje dependiendo de la posición de la mano. a) posición de la mano en la que afecta al eje X y b) posición de la mano en la que afecta al eje Z.

2.7. Conclusión

En la detección de gestos que se propone, a diferencia de los trabajos que se han realizado, se busca que sirvan como retroalimentación al reconocimiento de actividades. Estos gestos deben ser socialmente aceptables y realizados de forma natural y con el menor esfuerzo posible. Esto debido a que en los escenarios y situaciones en los que se puede encontrar la persona que realizaría el etiquetado de su actividad, los gestos deberían de pasar casi desapercibidos a terceros. Para lograr esto, se plantea hacer uso de la técnica del auto etiquetado, que permite tener movilidad y realizarse en cualquier momento, lo que no consiguen hacer las técnicas de etiquetado en ambientes controlados o bajo observación directa.

Capítulo 3. Propuesta de reconocimiento de gestos sutiles realizados con la mano

3.1. Posición y orientación de los sensores

En este trabajo se propone un método de detección de gestos sutiles realizados con la mano mediante el uso del giroscopio y acelerómetro ubicados en la muñeca. Se utiliza la muñeca por ser un lugar estratégico al ser accesible y cómodo y al hecho de que hay dispositivos vestibles para la muñeca, como es el caso de los relojes inteligentes, o smartwatches.

Tanto los sensores acelerómetro y el giroscopio reportan sus datos en tres dimensiones (ejes XYZ). Estos ejes deben ser orientados de tal forma que, sin importar la posición de la mano, los datos reportados en cada eje sean siempre congruentes con el gesto. Por la razón anterior, los dos sensores serán ajustados a la muñeca por medio de alguna banda ajustada a cada usuario, de tal manera que el eje X sea paralelo al antebrazo y su valor positivo debe estar en dirección hacia los dedos de la mano; el eje Y negativo deberá estar dirigido en dirección hacia el dedo pulgar de la mano y el eje Z negativo debe estar dirigido de la palma de la mano hacia el exterior, como se muestra en la Figura 6.

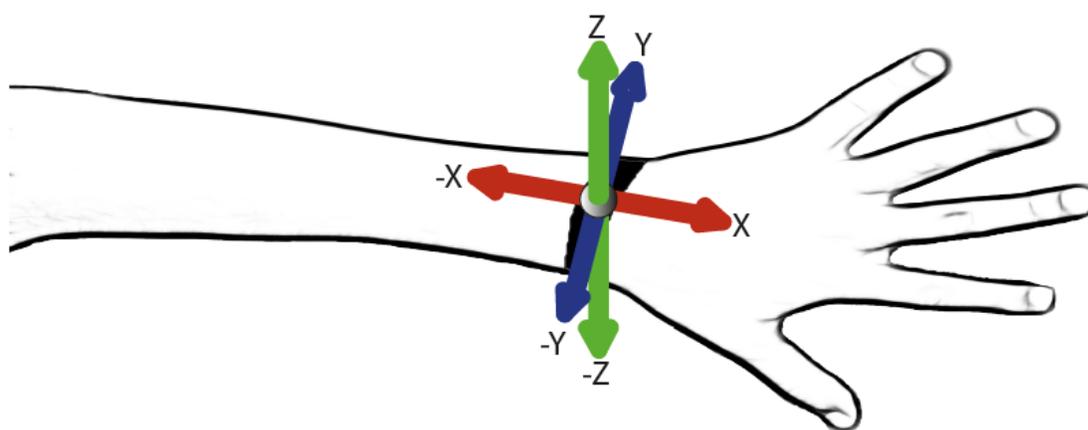


Figura 6: Orientación de los ejes de los sensores con respecto a la mano.

3.2. Detectar el inicio y el fin de un gesto sutil

Para realizar la lectura de los movimientos sutiles con la mano, se debe estimar el tiempo mínimo que una persona requiere para realizar el gesto. Este tiempo mínimo se estimará midiendo el tiempo máximo requerido para realizar el gesto más complejo. A este tiempo mínimo observado se le agregará un segundo y medio para compensar los retardos naturales en los tiempos de reacción de cada persona.

Una vez que se ha estimado el tiempo de la duración del gesto más complejo, se podrá programar por medio de una condición de inicio como un sonido o una vibración el tiempo en el cual los sensores serán abiertos para tomar las lecturas en sus ejes XYZ de cada sensor. Cabe señalar que la condición de inicio puede ser seleccionada por el usuario para su comodidad.

Dado que existirán gestos que se realicen en un tiempo menor al estimado y además habrá personas que realicen los gestos en un tiempo menor al observado, es necesario buscar en los valores de las lecturas de los ejes XYZ donde inicia y donde termina el gesto. Para poder definir el inicio y el fin de un gesto, se tomarán únicamente los datos provenientes del giroscopio, ya que como se vio en la sección anterior, la aceleración de la gravedad disminuye gradualmente y no de forma espontánea, por lo que no es una buena opción para buscar el inicio y el fin del gesto.

Para determinar el inicio y fin del gesto, se debe estar relativamente seguro de que los datos reportados por el giroscopio corresponden a un gesto y no a un movimiento involuntario o simplemente a ruido por realizar una actividad diferente a la indicada. Para lograr lo anterior los datos deberán ser sometidos a una comparación con uno de los gestos para determinar la aparente presencia de un gesto en la señal. Aun y cuando cada gesto tendrá su propia firma, existe cierta similitud en los movimientos que los hará similares. En este punto se propone utilizar el algoritmo de Distorsión Dinámica Temporal (DTW, por sus siglas en inglés) que se describe a continuación.

3.3. Distorsión Dinámica Temporal para determinar si en la señal se encuentra un gesto

El algoritmo DTW es popular por ser muy efectivo (Senin 2008) para detectar formas similares en series de tiempo y que pueden tener diferente longitud y fase. Aprovechando la característica de este

algoritmo, se busca discriminar aquellas señales que no se parezcan a alguno de los gestos, dado que la señal de los seis gestos es similar (cambiando de longitud y fase de usuario a usuario). Para lograr esto se debe configurar el algoritmo para que no sea tan estricto en la comparación y que además no sea demasiado benevolente para que discrimine gran parte de las señales que no son lo que se espera.

Como señal de referencia se utilizará la señal de uno de los seis gestos que esté libre de ruido y que sea lo más claro y estable posible. Para este efecto se usará la señal proveniente del giroscopio, ya que como se mencionó anteriormente, esta está libre de filtros que pudieran introducir ruido en algún momento de la atenuación de la señal, como ocurre con el acelerómetro. Las magnitudes de las señales proveniente de los 3 ejes (XYZ) se combinará de la siguiente forma $\sqrt{X^2 + Y^2 + Z^2}$ tanto en la señal de referencia como en la señal que se lea directamente del giroscopio al momento de realizar uno de los gestos, esto con el fin de compensar las diferentes orientaciones que pudieran tener los sensores (Wen et al. 2016), de esta misma forma se compara un vector de datos contra otro y no tres vectores contra otros 3.

La comparación de ambas señales se hará bajo los principios del algoritmo DTW, salvo una pequeña excepción que permitirá corregir señales de diferente tamaño. A continuación, se describe todo el proceso:

Dadas las secuencias de datos tratados de referencia establecida $X = (x_1, x_2, x_3, \dots, x_n)$ de tamaño $n \in N$ y del giroscopio $Y = (y_1, y_2, y_3, \dots, y_m)$ de tamaño $m \in N$, se fija un espacio de función denotado por F , donde X_n y $Y_m \in F$ para $n \in [1 : N]$ y $m \in [1 : M]$, donde para comparar dos elementos x y y , se necesita un costo local para una de las características (señal capturada y tratada directamente del giroscopio) definida por la siguiente función:

$$c : F \times F \rightarrow R > 0$$

En esta relación de costos, el costo $c(x, y)$ será más pequeño mientras x sea lo más próximo a y , en otro caso el costo será incremental. Al evaluar el costo local para cada par de elementos de la secuencia X y Y , se obtendrá la matriz de costos $C \in R^{N \times M}$ como se muestra en la Figura 7, y esta matriz está definida por $C(n, m) := c(x_n, y_m)$. El objetivo de este algoritmo es encontrar la alineación entre X y Y que haga que su costo total acumulado sea el mínimo posible, por lo que este alineamiento óptimo se extiende a lo largo de un camino de bajo costo dentro de la matriz de costos (Müller 2007), a esta ruta se le conoce como

ruta de distorsión. La Figura 8 muestra la ruta de costo acumulado mínimo generado por ambas secuencias.

Por definición una ruta de distorsión de N y M es una secuencia $p = (p_1, p_2, \dots, p_L)$ con $p_l = (n_l, m_l) \in [1: N] \times [1: M]$ para $l \in [1: L]$ satisface las siguientes condiciones:

- Condiciones de inicio y fin: $p_1 = (1, 1)$ y $p_L = (N, M)$. Es decir, la ruta siempre comenzará en la ubicación donde se encuentra el costo $C_{1,1}$ y de la misma forma siempre terminará en la celda donde se encuentre el costo $C_{n,m}$, que en el caso de la Figura 8 es la ubicación del costo $C_{7,5}$.
- Condición de monotonicidad: $n_1 \leq n_2 \leq \dots \leq n_L$ y $m_1 \leq m_2 \leq \dots \leq m_L$. Es decir, el crecimiento deberá ser siempre de forma incremental de manera que $n_k - n_{k-1} \geq 0$ y $m_k - m_{k-1} \geq 0$. La Figura 9 muestra una ruta de distorsión sin monotonicidad. Esta condición permite que las características no se repitan en la alineación de la ruta de distorsión.
- Condición del tamaño de paso: $p_{l+1} - p_l \in \{(1,0), (0,1), (1,1)\}$ para $l \in [1: L-1]$. Lo que significa que no debe haber puntos de ruptura en la ruta de distorsión. Esta condición permite que ningún elemento de X o Y sea omitido.

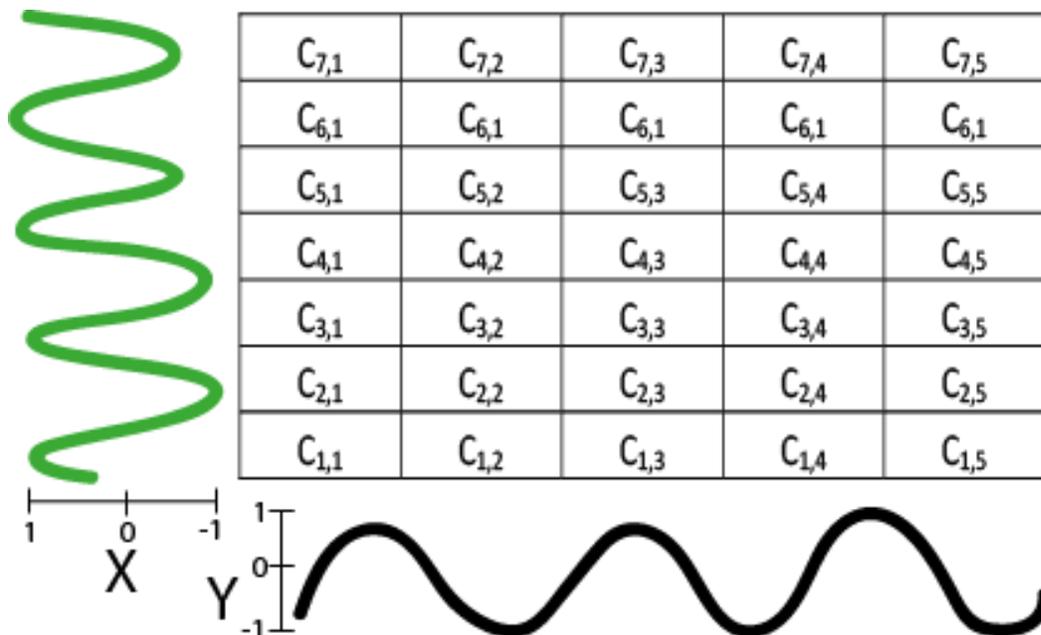


Figura 7: Matriz de costos de dos secuencias X (eje vertical) y Y (eje horizontal) generadas por el método DTW. En este caso, ambas secuencias son de distinto tamaño.

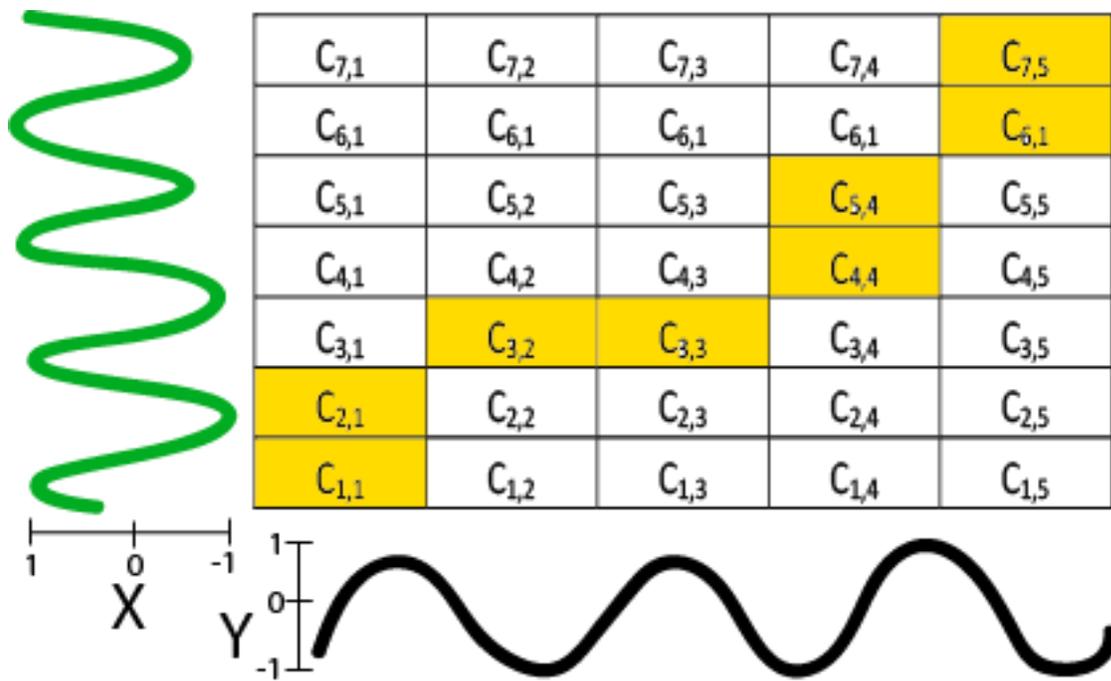


Figura 8: Ruta de distorsión con costo mínimo entre la secuencia X y Y, llamada ruta de distorsión.

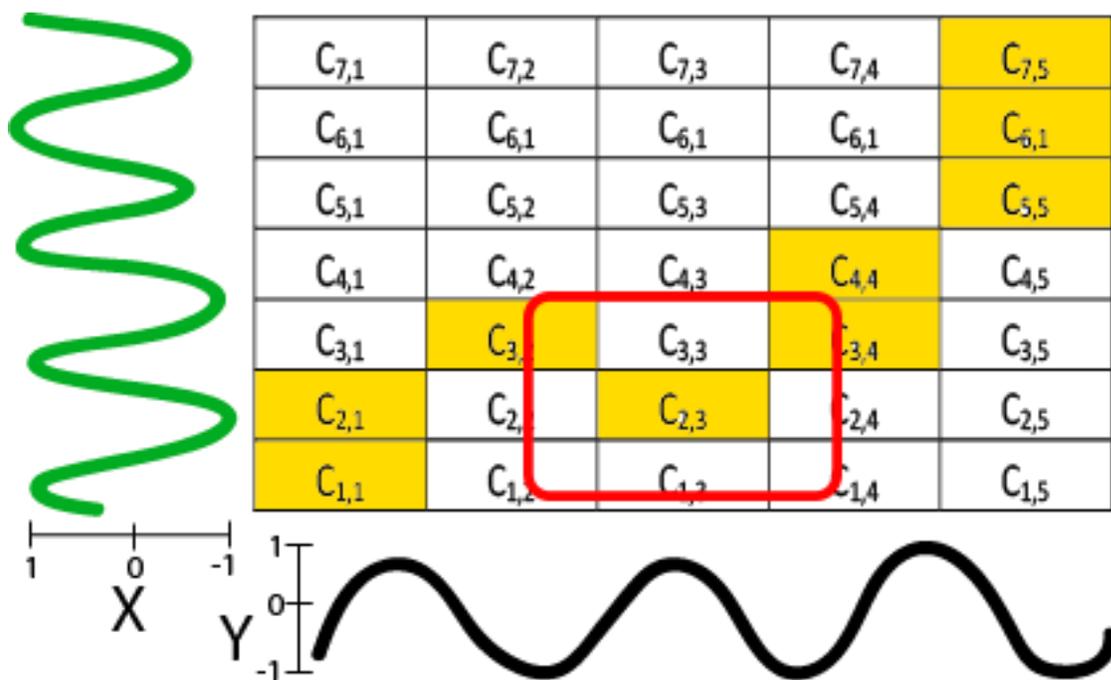


Figura 9: Ruta de distorsión que no cumple con la monotonicidad. En el rectángulo rojo aparece un elemento que no cumple con la regla $n_k - n_{k-1} \geq 0$, ya que $2 - 3 < 0$.

Al existir una cantidad exponencial de rutas de distorsión que cumplen con las condiciones anteriores, solo nos interesa aquella ruta que minimiza el costo total acumulado durante la trayectoria de dicha ruta. La ecuación (4) muestra la ecuación que se usará para calcular la ruta de distorsión con menor peso acumulado.

$$DTW(X, Y) = \min \left[\sqrt{\sum_{k=1}^k W_k} \right] \quad (4)$$

Resolver esta ecuación requiere un tiempo $O(nm)$ ya que se requiere calcular los costos en cada celda comenzando por (1,1) y moviéndose a sus adyacentes que cumplan con la monotonicidad, como se ilustra en la Figura 10. El costo W_k de cada celda k de la matriz de costos se puede calcular de dos formas, como el valor absoluto de la diferencia de ambas secuencias X y Y , o como el cuadrado de la diferencia de la secuencia tanto en X como en Y como se expresa en la ecuación (5) y (6). Para este estudio se utilizará la ecuación seis. Una vez calculada la matriz de costos, lo único que resta por hacer es seguir la ruta más corta iniciando desde el punto de término de la matriz (en el caso del ejemplo la celda con el costo $C_{7,5}$), conservando las condiciones de monotonicidad y del tamaño de paso (la comprobación de la ruta más corta no se explica por razones de simplificación, se puede observar el teorema 4.3 en (Müller 2007) así como su comprobación).

$$W_k = |X_k - Y_k| \quad (5)$$

$$W_k = (X_k - Y_k)^2 \quad (6)$$

Una vez que se tiene la ruta de distorsión, se verificará que esta ruta cumpla con las dos condiciones de igualdad entre ambas secuencias de datos (Vaughan & Gabrys 2016). Estas condiciones se ajustarán de tal forma que no sean tan estrictas, pero a su vez que tampoco sean inefectivas al dejar pasar cualquier señal. Las tres condiciones son las siguientes:

- Ventana de distorsión: Esta condición indica que la ruta de distorsión no debe vagar demasiado lejos de la diagonal principal, ya que mientras más cercana sea la ruta a esta diagonal, mayor similitud habrá entre ambas señales. El objetivo de esta condición es impedir que la alineación no intente omitir características diferentes y que se quede atascada en

características similares. Para que se cumpla esta condición el valor absoluto de la diferencia del k-esimo elemento de la ruta de distorsión de ambas secuencias debe ser menor a un ancho de ventana de tamaño r , con $r > 0$, de tal forma que $|X_k - Y_k| \leq r$. En la Figura 11 muestra más claramente la idea anterior. El cambio que se hizo a esta condición para que no fuera tan estricta (ya que se comparará con secuencias similares más no iguales) es que el 30% de la ruta, como máximo, puede estar fuera de esta ventana.

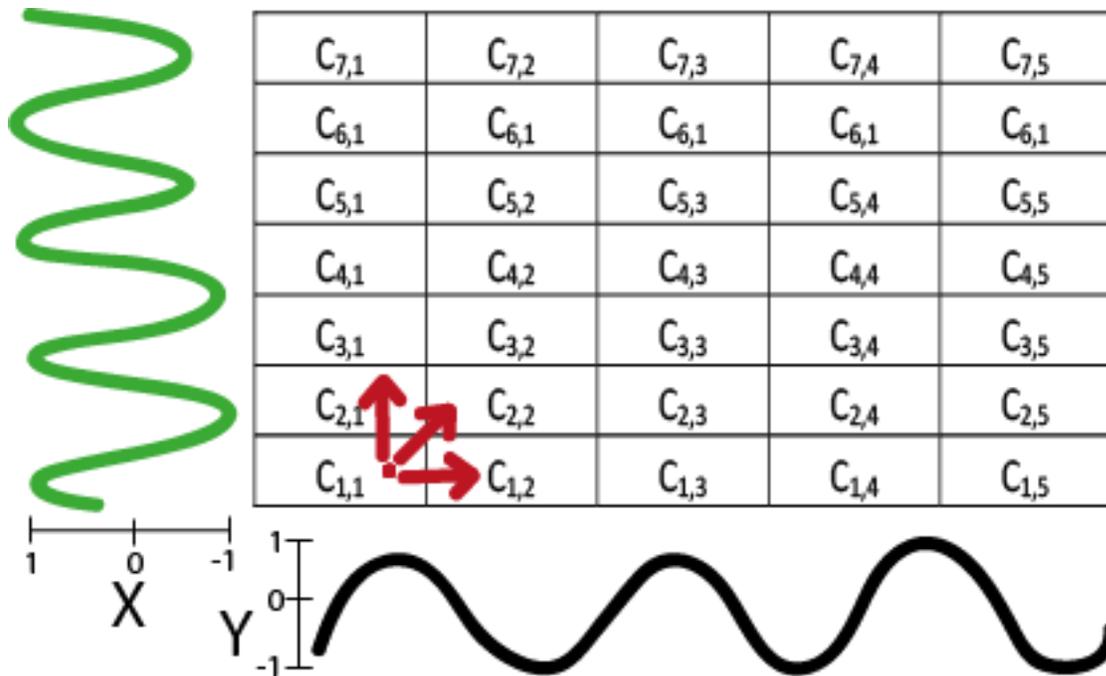


Figura 10: Movimiento para el cálculo de los costos de la matriz de costos, realizado en un tiempo $O(nm)$, siempre cumpliendo con las 3 condiciones.

- Restricción de pendiente: Esta condición implica que los crecimientos de forma horizontal y vertical no sean demasiado prolongados. La forma de evaluarlo es midiendo la pendiente de la trayectoria de la ruta de distorsión mediante las ecuaciones (7) y (8):

$$\frac{(Y_k - Y_0)}{X_k - X_0} \leq p \quad (7)$$

$$\frac{(X_k - X_0)}{Y_k - Y_0} \leq q \quad (8)$$

Donde $p \geq 0$ es el número de pasos permitidos en la dirección de Y y $q \geq 0$ es el número de pasos consecutivos en la dirección de X como se muestra en la Figura 12. Con esto se espera que después de q pasos exista uno o más pasos p y viceversa. Con esto se impide que la trayectoria sea demasiado empinada o sea demasiado superficial.

En esta restricción es donde más ajustes se realizan, dado que la matriz de costos es de tamaño $N \times M$, siempre habrá $|N-M|$ pasos más en la horizontal o en la vertical que pueden implicar violar esta regla. Como ejemplo esta la Figura 8, donde la matriz es de 7×5 , lo que implicaría que habrá forzosamente mínimo dos pasos más en dirección de X . En (Ratanamahatana & Keogh 2004) se establece que cuando la matriz es de $N \times M$, la ventana de distorsión debe ser interpolada para que sea funcional con este tipo de matrices, en lugar de simplemente no aplicarse esta restricción cuando se tiene una matriz de este tipo. Además de esto, las ecuaciones 7 y 8 están diseñadas para una línea recta donde $(x, y) \in (R, R)$ y para este caso $(x, y) \in (N, N)$, lo que provoca pérdida de información. En este estudio se compensará esta diferencia entre N y M al permitir $|N-M|$ pasos, por lo que restando este excedente de pasos la diferencia proporcional en los pasos horizontales y verticales no debe ser mayor al 35%, es decir, en ninguna de las dos direcciones debe haber más del 35% de pasos de forma continua.

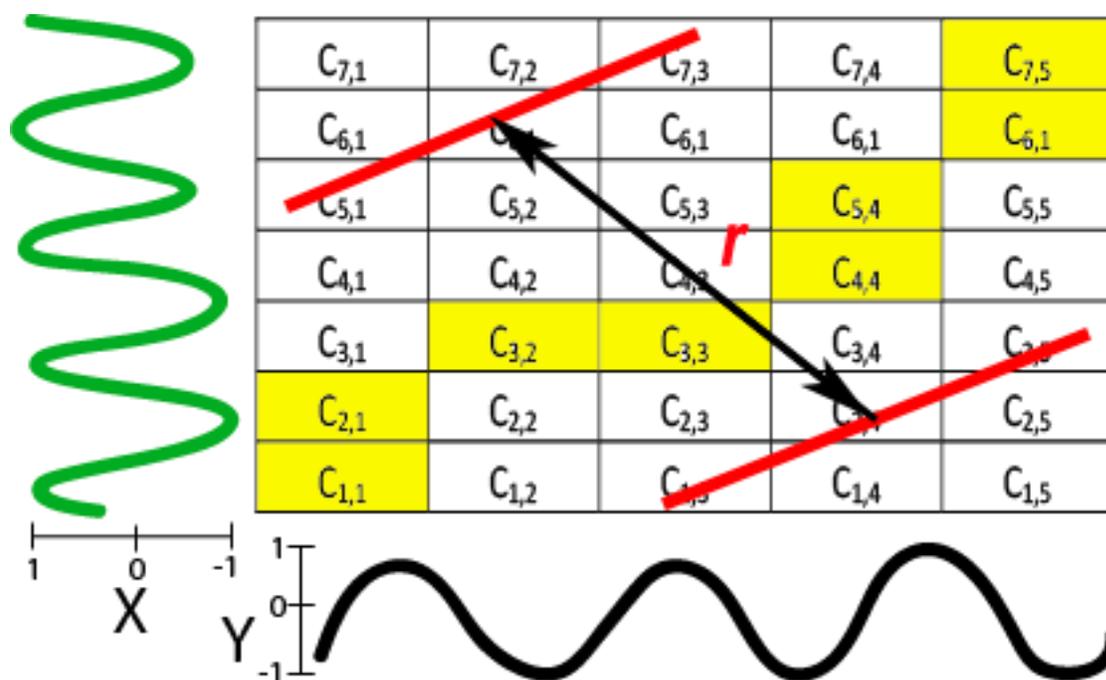


Figura 11: Ventana de distorsión de tamaño r . La ruta de distorsión debe mantenerse dentro de los límites establecidos por esta ventana.

De no cumplirse alguna de estas dos restricciones (asumiendo que la ruta de distorsión es válida o existe) la secuencia se desecha y se volverá a pedir al usuario capturar esta información, en caso contrario puede ser analizada para identificar el inicio y el fin del gesto.

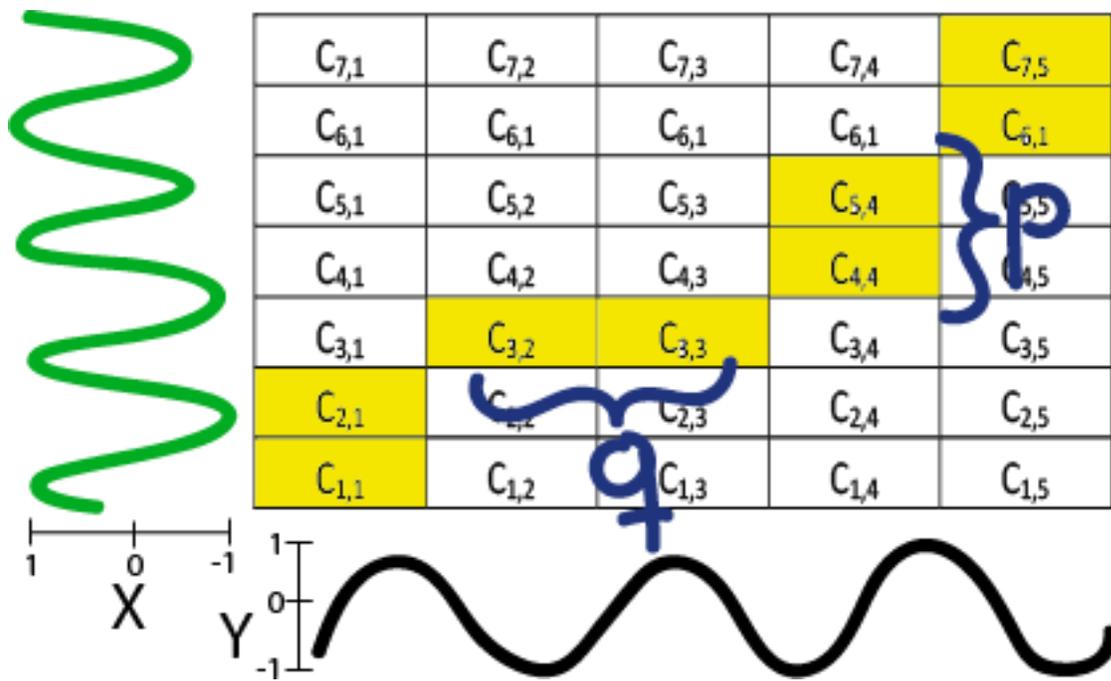


Figura 12: Restricción de pendiente donde la relación proporcional entre p y q no debe ser mayor al 20%, es decir, de cada 5 celdas que se acumulen en q debe de haber al menos 4 celdas previas en p y viceversa.

En la Figura 13 se puede observar la señal de referencia, que es el gesto de pellizco con el dedo índice, tomada como base por no contener ruido en exceso en su señal. En la figura *b*) se observa el gesto de pellizco doble con el anular que es reconocido por el DTW modificado como un posible gesto, según su ruta de distorsión, ya que cumple todas las reglas y tiene menos del 31% de la señal casi al finalizar la ruta (movimiento en q) según la Figura 14. En *c*) se observa la señal del movimiento de la mano al girarla en círculos sobre el eje de la muñeca, que es identificada como señal no perteneciente a un posible gesto dado que tiene más del 80% de los pasos en p y poco más del 50% de los pasos en q , además de sobrepasar la ventana de distorsión con más datos de los permitidos, Figura 15.

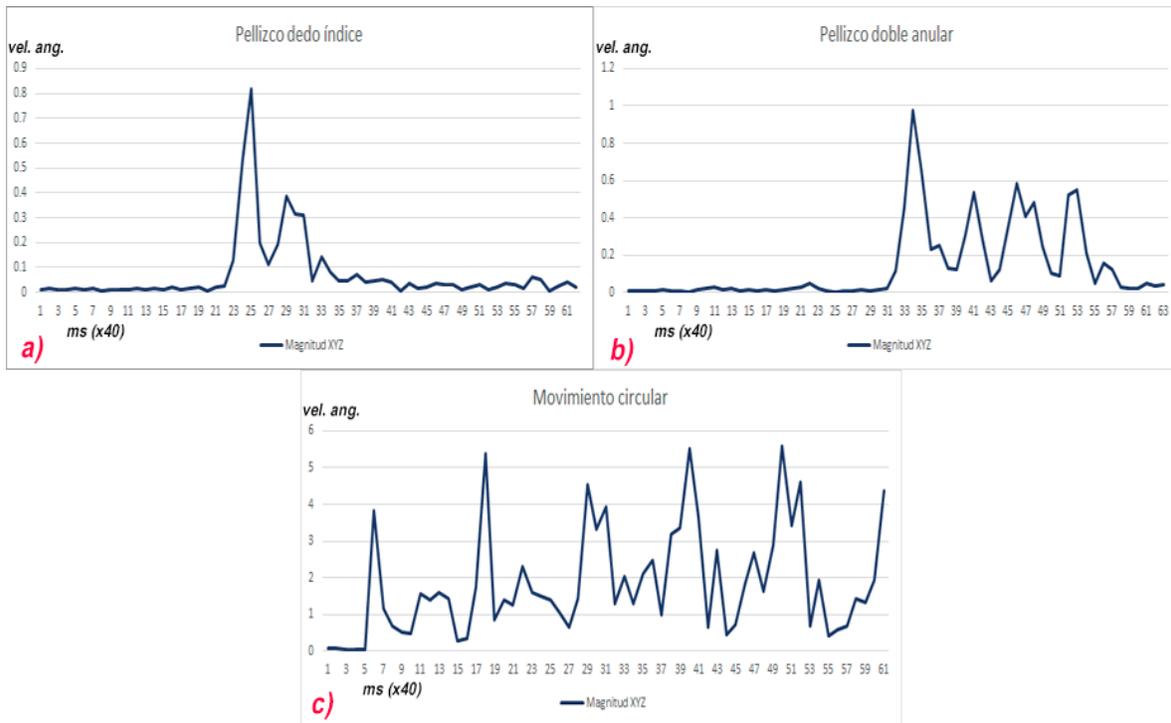


Figura 13: Ejemplo de (a) señal de referencia, (b) posible señal de un gesto y (c) señal del movimiento de la mano formando círculos manteniendo el antebrazo estático.

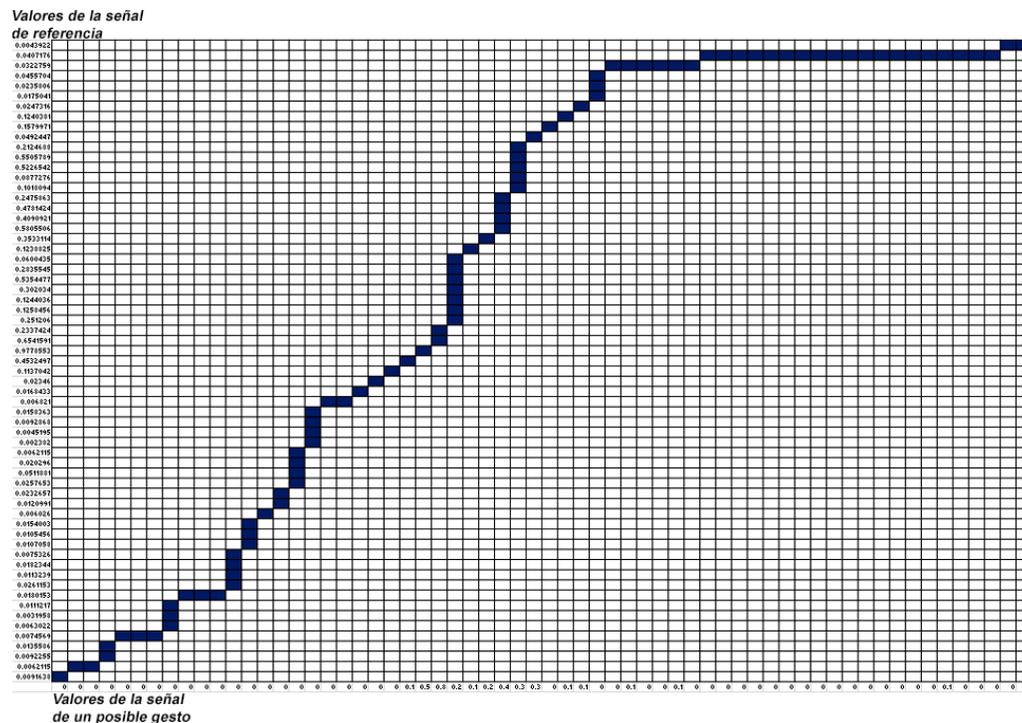


Figura 14: Ruta de distorsión generada a partir de la comparación de la señal de referencia y un posible gesto.

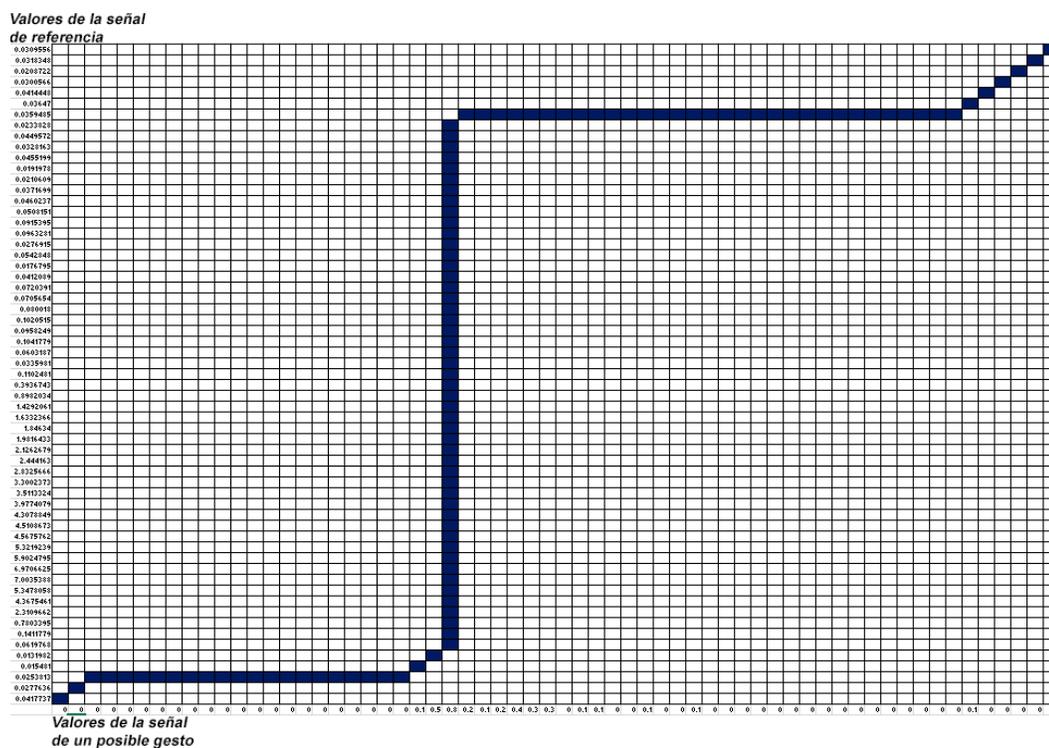


Figura 15: Matriz de distorsión generada a partir de la comparación de la señal de referencia y la señal con ruido que no pertenece a ninguno de los gestos sutiles.

3.4. Segmentación de la secuencia de datos

Una vez que se ha determinado que los datos provenientes del giroscopio se asemejan a uno de los gestos, mediante la comparación de secuencias de datos, se procede a buscar el inicio y el fin del gesto sutil.

Con la secuencia de datos que se utilizó en el proceso anterior (recordar que esta secuencia es resultado de la suma de las magnitudes de los ejes XYZ del giroscopio) se realizarán tantas divisiones como sean posibles, de seis elementos por división, salvo al final de la señal ya que de quedar dos elementos libres o menos, estos se unirán al segmento anterior, quedando este de siete u ocho elementos, si por el contrario el número de elementos sobrantes es mayor o igual a tres, estos constituirán el segmento final, como se muestra a la izquierda de la Figura 16, donde el segmento final se conformó de ocho elementos.

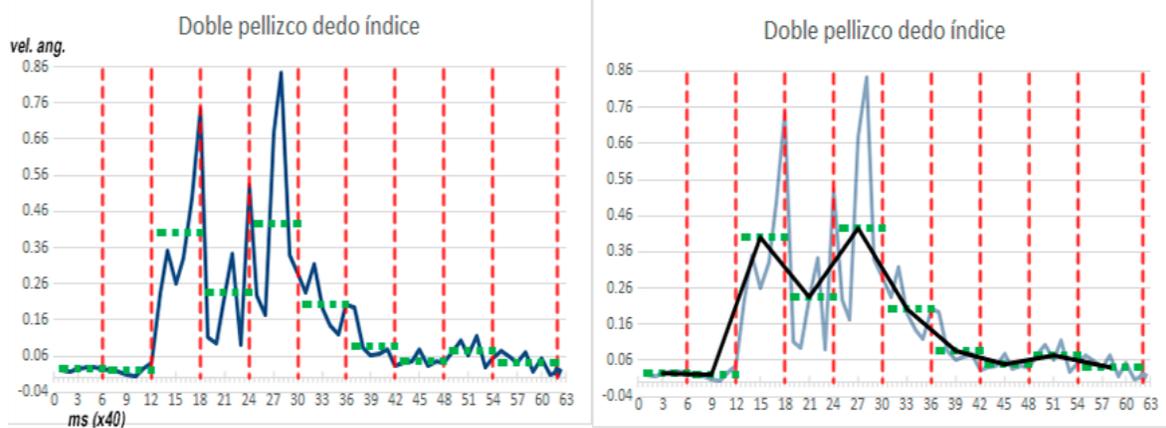


Figura 16: A la izquierda está la segmentación en secuencia de seis datos (salvo el final) provenientes del giroscopio para encontrar el segmento donde inicia y termina el gesto. En cada segmento se obtendrá el promedio (marcado por la línea verde punteada) para poder simplificar la señal y obtener una señal simplificada, como se muestra en la figura de la derecha, que permita eliminar datos con ruido.

Para cada segmento se calculará su promedio (línea verde punteada) para lograr una mejor representación de cada uno de los segmentos como se menciona en (Fu 2011). Si establecemos que la secuencia de datos está representada por $S = (s_1, s_2, \dots, s_m)$ y n es la dimensión de cada segmento (para este estudio se usarán seis datos por segmento), esta serie de tiempo comprimida establecida como $\hat{S} = (\hat{s}_1, \dots, \hat{s}_n)$ puede ser calculado mediante la ecuación 9.

$$\hat{S}_k = \frac{1}{e_k - f_k + 1} \sum_{i=f_k}^{e_k} S_i \quad (9)$$

Donde f_k y e_k representan la posición de inicio y fin respectivamente, del k -ésimo segmento en la secuencia o serie de tiempo S . Esta nueva serie de tiempo será el resultado mostrado en la parte derecha de la Figura 16, sobre la cual se empezará a buscar el inicio y fin del gesto.

Una vez que se tiene la nueva serie de tiempo \hat{P} , se define un umbral denominado umbral alto con valor escalar de 0.2 para determinar aquellas señales que podrían ser consideradas como un gesto, representada en la Figura 16 con una línea punteada verde. Para todo punto que se encuentre por encima

de este umbral, se buscará el inicio y el fin de esta señal, tomando en cuenta que el gesto inicia y termina en el reposo. Para saber dónde inicia y termina cada punto que está por encima del umbral alto, se define un nuevo umbral, denominado umbral bajo con valor escalar de 0.08 (en color rojo en la Figura 17) que permitirá establecer el inicio y el fin. El primer punto de menor o igual valor al umbral bajo, tanto del lado izquierdo (punto de inicio) como del lado derecho (punto final), será el inicio y el fin de ese punto en particular.

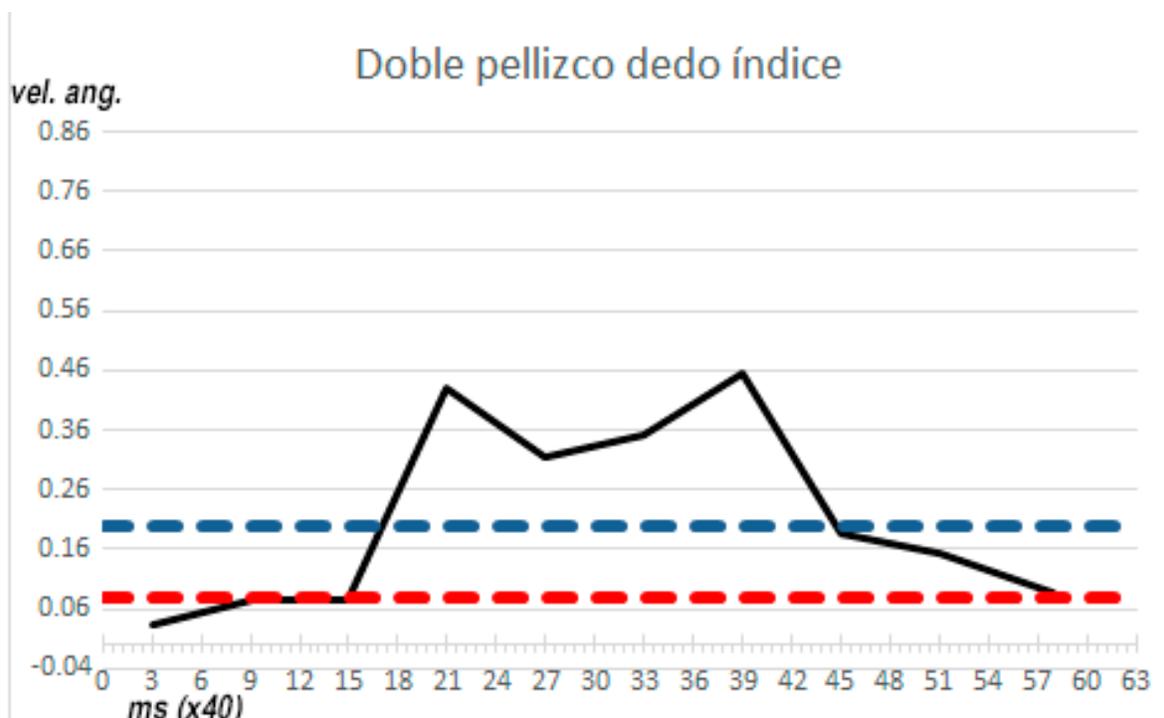


Figura 17: Buscando el inicio y el fin preliminares en la serie de tiempo proveniente de la segmentación.

Para el ejemplo mostrado en la Figura 17, se tienen tres puntos que superan el umbral alto, que están en la posición 15, 21 y 27 respectivamente. Para este caso en particular, los tres puntos tendrán como inicio el punto 9 y como fin el punto 45 con respecto al eje X, pertenecientes al segmento 2 (el punto 9) y al segmento 8 (el punto 45) respectivamente. Para este ejemplo, el inicio y el fin será el mismo para cada punto que supera el umbral, empezando a partir de la posición 6, ya que es aquí donde inicia el segmento 2, hasta la posición 48, que de igual manera es el final del segmento 8. Llamaremos trayectoria a cada punto que se encuentre entre el inicio y el fin de cada uno de los puntos que superen el umbral, y para nuestro caso de ejemplo serán tres trayectorias con un mismo inicio y un mismo fin.

Si el inicio de alguna de las trayectorias se detecta en el primer elemento de \hat{S} y este está por arriba del umbral bajo, se tomará como el inicio de la trayectoria esta posición inicial; de manera similar, si el fin de alguna trayectoria es detectado en el último elemento de \hat{S} , este será tomado como el fin de la trayectoria.

Cuando se tengan todas las trayectorias definidas, se deben conjuntar éstas hasta finalmente tener una sola trayectoria final, por lo que si tenemos la secuencia de datos representada por $S = (s_1, s_2, \dots, s_m)$, para conjuntar dos trayectorias que se unen en cierto punto y que ambas pertenecen a S , se debe cumplir lo siguiente:

- Si existen dos trayectorias T_1 y T_2 con punto de inicio d y f respectivamente, y punto final e y g , tal que $T_1 = (s_d, s_{d+1}, \dots, s_e)$ y $T_2 = (s_f, s_{f+1}, \dots, s_g)$ con $s_i \in S$, ambas trayectorias se unirán en una sola trayectoria $T_{12} = (s_i, s_{i+1}, \dots, s_j)$ con $i = \min(d, f)$ y $j = \max(e, g)$ siempre y cuando $d \leq f \leq e$ y/o $f \leq e \leq g$.

La nueva trayectoria T_{12} reemplazará a ambas trayectorias y esta nueva trayectoria se deberá conjuntar con las trayectorias restantes hasta que solo exista una trayectoria final T_{ij} con $1 \leq i < j \leq m$, donde $i-3$ y $j+3$ serán el inicio y el fin del gesto respectivamente, dado que estarán a la mitad del segmento. Si, por el contrario, al finalizar la conjunción de trayectorias existe más de una trayectoria, entonces se podrá decir que existe más de un movimiento al realizar la lectura del gesto, por lo que se desechará como un gesto válido y se solicitará al usuario que realice el gesto nuevamente.

Hasta este momento, el inicio y fin del gesto (de haberse encontrado) es parcial, ya que solamente se ha verificado que se tratara de un solo gesto y que este tenga el menor ruido posible. Por este motivo es necesario revisar la señal directamente en busca del inicio y fin del gesto, pero con la particularidad de que se buscará solo en el rango que se detectó anteriormente, es decir desde la posición i hasta j , y no desde la posición 1 hasta la posición m .

3.4.1. Establecer el inicio y el fin de un gesto sutil

Una vez con que se tiene acotado el inicio y el fin del gesto según la segmentación realizada, como se muestra en la Figura 18, nos regresaremos a la secuencia de datos generada por la suma de las

magnitudes de los tres ejes del giroscopio y se usará solamente de la posición inicial a la final establecida en el paso anterior. Esta secuencia acotada se recorrerá para buscar el momento que supere un umbral fijado en 0.19, debido a que a partir de este valor se estableció empíricamente que empieza un movimiento de la mano (establecido utilizando los datos de siete sujetos que realizaron los tres primeros gestos cinco veces cada uno) que no sea por efecto de movimientos espasmódicos que son casi imperceptibles como puede ser el temblor en la mano. De manera simultánea se buscará el momento en que la secuencia supere un umbral fijado en 0.17 (establecido por observaciones al igual que el anterior), pero recorriendo la secuencia desde el último dato hacia el primero y este será el nuevo fin del gesto. Si la distancia del inicio y final del gesto no supera los 200 milisegundos (tomar en cuenta que cada incremento en el eje horizontal equivale a 40 milisegundos) se desecha el gesto y se pedirá al sujeto que repita el gesto.

Con este nuevo inicio y fin del gesto se procederá a eliminar aquellos datos iniciales y finales que fueran tomados en cuenta debido a algún movimiento involuntario de la mano. Para eliminar estos pequeños picos involuntarios que pudieran haber pasado todos los filtros anteriores, se tiene que cumplir la regla de que el inicio y el fin del gesto no deben estar a más de 220 milisegundos de un valor igual o superior de 0.4 (establecido para garantizar que no fue producto de un movimiento brusco esporádico), como se muestra en la Figura 19. Si el nuevo inicio o fin de la secuencia no cumplen con esta regla los valores se recorrerán (hacia la derecha para el inicio y hacia la izquierda para el fin) hasta el siguiente valor más próximo que sea igual o superior a su umbral respectivo. Para el ejemplo mostrado en la Figura 19, como el inicio si cumple con la regla de los 220 ms no se hará el recorrido hacia la derecha. Por el contrario, el final de la secuencia no cumple esta regla, por lo que deberá recorrerse hacia la izquierda hasta que cumpla esta condición, como se muestra en la Figura 20. Este recorrido se hará hasta la posición 33, que tiene un valor de 0.18 y se encuentra a 200 ms del valor mínimo establecido en 0.4.

Así pues, para este ejemplo, el inicio y final definitivo se fijará en la posición 13 y 33 de la secuencia de datos con valores de 0.23 y 0.18 respectivamente. Solo restará realizar una última comprobación: el gesto no debe durar más de 1300 ms, de hacerlo será desechado y se pedirá al sujeto que repita el gesto. Para este ejemplo la duración es de 800 ms, por lo que el gesto se considera válido.

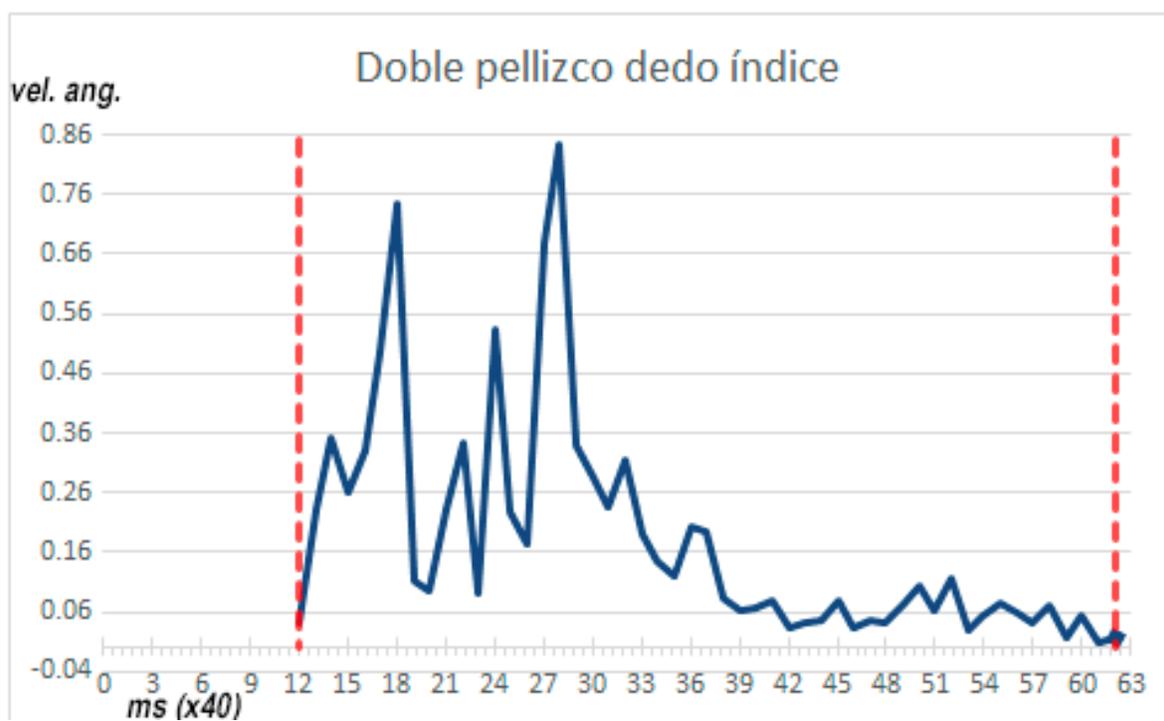


Figura 18: Secuencia de datos provenientes de la suma de las magnitudes de los ejes del giroscopio acotados según los umbrales aplicados a la segmentación.

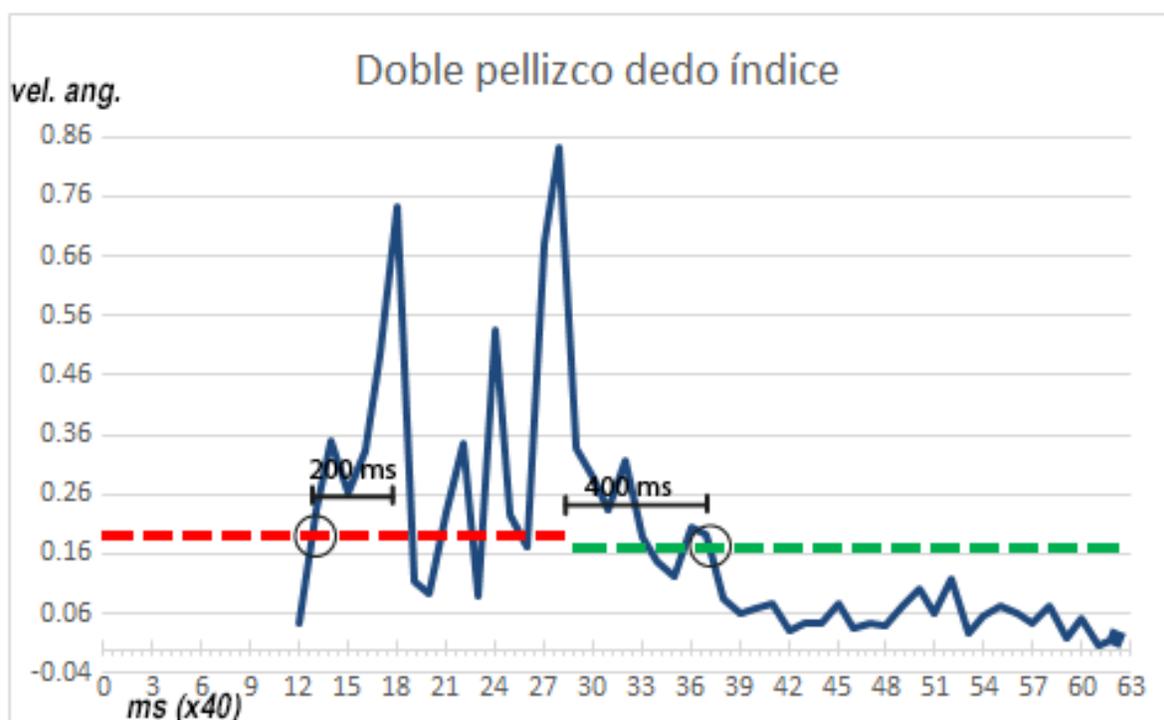


Figura 19: Búsqueda del inicio y fin del gesto mediante umbrales tanto de inicio, en color rojo, y de fin, en color verde, fijados en 0.19 y 0.17 respectivamente.

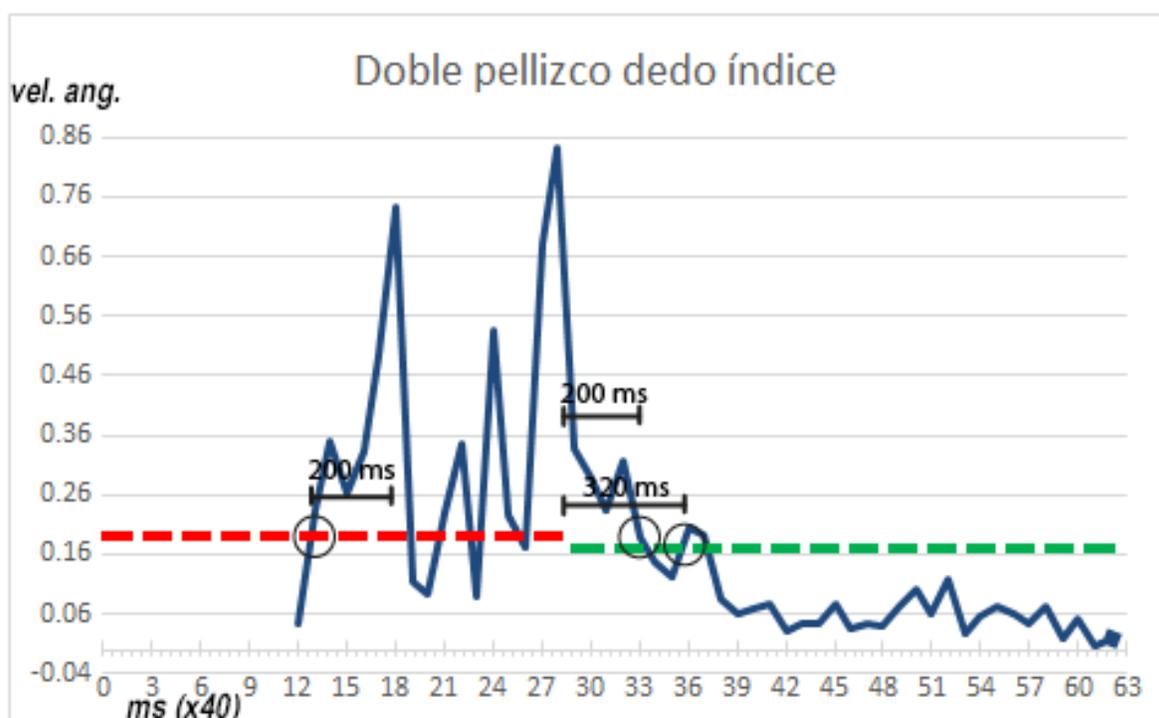


Figura 20: Recorrido hacia la izquierda por no cumplir la regla de estar a una distancia de 220 ms de un valor mínimo de 0.4. El inicio y final quedarían en la posición 13 y 33 respectivamente.

Con el inicio y fin establecidos, finalmente se tomará como inicio el punto inmediato anterior a este, de manera similar para el punto final se tomará el inmediato posterior. De no existir un punto posterior o anterior (que sean el inicio o el fin de la secuencia original) se quedará el punto tal como se obtuvo.

3.5. Extracción de características

Ahora que ya se tiene el inicio y fin de un posible gesto, se procederá a extraer las características que permitan identificar el gesto mediante el uso de minería de datos. Los valores que se tomarán en cuenta para la extracción de características serán los provenientes de los tres ejes de ambos sensores (acelerómetro y giroscopio), comprendidos entre la posición de inicio y fin preestablecida. En la Figura 21 se muestra las series de tiempo acotadas de los tres ejes de cada uno de los sensores.

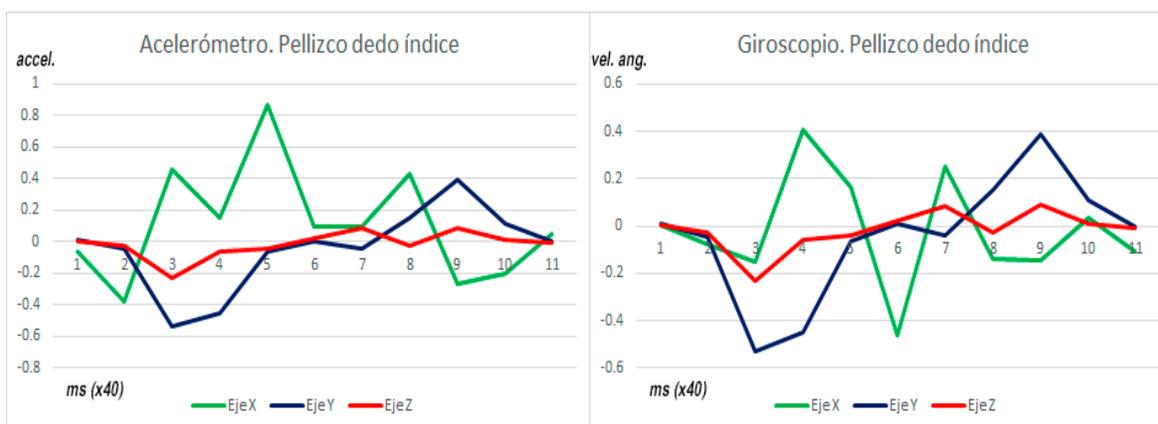


Figura 21: Series de tiempo de los tres ejes del acelerómetro (izq.) y giroscopio (der.) acotados por el inicio y fin preestablecido.

A partir de los ejes XYZ, tanto del acelerómetro como el giroscopio se obtendrán las combinaciones de las magnitudes de los tres ejes, como se especifican en la Tabla 2. El objetivo de estas combinaciones es poder compensar las orientaciones que podrían tener los sensores al realizar alguno de los gestos sutiles (Wen et al. 2016). Por orientaciones se debe entender, la posición que puede llegar a tener la mano como se explicó en el capítulo 3.3.

Tabla 2: Combinación de los ejes de cada sensor que permite crear nuevas características.

Acelerómetro/Giroscopio		
Ejes	Acrónimo	Operación
X	X	X
Y	Y	Y
Z	Z	Z
X y eje Y	XY	$\sqrt{X^2 + Y^2}$
X y eje Z	XZ	$\sqrt{X^2 + Z^2}$
Y y eje Z	YZ	$\sqrt{Y^2 + Z^2}$
X, eje Y y eje Z	XYZ	$\sqrt{X^2 + Y^2 + Z^2}$

De esta manera, en lugar de tener seis series de tiempo diferentes, una por cada eje de cada sensor, se tendrán 14 series de tiempo diferentes. A estas nuevas series de tiempo generadas se les aplicarán una serie de funciones estadísticas que permitan extraer una serie de características de cada serie de tiempo. Estos estadísticos serán el promedio, máximo, mínimo, desviación estándar y los tres cuartiles. Al combinar estos estadísticos con cada una de las series de tiempo generadas, se obtendrán 98 (14x7) características hasta el momento. Otras características que se tomaron en cuenta, pero que fueron desechadas, son la velocidad, velocidad angular y la duración. La razón por la que estas últimas características fueron desechadas es porque su razón de cambio entre cada gesto es muy baja, por ser movimientos muy parecidos.

Dado que cada serie de tiempo puede ser considerada como una función del tiempo, se puede buscar el área bajo la curva o la integral de la función. Dado que todos los puntos están separados a una distancia constante de 40 ms, se puede calcular el área bajo la curva de cada serie de tiempo sin que existan perturbaciones por el tiempo de lectura de cada punto. Ahora bien, la secuencia X, Y y Z pueden dar valores negativos, lo que causaría que de existir un área positiva y una negativa estas se contrarresten, tal como sucede al realizar la integral. Es por esto por lo que en estas tres series de tiempo se calculará el área positiva y la negativa por separado. Para las series de tiempo XY, XZ, YZ y XYZ no se realizará el cálculo del área bajo la curva. El resultado del cálculo de las diferentes áreas por cada serie de tiempo genera 12 características nuevas (seis por cada sensor), que sumadas a las 98 ya existen, nos dará un total de 110 características hasta el momento.

Al iniciar este estudio, se pensó en usar la transformada rápida de Fourier para crear un filtro pasa bajas, que permitiera reducir el ruido que pudiera provenir de cada una de las series de tiempo. Al observar las muestras de los siete sujetos para el análisis de la señal, se concluyó que los datos al ser muy pequeños, por tratarse de gestos sutiles, al aplicárseles este filtro pasa altas solo se obtendrían señales muy parecidas a una onda senoidal y por lo tanto prácticamente iguales en cada uno de los gestos. Por esta razón, se utilizó la transformada de Fourier para ser aplicada en cada una de las series de tiempo generadas de la combinación de sensores, y aplicarle a su vez los siete estadísticos propuestos. Con estas nuevas características añadidas tendríamos un total de 98 características más, que sumadas a las 110 ya generadas, nos darán un total de 208 características en total. La Tabla 3 muestra una vista simplificada de la estructura de las características extraídas de los datos provenientes del acelerómetro y giroscopio de los gestos sutiles.

Tabla 3: Representación simplificada de la extracción de 208 características a partir de los tres ejes tanto del acelerómetro como del giroscopio de los gestos sutiles.

Accel_X_Prom	Accel_X_Max	Accel_X_Min	Accel_X_D	Accel_X_Q1	Accel_X_Q2	Accel_X_Q3	...	Gyro_X_Prom	Gyro_X_Max	Gyro_X_Min	...	Accel_FFT Prom	...	Etiqueta
0.097811	1.330094	-1.581338	0.555059	-0.071609	0.415226	4.00972	...	1.063944	1.252639	2.053969	...	-0.1683	...	índice
0.072367	0.528059	-0.33529	0.206594	-0.083533	0.209251	2.601133	...	0.593658	0.298162	0.760125	...	-0.141149	...	índice
-0.015188	0.339446	-0.759814	0.23224	-0.09006	0.140396	1.552058	...	0.425493	0.686618	0.852984	...	-0.265349	...	medio
0.052106	1.189076	-0.685937	0.397189	-0.204513	0.271822	2.935822	...	0.743906	0.980658	1.442666	...	-0.447153	...	medio
0.238339	1.352288	-0.26882	0.419957	0.058562	0.285881	4.32058	...	0.994473	1.190135	1.565139	...	-0.207485	...	anular
0.084675	0.920418	-1.080098	0.54589	-0.126631	0.312022	4.15681	...	1.232085	0.951933	1.166285	...	-0.24916	...	anular
0.23409	1.978024	-0.734594	0.874873	-0.578759	0.833957	6.695059	...	2.034176	1.584207	3.25199	...	-0.48097	...	2índice
0.248937	2.253834	-1.106709	0.929791	-0.525776	0.91829	9.038885	...	2.187679	1.64938	2.596925	...	-0.107051	...	2índice
0.491688	1.376027	-0.78163	0.617181	0.102641	0.955071	9.140321	...	2.246388	0.499716	1.20347	...	-0.031767	...	2medio
0.577768	1.833708	-0.501343	0.686373	-0.147798	1.233162	11.867991	...	2.910451	0.775033	1.705943	...	-0.125331	...	2medio
0.014183	0.934908	-0.552613	0.349955	-0.242733	0.237221	1.700251	...	0.460709	0.688554	0.807878	...	-0.400517	...	2anular
-0.166639	0.280458	-0.648104	0.203923	-0.315104	-0.020065	3.29291	...	0.792021	0.270781	0.486886	...	-0.365185	...	2anular

3.6. Uso de aprendizaje máquina para establecer dos modelos en la detección de los gestos sutiles

El aprendizaje de máquina se ha usado desde la década de los 70's, cuando aparecieron los primeros algoritmos relacionados con este tema (Louridas & Ebert 2016) y ha sido aplicado en diferentes dominios como el análisis digital de imágenes, heurística de seguridad y reconocimiento de patrones. El objetivo del aprendizaje de máquina es que la computadora pueda distinguir un patrón de otro, en base a un conjunto de datos de entrenamiento y un conjunto de atributos asociados a una instancia a clasificar.

En la sección anterior se describe la manera en que se obtendrán los datos de entrenamiento para que un algoritmo clasifique los gestos sutiles con estos datos. Para la clasificación se usarán dos algoritmos o modelos que han probado ser eficientes. El primero es máquinas de soporte vectorial (SVM, por sus siglas en ingles) y el segundo es la Optimización Mínima Secuencial (SMO) (Samuel n.d.). El primer algoritmo se usará debido a que es uno de los algoritmos de clasificación con mayor influencia (Samuel n.d.) y ha demostrado ser tan bueno como las redes neuronales. El otro modelo ideal para usarse sería precisamente el de redes neuronales, sin embargo, en pruebas rápidas para determinar los tiempos de ejecución de los modelos, el algoritmo de redes neuronales presente en WEKA, llamado Perceptrón Multicapa, llevaba 10 minutos en ejecución cuando fue interrumpido por considerarse que era un tiempo demasiado largo. Por esta razón se decidió usar en este estudio SMO por hacer uso de máquinas de soporte vectorial, lo cual permitirá, de darse resultados aceptables y cercanos a SVM, generar los modelos directamente desde el reloj inteligente, lo que permitirá a su vez que cualquier usuario entrene al modelo y arroje mejores resultados. Lamentablemente, actualmente no se puede realizar el entrenamiento por

cualquier usuario ya que el algoritmo de SVM no es nativo de WEKA, por lo que la actual API de WEKA no permite el entrenamiento y generación del modelo al mismo tiempo para SVM.

3.6.1. Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial son algoritmos de aprendizaje supervisado, lo que significa que requieren de datos de entrenamiento. Además, SVM es un algoritmo de clasificación, lo que implica que se usan para predecir si un conjunto de características asociadas a una instancia pertenece a una clase en particular. Esto es lo que se necesita para detectar los gestos sutiles haciendo uso de las características que se extrajeron en la sección 3.5.

El objetivo de una máquina de soporte vectorial es encontrar el hiperplano de separación óptimo que maximice el margen de los datos de entrenamiento. Para entender qué significa esto, tomaremos dos características (aceleración promedio en el eje X del acelerómetro y velocidad promedio en el eje X del giroscopio) de las 208 extraídas anteriormente, de dos gestos solamente, el pellizco con el dedo índice y con el medio, con tres repeticiones cada uno, como se muestra en la Tabla 4.

Con los datos extraídos anteriormente se puede graficar una característica vs la otra, como se muestra en la Figura 22, con el objetivo de crear una línea que separa ambos gestos y que según su ubicación permita realizar la clasificación, como se muestra en la Figura 23.

Tabla 4: Datos promedio del acelerómetro (columna 1) y del giroscopio (columna 2) en su eje X, así como la etiqueta del gesto sutil al que pertenece. Estos datos son considerados de entrenamiento para los algoritmos de clasificación.

Acel. Eje X Promedio	Giros. Eje X Promedio	Etiqueta
0.02199	0.026023	índice
0.010771	-0.01627	índice
0.178271	-0.110329	índice
0.079895	0.004667	medio
0.072565	0.065159	medio
0.061183	-0.01459	medio

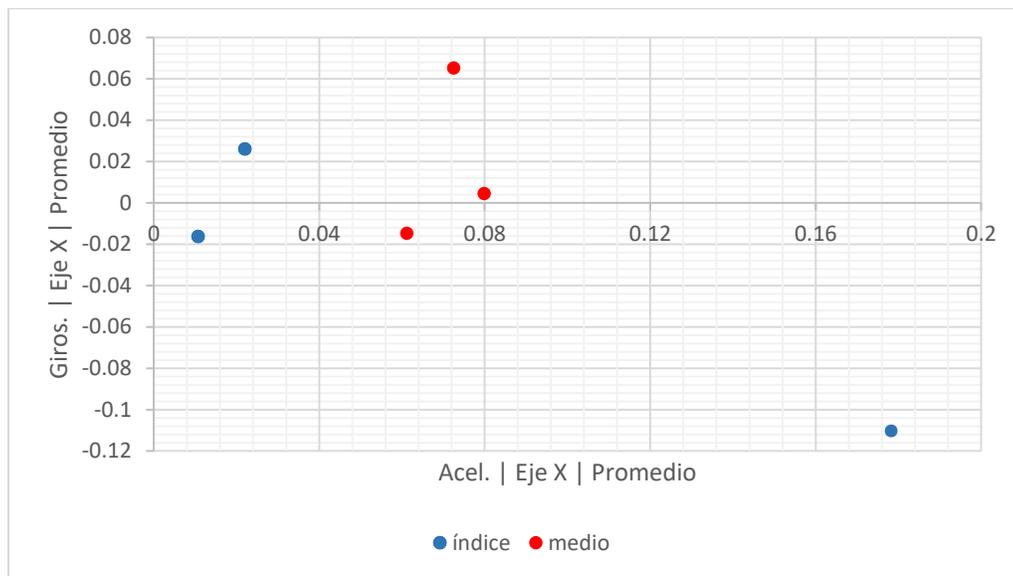


Figura 22: Eje X Promedio del acelerómetro vs Eje X promedio del giroscopio, generado a partir de la Tabla 4.

Esta línea generada es considerada un hiperplano en R^2 . Si se agregara una tercera característica más, el dominio del hiperplano sería R^3 y la gráfica resultaría de tres dimensiones, para este estudio el dominio está en R^{208} . El problema en SVM radica en encontrar aquel hiperplano que cumpla con el objetivo de este algoritmo, por lo que la cantidad de hiperplanos que se puede generar es prácticamente infinita, como se muestra en la Figura 24. Las máquinas de SV resuelven el problema de encontrar el hiperplano de separación óptima de tal forma que, este hiperplano se mantenga tan alejado como sea posible de los puntos de cada categoría. La separación máxima que hay entre una categoría y otra es la misma para ambos lados y a esta separación se le llamará costo y es uno de los principales parámetros a usar en el algoritmo de SVM.

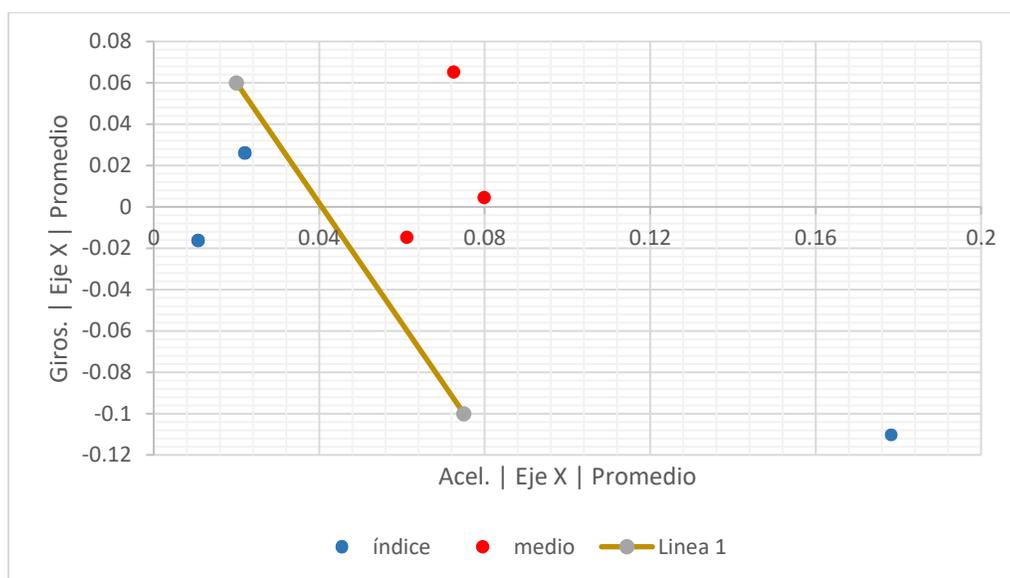


Figura 23: Línea que separa ambos gestos para lograr de esta forma su clasificación. Todo lo que esté a la izquierda de la línea será considerado como pellizco con el índice, de otra forma será pellizco con el dedo medio.

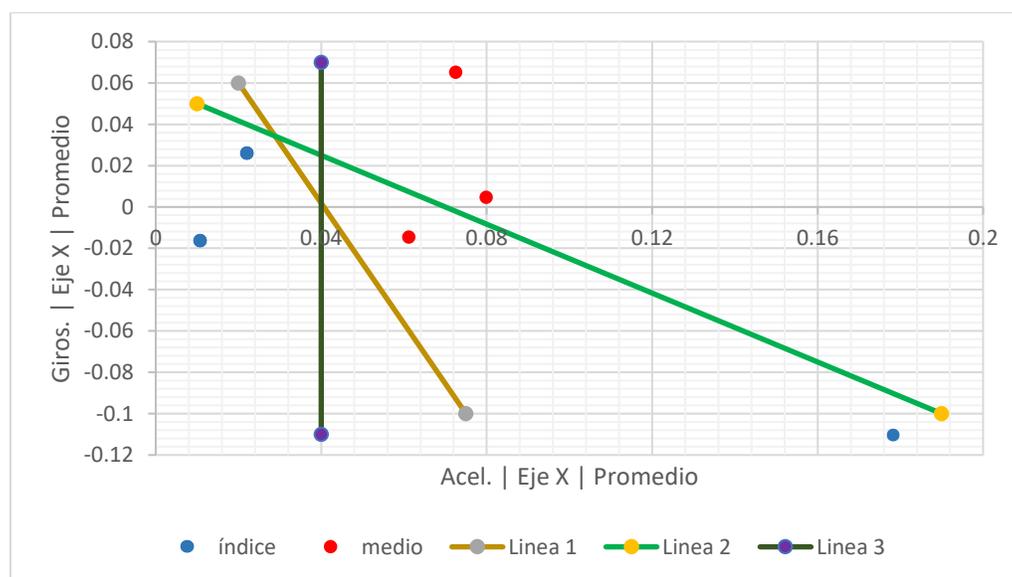


Figura 24: El número de hiperplanos que se pueden generar para separar ambos gestos llega a ser infinito.

Otro parámetro importante para tomar en cuenta es el valor de gama, que no es otra cosa que el ángulo que tomará el plano para separar los datos ya que, por ejemplo, la línea que separe ambas categorías podría ser una que tome una forma de un círculo, como se muestra en la Figura 25.

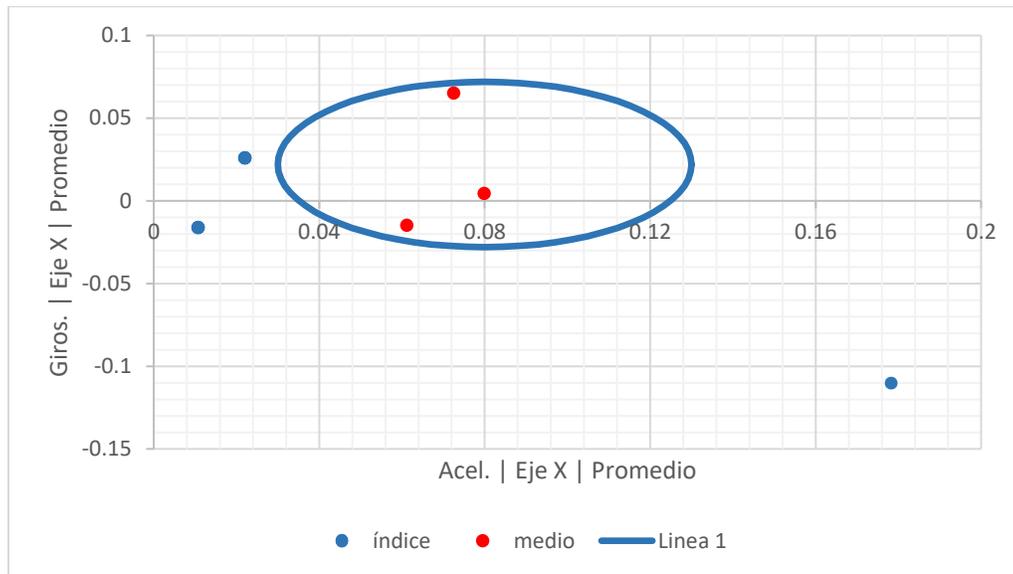


Figura 25: Hiperplano en forma de círculo debido a un valor de gama que así lo permite.

Un tercer parámetro por modificar para obtener mejor precisión al clasificar los gestos sutiles es el kernel, del cual los más usados son el lineal y radial basis function (RBF). El kernel lineal se usa por lo general cuando el número de características es demasiado grande (por lo general más de 1000) y el kernel RBF en caso contrario. En este estudio se usarán ambos kernel para determinar cuál presenta los mejores resultados.

3.6.2. Optimización Mínima Secuencial

Este algoritmo resuelve el problema de programación cuadrática (QP) en SVM al descomponerlo en sub-problemas QP y solucionar el menor problema de optimización posible, involucrando dos multiplicadores de Lagrange.

Tanto la descripción del funcionamiento de este algoritmo como los fundamentos y demostraciones matemáticas que lo fundamentan (incluido el algoritmo SVM) se puede consultar en (Platt 1998).

3.6.3. Escalamiento de datos

Tanto el algoritmo SVM como SMO necesitan de un escalamiento de los datos, ya que como se puede inferir por la corta explicación del funcionamiento de SVM, el algoritmo funciona realizando operaciones entre los vectores de datos que representan a cada una de las 208 características extraídas.

Como no se considera que alguna característica tenga mayor peso sobre cualquier otra, una o más características puede llegar a tomar ventaja sobre las otras debido a que su rango de valores puede ser mayor a otras. En la Tabla 4 se puede apreciar que los valores que toma el promedio del eje X en el acelerómetro es inferior a los que toma el tercer cuartil del eje X del acelerómetro.

La principal ventaja de escalar es evitar atributos en rangos numéricos mayores que dominan aquellos en rangos numéricos más pequeños, como estas dos características que se mencionaron. Otra ventaja es evitar dificultades numéricas durante el cálculo. Dado que los valores del kernel normalmente dependen de los productos internos de los vectores de características (Chih-Wei Hsu, Chih-Chung Chang 2008).

El valor escalado para todas las columnas se recomienda este entre [-1, 1] o bien entre [0, 1]. Para lograr esto se recomienda usar alguno de los factores de escalamiento de la Tabla 5.

Tabla 5: Escalamientos propuestos que permitirán que cada vector de características esté dentro del mismo rango de valores.

Escalamiento de datos

$valor\ escalado :=$	$(valor - promedio) / desviación\ estandar$
$valor\ escalado :=$	$(valor - mín) / (máx - mín)$
$valor\ escalado :=$	$(valor - (max - min) / 2) / ((max - min) / 2)$

Capítulo 4. Evaluación de estrategia propuesta para el reconocimiento de gestos sutiles

Se diseñó un experimento con el objetivo evaluar cuantitativamente la precisión de las técnicas de reconocimiento de gestos sutiles propuestas, que permitirán, entre otras muchas aplicaciones etiquetar los datos para el sensado participativo. El estudio cuantitativo se enfoca en la eficiencia de los dos modelos seleccionados (SVM y SMO) para determinar qué gesto realizó la persona.

El tipo de estudio fue intra-sujetos, es decir, un grupo de sujetos que decidieron participar voluntariamente, harán uso del reloj inteligente para realizar los seis gestos en tres posiciones diferentes. Los valores registrados por el acelerómetro, el giroscopio y el tiempo en que se desarrolla cada registro se grabaron en la memoria interna del reloj inteligente mediante un archivo de texto, separando cada columna de datos por medio del tabulador.

En las siguientes secciones se define la manera en que se llevó a cabo el experimento.

4.1. Selección del reloj inteligente

Dentro de la gama de dispositivos comerciales que se usan comúnmente para la detección de gestos, existen dos opciones principales: relojes inteligentes y Myo. Este último es un dispositivo portátil de control de gestos y movimientos que se coloca en el antebrazo y le permite tomar el control del teléfono, la computadora y otros dispositivos sin tocarlos. El reloj fue elegido dadas las grandes ventajas que tienen sobre Myo. Algunas de estas ventajas son: menor precio, se puede lograr el procesamiento de la señal directamente en el reloj, su ubicación es más estratégica y su uso es de forma habitual en la vida cotidiana, además de que su SDK (software development kit) es de libre acceso y muy bien documentado. En particular se utilizaron relojes inteligentes que utilizan el sistema operativo Android Wear, que está basado en Java.

Una vez definido el dispositivo a usar, se buscó que el reloj contara tanto con el sensor giroscopio y acelerómetro, junto con la suficiente capacidad de almacenamiento y procesamiento para llevar a cabo la toma de datos de los sensores, el procesamiento de la señal según lo estipulado en el capítulo tercero y finalmente la respuesta dada por los modelos seleccionados. Así pues, en este estudio se usaron dos

relojes inteligentes de diferentes marcas y características físicas (ver Figura 26). Uno de los relojes es un LG G watch W100, con correa de caucho que permite un agarre estable y cómodo a la muñeca. Este reloj tiene un peso de 63 gramos, tiene un procesador Qualcomm Snapdragon 400 a 1.2 GHz, 4 GB de memoria interna y 512 de RAM. El segundo reloj es un ASUS ZenWatch 2 con correa de piel sintética con agarre medio a la muñeca y un peso de 75 gramos. Este reloj cuenta con el mismo procesador, memoria externa y RAM que el modelo anteriormente descrito.

Con ambos dispositivos verificados para que pudieran ser fácilmente programables y que pudieran llevar a cabo los procesos que presentan los modelos de clasificación, se procede a verificar qué gestos son los que se realizarán y en que posiciones se harán.



Figura 26: Durante el desarrollo del experimento en esta tesis se utilizaron dos relojes inteligentes: a) LG G100; b) Asus ZenWatch.

4.2. Selección de gestos sutiles y posturas para realizar estos gestos

Dentro de los gestos que se revisaron en la literatura y según los movimientos naturales del brazo, antebrazo y mano, se buscó aquellos que fueran lo más sutiles y de mayor comodidad posibles para realizarlos y que no sean comúnmente utilizados para interactuar con relojes inteligentes, como el caso, por ejemplo, de girar la muñeca para subir o bajar de nivel en un menú. Dentro de la gran gama de gestos revisados, los más sutiles que se encontraron fueron los realizados únicamente con los dedos de las manos, algunos con poca o nula participación de la muñeca. En la Figura 27 se muestran los gestos a usarse; el primero de los gestos se propuso el trabajo reportado en (Dementyev & Paradiso 2014), los siguientes dos

gestos se realizan de forma similar al primero pero haciendo uso de diferente dedo. Los tres gestos restantes son variantes de los primeros tres, con la diferencia de ser con doble repetición. Se busca que estos seis gestos permitan retroalimentar el etiquetado de datos para el sentido participativo.

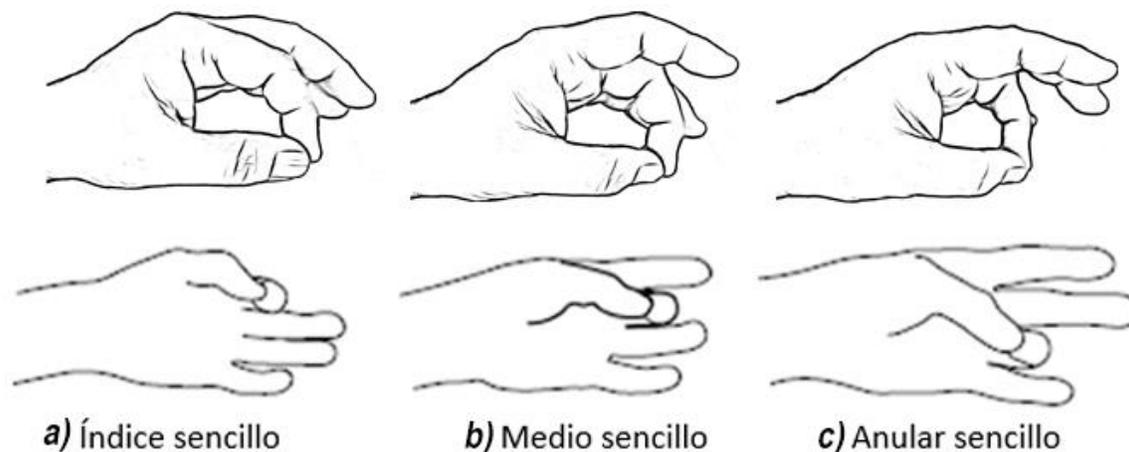


Figura 27: Gestos seleccionados para ser usados. a) es un pellizco con el dedo índice, b) pellizco con el dedo medio y c) pellizco con el dedo anular. Los restantes tres gestos son idénticos a los primeros, pero repitiendo dos veces el gesto.

Para el entrenamiento, cada uno de estos gestos será realizado en dos ocasiones en tres posturas diferentes. La primera postura será colocando el brazo sobre un soporte fijo, como una mesa, encima de un libro o una caja que permita mantener estabilidad en el brazo pero que además permita el libre movimiento de los dedos, el usuario deberá estar sentado al realizar los gestos. La segunda postura será estando el usuario de pie, el brazo deberá ser colocado al frente a la altura del pecho y con una pequeña inclinación hacia adentro, de tal forma que aparente que el usuario está leyendo la hora del reloj. La tercera postura será, de igual, estando parado el usuario, pero esta vez tendrá los brazos en descanso, de tal forma que el cuerpo adopte una postura natural cuando se encuentra de pie. **La Figura 28 muestra las tres posturas** en las cuales se realizarán los gestos para la toma de datos de entrenamiento.

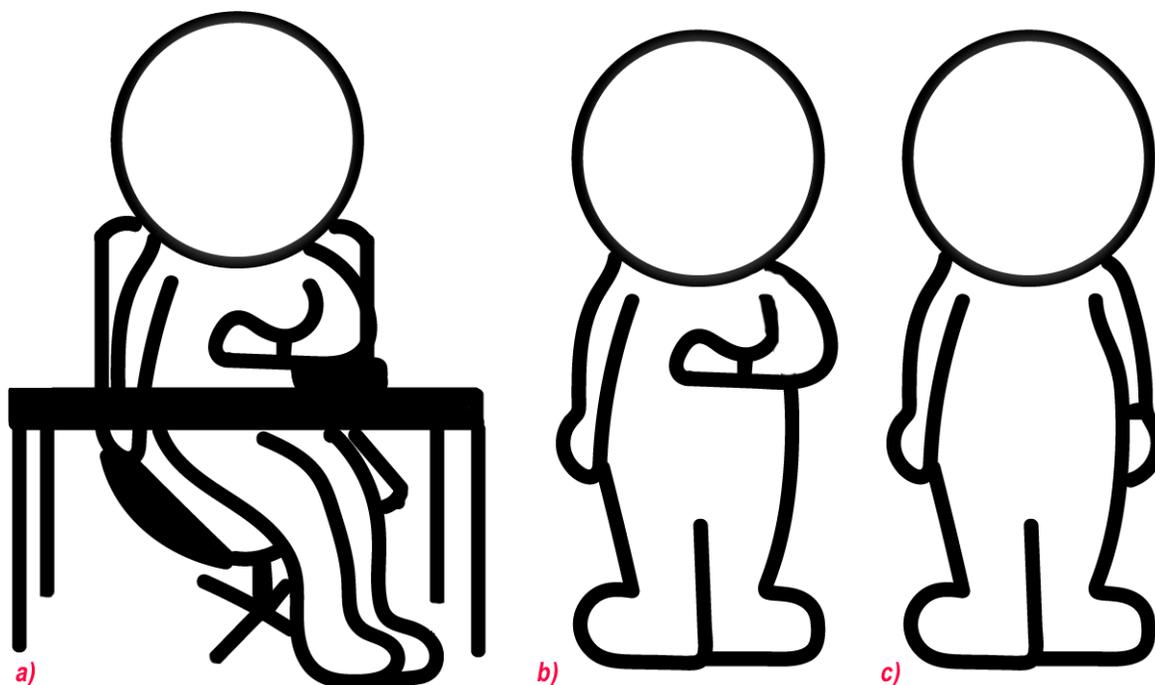


Figura 28: Diferentes posturas en los que se realiza cada uno de los gestos; a) el usuario sentado, con el brazo soportado por un libro a caja estando encima de una mesa; b) el usuario parado como si estuviera observando la hora en el reloj; y c) el usuario parado en descanso con los brazos a los costados.

4.3. Selección de sujetos para entrenamiento y pruebas de los algoritmos

En lo que respecta a los sujetos para recolectar los datos para el entrenamiento y pruebas, fueron seleccionados de forma arbitraria, cuidando que tuvieran una edad entre los 10 y los 60 años. La razón de la edad es que a los niños menores de 10 años el reloj no se ajustaría correctamente a su muñeca por tener el antebrazo muy delgado y en el caso de que fueran **mayores de 60 años, existe mayor probabilidad de algún tipo de enfermedad como Parkinson, que hace que se pierda el control en las manos.** En la Figura 29 se muestra la señal del giroscopio en los ejes XYZ del pellizco con el dedo índice de tres usuarios de diferente edad, un niño de 8 años, un adulto de 25 años y un adulto mayor de más de 70 años, para este ejemplo se puede observar el movimiento involuntario que puede presentarse en los adultos mayores.

Las personas seleccionadas deben de cubrir además los requisitos que se muestran a continuación:

- Saber leer (para poder seguir las ordenes que se mostraran en el reloj).

- No haber usado antes un reloj inteligente para evitar ventajas de uso sobre los otros usuarios.
- Que la mano dominante sea la derecha, para así poder hacer uso del reloj en la mano izquierda, ya que el estudio se realizó usando la mano izquierda y el cambio de mano implica cambio de signos de las señales y es un aspecto que no se contempló para este estudio.
- No tener problemas motores en las manos o padecimientos como TDAH o Alzheimer que lleguen a presentar una reducida velocidad de procesamiento de la respuesta.
- Tener una edad mínima de 10 años y máximo de **60** años.
- Tener un mínimo de 20 minutos disponibles para realizar las pruebas.
- Tener la voluntad de realizar las pruebas desinteresadamente.
- Pueda completar todas las pruebas, con las distintas posturas y gestos.

En la Figura 29 se puede observar la ejecución de uno de los gestos por parte de tres personas de edades diferentes. En el caso *a* (un niño de 8 años) en varias ejecuciones el reloj se movía de su lugar dado que tenía brazos muy delgados. Para el caso *b* (adulto mayor), quien dijo no tener enfermedades degenerativas diagnosticadas, se puede observar que la señal presenta una ligera vibración. Es posible que estas condiciones no se generalicen para los niños y adultos mayores de las edades mencionados, pero para evitar estas posibles situaciones se optó por realizar estas exclusiones.

Para el entrenamiento de los modelos de los algoritmos de SVM y SMO, se reclutaron 15 personas que participaron de forma voluntaria y que cumplieran con los criterios anteriormente expuestos, que se resume en la Tabla 6.

Tabla 6: Criterios de inclusión y exclusión para la selección de participantes en el estudio.

<i>Criterio</i>	<i>Descripción</i>
Inclusión	Personas entre los 10 y 50 años.
	Mano derecha dominante.
	Tener tiempo disponible para toma de datos.
Exclusión	Problemas motores en las manos.

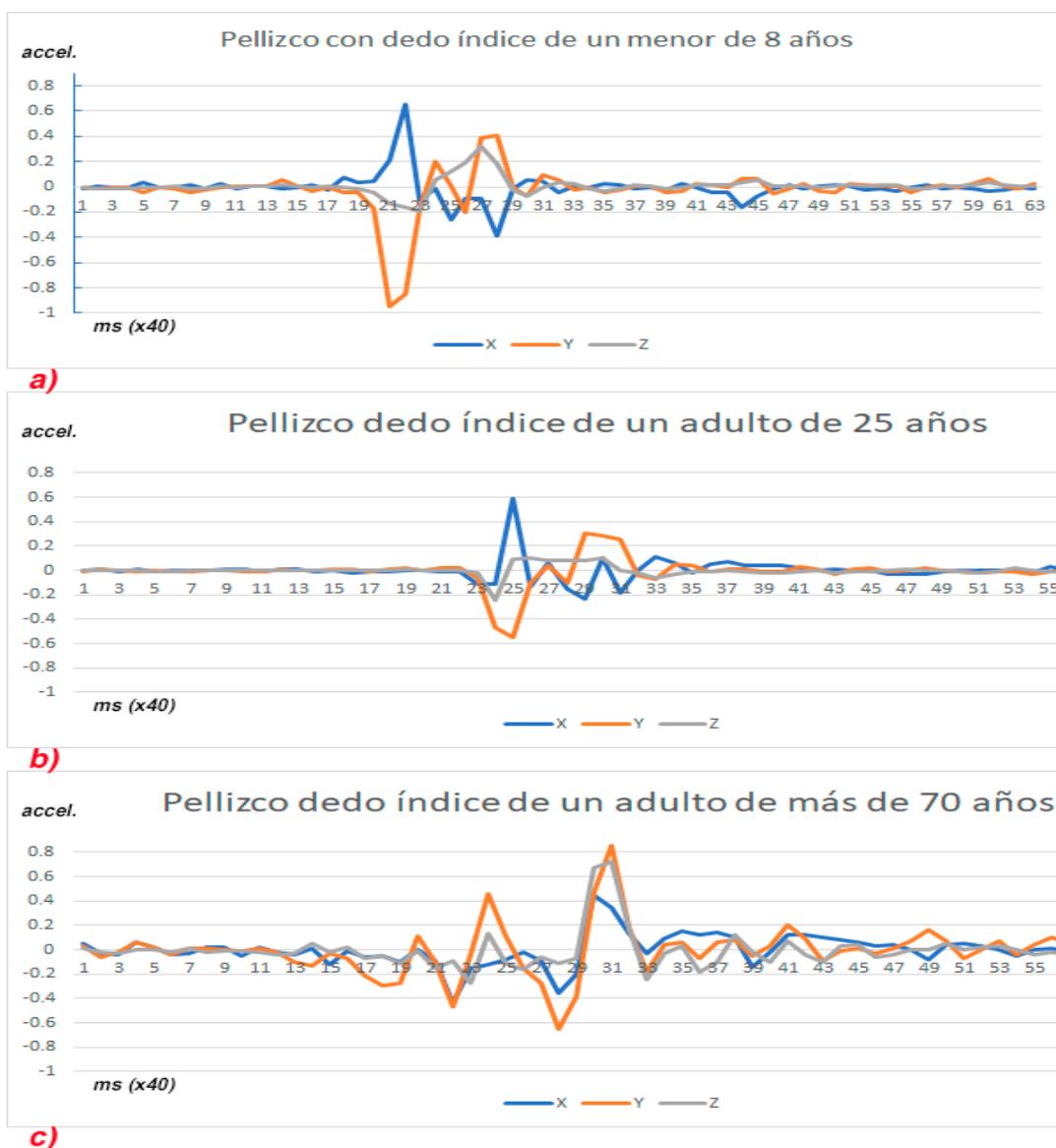


Figura 29: Gesto de pellizco de a) un niño de 8 años, b) un adulto de 25 años y c) un adulto mayor de más de 70 años. Se puede observar como las condiciones degenerativas naturales puede afectar el nivel de precisión en los adultos mayores (figura c).

Al inicio de la prueba se explicó a cada uno de los sujetos la forma en que debe colocar el reloj según lo especificado en la sección 3.1, los tipos de gestos que realizarían, el número de repeticiones por gesto, las posturas en que lo harían, además de tener los dedos de la mano extendidos de forma natural, es decir, que los dedos no estuvieran muy curvos tanto hacia adentro de la palma de la mano como hacia afuera. Los sujetos practicaron cada gesto para familiarizarse con la prueba.

4.4. Adquisición de datos.

Cada usuario realizó cada gesto dos veces de manera consecutiva en cada una de las tres posturas, para un total de 36 gestos por individuo y 540 gestos de todos los participantes. Las dos repeticiones de cada gesto tuvieron el propósito de garantizar que el modelo tenga el suficiente número de instancias para ser entrenado, ya que si no hay suficientes instancias para entrenar el modelo daría resultados pobres cuando sea usado en la prueba (Yadav & Shukla 2016). Una instancia es un set de datos que contienen todas las características

En la toma de datos entre gesto y gesto, al usuario se le notificaba en que momento debería de realizar el gesto correspondiente mediante vibración o un sonido, con una duración de un segundo cada una. De forma predeterminada se usará el sonido, ya que de esta manera el gesto se puede realizar inmediatamente después de que se escuche el sonido. En el caso de la vibración, el usuario debe esperar a que esta termine para poder realizar el gesto, dado que si se toma la muestra mientras se está llevando a cabo la vibración, esta genera movimiento que es captado por la señal de ambos sensores y que se consideraría ruido innecesario.

Para evitar que el usuario tenga que estar indicando cuando iniciar la captura de un nuevo gesto, se programó la aplicación en el reloj inteligente de modo que reciba instrucciones desde un teléfono inteligente vinculado al reloj por medio de bluetooth. De esta forma, la persona responsable de la toma de datos tendrá el control en todo momento de lo que deberá hacer el usuario y en qué momento lo hará.

La Figura 30 muestra la arquitectura de la aplicación con la comunicación entre el reloj y el teléfono por medio de bluetooth y haciendo uso de la Wearable Data Layer API (encargada de establecer todos los protocolos de comunicación entre ambos dispositivos). Coordinar el momento exacto para hacer cada uno de los gestos implica una comunicación bidireccional entre ambos dispositivos, que permita que el reloj le informe al teléfono en qué momento concluyó la toma actual de la muestra, para que el reloj espere la orden de realizar un nuevo gesto.

La persona que controla el teléfono es quien indica que gesto se realizará, con cuantas repeticiones (para este estudio, dos), si se usará sonido o vibración, en qué posición realizarán estos gestos, entre otras. **La Figura 31 muestra la vista principal** de cómo se despliegan estas opciones en el teléfono.

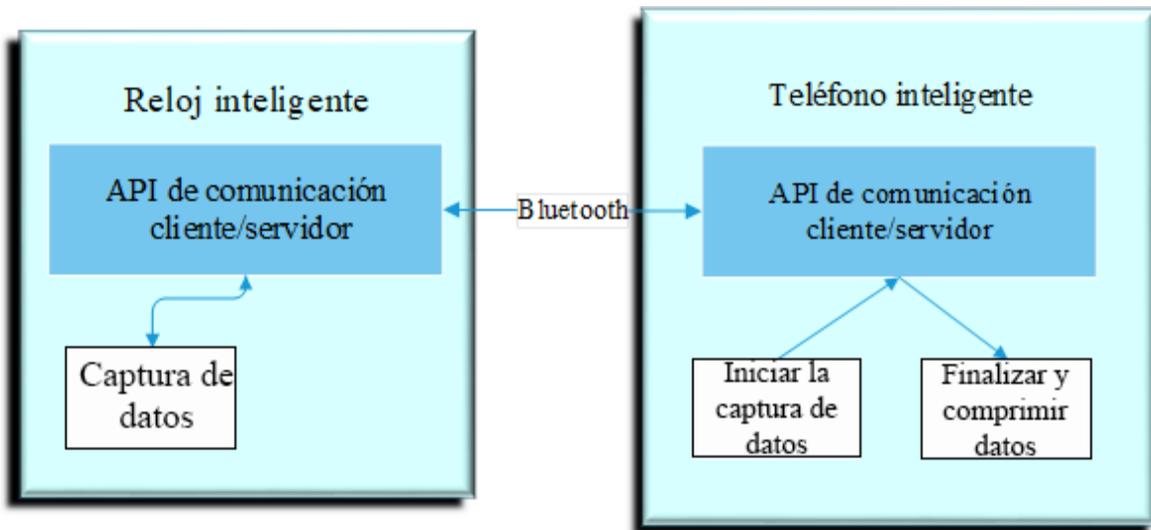


Figura 30: Comunicación entre teléfono-reloj para coordinar la toma de los datos de los sensores en cada gesto. Al finalizar la toma de datos, todos los datos son comprimidos para ser recuperados posteriormente.

Una vez que se han capturado los datos de los quince participantes en el reloj inteligente y que se calcularon todas las características, según el procedimiento descrito en el capítulo 3, se procedió a crear los modelos del reconocimiento de gestos haciendo uso de validación cruzada para encontrar los mejores parámetros posibles para cada modelo.

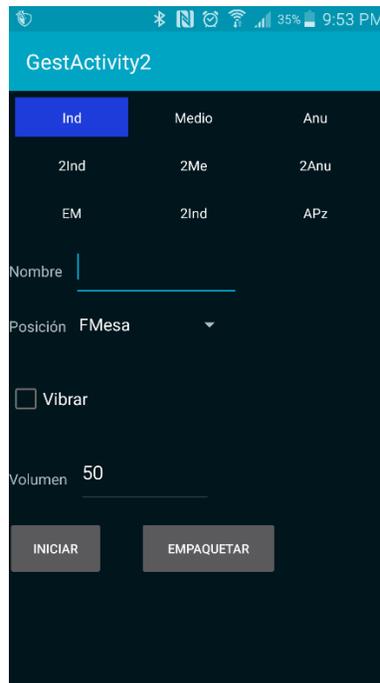


Figura 31: Captura de pantalla de la app en el teléfono que controla la captura de gestos en el reloj.

4.5. Validación cruzada para determinación de parámetros

Tanto los modelos de regresión (trabaja con datos continuos) como los de clasificación (trabaja con clases) pueden presentar buena capacidad de predicción en los datos de entrenamiento, aunque de la misma manera podrían no ser tan efectivos al realizar esta tarea [47,48]. Para valorar el rendimiento predictivo de un modelo, se utiliza comúnmente la técnica de validación cruzada (VC) (Yadav & Shukla 2016).

El funcionamiento de la VC consiste en dividir la muestra de tamaño n en una submuestra de entrenamiento de tamaño $n-1$ y una submuestra de prueba de tamaño 1 y evaluar el modelo. Después se hace la misma división de la muestra, pero ahora la submuestra de prueba será otro elemento diferente y a continuación se evalúa el modelo, para obtener al final n evaluaciones y como resultado final la efectividad del modelo es la media aritmética de las n evaluaciones. Este tipo de validación cruzada se llama validación cruzada dejando uno fuera (Yadav & Shukla 2016).

Existen varias formas de dividir los datos de entrenamiento y prueba, y según la forma en que se realice esta división de datos será el nombre que reciban, como por ejemplo VC aleatoria con k -iteraciones

(se realiza la división de datos de forma aleatoria, algunos datos podrían no evaluarse), VC dejando uno fuera (anteriormente explicado y se realiza n veces) y VC con k -iteraciones, en el cual los datos se dividen en k subconjuntos de prueba y el resto ($k-1$ subconjuntos) es usado para entrenamiento.

El tipo de VC que usualmente se utiliza es el de k -iteraciones. La Figura 32 muestra de forma más clara cómo se llevaría a cabo la validación cruzada en 10 instancias, con 5 iteraciones. Como corolario, se puede observar que, de hacer el mismo procedimiento, pero con 10 iteraciones, se trataría de la validación dejando uno afuera, por lo que podemos decir que esta última es un caso especial de la VC con k -iteraciones.

Debido a que no existe un número de iteraciones que sea considerada como la mejor, de forma empírica se podría decir que el número de iteraciones que mejor resultados obtiene es de cinco o diez iteraciones (Yadav & Shukla 2016). Para este estudio se usará una validación cruzada de 5 y 10 iteraciones para buscar los mejores parámetros para cada uno de los dos modelos a evaluar.

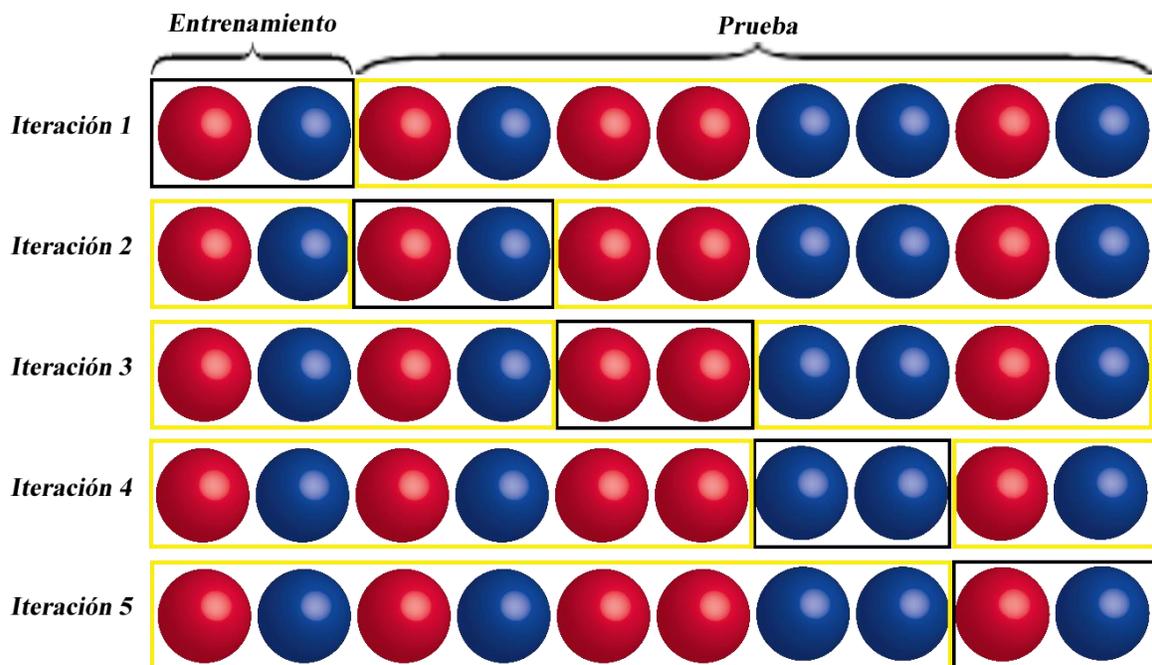


Figura 32: Validación cruzada con 5-iteraciones. Las instancias encerradas en el recuadro negro son consideradas como entrenamiento y las que están encerradas en recuadro amarillo son de prueba.

4.6. Uso de herramienta de minería de datos WEKA

La herramienta de minería de datos WEKA, acrónimo de Waikato Environment for Knowledge Analysis, es un software de distribución libre que permite el análisis de datos para aplicar, analizar y evaluar las técnicas más relevantes sobre cualquier conjunto de datos (Jiménez & Sierra 2003).

Mediante el uso de la herramienta Weka se llevará a cabo la aplicación, análisis y evaluación de los dos algoritmos de clasificación que se usarán en este estudio. La herramienta de WEKA en su versión 3.9.1 cuenta con dos sub-herramientas que se usarán en este estudio: Explorer (Explorador) y KnowledgeFlow (Flujo de conocimiento).

La opción Explore permite llevar a cabo la ejecución de los algoritmos de análisis implementados sobre los archivos de entrada, pudiendo realizar una ejecución independiente por cada prueba. En esta opción se encuentran varias pestañas, como se muestra en la Figura 33 y Figura 34, que se usarán serán la pestaña de **Preprocess** (izquierda) que permite introducir los 540 datos de entrenamiento, así como poderle aplicar filtros a estos datos. La pestaña **Classify** (derecha) nos permitirá evaluar la eficiencia de los parámetros de cada modelo para encontrar los mejores, así como la generación del modelo que se usará en la detección de los gestos desde el reloj inteligente.



Figura 33: Ventana de Explore, con las pestañas de Preprocess y Classify, que son las opciones que más serán utilizadas.

En la pestaña de Preprocess se introduce la información obtenida de las 208 características en un formato denominado *Attribute-Relation File Format* (extensión arff), que se muestra en la Tabla 7. Básicamente el archivo está constituido por 3 partes que son: **@relation** donde se indica el nombre representativo de la relación de los atributos y la clase, **@attribute** se especificarán todos los atributos (en

este caso los 208) junto con su tipo de dato que puede ser string, numeric, date o datos nominales que son un conjunto de valores que puede tomar el atributo y que están encerrados entre corchetes, y finalmente **@data** que contendrá los valores de los atributos organizados por columnas y separados normalmente por coma o tabulación. Cada una de las filas que se forma en **@data** se llama instancia.

Tabla 7: Estructura de los datos del estudio según el formato arff.

```

@relation gest_features

@attribute Accelerometer_X_Prom numeric
@attribute Accelerometer_X_Max numeric
@attribute Accelerometer_X_Min numeric
.
.
.
@attribute label {indice,medio,anular,2indice,2medio,2anular}

@data
0.097811    1.330094    -1.581338    ... indice
0.072367    0.528059    -0.33529     ... medio
-0.015188   0.339446    -0.759814    ... anular
0.052106    1.189076    -0.685937    ... 2indice
0.238339    1.352288    -0.26882     ... 2medio
0.248937    2.253834    -1.106709    ... 2anular

```

En la pestaña Classify se selecciona el algoritmo de clasificación que se desee utilizar (para el estudio SVM y SMO), las opciones de las pruebas (validación cruzada, división porcentual, entro otras) y la variable que se clasificara (en este caso el atributo *label*). Las salidas arrojadas en esta pestaña son la información de los datos, instancias correctamente clasificadas y no clasificadas, la precisión del algoritmo por clase y la matriz de confusión. La opción **KnowledgeFlow** permite realizar lo anterior, pero por medio diagrama de flujo, lo que es muy útil para comparar los dos modelos al mismo tiempo y bajo las mismas condiciones.

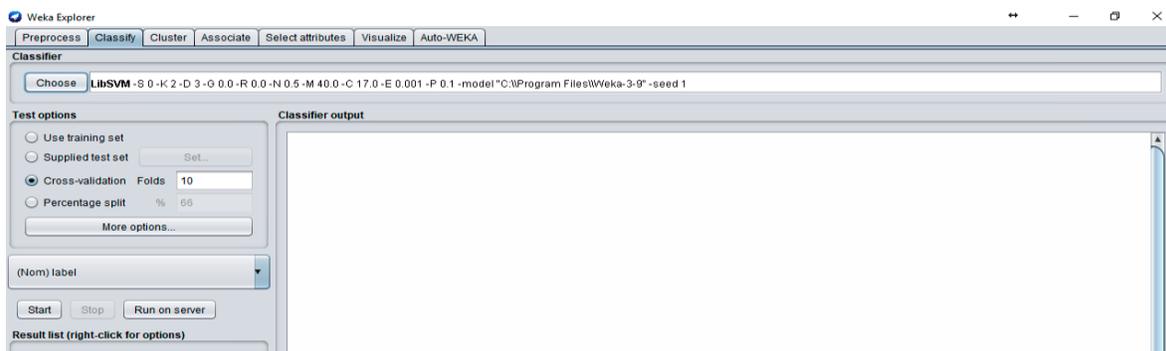


Figura 34: Pestaña de clasificación donde se puede realizar la validación cruzada para encontrar los mejores parámetros y evaluar el algoritmo.

4.7. Filtrado de datos

Una de las características destacables con las que cuenta WEKA, es el filtrado de datos. Los filtros se pueden aplicar tanto a las características como a las instancias y pueden ser aplicados en cascada, es decir, la entrada de datos de un filtro puede ser la salida de un filtro anteriormente aplicado. Para este estudio se aplicará el filtro llamado *RemoveMisclassified*, que se aplica sobre las instancias de los datos de entrenamiento. Como parámetros este filtro requiere un clasificador, por lo que se usará SVM como clasificador, con 10 repeticiones. Este filtro eliminará aquellas instancias que se clasifiquen de forma incorrecta. Este filtro es aplicado solamente para estimar de forma rápida la eficiencia en la clasificación de los gestos sutiles.

4.8. Implementación de modelos

Ambos modelos se crean a partir de los datos de entrenamiento escalados según lo indicado en el apartado 3.6.3, donde se busca obtener la mejor función de escalamiento de entre las propuestas. El archivo con formato *arff* se creó conforme lo estipulado en el apartado 4.6. A estos datos se les aplicaron los filtros mencionados en el apartado 4.7 con los parámetros que ahí mismo se especifican. Mediante la validación cruzada, se buscan obtener los mejores parámetros para ambos modelos.

La validación cruzada se realizó con 10 repeticiones. Para encontrar los mejores parámetros, en el caso de SVM se empezó con valores de crecimiento en forma exponencial para el costo (C) y el valor de gama (γ), como se sugiere en (Chih-Wei Hsu, Chih-Chung Chang 2008) con $C = \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^{15}\}$ y $\gamma = \{0, 2^{-15}, 2^{-14}, \dots, 2^6\}$. Cuando se detecte entre que valores de C y γ se obtienen los mejores resultados, se realizará el acercamiento por división de valores, es decir, si por ejemplo con un costo entre 23 y 24 se obtienen mejores resultados, se procederá a buscar el nuevo valor entre el mejor de ambos y la mitad de ambos, lo que podría ser entre 8 y 12 o entre 12 y 16. Se realizará de forma sucesiva este procedimiento hasta encontrar el mejor resultado que no cambie durante las tres últimas divisiones.

El kernel que se usará en SVM será el de Radial Basis Function (RBS) y para SMO será Polykernel. La función del kernel, de forma general, es la de mapear una secuencia de datos sobre otra busca de similitudes (Hofmann et al. 2008).

Con los modelos calibrados con los mejores parámetros, se crea el archivo binario que contiene el modelo, obteniéndose un archivo por modelo. Este archivo se utiliza en el reloj para ser usado al detectar los gestos realizados por los sujetos. Para la creación de este archivo se da clic con el botón secundario del ratón sobre la lista de resultados, y en el menú contextual que aparece se selecciona la opción *Save model* se guarda el archivo en la ruta seleccionada como se muestra en la Figura 35.

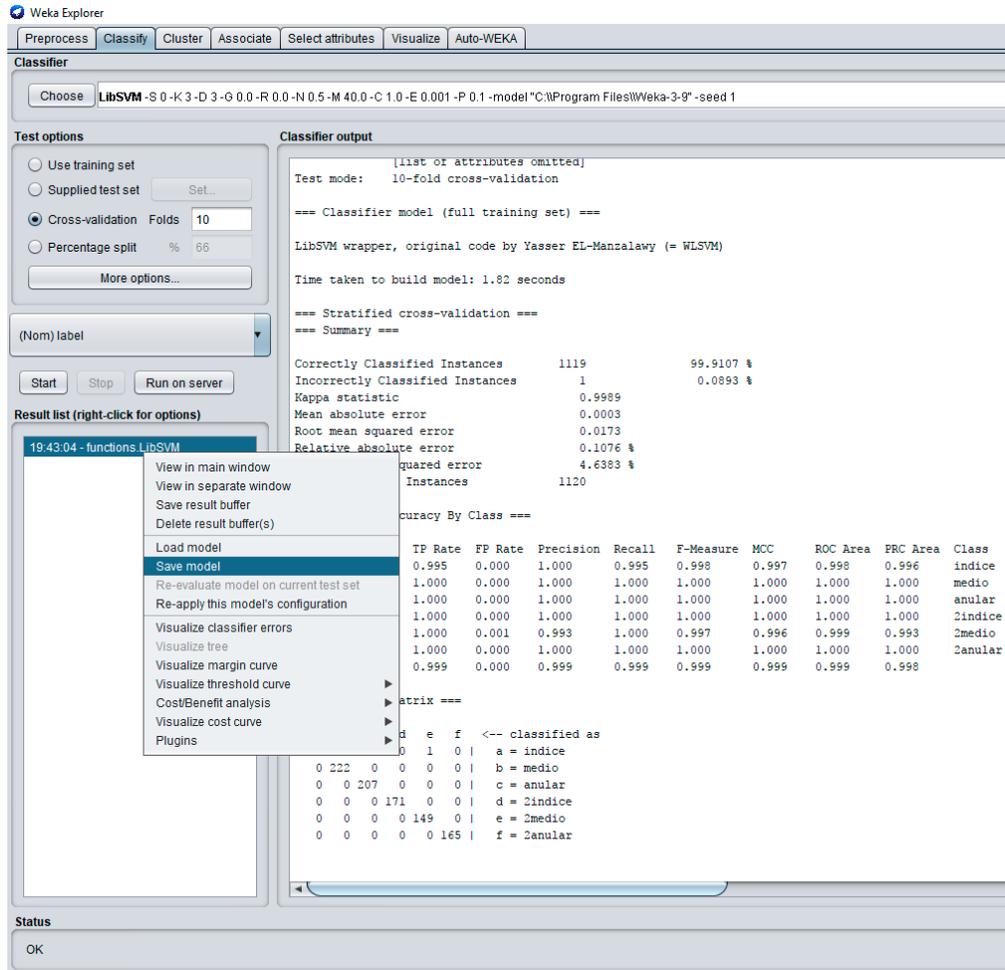


Figura 35: Guardando el archivo binario que contendrá el modelo para la detección de los gestos sutiles.

4.9. Evaluación de la detección de los gestos sutiles

La evaluación de la efectividad de los gestos se realizó en dos etapas. En la primera de ellas, se usó la herramienta WEKA con la opción Supplied test set (Conjunto de pruebas suministradas), en lugar de validación cruzada. Esta opción permite evaluar los datos de prueba con un conjunto de datos externo al conjunto de datos de entrenamiento. Este conjunto de datos externo será generado a partir de cinco usuarios diferentes a los que se utilizaron para entrenar el modelo. Cada usuario realizó los seis gestos en las posiciones *a* y *b* mostrados en la Figura 28. Este set de datos se clasificó según el modelo que se obtuvo

anteriormente. Para verificar la efectividad del modelo se tomó en cuenta la exactitud del modelo (gesto bien clasificados contra el número de gestos), la matriz de confusión y la curva ROC² del modelo por gesto.

La otra evaluación será en tiempo real desde el reloj. La evaluación se realizará a partir de 10 preguntas que se hicieron al usuario, en las que debe contestar con una de tres posibles respuestas. Cada una de las respuestas corresponde a cada uno de los tres primeros gestos. Estas respuestas son muy fáciles de responder, ya que lo que se busca es verificar la efectividad del modelo y no los conocimientos generales del usuario. Alternadamente a estas 10 preguntas, al usuario se le pide al azar que haga uno de los tres gestos restantes para poder evaluar así los seis gestos. La razón de usar solo preguntas de tres respuestas y preguntar alternadamente que se realice uno de los tres gestos restantes es debido a que tener preguntas de seis respuestas diferentes, puede tomarle al usuario demasiado tiempo para analizar las posibles respuestas o bien confundirlo por la cantidad de respuestas.

Cada respuesta a cada pregunta iniciará después de 10 segundos de haber aparecido la pregunta, esto con el objeto de darle tiempo al usuario de que lea tanto la pregunta como las respuestas. La lectura de datos de los sensores será controlada mediante una vibración de un segundo o bien por medio de un sonido con el mismo tiempo de duración. El tiempo de duración de la lectura para realizar el gesto es de 2.7 segundos. Cada gesto deberá ser realizado después del aviso por medio de la vibración o durante la reproducción del sonido (según las condiciones del recinto donde se lleve a cabo la toma de muestras se usará uno u otro).

4.10. Conclusión

El diseño de este experimento permite capturar en un ambiente controlado secuencias de datos del acelerómetro y giroscopio de un reloj inteligente durante 2.7 segundos, con la finalidad de extraer el momento en que el usuario realizó un gesto sutil.

Esta señal es analizada para tratar de encontrar si se realizó uno de los gestos esperados y en qué momento en que se realizó el gesto o bien si no se realizó uno de los gestos esperados. En el siguiente capítulo se presentan los resultados obtenidos con 15 sujetos realizando dos repeticiones de los gestos en

² Representación gráfica de la razón de verdaderos positivos Vs falsos positivos.

cada una de las tres posiciones, así como los resultados obtenidos en la detección de un posible gesto en la señal obtenida con el reloj inteligente y que es usada para entrenar los dos clasificadores para lograr la detección de estos gestos.

La Figura 36 **Error! No se encuentra el origen de la referencia.** resume el proceso desarrollado que permitió la elaboración de esta tesis. La **selección de gestos** incluye todo el proceso de análisis de la literatura relacionada y el establecimiento de los 6 gestos sutiles a usarse, esto incluyó la eliminación de dos posibles gestos por ser muy parecidos a otros gestos. En el **establecimiento de posturas y la señal de acelerómetro y giroscopio de 15 usuarios** corresponde a la búsqueda de posiciones poco intrusivas adecuadas a los gestos sutiles y al establecimiento de las características deseadas de los usuarios que participarían en el experimento. La **búsqueda de un posible gesto y segmentación de la señal** comprende todo el análisis realizado a la señal de ambos sensores, así como su tratamiento para poder obtener la **extracción de características** que alimentarán a dos algoritmos de clasificación. Los pasos restantes corresponden a la clasificación de los gestos.

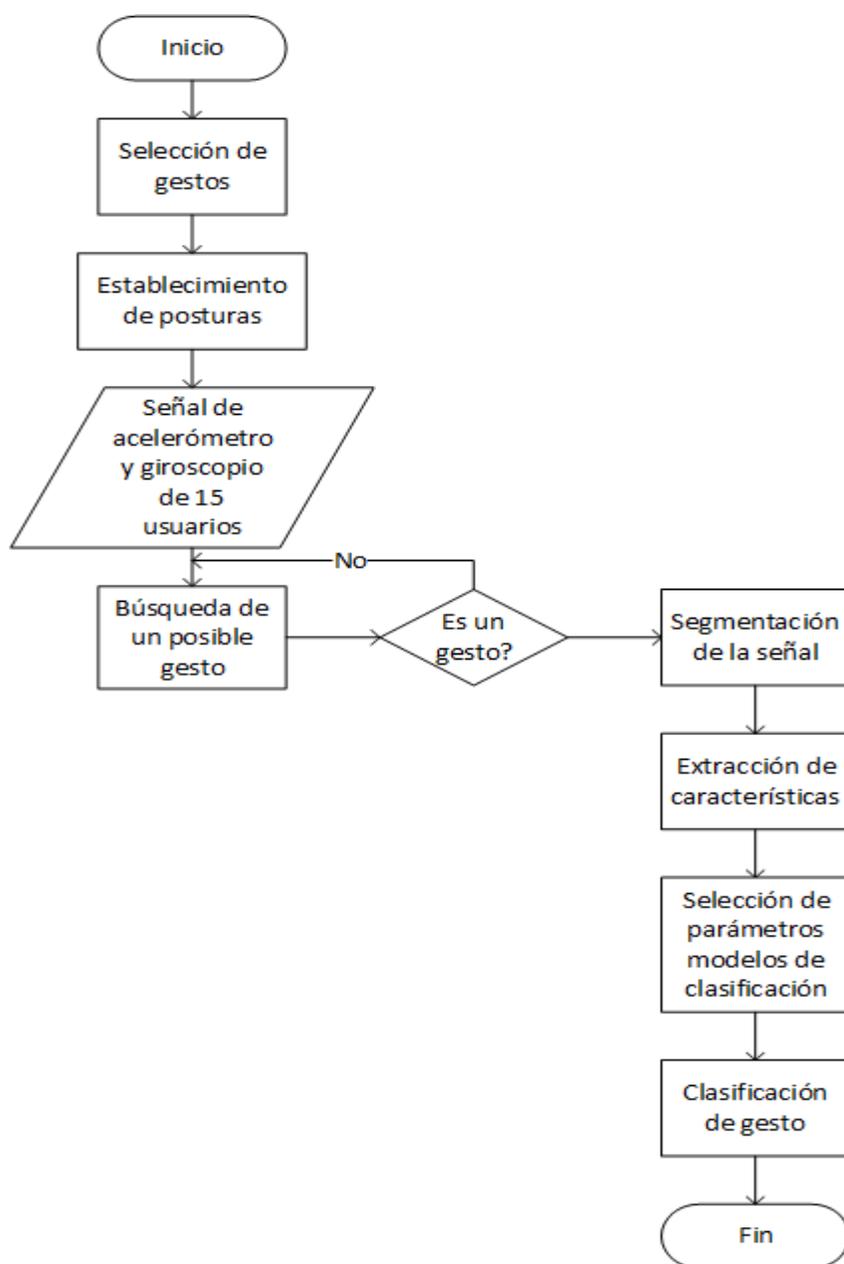


Figura 36: Diagrama de flujo que representa de forma general el trabajo realizado en esta tesis.

Capítulo 5. Resultados

Un total de 15 personas participaron de forma voluntaria en la evaluación, 8 mujeres y 7 hombres con edad promedio de 28 años (desviación estándar de 7.1 años). Los participantes recibieron instrucciones verbales sobre la actividad a realizar y manifestaron verbalmente no tener impedimento físico-motor en las manos que les impediría realizar los gestos requeridos. Además, ninguno de los 15 participantes posee un reloj inteligente o ha usado alguno de ellos de forma cotidiana.

5.1. Desarrollo de los gestos sutiles

Para realizar los gestos sutiles, a los participantes se les instruyó sobre la forma de ejecutar cada uno de los gestos en las posturas que se estipulan en la sección 4.2. Cada participante realizó un ejercicio previo de todos los gestos en cada una de las posturas donde se les hacían observaciones en caso de que lo realizaran incorrectamente. La mayoría de los participantes entendió la posición en la que tenía que mover los dedos para realizar cada gesto, desde la primera postura en la que realizarían los gestos, solamente con una persona fue necesario que el instructor realizara una demostración rápida de cada gesto en cada posición, con lo que consiguió hacerlo sin problemas.

Algunos de los problemas que se presentaron fueron que algunos sujetos curvaban de forma poco natural los dedos al realizar los gestos, lo que fue solucionado con el entrenamiento previo. Al momento de realizar cada gesto, el investigador anotaba en una hoja el gesto realizado por el participante, para verificar que lo estuviera realizando correctamente, ya que 3 de los 15 usuarios olvidaron el gesto que seguía en la secuencia que se muestra en la Figura 27. Se les pidió realizar la secuencia completa de gestos en dos ocasiones, con un descanso de 10 segundos entre cada una de ellas. Si el participante realizaba otro gesto diferente al que en ese momento debía realizarse, toda la secuencia era desechada y se repetía nuevamente. Solamente con uno de los participantes fue necesario desechar la secuencia de la segunda ronda de la repetición de los gestos cuando estaba en la postura c) de la Figura 28.

5.2. Uso de la segmentación de la secuencia de datos

Cuando el ventaneo de la secuencia de datos se concluyó, se tomó una muestra al azar del 10% de los datos de cada uno de los gestos. El 10% de los datos fue seleccionado haciendo uso de la función

ALEATORIO.ENTRE(0,91), que es una función de EXCEL que genera un número al azar entre 0 y 91. Esta función se actualizaba en 9 ocasiones (para así tener el 10% de los datos) para cada uno de los gestos.

De esta muestra de datos aleatorios que se revisó, en todos los casos se pudo apreciar visualmente que el inicio y el fin del gesto obtenido, por medio del ventaneo, correspondía con el inicio y el fin del gesto. En la Figura 37 se muestra uno de los gestos realizados por uno de los participantes, donde se discrimina un pequeño movimiento de la mano de aproximadamente 120 milisegundos de duración. A la derecha de la Figura 37 se muestra la señal del gesto seleccionada por medio del ventaneo, donde se ha discriminado el ruido.

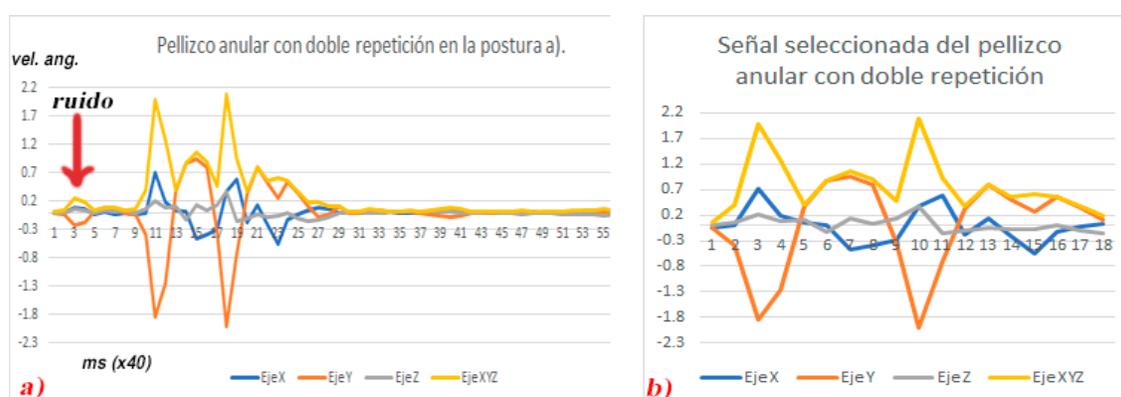


Figura 37: Gesto anular con doble repetición realizado por uno de los participantes donde existe ruido en la señal, mismo que es discriminado al seleccionar, mediante el ventaneo, solo la señal del gesto mostrada en b). La duración total del gesto es de aproximadamente 720 milisegundos.

La cantidad de gestos rechazados por no cumplir con lo dispuesto en la sección 3.3, por no ser considerados como un gesto válido, se muestra en la Tabla 8. Es de destacar que no hay mucha diferencia entre el porcentaje de gestos rechazados con los distintos dedos, donde en todos los casos esta entre el 3.33 y 6.66% del total de gestos.

Tabla 8: Porcentaje de gestos rechazados por no cumplir lo predispuesto en la sección 3.3.

Gesto	# de gestos rechazados	Total	% de gestos rechazados
Índice	3	90	6.66%
Medio	6	90	6.66%

Anular	4	90	4.44%
Índice con repetición	3	90	3.33%
Medio con repetición	4	90	4.44%
Anular con repetición	5	90	5.55%

5.3. Detección de posibles gestos por medio del algoritmo DTW

En la detección de un posible gesto por parte del algoritmo DTW, en promedio el 94.74% de los 515 gestos restantes son reconocidos como un gesto válido. La Tabla 9 muestra el porcentaje de cada uno de los gestos reconocidos por dicho algoritmo.

Tabla 9: Porcentaje de reconocimiento de un posible gesto por parte del algoritmo DTW.

Gesto	Posible gesto	Total	% de posibles gestos
Índice	85	87	97.70%
Medio	79	84	94.04%
Anular	82	86	95.34%
Índice con repetición	83	87	95.40%
Medio con repetición	81	86	94.18%
Anular con repetición	78	85	91.76%

5.4. Selección de parámetros para SVM

Para la selección de los mejores parámetros del algoritmo SVM, siguiendo lo estipulado en la sección 4.8, se muestra en la Tabla 10 los parámetros del costo y gamma además de las instancias correctamente clasificadas. En esta tabla se puede observar que el mayor número de gestos correctamente clasificados se obtiene con un costo $C = 4$ y $\gamma = 0.0078125$, con un 81.55% de los gestos correctamente clasificados haciendo uso de la validación cruzada con 10 repeticiones.

5.5. Efectividad de cada uno de los gestos en SVM

La Tabla 11 muestra la matriz de confusión con los gestos que se clasifican tanto correcta como incorrectamente. El gesto que tiene un mayor porcentaje de gestos correctamente clasificados es el pellizco con el dedo índice con 76 de los 85 gestos correctamente clasificados, para un 89.4% de efectividad.

Tabla 10: Valores del costo y gamma para el algoritmo SVM según lo estipulado en la sección 4.8

Costo (valor propuesto)	Gamma (valor propuesto)	Costo (valor real)	Gamma (valor real)	% de instancias clasificadas correctamente
2^{-5}	0	0.03125	0	28.07
2^{-4}	2^{-15}	0.0625	3.05176E-05	17.41
$(2^{-5}+2^{-4})/2$	$(0+2^{-15})/2$	0.046875	1.52588E-05	17.41
2^{-5}	$(0+2^{-15})/2$	0.03125	1.52588E-05	17.41
$(2^{-5}+2^{-4})/2$	0	0.046875	0	42.62
2^{-4}	0	0.0625	0	53.48
2^{-4}	2^{-15}	0.0625	3.05176E-05	17.41
2^{-3}	0	0.125	0	57.99
2^{-2}	0	0.25	0	61.88
2^{-1}	0	0.5	0	68.03
2^0	0	1	0	72.54
2^1	0	2	0	76.22
2^2	0	4	0	80.94
2^3	0	8	0	80.32
$(2^2+2^3)/2$	0	6	0	80.53
$(4+6)/2$	0	5	0	80.53
$(5+4)/2$	0	4.5	0	80.94
2^4	0	16	0	80.53
2^5	0	32	0	80.73
2^6	0	64	0	80.73
2^2	2^{-15}	4	3.05176E-05	46.51
2^2	$(2^{-15}+0)/2$	4	1.52588E-05	37.5
2^2	2^{-14}	4	6.10352E-05	54.5
2^2	2^{-13}	4	0.00012207	60.04
2^2	2^{-12}	4	0.000244141	63.93
2^2	2^{-10}	4	0.000976563	73.56
2^2	2^{-8}	4	0.00390625	79.09

2^2	2^{-7}	4	0.0078125	81.55
$(5+4)/2$	2^{-7}	4.5	0.0078125	81.14
2^2	2^{-6}	4	0.015625	78.27
2^2	$2^{-7.5}$	4	0.005524272	79.92
2^2	$2^{-7.25}$	4	0.006569503	79.71

El pellizco con el dedo índice con dos repeticiones tiene el menor porcentaje de gestos clasificados correctamente con 64 de los 83 gestos, para un 77.10% de efectividad. El resto de los gestos en promedio ronda el 80% de efectividad.

Tabla 11: Matriz de confusión obtenida a partir de la validación cruzada con 10 repeticiones de la clasificación de los seis gestos. Los gestos que son más frecuentemente confundidos es el índice con repetición con el medio con repetición y el medio con repetición con el anular con repetición.

	a	b	c	d	e	f	← Clasificado como
76	0	2	5	2	0	a=	índice
1	64	2	1	5	6	b=	medio
2	2	66	3	7	2	c=	anular
3	0	1	64	12	3	d=	índice repetición
0	2	0	2	65	12	e=	medio repetición
0	1	2	2	10	63	f=	anular repetición

En el caso del gesto de pellizco con el dedo índice, de los 85 gestos, se clasificaron 9 incorrectamente (9 falsos negativos) y 6 gestos fueron clasificados como gesto de índice cuando en realidad pertenecían a otro tipo de gesto (6 falsos positivos). El valor predictivo positivo (probabilidad de que un gesto sea índice si es clasificado como tal) es de 0.92, resultado de dividir el total de gestos con el índice correctamente clasificados contra el total de gestos clasificados como pellizco con el índice (76/82).

Finalmente, los valores de las características no se escalaron, ya que en ningún momento se logró tener un porcentaje de gestos correctamente clasificados superior a un 75% al realizar el escalamiento. Como ejemplo, se observó que escalando los datos con la segunda opción del escalamiento mostrado en la Tabla 5, con un Costo $C = 50$ y $\gamma = 0.022$ se logró obtener solo un 73.56% de los gestos correctamente clasificados, siguiendo los pasos para obtener los mejores parámetros para SVM.

5.6. Selección de parámetros para SMO

Al seguir los mismos pasos que se siguieron con el clasificador SVM, descritos en la sección 5.4, se obtuvo una precisión en la clasificación de los gestos del 81.15%, con un costo $C = 11$ y haciendo uso del kernel Polykernel, y sin aplicar el filtro para normalizar y estandarizar los datos, manteniendo el resto de sus valores por omisión. Al igual que en SVM, en SMO se obtuvieron valores más bajos en los datos escalados de tan solo un 75% de los gestos correctamente clasificados, pero en esta ocasión normalizando y estandarizando los datos.

5.7. Efectividad de cada uno de los gestos en SMO

El porcentaje de gestos correctamente clasificados (precisión) se muestra en la Tabla 12. Como ocurre en SVM, el pellizco con el dedo índice es el que mejor clasificado resulta y el pellizco con el dedo índice con doble repetición es el que tiene el más bajo porcentaje de clasificación, con un 77.10%.

Tabla 12: Porcentaje de gestos correctamente clasificados (también llamado recall) del total de cada uno de los gestos.

Gesto	Gesto clasificado correctamente	Total de gestos	% de gestos clasificados
Índice	76	85	89.41%
Medio	64	79	81.01%
Anular	64	82	78.04%
Índice con repetición	64	83	77.10%
Medio con repetición	64	81	79.01%
Anular con repetición	64	78	82.05%

5.8. Clasificación de los tres primeros gestos

La eficiencia promedio al clasificar únicamente los tres primeros gestos sutiles es de 91.46%, es decir 225 gestos correctamente clasificados de 246. Estos gestos son suficientes para problemas que requieran dar respuestas a pregunta ternarias del tipo *Si, No, No lo sé* o bien como el inicio y el fin del etiquetado de datos para el sensado participativo. En la Tabla 13 se presentan los resultados de la evaluación de los tres primeros gestos de pellizco con el dedo índice, medio y anular, usando SVM y manteniendo los mismos parámetros encontrados en la sección 5.4.

Tabla 13: Evaluación de los tres primeros gestos como una opción a usarse como respuestas a preguntas ternarias como Si, No, No lo sé.

Gesto	Gesto clasificado correctamente	Total de gestos	% de gestos clasificados
Índice	80	85	94.11%
Medio	70	79	88.60%
Anular	75	82	91.46%

En la Tabla 14 se muestra la matriz de confusión de los tres primeros gestos, donde el pellizco con el dedo índice tiene la mayor cantidad de gestos clasificados correctamente con 80, seguido del pellizco con el dedo anular con 75.

Tabla 14: Matriz de confusión de los tres primeros gestos sutiles.

a	b	c	←-- Clasificado como
80	0	5	a= índice
1	70	8	b= medio
1	6	75	c= anular

5.9. Clasificación de los últimos tres gestos

De manera similar a lo que se mostró en la sección 5.8, se muestran los resultados para la clasificación de los tres últimos gestos para evaluar la pertinencia de que sean usados como respuestas a preguntas ternarias y/o como inicio y fin del etiquetado de datos para el sensado participativo.

El porcentaje promedio de los gestos correctamente clasificados fue de 80.57%. Ninguno de los tres gestos fue claramente mejor que otro, según lo que muestra la Tabla 15 obtenida del análisis de WEKA.

Tabla 15: Resultados obtenidos de WEKA con referencia a los tres últimos gestos sutiles; Matriz de confusión (izq.) y la precisión (der.).

	a	b	c	Clasificado como	Precisión
	68	13	2	a= índice	0.883
	5	66	10	b= medio	0.717
	4	13	61	c= anular	0.836

5.10. Clasificación de un conjunto de datos suministrado

Al clasificar 93 instancias suministradas de forma independiente a los datos de prueba y usando el modelo generado con SVM se obtiene un 80.64% de los gestos correctamente clasificados. La Tabla 16 muestra la matriz de confusión de la clasificación de las 93 instancias, así como la precisión que obtuvo cada uno de los seis gestos.

Con el uso del modelo generado con SMO y al clasificar las mismas 93 instancias suministradas de forma independiente a los datos de prueba, se obtiene un 69.89% de los datos correctamente clasificados.

Tabla 16: Matriz de confusión (izq.) de las muestras de los seis gestos sutiles y la precisión (der.) de cada uno de ellos, haciendo uso del clasificador SVM.

						Precisión		
a	b	c	d	e	f	←--- Clasificado como		
14	0	1	1	0	0	a=	índice	0.875
0	14	1	0	0	0	b=	medio	1.000
2	0	12	0	1	1	c=	anular	0.750
0	0	0	16	0	0	d=	índice repetición	0.889
0	0	2	1	6	6	e=	medio repetición	0.667
0	0	0	0	2	13	f=	anular repetición	0.650
								0.806

5.11. Discusión

Durante el desarrollo del experimento se observó que con el uso del algoritmo DTW, para determinar si la señal contenía un posible gesto o no, pareciera que no funcionaba correctamente, ya que el porcentaje de gestos rechazados fue poco, por lo que se requirió comprobar su funcionamiento con otras señales diferentes a un gesto sutil, corroborando que si discriminaba correctamente este tipo de señales. El algoritmo DTW fue implementado con la finalidad de detectar señales que no correspondiera a un posible gesto, ya que la señal podría provenir de un movimiento brusco del mano causado por correr, caminar, brincar o simplemente agitar la mano. Este procedimiento resultaría muy útil en pruebas de campo realizadas en ambientes no controlados, donde la respuesta del usuario podría no ser uno de los gestos sutiles. Al detectar que la señal no corresponde a un gesto sutil, se le podría realizarse nuevamente la pregunta que se le hizo al usuario.

Los resultados obtenidos en el filtrado de la señal por medio del ventaneo, derivó en características que permitieron una buena diferenciación entre cada uno de los gestos. A priori, se tenía la idea de que la clasificación de los tres primeros gestos sería más precisa que utilizando los seis gestos. El porcentaje de la clasificación haciendo uso de los tres primeros gestos fue considerablemente alta (91%) y un poco menor al hacer uso de los seis gestos juntos (81%), lo que corrobora la hipótesis inicial. Este hecho permite hacer uso de los tres primeros gestos en preguntas que impliquen 3 posibles respuestas obteniéndose una alta certidumbre de que la respuesta es la que el usuario realmente dio.

Al hacer uso de los seis gestos al etiquetar los datos, se tendría aproximadamente 1 de cada 5 respuestas incorrectamente etiquetadas. De considerar que la respuesta tiene que tener mayor precisión,

podría preguntársele al usuario 2 veces en lapsos de tiempo cortos y si las respuestas son consistentes, se toma como verdadera. Los tiempos entre pregunta y pregunta deben ser lo más cortos posibles, ya que el usuario podría estar haciendo otra actividad de la que originalmente estaba realizando.

Con los resultados obtenidos por el clasificador SVM fueron mejores con respecto a SMO, tanto en la clasificación de los tres primeros gestos como con los seis gestos juntos, puede ser usado solamente este clasificador. La implicación de usar SMO, además de lo expuesto en este estudio, era que este algoritmo es nativo del software WEKA y de haber resultado tan bueno como SVM en ambas pruebas podría ser usado para que el usuario entrenara al modelo previo a los ejercicios de prueba para obtener posiblemente mejores resultados. Lamentablemente con la versión actual de la librería de SVM (v1.0.10) y la API de WEKA (v3.9), no es posible entrenar al modelo y clasificar las señales al mismo tiempo desde el reloj inteligente. Con versiones anteriores de la librería de SVM (v1.0.6) y la API de WEKA (v3.8) si es posible realizar ambas acciones de manera simultánea.

El obtener resultados de precisión más bajos en los datos escalados que en los no escalados, puede deberse a que al ser gestos que tienen poca diferencia entre la señal de uno y otro gesto, al hacer el escalamiento se vuelve aún menor y puede provocar que el algoritmo se confunda al clasificar correctamente cada uno de los gestos.

En cuanto al tipo de error que se puede presentar (tipo I y II), partiendo de la hipótesis de que los gestos pueden ser reconocidos eficientemente, los errores de tipo II o falso negativo no resultan tan graves en esta aplicación, ya que al existir un gesto que no fue reconocido como tal simplemente se le pide al usuario que lo realice nuevamente. En el caso de los errores de tipo I o falsos positivos, al detectar un gesto que no realmente no era tal, al tener un porcentaje de detección de los seis gestos muy bueno, se puede aplicar en circunstancias donde el tener uno de cada diez datos incorrectamente etiquetado puede no ser gran problema, pero si se aplican los gestos en aplicaciones críticas como enfermedades peligrosas o de cuidado, podría no resultar suficiente este nivel de detección de los gestos.

Capítulo 6. Conclusiones y trabajo futuro

6.1. Conclusiones

Los gestos son un medio para interactuar naturalmente con las demás personas, actualmente también nos sirven para interactuar con dispositivos que nos permiten realizar tareas de forma más fácil, cómoda o segura. Existe una gran variedad de dispositivos que utilizan el reconocimiento de gestos como medio de interacción, pero los más comunes son los teléfonos inteligentes. Hacer uso de este tipo de dispositivos permite una interacción casi en cualquier momento de nuestro día, pero su uso se ve limitado en ciertas situaciones donde es imposible el poder usarlos, como por ejemplo al momento de manejar, al tomar un baño, en una reunión de trabajo, entre otras. Una alternativa a este tipo de dispositivos son los relojes inteligentes, que pueden ser usados casi en cualquier momento, incluso al momento de tomar una ducha.

Los sensores presentes en los relojes inteligentes y los algoritmos de clasificación usados en este estudio, demostraron tener la capacidad de poder detectar gestos que impliquen poco esfuerzo, con una precisión del 81%, lo que es apropiado para aplicaciones como mandos a distancia, seleccionar alguna opción de un menú, establecer el inicio de una actividad, mini juegos, entre otras que son tolerantes al porcentaje de error obtenido. A continuación, se presentan las contribuciones que arrojó este estudio, atado a las limitaciones expuestas que dejan campo de estudio abierto para trabajo futuro que complemente el estudio.

6.2. Contribución

Las principales contribuciones de este estudio se indican a continuación:

- **Gestos sutiles con poco esfuerzo:** En este estudio se mostró la factibilidad de reconocer gestos realizados que requieran poco esfuerzo, utilizando un dispositivo vestible.
- **Modelo para filtrado:** Se generó un modelo para el filtrado de la señal que permite el preprocesamiento de la misma, para determinar si la señal contiene un gesto de interés, esto permite disminuir el número de falsos positivos o errores tipo I.

- **Modelos para clasificación:** Se propusieron un conjunto de características de las señales del acelerómetro y giroscopio que pueden ser implementadas en modelos de clasificación, para obtener buenos resultados en la detección de los gestos sutiles, donde las señales provenientes de diferentes gestos pueden llegar a ser muy parecidas.
- **Selección de parámetros:** Se presentó un procedimiento para determinar de forma rápida y con buenos resultados los mejores parámetros que un clasificador requiere haciendo uso de la validación cruzada y de aproximaciones.

6.3. Limitaciones

El utilizar un algoritmo de clasificación no nativo del software WEKA, impidió implementar en este estudio el entrenamiento directamente del usuario al modelo y hacer la clasificación de los gestos al mismo tiempo sin la intervención del investigador, esto debido a que se tienen que usar dos API diferentes de WEKA, una para el entrenamiento y otra para la clasificación.

Otra limitación son las condiciones controladas en las que se realizó el estudio, ya que en condiciones naturales se pueden presentar condiciones que generen ruido en exceso en la señal del giroscopio y acelerómetro, aun y cuando estos gestos están pensados para realizarse en condiciones pasivas o de poco movimiento.

Por otra parte, se utilizaron solamente dos tipos de relojes inteligentes, que eran los que se tenían disponibles, lo que limita la generalización de los resultados.

6.4. Trabajo futuro

Algunas propuestas de trabajo futuro que contribuirían en el trabajo realizado en esta tesis son:

- **Aplicación:** Desarrollar una aplicación que implemente los gestos sutiles para el etiquetado de actividad, para realizar comandos rápidos por personas que tienen poca movilidad, entre otras.

- **Evaluación en condiciones más diversas:** Realizar evaluaciones más exhaustivas en condiciones naturales tanto de los umbrales para el filtrado de la señal por medio del ventaneo como para la correcta clasificación de los gestos.
- **Entrenamiento de los modelos:** Implementar de forma automática el entrenamiento del modelo predictivo por parte del usuario sin la intervención del investigador.
- **Plataformas vía internet:** Hacer uso, por medio de la internet, de plataformas públicas como Fiware, que permitan el procesamiento de un modelo más general a partir del entrenamiento que registre cada usuario y que se vea reflejado en cada reloj inteligente.
- **Ensamblados:** Explorar la posibilidad de usar varios modelos con los mejores porcentajes de precisión al mismo tiempo, para realizar un consenso entre todos para seleccionar la mejor opción buscando mejorar la precisión en la detección de los gestos sutiles. Esto gracias al poder de procesamiento con el que cuentan los relojes inteligentes actualmente.
- **Variantes del gesto:** Realizar estudios de variantes de los gestos que permitan complementar los propuestos o reemplazar aquellos que tuvieron resultados más bajos.

Literatura citada

- Alavi, S., Arsenault, D. & Whitehead, A., 2016. Quaternion-Based Gesture Recognition Using Wireless Wearable Motion Capture Sensors. *Sensors*, 16(5), p.605. Available at: <http://www.mdpi.com/1424-8220/16/5/605>.
- Aoyama, S., Shizuki, B. & Tanaka, J., 2016. ThumbSlide: An Interaction Technique for Smartwatches using a Thumb Slide Movement. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, pp.2403–2409. Available at: <http://dl.acm.org/citation.cfm?id=2851581.2892435>.
- Avgerinakis, K., Briassouli, A. & Kompatsiaris, I., 2013. Activity detection and recognition of daily living events. *Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare - MIIRH '13, (OCTOBER)*, pp.3–10. Available at: <http://dl.acm.org/citation.cfm?doid=2505323.2505327> [Accessed August 22, 2016].
- Bobeth, J. et al., 2012. Evaluating performance and acceptance of older adults using freehand gestures for TV menu control. *Proceedings of the 10th European conference on Interactive tv and video*, pp.35–44. Available at: <http://doi.acm.org/10.1145/2325616.2325625> http://dl.acm.org/ft_gateway.cfm?id=2325625&type=pdf.
- Bojórquez, D.A.M., 2015. *Detección de ansiedad por medio de cómputo vestible*. CICESE.
- Chih-Wei Hsu, Chih-Chung Chang, and C.-J.L., 2008. A Practical Guide to Support Vector Classification. *BJU international*, 101(1), pp.1396–400. Available at: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Costante, G. et al., 2014. Personalizing a smartwatch-based gesture interface with transfer learning. *European Signal Processing Conference*, pp.2530–2534.
- Dementyev, A. & Paradiso, J.A., 2014. WristFlex: Low-Power Gesture Input with Wrist-Worn Pressure Sensors. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*, pp.161–166. Available at: <http://dl.acm.org/citation.cfm?doid=2642918.2647396>.
- Fu, T.C., 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1), pp.164–181. Available at: <http://dx.doi.org/10.1016/j.engappai.2010.09.007>.
- Hayashi, K. & Takai, K., 2015. Finite-sample analysis of impacts of unlabelled data and their labelling mechanisms in linear discriminant analysis. *Communications in Statistics - Simulation and Computation*, 46(June), pp.00–00. Available at: <http://www.tandfonline.com/doi/full/10.1080/03610918.2014.957847>.
- Hofmann, T., Scholkopf, B. & Smola, A.J., 2008. Kernel methods in machine learning. *Annals of Statistics*, 36(3), pp.1171–1220.
- Hong, F. et al., 2016. MGRA: Motion Gesture Recognition via Accelerometer. *Sensors (Basel, Switzerland)*, 16(4). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27089336> [Accessed July 8, 2016].

- Jain, A.K., Duin, R.P.W. & Mao, J., 2000. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), pp.4–37. Available at: http://ieeexplore.ieee.org/ielx5/34/17859/00824819.pdf?tp=&arnumber=824819&isnumber=17859%5Cnhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=824819.
- Jiménez, M.G. & Sierra, a. Á., 2003. Análisis de Datos en WEKA—Pruebas de Selectividad. *It.Uc3M.Es*, pp.1–9. Available at: <http://www.it.uc3m.es/jvillena/irc/practicass/06-07/28.pdf>.
- Kalantarian, H., Alshurafa, N. & Sarrafzadeh, M., 2015. Detection of Gestures Associated with Medication Adherence using Smartwatch-based Inertial Sensors. *IEEE Sensors Journal*, 16(c), pp.1–1. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7315013>.
- Kellogg, B., Talla, V. & Gollakota, S., 2014. Bringing Gesture Recognition To All Devices. *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pp.303–316. Available at: <https://www.usenix.org/system/files/conference/nsdi14/nsdi14-paper-kellogg.pdf>.
- Ko, A.J., 2017. A three-year participant observation of software startup software evolution. *Proceedings - 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track, ICSE-SEIP 2017*, pp.3–12.
- Korpela, J. et al., 2015. An energy-aware method for the joint recognition of activities and gestures using wearable sensors. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC '15*. New York, New York, USA: ACM Press, pp. 101–108. Available at: <http://dl.acm.org/citation.cfm?doid=2802083.2808400> [Accessed August 17, 2016].
- Korpela, J. et al., 2016. Tree-structured classifier for acceleration-based activity and gesture recognition on smartwatches. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, pp. 1–4. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7457053> [Accessed July 8, 2016].
- Lane, N.D., Choudhury, T. & Zhao, F., 2011. Mobile Sensing: Challenges, Opportunities and Future Directions. *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, p.637. Available at: <http://dl.acm.org/citation.cfm?doid=2030112.2030253> [Accessed August 20, 2016].
- Liu, Z. et al., 2014. Visual Gesture Recognition for Human Robot Interaction Using Dynamic Movement Primitives.
- Louridas, P. & Ebert, C., 2016. Machine Learning. *IEEE Software*, 33(5), pp.110–115. Available at: <http://ieeexplore.ieee.org/document/7548905/> [Accessed August 30, 2017].
- Lu, Z. et al., 2014. A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices. *IEEE Transactions on Human-Machine Systems*, 44(2), pp.293–299. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6748952> [Accessed August 20, 2016].
- Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), pp.81–97. Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/h0043158>.

- Müller, M., 2007. *Information Retrieval for Music and Motion*, Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: <http://link.springer.com/10.1007/978-3-540-74048-3>.
- Neto, P., Pires, J.N. & Moreira, A.P., 2009. Accelerometer-based control of an industrial robotic arm. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp.1192–1197.
- Platt, J.C., 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. *Advances in kernel methods*, pp.185–208. Available at: http://www.bradblock.com/Sequential_Minimal_Optimization_A_Fast_Algorithm_for_Training_Support_Vector_Machine.pdf.
- Porzi, L. et al., 2013. A Smart Watch-based Gesture Recognition System for Assisting People with Visual Impairments Categories and Subject Descriptors. *IMMPD '13 Proceedings of the 3rd ACM international workshop on Interactive multimedia on mobile & portable devices*, pp.19–24.
- Ratanamahatana, C. & Keogh, E., 2004. Everything you know about dynamic time warping is wrong. *Third Workshop on Mining Temporal and Sequential Data*, pp.22–25. Available at: [http://spoken-number-recognition.googlecode.com/svn/trunk/docs/Dynamic time warping/DTW_myths.pdf](http://spoken-number-recognition.googlecode.com/svn/trunk/docs/Dynamic%20time%20warping/DTW_myths.pdf).
- Rico, J. & Brewster, S., 2010. Usable Gestures for Mobile Interfaces : Evaluating Social Acceptability. , pp.887–896.
- Salomoni, P. et al., 2012. S2S: a peer to peer protocol for participative sensing. *Proceedings of the 27th Annual ACM Symposium on Applied Computing - SAC '12*, p.634. Available at: <http://dl.acm.org/citation.cfm?id=2245276.2245398>.
- Samuel, H.W., Fastest Association Rule Mining Algorithm Predictor Categories and Subject Descriptors.
- Senin, P., 2008. Dynamic Time Warping Algorithm Review. *Science*, 2007(December), pp.1–23. Available at: <http://129.173.35.31/~pf/Linguistique/Treillis/ReviewDTW.pdf>.
- Shoab, M. et al., 2016. Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors. *Sensors*, 16(4), p.426. Available at: <http://www.mdpi.com/1424-8220/16/4/426>.
- Song, L. et al., 2015. Health sensing by wearable sensors and mobile phones: A survey. *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services, Healthcom 2014*, pp.453–459.
- Spanias, J.A., Perreault, E.J. & Hargrove, L.J., 2014. A strategy for labeling data for the neural adaptation of a powered lower limb prosthesis. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, pp.3090–3093.
- Stisen, A. et al., 2015. Smart Devices are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition. *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems - SenSys '15*, pp.127–140. Available at: <http://dl.acm.org/citation.cfm?doid=2809695.2809718>.
- Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, 36(2), pp.111–147.
- Sun, Y., Lank, E. & Terry, M., 2017. Label-And-Learn: Visualizing the Likelihood of Machine Learning

- Classifier's Success During Data Labeling. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*. New York, New York, USA: ACM Press, pp. 523–534. Available at: <http://dl.acm.org/citation.cfm?doid=3025171.3025208>.
- Takahashi, T. & Kishino, F., 1991. Hand gesture coding based on experiments using a hand gesture interface device. *ACM SIGCHI Bulletin*, 23(2), pp.67–74. Available at: <http://portal.acm.org/citation.cfm?doid=122488.122499> [Accessed August 20, 2016].
- Vaughan, N. & Gabrys, B., 2016. Comparing and Combining Time Series Trajectories Using Dynamic Time Warping. *Procedia Computer Science*, 96(September), pp.465–474. Available at: <http://dx.doi.org/10.1016/j.procs.2016.08.106>.
- Wen, H., Ramos Rojas, J. & Dey, A.K., 2016. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pp.3847–3851. Available at: <http://dl.acm.org/citation.cfm?doid=2858036.2858466>.
- Wenhui, W. et al., 2009. Dynamic gesture recognition based on multiple sensors fusion technology. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2009*, pp.7014–7017. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19964189> [Accessed July 8, 2016].
- Wilhelm, M. et al., 2015. eRing: multiple finger gesture recognition with one ring using an electric field. *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction - WOAR '15*, pp.1–6. Available at: <http://dl.acm.org/citation.cfm?doid=2790044.2790047>.
- Wu, Y., Chen, K. & Fu, C., 2016. Natural Gesture Modeling and Recognition Approach Based on Joint Movements and Arm Orientations. , 16(21), pp.7753–7761.
- Xu, R., Zhou, S. & Li, W., 2011. MEMS Accelerometer Based Non-Specific-User Hand Gesture Recognition. *IEEE Sensors Journal*, 2(c), pp.1838–1843. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6009159>.
- Yadav, S. & Shukla, S., 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, (Cv), pp.78–83.
- Yang, G. et al., 2015. HIGH-ORDER INFORMATION FOR ROBUST IRIS RECOGNITION UNDER LESS CONTROLLED CONDITIONS Dalian University of Technology , 2 The Hong Kong Polytechnic University. , pp.4535–4539.
- Zeni, M., Zaihrayeu, I. & Giunchiglia, F., 2014. Multi-device Activity Logging. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, pp.299–302. Available at: http://doi.acm.org/10.1145/2638728.2638756%5Cnhttp://dl.acm.org/ft_gateway.cfm?id=2638756&type=pdf.
- Zhao, Y. et al., 2015. Demo : Finger and Hand Gesture Recognition using Smartwatch. *MobiSys*, p.95616.