



**INSTITUTO POLITÉCNICO NACIONAL**  
**CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN.**

## **T E S I S**

**“Monitoreo y análisis de variables fisiológicas con  
un enfoque hacia los dispositivos móviles”**

PARA OBTENER EL GRADO DE:  
**Maestría en Ciencias de la Computación.**

P R E S E N T A:

**Ing. Ana Rosa Acosta Arenas**

Directores de Tesis:

Dr. Miguel Jesús Torres Ruiz

Dr. José Giovanni Guzman Lugo



Ciudad de México

Enero 2018.



SIP-14 bis

# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de           México           siendo las   14:00   horas del día   08   del mes de   septiembre   de   2017   se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

**Centro de Investigación en Computación**

para examinar la tesis titulada:

**“Monitoreo y análisis de variables fisiológicas con un enfoque hacia los dispositivos móviles”**

Presentada por el alumno:

<b>ACOSTA</b> Apellido paterno	<b>ARENAS</b> Apellido materno	<b>ANA ROSA</b> Nombre(s)
		Con registro: <b>B 1 5 1 1 3 3</b>

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA Directores de Tesis

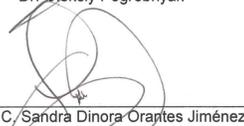
  
\_\_\_\_\_  
Dr. Miguel Jesús Torres Ruiz

  
\_\_\_\_\_  
Dr. José Giovanni Guzmán Lugo

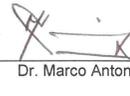
  
\_\_\_\_\_  
Dr. Sergio Suárez Guerra

  
\_\_\_\_\_  
Dr. Ofeksiy Pogrebnyak

  
\_\_\_\_\_  
Dr. Marco Antonio Moreno Ibarra

  
\_\_\_\_\_  
M. en C. Sandra Dinora Orantes Jiménez

### PRESIDENTE DEL COLEGIO DE PROFESORES

  
\_\_\_\_\_  
Dr. Marco Antonio Ramírez Salinas





**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESIÓN DE DERECHOS*

En la Ciudad de México el día 04 del mes de Diciembre del año 2017, el (la) que suscribe Ana Rosa Acosta Arenas alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro B151133, adscrito al Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección del Dr. Miguel Jesús Torres Ruiz y Dr. José Giovanni Guzmán Lugo y cede los derechos del trabajo intitulado Monitoreo y análisis de variables fisiológicas con un enfoque hacia los dispositivos móviles, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección ana.acosta.arenas@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

*Ana Rosa Acosta Arenas*

Ana Rosa Acosta Arenas.

# Resumen

La enfermedad de Parkinson (PD) es un trastorno neurodegenerativo progresivo que se caracteriza por síntomas motores. Alrededor de 10 millones de personas en todo el mundo están viviendo con la enfermedad de Parkinson. Los síntomas pueden ser tratados con medicación o con estimulación cerebral profunda (DBS). Sin embargo, existen algunos problemas reconocidos en el diagnóstico y tratamiento de la enfermedad.

Actualmente no existen pruebas de sangre o de laboratorio que diagnostiquen la enfermedad de Parkinson. Por ello, el diagnóstico se basa en la historia clínica y en un examen neurológico. Hay muchos síntomas y características de la enfermedad de Parkinson que pueden ser objetivamente medidos y controlados utilizando los sensores de los dispositivos móviles o smartphones. Con el uso de aplicaciones destinadas a la salud, los participantes pueden monitorear su estado de salud, sin la necesidad de visitar un hospital o un entorno de laboratorio.

En este trabajo se propone una metodología que permita monitorear los síntomas relacionados con la enfermedad de Parkinson, mediante la recopilación de datos por medio de un dispositivo móvil, con el propósito de evaluar la información obtenida utilizando técnicas de aprendizaje automático, para su interpretación por parte de un especialista médico. Esta metodología incluye el 1) El monitoreo de los síntomas de la enfermedad de Parkinson a través de la aplicación móvil, 2) Una etapa de preprocesamiento de la información donde se conjuntan dos diferentes fuentes de información así como técnicas de procesamiento digital de señales para la extracción de características. 3) Una etapa de clasificación utilizando técnicas de aprendizaje automático para la clasificación de participantes con diagnóstico de Enfermedad de Parkinson y sin diagnóstico y finalmente 4) Una etapa de evaluación del desempeño de los algoritmos de clasificación.

# Abstract

The Parkinson's disease (PD) is a progressive neurodegenerative disorder that is characterized by motor symptoms. These symptoms can be relieved with medication or with deep brain stimulation (DBS). More than 10 million people in the worldwide are living with the Parkinson's disease. However, there are well-recognized problems in the diagnostics and treatment of the PD.

Nowadays, there are no blood or laboratory tests to diagnose the Parkinson's disease. Therefore, the diagnosis is based on the clinical history and a neurological examination. There are many symptoms and characteristics of the Parkinson's disease that can be objectively measured and controlled using the sensors of mobile devices or smartphones. Thus, with the use of applications aimed at health, participants can monitor their health status, without the need to visit a hospital or a specific laboratory environment.

Therefore, this dissertation proposes a methodology that allows monitoring the symptoms related to the Parkinson's disease, through the data collection by means of a mobile device, with the purpose of evaluating information obtained from embedded sensor, using a set of automatic learning techniques, for its interpretation by a medical specialist. The proposed methodology includes the following stages: 1) The monitoring of the symptoms of the Parkinson's disease through the mobile application, 2) A preprocessing task of information where two different sources of information are combined as well as techniques of digital signal processing for feature-extraction. 3) A classification task by using machine learning approaches for classifying participants diagnosed with the Parkinson's disease and without diagnosis, and 4) A evaluation process with respect to the performance of the classification algorithms.

# Agradecimientos

A mis padres, por su esfuerzo dedicación y apoyo incondicional.

A mis asesores, Dr. Miguel Torres Ruiz y Dr. Giovanni Guzmán Lugo, por su apoyo continuo durante mis estudios, por su paciencia motivación y conocimiento.

Al Dr. Alfonso Arellano y la Dr. Mayela Rodriguez Violante por las facilidades otorgadas para realizar este estudio de investigación, por su conocimiento y enseñanzas.

Al Instituto Politécnico Nacional (IPN) y al Consejo Mexicano de Ciencia y Tecnología (Conacyt) por el apoyo financiero y académico en la realización de este trabajo.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	1
1.2. Planteamiento del Problema . . . . .	2
1.3. Justificación . . . . .	4
1.4. Hipótesis . . . . .	5
1.5. Objetivo General . . . . .	6
1.6. Metas . . . . .	6
1.7. Aportaciones . . . . .	6
1.8. Organización de la Tesis . . . . .	6
<b>2. Trabajos Relacionados</b>	<b>8</b>
2.1. Aplicaciones móviles enfocadas en el cuidado de la salud . . . . .	8
2.2. Aplicaciones móviles en la enfermedad de Parkinson . . . . .	10
2.2.1. Minería de datos en diagnóstico de Enfermedad de Parkinson. . . . .	13
2.3. Aplicaciones comerciales . . . . .	15
2.3.1. Aplicación Móvil mPower . . . . .	16
2.3.2. Aplicación Móvil Roche App Parkinson's Disease . . . . .	16
<b>3. Marco Teórico</b>	<b>18</b>
3.1. Introducción . . . . .	18
3.2. Síntomas de la enfermedad de Parkinson y diagnóstico . . . . .	19
3.2.1. Escala unificada para la evaluación de la enfermedad de Parkinson . . . . .	20
3.3. Plataforma Research Kit Apple . . . . .	22
3.4. Algoritmo SMOTE Synthetic Minority Over-sampling TEchnique . . . . .	24
3.5. Aprendizaje de Máquina . . . . .	25

3.5.1. Regresión Logística . . . . .	25
3.5.2. Árbol de decisión (modelo de clasificación ID3) . . . . .	27
3.5.3. Redes Neuronales Artificiales . . . . .	29
3.6. F1 Score . . . . .	33
3.7. Máquinas de Soporte Vectorial . . . . .	33
3.8. Random forest o Bosques Aleatorios . . . . .	34
<b>4. Metodología Propuesta</b>	<b>38</b>
4.1. Introducción . . . . .	38
4.2. Etapa de Monitoreo . . . . .	39
4.2.1. Destreza Manual . . . . .	40
4.2.2. Caminata . . . . .	41
4.2.3. Memoria Espacial . . . . .	43
4.2.4. Cuestionario Calidad de Vida . . . . .	44
4.3. Preprocesamiento de los datos . . . . .	45
4.3.1. Integración de las bases de datos y limpieza de los datos	46
4.3.2. Transformación de los datos . . . . .	48
4.3.3. Algoritmo SMOTE para desbalance entre clases . . . . .	50
4.3.4. Aprendizaje Supervisado . . . . .	52
4.4. Clasificador . . . . .	53
4.4.1. Regresión Logística. . . . .	54
4.4.2. Boosted Decision Tree. . . . .	55
4.4.3. Máquina de Soporte Vectorial. . . . .	56
4.4.4. Redes neuronales. . . . .	57
4.5. Evaluación del Desempeño . . . . .	58
<b>5. Experimentos y Resultados</b>	<b>61</b>
5.1. Descripción de la Aplicación Móvil . . . . .	61
5.1.1. Formulario de Consentimiento Informado . . . . .	63
5.1.2. Registro . . . . .	67
5.2. Actividades del Estudio . . . . .	67
5.2.1. Prueba de Destreza Manual . . . . .	68
5.2.2. Actividad de Caminata . . . . .	69
5.2.3. Actividad de Memoria . . . . .	71
5.2.4. Evaluación de los Síntomas . . . . .	74
5.2.5. Tarjeta de Cuidados . . . . .	82
5.2.6. Ejercicio Manual. Prueba de las 3 Estacas . . . . .	85
5.3. Recolección de Datos . . . . .	87

5.3.1. Estructura de los Datos . . . . .	88
5.4. Diseño del Estudio . . . . .	89
5.4.1. Clasificación de los participantes . . . . .	91
5.4.2. Análisis de Resultados Obtenidos . . . . .	92
5.4.3. Tiempo de ejecución de los clasificadores. . . . .	94
<b>6. Conclusiones y Trabajo a Futuro</b>	<b>96</b>
6.1. Conclusiones . . . . .	96
6.2. Limitaciones . . . . .	97
6.3. Trabajo a Futuro . . . . .	97
<b>Referencias</b>	<b>99</b>
6.4. Requerimientos Funcionales . . . . .	110
6.5. Requerimientos No Funcionales . . . . .	111
6.6. Arquitectura de la Aplicación . . . . .	111
6.7. Análisis de Casos de Uso . . . . .	112
6.8. Descripción Detallada de los Casos de Uso . . . . .	113
6.9. Diagrama de Actividades . . . . .	116
6.10. Diagrama de Clases . . . . .	117
6.11. Diagrama de Secuencia . . . . .	118

# Índice de cuadros

4.1. Configuración del algoritmo SMOTE . . . . .	52
4.2. Características utilizadas en el modelo de aprendizaje automático . . . . .	59
5.1. Parámetros resultantes de la prueba de destreza manual . . . . .	69
5.2. Parámetros resultantes de la Actividad de Caminata . . . . .	71
5.3. Resultados por score actividad de Memoria participantes con Diagnóstico de EP. . . . .	73
5.4. Resultados por score actividad de Memoria participantes sin Diagnóstico de EP. . . . .	74
5.5. Resultados evaluación de los síntomas. . . . .	82
5.6. Número de Participantes por Actividad . . . . .	88
5.7. Características de los participantes del estudio . . . . .	89
5.8. Desempeño de los Clasificadores . . . . .	92
5.9. Características del sistema de cómputo . . . . .	94
6.1. Requerimientos Funcionales . . . . .	110
6.2. Requerimientos No Funcionales . . . . .	111
6.3. Descripción del actor Usuario . . . . .	112
6.4. Caso de Uso Iniciar Sesión . . . . .	113
6.5. Caso de Uso Registrar Nuevo Usuario . . . . .	113
6.6. Caso de Uso Cerrar Sesión . . . . .	113
6.7. Caso de Uso Actividad de Destreza Manual . . . . .	114
6.8. Caso de Uso Actividad de Caminata . . . . .	114
6.9. Caso de Uso Actividad de Memoria 1 . . . . .	115
6.10. Caso de Uso Actividad de Memoria 2 . . . . .	115

# Índice de figuras

1.1. Diagrama a Bloques del planteamiento del problema . . . . .	3
2.1. Aplicación móvil mPower . . . . .	16
2.2. Aplicación móvil Roche . . . . .	17
3.1. Plataforma Research Kit. . . . .	22
3.2. Ejemplo de clasificación de SVM para dos clases . . . . .	34
4.1. Metodología Propuesta . . . . .	39
4.2. Etapa de Monitoreo. . . . .	40
4.3. Actividad de Destreza Manual . . . . .	41
4.4. Actividad de Caminata . . . . .	42
4.5. Preprocesamiento de la Información. . . . .	46
5.1. Diagrama funcional de la Aplicación Móvil. . . . .	62
5.2. Arquitectura del Framework ResearchKit de Apple. . . . .	62
5.3. Consentimiento Informado . . . . .	64
5.3. Consentimiento Informado . . . . .	65
5.3. Consentimiento Informado . . . . .	66
5.4. Registro de la aplicación móvil Parkinson App . . . . .	67
5.5. Prueba de Destreza Manual . . . . .	68
5.6. a) Fragmento de datos de Prueba de Destreza Manual b) Fragmento de archivo tappingResults.json de la primera tabla. . . . .	69
5.7. Actividad de Caminata . . . . .	70
5.8. Fragmento de datos de Actividad de Caminata . . . . .	70
5.9. Fragmento de archivo deviceMotion.json de la actividad de Caminata . . . . .	71
5.10. Actividad de Memoria . . . . .	72
5.11. Juego de Memoria . . . . .	73

## ÍNDICE DE FIGURAS

---

5.12. Resultados de la Actividad de Memoria a) Participantes sin diagnóstico de EP b) Participantes con diagnóstico de EP. . .	74
5.13. Cuestionario Evaluación de los Síntomas . . . . .	76
5.13. Cuestionario Evaluación de los Síntomas . . . . .	77
5.13. Cuestionario Evaluación de los Síntomas . . . . .	78
5.14. Resultados de las encuestas de acuerdo a la intensidad del síntoma . . . . .	79
5.14. Resultados de las encuestas de acuerdo a la intensidad del síntoma . . . . .	80
5.14. Resultados de las encuestas de acuerdo a la intensidad del síntoma . . . . .	81
5.15. Tarjeta de Cuidados . . . . .	84
5.16. Prueba de las Estacas . . . . .	86
5.17. Resultados del porcentaje de la Actividad Completada, bajo efecto de medicación. . . . .	87
5.18. Coordenadas Actividad de Destreza Manual . . . . .	89
5.19. Componentes de Aceleración de la Prueba de Caminata . . .	90
5.19. Componentes de Aceleración de la Prueba de Caminata . . .	91
5.20. Método de k-fold cross validation con 10 iteraciones . . . . .	92
5.21. Curvas ROC para los diferentes métodos de clasificación . .	93
5.22. Gráfica tiempo de ejecución en segundos. . . . .	95
6.1. Diagrama Casos de Uso Usuario . . . . .	112
6.2. Diagrama de Actividades del Sistema . . . . .	116
6.3. Diagrama de Clases del Sistema . . . . .	117
6.4. Diagrama de Secuencia - Registrar Usuario. . . . .	118
6.5. Diagrama de Secuencia - Actividad Caminata . . . . .	118
6.6. Diagrama de Secuencia - Actividad de Destreza Manual . . .	119
6.7. Diagrama de Secuencia - Actividad Juego de Memoria 1 . .	119
6.8. Diagrama de Secuencia - Actividad Juego de Memoria 2 . .	120

# Capítulo 1

## Introducción

### 1.1. Antecedentes

En la actualidad gracias a la infraestructura tecnológica, cientos de millones de personas alrededor del mundo tienen un smartphone en el bolsillo equipado con poderosos procesadores y avanzados sensores que pueden realizar funciones como registrar el movimiento, hacer mediciones y guardar información, los cuales se han convertido en una herramienta para realizar estudios médicos. La enorme cantidad de dispositivos de este tipo abre un mundo de posibilidades relacionadas al cuidado de la salud, aprovechando la funcionalidad de los smartphones para recolectar y procesar los datos y utilizando a los humanos como sensores en un entorno colaborativo [Boulos y cols. (2011)].

Por otra parte, la enfermedad de Parkinson (PD) es un trastorno neurodegenerativo progresivo que se caracteriza por síntomas motores que afecta en su mayoría a la población adulta. Sin embargo, los métodos actuales para cuantificar el deterioro motor de Parkinson y los efectos del tratamiento no son muy precisos, debido a las dificultades del diagnóstico [de Lau y Breteler (2006)].

Hay muchos síntomas y características de la enfermedad de Parkinson que pueden ser objetivamente medidos y controlados mediante smartphones, con la finalidad de permitir un mejor monitoreo y análisis de la progresión de la enfermedad.

## 1.2. Planteamiento del Problema

Actualmente, no existen pruebas de sangre o de laboratorio que diagnostiquen la enfermedad de Parkinson. Por ello, el diagnóstico se basa en la historia clínica y en un examen neurológico. En algunos casos, la enfermedad de Parkinson puede ser difícil de diagnosticar con precisión al inicio de la enfermedad. Los signos y síntomas tempranos de la enfermedad de Parkinson a veces se pueden descartar como los efectos del envejecimiento normal, de este modo, los médicos piden estudios de imágenes cerebrales o pruebas de laboratorio con el fin de descartar otros trastornos neurodegenerativos. Sin embargo, las exploraciones cerebrales por tomografía computarizada e imágenes por resonancia magnética de las personas con la enfermedad de Parkinson, generalmente parecen normales [Samà y cols. (2012)].

Las medidas actuales para el diagnóstico de la salud y la enfermedad de Parkinson son a menudo insensibles, esporádicas y subjetivas. Además, estas medidas generalmente no están diseñadas para proporcionar una retroalimentación significativa a los individuos.

Con la finalidad de apoyar en el diagnóstico objetivo, utilizando smartphones como sensores, se propone el siguiente sistema [Figura 1.1], el cual permite obtener y clasificar los síntomas relacionados con la Enfermedad de Parkinson y consta de tres etapas principales:

## 1.2. PLANTEAMIENTO DEL PROBLEMA

---

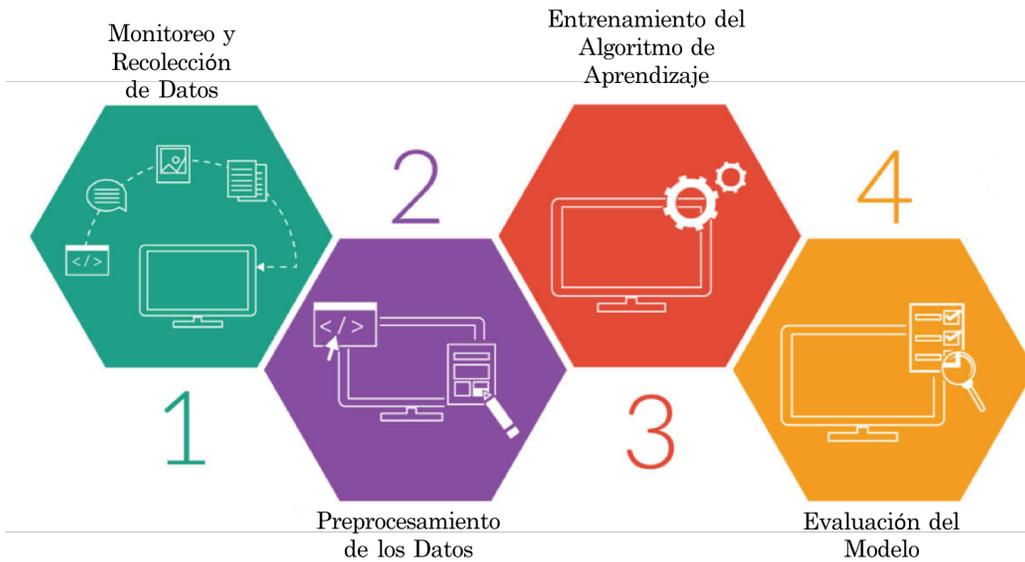


Figura 1.1: Diagrama a Bloques del planteamiento del problema

1. Datos de entrada a través de una aplicación móvil que permite realizar las siguientes pruebas: una prueba de destreza manual, una prueba de balance y ciclo de marcha y una prueba de memoria.
  - La prueba de destreza manual permite medir el nivel de temblor de las manos de un paciente mediante una prueba que consiste en presionar dos botones sistemáticamente.
  - La prueba de memoria, permite medir el nivel de deterioro de la memoria, falta de criterio y confusión, mediante un juego de memoria.
2. La prueba de balance y ciclo de marcha, permite evaluar el balance mientras las personas realizan actividad de ciclo de marcha en un tiempo determinado.
3. Un algoritmo de aprendizaje automático que permite identificar a pacientes con Parkinson.
4. Un modelo de verificación que permite evaluar la eficiencia del sistema propuesto.

### 1.3. Justificación

A medida que las nuevas tecnologías mejoran los niveles de vida, nos enfrentamos al aumento del envejecimiento de la población, un proceso que aumenta la proporción de personas de edad avanzada dentro de la población total y es uno de los principales problemas de este siglo. Afecta o afectará tanto a los países desarrollados como a los países en desarrollo [Dang T. (2001)]. Por lo tanto, las enfermedades asociadas con el envejecimiento, como la enfermedad de Parkinson van en aumento. Los efectos de la enfermedad de Parkinson se presentan en una de cada quinientas personas, la mayoría de estas personas son mayores de 50 años, por lo que los teléfonos inteligentes representan una herramienta útil para recopilar información sobre los síntomas cambiantes de la enfermedad de Parkinson, sin necesidad de acudir a un hospital.

En la actualidad, una de las desventajas de participar en un estudio médico que permita proporcionar nuevas claves en el diagnóstico de la enfermedad, es que se requiere la participación activa de los pacientes en viajar al hospital o institución médica para realizarse las pruebas médicas y llenar los cuestionarios destinados. El uso de smartphones para monitorear el estado de salud de las personas representa una ventaja para los pacientes, pues es posible generar los datos en donde quiera que se encuentre, proporcionando información objetiva.

Otro inconveniente que enfrenta la investigación médica es la falta de continuidad en los datos, es decir, los participantes de un estudio deben ir a los centros médicos una vez cada tres meses o más, dependiendo de la disponibilidad del hospital para realizarse pruebas, aunque sus síntomas cambien constantemente. Debido a esto la información obtenida resulta imprecisa. Con el uso de aplicaciones destinadas a la salud, los pacientes registran sus datos día a día realizando sus actividades cotidianas, sin la necesidad de visitar un hospital, además de que pueden complementarse con encuestas y seguimientos de los síntomas.

Lo más importante de estas aplicaciones es la colaboración activa de las personas que facilita a los investigadores obtener posibles relaciones entre los síntomas y los hábitos diarios de sus pacientes, además de que se potencializa la probabilidad de obtener datos necesarios para las técnicas de aprendizaje

automático, *Machine Learning*, en donde es necesario generar la mayor cantidad de datos para desarrollar tratamientos individuales o de alcance mayor.

### 1.4. Hipótesis

Las manifestaciones de la enfermedad de Parkinson pueden incluir temblores, cambios en el ciclo marcha o bloqueos, bradiquinesia, rigidez e inestabilidad postural [Samà y cols. (2012)], aunque la enfermedad de Parkinson no afecta a todas las personas de la misma manera, la tasa de progresión y los síntomas particulares difieren entre los individuos, por esta razón su diagnóstico y medición de la progresión resulta una tarea desafiante.

Actualmente, no existe una cura conocida para la enfermedad, por consiguiente, los tratamientos están enfocados a mitigar los efectos de estos síntomas. En promedio, los pacientes visitan a su médico cada 4 o 6 meses. Sin embargo, la variabilidad de los síntomas en la vida cotidiana y los efectos de los tratamientos propuestos ha sido poco estudiada y puede revelar una oportunidad para mejorar la calidad de vida de los pacientes con enfermedad de Parkinson [Schapira (2005)] [Trister y cols. (2016)].

La hipótesis de este trabajo está basada en utilizar los teléfonos móviles o smartphones, los cuales se encuentran equipados con sensores como acelerómetros y giroscopios además de un micrófono integrado, para proporcionar sustitutos cuantitativos de síntomas de la enfermedad que apoyen al diagnóstico de la enfermedad de Parkinson de forma no invasiva.

De esta forma, las evaluaciones cuantitativas periódicas sin necesidad de acudir a un hospital, podrían conducir a una mejor comprensión de los síntomas y progresión de la enfermedad de Parkinson, utilizando algoritmos de aprendizaje automático y proporcionando un beneficio único a las personas que viven con la enfermedad. Adicionalmente, el monitoreo puede proporcionar mayor información a los médicos e investigadores para el seguimiento de la evolución de la enfermedad.

## 1.5. Objetivo General

Desarrollar una metodología que permita monitorear a los individuos que manifiestan síntomas relacionados con la enfermedad de Parkinson, mediante la recopilación de datos por medio de un dispositivo móvil, con el propósito de evaluar la información obtenida utilizando técnicas de aprendizaje automático, para su interpretación por parte de un especialista médico.

## 1.6. Metas

- Definir una metodología que permita monitorear los síntomas relacionados a la enfermedad de Parkinson mediante los sensores de un dispositivo móvil.
- Desarrollar una aplicación móvil que sea capaz de monitorear cuantitativamente los síntomas relacionados a la enfermedad de Parkinson.
- Diseñar un algoritmo basado en técnicas de aprendizaje automático que permita clasificar los datos proporcionados por la aplicación móvil y distinguir a los pacientes con la enfermedad de Parkinson.

## 1.7. Aportaciones

- Una metodología para la estimación de la enfermedad de Parkinson, con base en información geográfica voluntaria.
- Una aplicación móvil para dispositivos con sistema operativo iOS.
- Un algoritmo que permite clasificar los datos proporcionados por la aplicación móvil y distinguir a los pacientes con la enfermedad de Parkinson.

## 1.8. Organización de la Tesis

El resto del documento está organizado de la siguiente manera: En el Capítulo 2 se presenta el estado del arte, resultado de una investigación que pretende presentar los trabajos relacionados con esta investigación, y que sirven como base para la motivación de esta tesis. En el Capítulo 3 se muestran

los fundamentos teóricos que enmarcan las posteriores aplicaciones; asimismo, se describen los conceptos relevantes para el presente trabajo. En el Capítulo 4 se describe la metodología propuesta para el análisis de datos fisiológicos relacionados con la enfermedad de Parkinson. En el Capítulo 5 se analizan y discuten los resultados obtenidos con los experimentos realizados. El Capítulo 6 presenta las conclusiones de este trabajo, así como propuestas de trabajos de investigación a futuro. Adicionalmente, se cuenta con un conjunto de anexos que presentan la modelación de la aplicación móvil.

# Capítulo 2

## Trabajos Relacionados

### 2.1. Aplicaciones móviles enfocadas en el cuidado de la salud

A medida que la población de mayor edad crece en todo el mundo, también lo hacen las enfermedades crónicas y la necesidad de una atención más adecuada y constante que requieren de los médicos y sus familiares. En todo el mundo, esta atención médica constante puede ser demasiado costosa o inaccesible. Para una vasta población con o sin enfermedades crónicas, la necesidad de una técnica de atención médica asequible para el monitoreo constante es primordial a través de un método completamente diferente de la atención médica tradicional centrada en el hospital [J. I. Pan y Huang (2015)].

En la actualidad, debido a que la capacidad del hardware de los teléfonos inteligentes y otros dispositivos móviles ha mejorado, un creciente número de funcionalidades de *e-health* se han puesto a disposición en plataformas móviles, por lo que el campo de las aplicaciones relacionadas a la salud es un tema que se encuentra en gran auge.

*Electronic-health (e-health)* Es un término utilizado en las aplicaciones de software relacionadas al cuidado de la salud que proporcionan herramientas, procesos y medios de comunicación para apoyar la práctica de la salud electrónica. [Yan y cols. (2010)].

## 2.1. APLICACIONES MÓVILES ENFOCADAS EN EL CUIDADO DE LA SALUD

---

*m-Health* es un término general que abarca las áreas de redes, informática móvil, sensores médicos y otras tecnologías dentro de la asistencia sanitaria. El concepto de *m-health* se refiere a informática móvil, sensores médicos, comunicaciones y tecnologías para el cuidado de la salud [Istepanian y cols. (2006)].

Los teléfonos inteligentes se han convertido en el equipo estándar de la vida diaria y ofrecen una interesante alternativa para el desarrollo de aplicaciones médicas *m-Health*, evitando así que los usuarios tengan que comprar hardware adicional. Además, las personas se sienten cómodas utilizando un teléfono inteligente que esté monitoreando sus actividades diarias, en lugar de un dispositivo más grande e incómodo, llamando la atención sobre su entorno social [Figueras y cols. (2015)]. Existen aplicaciones móviles centradas que ayudan a controlar las enfermedades crónicas, manejo del estilo de vida, dejar de fumar e incluso el autodiagnóstico [Boulos y cols. (2014)].

La epidemia de diabetes mellitus se refleja en la cantidad de aplicaciones dirigidas a pacientes diabéticos. Solo en la plataforma de Android, más de 80 aplicaciones de diabetes ofrecen una variedad de funciones, que incluyen autocontrol de la grabación de glucosa en sangre, medicación o registros de insulina y calculadoras de dosis de insulina [Demidowich y cols. (2012)]. Otra aplicación de intervención de diabetes integra la comunicación entre los pacientes y un proveedor de atención médica. El paciente registra el azúcar en la sangre en ayunas, sus hábitos alimentarios diarios, el cumplimiento de la medicación, la actividad física y sus emociones [Boulos y cols. (2014)].

En pacientes con enfermedades crónicas que amenazan la vida, las aplicaciones móviles pueden permitirles rastrear e incluso informar los síntomas. La aplicación *m.Carat* fue desarrollada en la Facultad de Medicina de la Universidad de Porto, Portugal, para pacientes con rinitis alérgica y asmática para registrar sus exacerbaciones, síntomas, medicamentos, pruebas de función pulmonar y visitas al médico o al hospital. Los usuarios también reciben información sobre enfermedades, medicamentos, notificaciones de tareas y es posible sincronizar registros con una base de datos en línea para el control de los síntomas [Burnay y cols. (2013)].

Una aplicación desarrollada para pacientes con demencia, *iWander*, ayuda a los pacientes en su vida diaria al ofrecer indicaciones audibles para

dirigir al paciente a casa, enviar notificaciones y coordenadas GPS a los cuidadores o llamar servicio de emergencia [Sposaro y cols. (2010)].

## 2.2. Aplicaciones móviles en la enfermedad de Parkinson

Algunos dispositivos pueden ser utilizados para medir los síntomas de Parkinson relacionados a anomalías en el ciclo de marcha así como en detectar temblores, un acelerómetro es un buen candidato. Éste es un dispositivo electrónico que puede medir aceleración dinámica (vibración) y estática (gravedad). Su arquitectura interna consiste en una barra móvil suspendida en unos micro-resortes que proveen resistencia en contra de la aceleración. La deflexión de la barra se convierte entonces en un medidor de aceleración [Resnick y cols. (2002)]. Tres acelerómetros pueden ser incorporados en un dispositivo único que provee información en los tres movimientos dimensionales (acelerómetro triaxial). Su diseño pequeño permite utilizarse en dispositivos electrónicos portables como son los smartphones.

[J. I. Pan y Huang (2015)], desarrollaron una aplicación móvil en Android con la finalidad de prevenir caídas en pacientes con Parkinson. La primera fase consiste en detectar y reconocer mediante un microcontrolador la postura encorvada y los cambios en el ciclo de la marcha que producen que el paciente se quede bloqueado «congelado» al caminar, pateando con muy breves pasos, con una especie de «tartamudeo» de la marcha sin poder avanzar. La siguiente fase se compone de una alarma o señal de audio para romper los síntomas de congelamiento del paciente, establece una red de sensores basada en acelerómetros que reportan información acerca de los movimientos.

Los datos registrados se transfieren un smartphone por medio del protocolo de comunicación Bluetooth. Las señales de entrada son reconocidas y clasificadas. Las principales ventajas de este enfoque propuesto incluyen seguridad para detectar la postura encorvada y la congelación de la marcha, así como el producir la señal de audio para ayudar a los pacientes a romper el estado y su portabilidad [J. I. Pan y Huang (2015)].

## 2.2. APLICACIONES MÓVILES EN LA ENFERMEDAD DE PARKINSON

---

[LeMoyne y cols. (2010)], proponen la implementación de una aplicación para iPhone que activa el dispositivo para servir como un sistema de acelerómetro inalámbrico funcional y detectar temblores en la enfermedad de Parkinson. Los datos pueden ser grabados en un lugar público o privado y la señal del acelerómetro resultante del temblor puede ser transmitida a través de correo electrónico a una ubicación remota para el post-procesamiento de datos. La prueba inicial y la evaluación de la aplicación iPhone reduce la presión sobre los recursos médicos limitados y altamente especializados.

[Figueras y cols. (2015)], proponen una aplicación móvil basada en una arquitectura Android que permite la caracterización cuantitativa de los temblores presentes en la enfermedad de Parkinson durante la ejecución de ciertos movimientos, para ser utilizado en el diagnóstico de la enfermedad. El proceso de adquisición no requiere la presencia de personal altamente calificado y permite el registro de los desplazamientos durante largos períodos de tiempo en ambientes del hogar en cualquier situación común.

[Kim y cols. (2015)], proponen un sistema basado en smartphones para detectar los síntomas de congelamiento en la marcha en una forma no restringida. La viabilidad para detectar el bloqueo de la marcha fue probada en diversas posiciones del cuerpo, tales como el tobillo, el bolsillo del pantalón, la cintura y bolsillo del pecho. Utilizando los datos medidos del acelerómetro y el giroscopio del smartphone, se aplicaron algoritmos de aprendizaje automático para clasificar episodios de bloque o congelamiento de la marcha normal.

[Arora y cols. (2016)], desarrollaron una app en Android compuesta por diferentes test (voz, postura, marcha, tapping, test de reacción) en un grupo de personas con EP y un grupo control. Un grupo de diez participantes sanos y 10 participantes con enfermedad de Parkinson, realizaron las diferentes pruebas cinco veces al día, durante un mes. Una vez a la semana un especialista realizó una encuesta UPRDS, utilizando análisis estadístico se discriminó si el participante tenía diagnóstico de Parkinson. El sistema mostró una sensibilidad del 96.2% y una especificidad del 96% para distinguir entre los participantes sanos y aquellos con EP.

[Broen y cols. (2016)], evaluaron la capacidad de un smartphone para llevar a cabo un método de recogida de datos (ESM). Su método estudia

la relación compleja entre síntomas motores, estados afectivos y factores contextuales. Al usar un dispositivo móvil con una aplicación de propósito especial (aplicación), los síntomas motores, los estados afectivos y los factores contextuales se evalúan repetidamente en momentos aleatorios del flujo de la vida diaria, produciendo una serie de síntomas y experiencia intensiva en el tiempo. El objetivo de este estudio fue estudiar la viabilidad de este método. El método de recogida de datos mostró capacidad para detectar fluctuaciones motoras durante el día mediante los datos obtenidos de las preguntas.

[Ellis y cols. (2015)], evaluaron un sistema de valoración y tratamiento de la marcha mediante una aplicación móvil para iOS, en comparación con las mediciones obtenidas mediante sensores y mediante un tapiz, en diferentes situaciones. Se apreciaron diferencias significativas entre grupos de personas con EP y grupo de personas sanas en los parámetros de la marcha en todas las herramientas de medición.

[D. Pan y cols. (2015)], diseñaron, desarrollaron y evaluaron un prototipo de aplicación *mHealth* basada en la nube y denominada "PD Dr", la cual recopila información cuantitativa y objetiva sobre la EP que permitiría la evaluación en el hogar y la monitorización de los principales síntomas de la enfermedad de Parkinson. Por medio de una aplicación móvil en la plataforma Android se recopilan datos de movimiento relacionados con la EP, utilizando el acelerómetro 3D del smartphone y se envían los datos a un servicio en la nube. Para evaluar este sistema, se recogieron datos del sistema de 40 pacientes con EP y se compararon con la calificación de los expertos en escalas de calificación estandarizadas UPRDS. La evaluación mostró que "PD Dr" podría capturar de manera efectiva las características de movimiento importantes que diferencian la gravedad de la EP e identifica los síntomas críticos.

[Printy y cols. (2014)], propusieron una aplicación móvil en Android para la valoración de la bradicinesia en sujetos con EP, en comparación con la puntuación del aspecto motor de la UPDRS. La aplicación fue capaz de clasificar los sujetos con EP según la afectación en leve o severa con una precisión del 94,5 %.

[Mandal y Sairam (2014)], presentaron un modelo de clasificación para el

## *2.2. APLICACIONES MÓVILES EN LA ENFERMEDAD DE PARKINSON*

---

diagnóstico de la enfermedad de Parkinson (EP), utilizando un conjunto de datos de voz. Los métodos utilizados incluyen la regresión logística multinomial dispersa, árboles de decisión y máquinas de soporte vectorial combinado con análisis de componentes principales y redes neuronales artificiales. La precisión más alta obtenida por regresión logística lineal y regresión logística multinomial dispersa fue de 100 % y la sensibilidad, especificidad de 0.983 y 0.996, respectivamente.

### **2.2.1. Minería de datos en diagnóstico de Enfermedad de Parkinson.**

Se han realizado diversos trabajos relacionados con minería de datos en el diagnóstico de la Enfermedad de Parkinson, algunos de ellos incluyen el monitoreo de diferentes conjuntos de datos entre los que se encuentran, registros de voz, registros de caminata o de actividades de destreza manual. [Ramani y Sivagami (2011)] analizaron un conjunto de datos compuesto por registros de voz de 31 personas con 23 características. A este conjunto de datos de registros de voz se lleva a cabo un estudio comparativo de diferentes métodos de clasificación mediante la aplicación del análisis de relevancia de características y Análisis de precisión para obtener la mejor regla de clasificación con la finalidad es clasificar participantes con diagnóstico de enfermedad de Parkinson, se obtiene una exactitud global del 100 % utilizando árboles de decisión.

[D. Gil y Johnsson (2009)] evaluaron el rendimiento de un clasificador para registros de voz construido mediante ANN y SVM. Utilizando Perceptrones Multicapa (MLP) y SVM con dos tipos de kernel, obtienen un nivel de precisión alta de la matriz de confusión con respecto a los diferentes parámetros de medición como precisión, sensibilidad, especificidad. Concluyeron mostrando un alto grado de certeza superior al 90 %.

[Salvatore y cols. (2014)] proponen un algoritmo de aprendizaje automático supervisado basado en Análisis de Componentes Principales como técnica de extracción de características y Máquinas de Vector de Apoyo para predecir el diagnóstico de la enfermedad de Parkinson (PD) y la parálisis supranuclear progresiva (PSP) para imágenes de resonancia magnética (conjunto de datos de MRI). Predicción de la enfermedad de Parkinson (PD) versus controles, parálisis supranuclear progresiva (PSP). La exactitud global

(especificidad / sensibilidad) fue de 83.2 % para grupos etiquetados binarios, respectivamente.

[Butt y cols. (2017)] utilizaron sensores inerciales colocados en los pies de los pacientes para el diagnóstico de Enfermedad de Parkinson. Los resultados entrenaron en máquina de vectores de soporte, regresión logística y red neuronal para clasificar los dos grupos de pacientes. La mayor precisión de clasificación obtenida fue utilizando máquinas de soporte vectorial con 79.66 % y AUC= 0.8790. Se obtuvo una precisión de clasificación del 76.2 % AUC= 0.7832 mediante regresión logística y utilizando redes neuronales se obtuvo una precisión del 83.10 % AUC= 0.889.

[Saad y cols. (2013)] propusieron una *Bayesian Belief Network (BBN)* para encontrar congelación de pacientes con enfermedad de Parkinson y utilizaron un conjunto de datos de video disponible en línea extraído de pacientes con enfermedad de Parkinson. Cada archivo estaba compuesto por una matriz de tres sensores en las direcciones  $x$ ,  $y$  y  $z$ . Los datos a su vez estaban etiquetados de forma que registraron el congelamiento de la marcha. Estas anotaciones se etiquetaron sincronizando los datos mediante un video que registraba cada ejecución del paciente y los resultados aparecían al probar el clasificador *Bayesian Naïve Classifier (BNC)*.

En el Centro de Investigación en Computación del Instituto Politécnico Nacional se han realizado algunos trabajos relacionados con el diagnóstico de la enfermedad de Parkinson, [GARCÍA (2016)] propone un modelo basado en lógica difusa capaz de cuantificar el estado de una persona con dicha enfermedad, aplicando conceptos de cinemática a las señales obtenidas mediante Sensores de Movimiento Inercial (IMU por sus siglas en inglés) y calculando atributos particulares como la amplitud o frecuencia de sus temblores durante la exploración motora. [Rodríguez (2017)] propone un modelo computacional para cuantificar de manera precisa y objetiva ocho características biomecánicas utilizando técnicas avanzadas de procesamiento digital de señales, las cuales son obtenidas por medio de sensores inerciales de pacientes con Parkinson durante los ejercicios de pronación/supinación de ambas manos.

## 2.3. Aplicaciones comerciales

[del Rey y cols. (2017)], realizaron una revisión en los principales mercados de aplicaciones móviles, seleccionando 103 aplicaciones relacionadas con la EP. De ellas, 49 se encontraron para dispositivos Apple y 39 para dispositivos basados en Android y 6 para dispositivos con Windows Phone. A su vez, 11 aplicaciones se encontraron disponibles tanto en dispositivos Android como en Apple y una en los tres mercados.

Respecto al precio de las mismas, 74 apps eran gratuitas y 29 de pago. Teniendo en cuenta el tipo de usuario al que estaban dirigidas, 45 de ellas estaban pensadas para su uso por pacientes de EP, 34 para el uso por profesionales (sin presentar algún tipo de filtro o identificación que así lo constatará, a modo de limitación para los no profesionales) y 24 para su uso tanto por profesionales como por pacientes.

En cuanto a su clasificación, se dividieron en las siguientes categorías: un total de 54 aplicaciones se considerarían con utilidad en la EP, aunque no estén específicamente diseñadas para esta enfermedad y 49 serían específicas para la EP. En el primer grupo encontramos 23 apps que ofrecen información sobre diferentes enfermedades neurológicas, 15 aplicaciones diseñadas para la mejora del habla, 7 centradas en pruebas para analizar la función de miembro superior, 2 basadas en el automanejo de la medicación y monitorización de los síntomas y 2 que permitían utilizar el teléfono para administrar diferentes escalas utilizadas en neurología.

Las restantes se basaban en mejorar la facilidad de uso del teléfono, mejorar el manejo de la salivación, realizar ejercicios cognitivos, llevar a cabo un seguimiento de pacientes y servir como herramienta en el diseño mediante impresión en 3 dimensiones de objetos útiles para personas con enfermedades neurológicas.

Se debe destacar que grandes empresas como la compañía Apple, con la aplicación *mPower* y la farmacéutica Roche también están llevando cabo proyectos que incluyen apps para la valoración de la EP. Todo ello apunta a que en los próximos años la evidencia acerca de este tipo de herramientas será mucho mayor y, por lo tanto, será necesario un análisis de la misma para confirmar la utilidad de las apps [del Rey y cols. (2017)].

### 2.3.1. Aplicación Móvil mPower

La aplicación móvil **mPower** desarrollada para iOS, recoge datos en tiempo real de los pacientes de Parkinson, en un esfuerzo por comprender mejor la enfermedad y cómo afecta la vida diaria. Permite a los pacientes realizar un seguimiento de cómo los síntomas y el tratamiento está impactando en la progresión de la enfermedad. A través del micrófono, la aplicación puede medir el estado general de la enfermedad mediante el análisis de los cambios sutiles en la voz de los pacientes. Además, utiliza otras funciones del iPhone como la pantalla táctil y los sensores de movimiento para medir destreza, equilibrio y patrón de marcha y memoria [Bot y cols. (2016)].

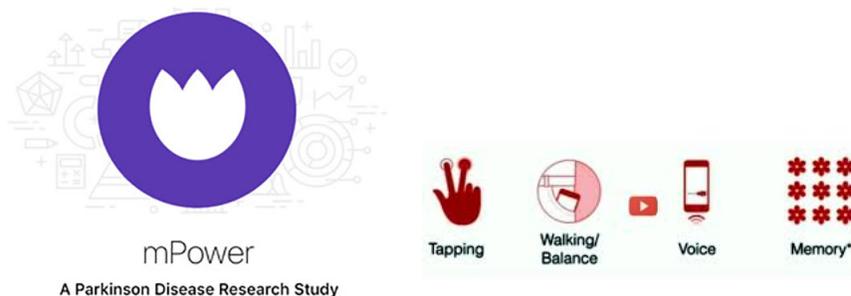


Figura 2.1: Aplicación móvil mPower

### 2.3.2. Aplicación Móvil Roche App Parkinson's Disease

La aplicación ha sido desarrollada para Android y está siendo utilizada en un ensayo de *Fase I* a cargo de Prothena en colaboración con Roche [del Rey y cols. (2017)]. Esta permite la medición continua de la fluctuación de la enfermedad de Parkinson cada día y durante todo el día, las pruebas que se permiten realizar en la aplicación se muestran en la (Figura 2.2).

La aplicación solicita a los pacientes seguir una rutina diaria, la cual consiste en seis pruebas activas seguidas de vigilancia pasiva. Estas evaluaciones están diseñadas para proporcionar información sobre los síntomas de una persona, su progreso y su impacto en la vida diaria.

### 2.3. APLICACIONES COMERCIALES

---

Los datos recogidos se utilizan para predecir la UPDRS, *Unified Parkinson's Disease Rating Scale*, (Escala unificada para la evaluación de la Enfermedad de Parkinson). Las puntuaciones y fluctuaciones de los síntomas durante el estudio se registran en la aplicación.

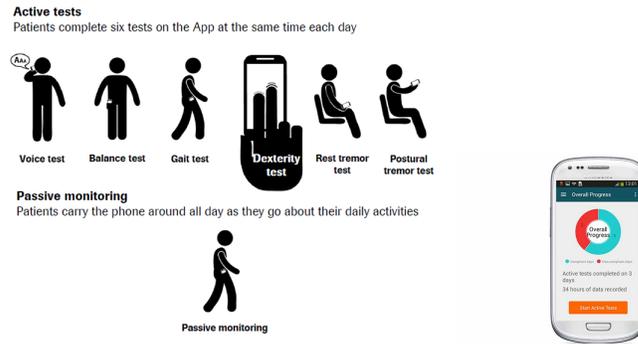


Figura 2.2: Aplicación móvil Roche

# Capítulo 3

## Marco Teórico

### 3.1. Introducción

La enfermedad de Parkinson (EP) afecta a 10 millones de personas en todo el mundo. Los afectados por la enfermedad se encuentran en el rango de 40 a 60 años de edad. Sin embargo, otras fuentes han confirmado que incluso los jóvenes de 20 a 30 años, también pueden estar en riesgo de desarrollar síntomas de la enfermedad. Esto tiene que ver con la genética involucrada durante el crecimiento. Los tratamientos para la EP todavía están bajo investigación y como no se ha establecido aún ningún tratamiento completo. Las personas con EP desarrollan dolores crónicos que impiden el movimiento normal y la funcionalidad del cuerpo [Foundation (2016)].

La enfermedad de Parkinson (EP) es un trastorno crónico y progresivo del movimiento, lo que significa que los síntomas continúan y empeoran con el tiempo. La causa es desconocida, y aunque actualmente no hay cura, existen opciones de tratamiento como la medicación y la cirugía para controlar los síntomas [Eskofier y cols. (2016)].

El Parkinson implica el mal funcionamiento y muerte de las células nerviosas vitales en el cerebro, llamadas *neuronas*. Afecta principalmente neuronas en un área del cerebro llamada *sustancia negra*. Algunas de estas neuronas moribundas producen dopamina, una sustancia química que envía mensajes a la parte del cerebro que controla el movimiento y la coordinación. A medida que avanza la enfermedad, la cantidad de dopamina producida en

### 3.2. SÍNTOMAS DE LA ENFERMEDAD DE PARKINSON Y DIAGNÓSTICO

---

el cerebro disminuye, dejando a una persona incapaz de controlar el movimiento normalmente [Gröger y cols. (2014)].

## 3.2. Síntomas de la enfermedad de Parkinson y diagnóstico

Aunque la etiología de la enfermedad de Parkinson aún no es muy clara, las recientes investigaciones coinciden en que la edad, susceptibilidad genética, la exposición a toxinas ambientales entre otros factores, pueden ser causas de este padecimiento [DeMaagd y Philip (2015)].

Los síntomas que experimentan las personas con EP varían de persona a persona. Los signos motores primarios de la enfermedad de Parkinson incluyen los siguientes:

- Temblor de manos, brazos, piernas, mandíbula y cara.
- Bradicinesia o lentitud de movimiento.
- Rigidez de los miembros y el tronco.
- Inestabilidad postural o alteración del equilibrio y coordinación. [Foundation (2016)]

Actualmente no existen pruebas de sangre o de laboratorio que diagnostiquen la enfermedad de Parkinson. Por ello, el diagnóstico se basa en la historia clínica y en un examen neurológico, normalmente siguiendo la escala *UPRDS*. En algunos casos, la enfermedad de Parkinson puede ser difícil de diagnosticar con precisión al inicio de la enfermedad. Los signos y síntomas tempranos de la enfermedad de Parkinson a veces se pueden descartar como los efectos del envejecimiento normal. En apoyo al diagnóstico de la enfermedad, se realizan estudios de imágenes cerebrales o pruebas de laboratorio con el fin de descartar otros trastornos. Sin embargo, las exploraciones cerebrales por tomografía computarizada e imágenes por resonancia magnética de las personas con la enfermedad de Parkinson, generalmente parecen normales, ya que muchas otras enfermedades tienen características similares pero requieren tratamientos diferentes por lo cual es fundamental

hacer un diagnóstico preciso para que las personas puedan recibir el tratamiento adecuado [DeMaagd y Philip (2015)] [Goetz y cols. (2008)] [Trister y cols. (2016)].

### 3.2.1. Escala unificada para la evaluación de la enfermedad de Parkinson

La escala unificada para la evaluación de la enfermedad de Parkinson (*Unified Parkinson's Disease Rating Scale*), abreviada UPDRS es una escala de estratificación usada para seguir el avance de la Enfermedad de Parkinson [Goetz y cols. (2008)]. Está compuesta por las siguientes secciones:

#### Parte I: Estado mental, de comportamiento y humor

- Deterioro intelectual.
- Trastornos del pensamiento.
- Depresión.
- Motivación, iniciativa.

#### Parte II: Actividades de la vida diaria

- Habla.
- Salivación.
- Deglución.
- Escritura.
- Cortar comida y sostener cubiertos.
- Vestirse.
- Higiene.
- Rodar sobre la cama.
- Caídas (Sin relación a la congelación).
- Congelación al caminar.
- Marcha.
- Temblor.
- Quejas sensoriales relacionadas con la enfermedad.

### *3.2. SÍNTOMAS DE LA ENFERMEDAD DE PARKINSON Y DIAGNÓSTICO*

---

#### **Parte III: Examen motor (por el médico)**

- Habla.
- Expresión facial.
- Temblor en reposo.
- Temblor postural de las manos.
- Rigidez.
- Golpeteo de los dedos (El paciente golpetea el pulgar con el índice en sucesión rápida).
- Movimiento de las manos (El paciente abre y cierra las manos rápida y constantemente).
- Movimientos rápidos alternados de manos (Pronación-Supinación, vertical-horizontal).
- Agilidad en las piernas.
- Levantarse de una silla.
- Postura.
- Manera de caminar.
- Estabilidad postural.
- Bradicinesia e hipocinesia corporal.

#### **Parte IV: Complicaciones de la terapia (en la semana pasada)**

- Discinesias.
- Fluctuaciones clínicas.
- Otras complicaciones.

#### **Parte V: Escala modificada de Hoehn y Yahir**

- Severidad de la enfermedad.

#### **Parte VI: Escala de Schwab y England**

- Evaluación de actividades cotidianas. Esos puntos son evaluados mediante la entrevista y la exploración física.

Médicos e investigadores utilizan la escala UPDRS, en especial la sección de evaluación para seguir la progresión de la enfermedad de Parkinson en una persona. Los investigadores clínicos lo usan para medir los beneficios de una terapia en un sistema de clasificación más unificado y aceptado. Los neurólogos también lo utilizan en la práctica clínica para seguir la progresión de los síntomas de sus pacientes de una manera más objetiva.

Existen otras escalas de evaluación para la enfermedad de Parkinson como la escala de Hoehn y Yahr y la escala de Schwab y England, aunque ambas medidas están incluidas dentro de la UPDRS en formato modificado [Parkinson y Society (2016)].

### 3.3. Plataforma Research Kit Apple



Figura 3.1: Plataforma Research Kit.

*ResearchKit* es un proyecto de código abierto para la colaboración global con la investigación médica, el cual es dispuesto por Apple. El mayor desafío que enfrentan los investigadores médicos es reclutar participantes para un estudio; a menudo, sólo se consiguen unas cuantas personas. Pero gracias al gran número de usuarios de iPhone en todo el mundo, las *apps* basadas en *ResearchKit* pueden inscribir participantes y reunir datos en una mayor proporción. La información puede recopilarse de forma mucho más sistemática, cada día o incluso cada hora, mientras las personas siguen con su vida habitual. El resultado es un grupo de estudio mucho más grande y variado, datos más frecuentes y una muestra más precisa de la población [Inc. (2017b)][Bot y cols. (2016)].

La arquitectura de código abierto permite a los investigadores aprovechar los módulos iniciales de *ResearchKit* para estudiar la salud y el bienestar y tener una mejor comprensión de la enfermedad. A partir del *framework* es posible crear nuevos módulos basados en el código de fuente abierta y contribuir a *ResearchKit*. Los módulos iniciales abordan los elementos más comunes que se encuentran en los estudios de investigación, el consentimiento del participante, las encuestas y las tareas activas [MV y cols. (2017)].

El *consentimiento del participante*: Es un elemento fundamental para los estudios de investigación, se trata de una plantilla de consentimiento electrónico visual que se puede adaptar para explicar los detalles del estudio y obtener las firmas de los participantes. Este módulo incluye elementos como segmentos de vídeo que explican el estudio y un cuestionario interactivo para confirmar que el participante entiende de qué se trata (<https://github.com/ResearchKit/AppCore>).

*Encuestas*: Este módulo proporciona una interfaz de usuario que facilita la personalización de las preguntas y respuestas de manera que los participantes del estudio las completen y compartan inmediatamente con los investigadores.

*Tareas activas*: El módulo de tareas activas permite a los investigadores obtener datos más específicos para su estudio invitando a los participantes a realizar actividades que generen datos, utilizando los sensores avanzados del iPhone. Los módulos de tareas activas iniciales incluyen tareas para medir las actividades motoras, estado físico, la cognición y la voz. La plataforma de código abierto permite compartir patrones de diseño, contenidos éticos y buenas prácticas en el desarrollo de aplicaciones móviles para el cuidado de la salud.

Cuando el participante concede el permiso, las aplicaciones de *ResearchKit* pueden acceder a la información obtenida de los sensores del iPhone como el acelerómetro, giróscopio, micrófono y GPS para conocer más detalladamente los niveles de actividad del participante, deficiencias motoras, memoria y más [Inc. (2017b)].

*ResearchKit* funciona también con la plataforma *HealthKit*, la cual es un marco de trabajo basado en software que sirve para proporcionar a los

programadores la posibilidad de comunicar entre sí las aplicaciones de salud y de aptitud física. Con el permiso del participante, las aplicaciones de *ResearchKit* pueden acceder a los datos de la Aplicación de Salud como son el peso, presión arterial, los niveles de glucosa y el uso de inhaladores de asma, que se miden con dispositivos de terceros y aplicaciones, para ser utilizados [Inc. (2017a)].

### 3.4. Algoritmo SMOTE Synthetic Minority Over-sampling TEchnique

La técnica de Synthetic Minority Oversampling aumenta el número de casos pertenecientes a la minoría en su conjunto de datos, mediante una técnica estadística que construye nuevos casos de las características existentes, en lugar de simplemente replicar los casos existentes [Nitesh V. Chawla (2002)]

Normalmente se utiliza SMOTE en conjuntos de datos que están desequilibrados, lo que significa que la clase que desea analizar está subrepresentada. La técnica de muestreo de minorías sintéticas fue desarrollada por investigadores de aprendizaje automático, como una forma de aumentar el número de casos raros sin aumentar el sesgo. El módulo devuelve un conjunto de datos que contiene los casos originales, más un número adicional de casos de minorías sintéticas [Nitesh V. Chawla (2002)][Provost y Fawcett (2001)].

La técnica funciona generando nuevas instancias de los casos existentes que suministra como entrada. Las nuevas instancias se crean tomando muestras del espacio de características para cada clase de destino y sus vecinos más cercanos, así como generando nuevos ejemplos que combinan las características del caso de destino con las características de sus vecinos. Este enfoque aumenta las características disponibles para cada clase y hace que las muestras sean más generales. SMOTE toma todo el conjunto de datos como una entrada y aumenta el porcentaje de casos minoritarios solamente.

En [Algorithm 1] se puede encontrar el pseudocódigo que permite generar muestras sintéticas basadas en la Técnica SMOTE.

Ejemplo de generación de muestras sintéticas con SMOTE

Se considera una muestra  $(6, 4)$  y sea  $(4, 3)$  su vecino más cercano.  $(6, 4)$  es la muestra para la cual se están identificando  $k$ -vecinos más cercanos y  $(4, 3)$  es uno de los vecinos más cercanos a  $k$ .

$$\begin{aligned} f_{11} &= 6 & f_{21} &= 4 & f_{21} - f_{11} &= -2 \\ f_{12} &= 4 & f_{22} &= 3 & f_{22} - f_{12} &= -1 \end{aligned} \tag{3.1}$$

Las muestras nuevas se generarían como sigue:

$$(f1', f2') = (6, 4) + rand(0, 1) * (-2, -1) \tag{3.2}$$

$rand(0, 1)$  Genera un número aleatorio entre 0 y 1. [Nitesh V. Chawla (2002)]

## 3.5. Aprendizaje de Máquina

El Aprendizaje de Máquina (ML) se ha utilizado con éxito en el campo médico, por ejemplo en la clasificación de arritmias utilizando datos de electrocardiogramas o regresión de variables clínicas utilizando datos de imágenes médicas [Guvenir y cols. (1997)]. Las ventajas del ML en estas y otras aplicaciones son los análisis objetivos, basados en datos independientes de la situación [Ying Wang (2009)]. Particularmente, ML también se ha aplicado en el análisis de sensores inerciales (IMU) para la evaluación automática de los trastornos del movimiento como la enfermedad de Parkinson [Eskofier y cols. (2016)].

### 3.5.1. Regresión Logística

En estadística, la regresión logística es un tipo de análisis de regresión, utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras [Hosmer (2000)]. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de los Modelos Lineales Generalizados (GLM por sus siglas en inglés) que usa como función de enlace la función *logit* [J y WJ (2016)]. Las probabilidades que describen el posible resultado de un único ensayo se modelan como una función de

variables explicativas, utilizando una función logística.

La regresión logística es usada extensamente en las ciencias médicas y sociales. Otros nombres para regresión logística utilizado en varias áreas de aplicación incluyen el modelo logístico, modelo *logit*, y el clasificador de máxima entropía [Vinuesa (2017)].

La regresión logística analiza datos distribuidos binomialmente de la forma siguiente:

$$Y_i \sim B(p_i, n_i), \text{ para } i = 1, \dots, m, \quad (3.3)$$

Donde los números de ensayos de *Bernoulli* no son conocidos y las probabilidades de éxito  $p_i$  son desconocidas. Un ejemplo de esta distribución es el porcentaje de semillas ( $p_i$ ) que germinan después de que no son plantadas [Hosmer (2000)].

El modelo es entonces obtenido con base en que cada ensayo (con valor de  $i$ ) y el conjunto de variables explicativas/independientes puedan informar acerca de la probabilidad final. Estas variables explicativas pueden pensarse como un vector  $X_i$  – *dimensional* y el modelo toma entonces la forma siguiente:

$$p_i = E\left(\frac{Y_i}{n_i} \middle| X_i\right). \quad (3.4)$$

Los *logits* de las probabilidades binomiales desconocidas (i.e., los logaritmos de la razón de momios) son modeladas como una función lineal de los  $X_i$ .

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}. \quad (3.5)$$

Cabe señalar que un elemento particular de  $X_i$  puede ser ajustado a 1 para toda  $i$ , obteniéndose una constante independiente en el modelo. Los parámetros desconocidos  $j$  son usualmente estimados a través de máxima verosimilitud [Vinuesa (2017)].

La interpretación de los estimados del parámetro  $j$ , es como los efectos aditivos en el logaritmo de la razón de momios para una unidad de cambio

en la  $j$ -ésima variable explicativa. En el caso de una variable explicativa dicotómica, por ejemplo género,  $e^\beta$  es la estimación de la razón de momios (*odds ratio*) de tener el resultado. Por ejemplo, hombres comparados con mujeres. El modelo tiene una formulación equivalente dada por:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}} \quad (3.6)$$

Esta forma funcional es comúnmente identificada como un *perceptrón* de una capa simple o red neuronal artificial de una sola capa. Una red neuronal de una sola capa calcula una salida continua en lugar de una función definida a trozos. La derivada de  $p_i$  con respecto a  $X = x_1 \dots x_k$  se calcula de la forma general:

$$y = \frac{1}{1 + e^{-f(X)}} \quad (3.7)$$

donde  $f(X)$  es una función analítica en  $X$ . Con esta se elige, la red de capa simple que es idéntica al modelo de regresión logística. Esta función tiene una derivada continua, la cual permite ser usada en propagación hacia atrás. Esta función también es preferida, ya su derivada se calcula de una manera más simple:

$$y' = y(1 - y) \frac{df}{dX} \quad (3.8)$$

### 3.5.2. Árbol de decisión (modelo de clasificación ID3)

El algoritmo ID3 se apoya en técnicas matemáticas y probabilísticas; introduce el concepto de entropía, la cual es una medida de incertidumbre o de desorden, y es usado para ayudar a decidir qué atributo debe ser el siguiente en seleccionarse. En general, un atributo que puede ayudar a discriminar más objetos, tiende a reducir más la entropía, y por tal motivo, debe ser seleccionado como un nodo de prueba o de selección para la siguiente subdivisión [Quinlan (1986)].

Para calcular la entropía de  $n$  clases se utiliza la fórmula siguiente:

$$\text{Entropia}(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (3.9)$$

Donde:

$S$ : es una colección de objetos.

$P_i$ : es la probabilidad de los posibles valores.

$i$ : las posibles respuestas de los objetos.

Para una muestra homogénea la entropía es igual a cero. Por el contrario, si la distribución de objetos que pertenecen a una clase es la misma, entonces la entropía es igual a uno, o sea es la máxima incertidumbre.

Si en el cálculo de la entropía se presentara una proporción cero, entonces se tomará la expresión  $\log_2$  como igual a cero.

También se introduce el concepto de ganancia de información, ésta es una medida de discriminación, un indicador del siguiente atributo a ser seleccionado para continuar con el proceso de división, discriminando el atributo seleccionado entre los demás atributos aún no clasificados y se calcula utilizando la fórmula siguiente:

$$\text{Gan Inf}(S, A) = \text{Entropia}(S) - \sum_{v \in V(A)} \frac{|Sv|}{|S|} \text{Entropia}(Sv) \quad (3.10)$$

Donde:

$S$ : es una colección de objetos.

$A$ : son los atributos de los objetos.

$V(A)$ : Conjunto de valores que  $A$  puede tomar.

Ross Quilian [Quinlan (1986)] define el esquema del algoritmo ID3, el cual consiste de los siguientes pasos:

- 1. Calcular la entropía para todas las clases.
- 2. Seleccionar el mejor atributo basado en la reducción de la entropía.
- 3. Iterar hasta que todos los objetos sean clasificados.

Diversos problemas pueden ser modelados y resueltos eficientemente con el uso de los árboles de decisión, para ello dichos problemas deben satisfacer las siguientes características [Rokach y Maimon (2008)].

- La descripción de los objetos debe efectuarse en términos de parejas, el atributo y su valor.
- Cada salida debe ser distinta, es decir, los nodos terminales deben contener una respuesta única.
- Debe haber alternativas entre 2 o más cosas, para hacer posible la clasificación.
- El entrenamiento debe estar libre de errores, es decir, el problema a clasificar debe contener la información correcta, para que el resultado sea eficiente.
- El entrenamiento debe contener información suficiente para poder hacer la clasificación.

Cabe mencionar que este modelo permite predecir qué actividad se tomará sobre la base de la información que se tiene, además posee una gran exactitud en su aplicación. [Rokach y Maimon (2014)]

#### 3.5.3. Redes Neuronales Artificiales

Una red neuronal se compone de unidades llamadas *neuronas*. Cada neurona recibe una serie de entradas a través de interconexiones y emite una salida. Esta salida viene dada por tres funciones principales [Zell (2014)]

1. Una función de propagación (también conocida como función de excitación), que por lo general consiste en la sumatoria de cada entrada multiplicada por el peso de su interconexión (valor neto). Si el peso es positivo, la conexión se denomina *excitatoria*; si es negativo, se denomina *inhibitoria*.
2. Una función de activación que modifica a la anterior. Puede no existir, siendo en este caso la salida la misma función de propagación.
3. Una función de transferencia que se aplica al valor devuelto por la función de activación. Se utiliza para acotar la salida de la neurona y generalmente viene dada por la interpretación que se quiere dar a dichas salidas. Algunas de las más utilizadas son la función *sigmoidea* y *tangente hiperbólica*.

Con un paradigma convencional de programación en Ingeniería de Software, el objetivo del programador es modelar matemáticamente (con distintos grados de formalismo) el problema en cuestión y posteriormente formular una solución (programa) mediante un algoritmo codificado que tenga una serie de propiedades que permitan resolver dicho problema. En contraposición, la aproximación basada en las RNA parte de un conjunto de datos de entrada suficientemente significativo y el objetivo es conseguir que la red aprenda automáticamente las propiedades deseadas [Zell (2014)].

Por ejemplo en una red que se aplica al diagnóstico de imágenes médicas; durante la fase de entrenamiento el sistema recibe imágenes de tejidos que se sabe que son cancerígenos y tejidos que se sabe que son sanos, así como las respectivas clasificaciones de dichas imágenes. Si el entrenamiento es el adecuado, una vez concluido, el sistema podrá recibir imágenes de tejidos no clasificados y obtener su clasificación sano/no sano con un buen grado de seguridad. Las variables de entrada pueden ser desde los puntos individuales de cada imagen hasta un vector de características de las mismas que se puedan incorporar al sistema (por ejemplo, procedencia anatómica del tejido de la imagen o la edad del paciente al que se le extrajo la muestra) [N. Sáenz Bajoa (2012)][Ahmed (2005)].

## Estructura

La mayoría de los científicos coinciden en que una RNA es muy diferente en términos de estructura de un cerebro animal. Al igual que el cerebro, una RNA se compone de un conjunto masivamente paralelo de unidades de proceso muy simples y es en las conexiones entre estas unidades donde reside la inteligencia de la red. Sin embargo, en términos de escala, un cerebro es muchísimo mayor que cualquier RNA creada hasta la actualidad y las neuronas artificiales también son más simples que su contrapartida animal [Zell (2014)].

Biológicamente, un cerebro aprende mediante la reorganización de las conexiones sinápticas entre las neuronas que lo componen. De la misma manera, las RNA tienen un gran número de procesadores virtuales interconectados que de forma simplificada simulan la funcionalidad de las neuronas biológicas. En esta simulación, la reorganización de las conexiones sinápti-

cas biológicas se modelan, mediante un mecanismo de pesos, que son ajustados durante la fase de aprendizaje. En una RNA entrenada, el conjunto de los pesos determina el conocimiento de esa RNA y tiene la propiedad de resolver el problema para el que la RNA ha sido entrenada. [Ahmed (2005)][N. Sáenz Bajoa (2012)].

Por otra parte, en una RNA además de los pesos y las conexiones, cada neurona tiene asociada una función matemática denominada *función de transferencia*. Dicha función genera la señal de salida de la neurona a partir de las señales de entrada. La entrada de la función es la suma de todas las señales de entrada por el peso asociado a la conexión de entrada de la señal. Algunos ejemplos de entradas son la función *Escalón de Heaviside*, la *lineal* o *mixta*, la *sigmoide* y la *función Gaussiana*, recordando que la función de transferencia es la relación entre la señal de salida y la de entrada [N. Sáenz Bajoa (2012)].

#### Ventajas

Las redes neuronales artificiales (RNA) tienen muchas ventajas debido a que están basadas en la estructura del sistema nervioso, principalmente el cerebro.

- *Aprendizaje*: Las RNA tienen la habilidad de aprender mediante una etapa que se llama etapa de aprendizaje. Esta etapa consiste en proporcionar a la RNA datos como entrada a su vez que se le indica cuál es la salida (respuesta) esperada.
- *Auto-organización*: Una RNA crea su propia representación de la información en su interior, quitándole esta tarea al usuario.
- *Tolerancia a fallos*: Debido a que una RNA almacena la información de forma redundante, ésta puede seguir respondiendo de manera aceptable aún si se daña parcialmente.
- *Flexibilidad*: Una RNA puede manejar cambios no importantes en la información de entrada, como señales con ruido u otros cambios en la entrada (por ejemplo, si la información de entrada es la imagen de

un objeto, la respuesta correspondiente no sufre cambios si la imagen cambia un poco su brillo o el objeto cambia ligeramente).

- *Tiempo real*: La estructura de una RNA es paralela, por lo cual si esto es implementado con computadoras o en dispositivos electrónicos especiales, se pueden obtener respuestas en tiempo real. [Farley y Clark (1954)].

### Algoritmos de Aprendizaje

El entrenamiento de un modelo de red neuronal significa esencialmente seleccionar un modelo del conjunto de modelos permitidos (o, en un marco bayesiano, determinar una distribución sobre el conjunto de modelos permitidos) y que minimice el criterio de costo [ R. V. Gil y Páez (2007)]. Existen numerosos algoritmos disponibles para la formación de modelos de redes neuronales. La mayoría de ellos pueden ser vistos como una aplicación directa de la teoría de la optimización y la estimación estadística [Pineda (2017)].

La mayoría de los algoritmos utilizados en el entrenamiento de redes neuronales artificiales emplean alguna forma de descenso de gradiente, usando retropropagación para calcular los gradientes reales. Esto se realiza simplemente tomando la derivada de la función de costo con respecto a los parámetros de red y luego cambiando esos parámetros en una dirección relacionada con el gradiente. Los algoritmos de entrenamiento de *backpropagation* se clasifican generalmente en tres categorías:

1. *Steepest Descent* (con tasa de aprendizaje variable, con tasa de aprendizaje variable y momentum, backpropagation resiliente);
2. *Cuasi-Newton* (*Broyden-Fletcher-Goldfarb-Shanno*, un paso secante);
3. *Levenberg-Marquardt y gradiente conjugado* (actualización de Fletcher-Reeves, actualización de Polak-Ribière, reinicio de Powell-Beale, gradiente conjugado escalado).

Los métodos evolutivos, la programación de expresión genética, el recocido simulado, la optimización de la expectativa, los métodos no paramétricos y la optimización del enjambre de partículas son otros métodos para entrenar redes neuronales Forouzanfar y cols. (2010).

## 3.6. F1 Score

El Valor F1-score o medida-F en estadística es la medida de precisión que tiene un test o prueba. Se emplea en la determinación de un valor único ponderado de la precisión y la exhaustividad. Se suele emplear en la fase de pruebas de algoritmos de búsqueda y recuperación de información y clasificación de documentos [Powers (2011)].

El valor F se considera como una media armónica que combina los valores de la precisión (*precision*) y de la exhaustividad (*recall*). De tal forma que:

$$F_1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad F_1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad (3.11)$$

## 3.7. Máquinas de Soporte Vectorial

Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) se pueden etiquetar las clases y entrenar una máquina de soporte vectorial (SVM) para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de una muestra en el espacio, separando las clases a 2 espacios, los más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases más cercanas al que se le denomina como *vector soporte*. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o a la otra clase [Scholkopf y Smola (2001)].

Formalmente, una SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta [Scholkopf y Smola (2001)]

En el siguiente ejemplo idealizado para 2-dimensiones, la representación de los datos a clasificar se realiza en el plano x-y. El algoritmo SVM trata de encontrar un hiperplano 1-dimensional (en el ejemplo que nos ocupa es una

línea), que une a las variables predictoras y constituye el límite que define si un elemento de entrada pertenece a una categoría o a la otra.

Existe un número infinito de posibles hiperplanos (líneas) que realizan la clasificación pero, ¿cuál es la mejor y cómo la definimos? Hay hiperplanos infinitos posibles,  $H_1$  no separa las clases.  $H_2$  las separa, pero solo con un margen pequeño.  $H_3$  las separa con el margen máximo. La mejor solución es aquella que permite un margen máximo entre los elementos de las dos categorías. Se denominan vectores de soporte a los puntos que conforman las dos líneas paralelas al hiperplano, siendo la distancia entre ellas (margen) la mayor posible.

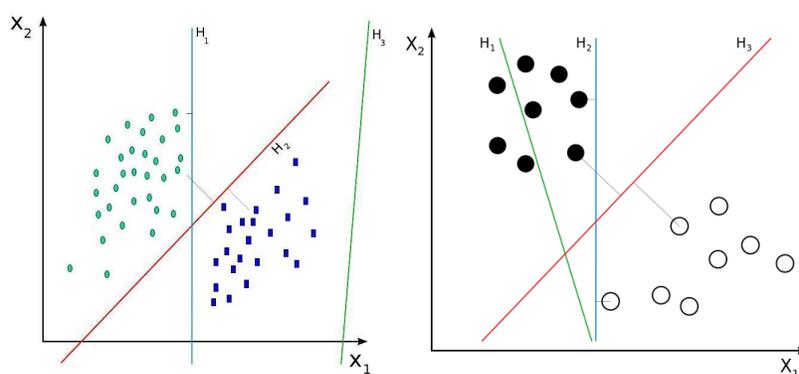


Figura 3.2: Ejemplo de clasificación de SVM para dos clases

### 3.8. Random forest o Bosques Aleatorios

Es una combinación de árboles predictores, tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de *bagging* que construye una larga colección de árboles no correlacionados y luego los promedia [Breiman (2017)]

El algoritmo para inducir un *random forest* fue desarrollado por Leo Breiman<sup>1</sup> y Adele Cutler, por lo que *Random Forest* es su marca de fábrica. El término aparece en la primera propuesta de *Random decision forests*, hecha por Tin Kam Ho de Bell Labs en 1995. El método combina la idea

de *bagging* de Breiman y la selección aleatoria de atributos, introducida independientemente por Ho, Amit y Geman [Breiman (2017)] para construir una colección de árboles de decisión con variación controlada.

La selección de un subconjunto aleatorio de atributos es un ejemplo del método *random subspace*, el que, según la formulación de Ho, es una manera de llevar a cabo la discriminación estocástica propuesta por Eugenio Kleinberg. [Diener (2017)]

En muchos problemas el rendimiento del algoritmo *random forest* es muy similar al algoritmo de *boosting*, y es más simple de entrenar y ajustar. Como una consecuencia el *random forest* es más popular y es ampliamente utilizado [Cutler y cols. (2007)].

La idea esencial del *bagging* es promediar muchos modelos ruidosos pero aproximadamente imparciales, y por tanto reducir la variación. Los árboles son los candidatos ideales para el *bagging*, dado que ellos pueden registrar estructuras de interacción compleja en los datos y si crecen suficientemente profundo, tienen relativamente baja parcialidad. Producto de que los árboles son notoriamente ruidosos, ellos se benefician grandemente al promediar [Breiman (2017)].

Cada árbol es construido usando el siguiente algoritmo:

Sea  $N$  el número de casos de prueba,  $M$  es el número de variables en el clasificador. Sea  $m$  el número de variables de entrada a ser usado para determinar la decisión en un nodo dado;  $m$  debe ser mucho menor que  $M$ . Elegir un conjunto de entrenamiento para este árbol y usar el resto de los casos de prueba para estimar el error. Para cada nodo del árbol, elegir aleatoriamente  $m$  variables en las cuales se basa la decisión. Calcular la mejor partición del conjunto de entrenamiento a partir de las  $m$  variables. Para la predicción un nuevo caso es empujado hacia abajo por el árbol. Luego se le asigna la etiqueta del nodo terminal donde termina. Este proceso es iterado por todos los árboles en el ensamblado y la etiqueta que obtenga la mayor cantidad de incidencias se reporta como la predicción [Diener (2017)].

#### **Ventajas**

1. Es uno de los algoritmos de aprendizaje más certeros que hay disponible. Para un conjunto de datos lo suficientemente grande produce un

clasificador muy certero.

2. Se ejecuta eficientemente en bases de datos grandes. Puede manejar cientos de variables de entrada sin excluir ninguna.
3. Proporciona estimados de qué variables son importantes en la clasificación.
4. Tiene un método eficaz para estimar datos perdidos y mantener la exactitud cuando una gran proporción de los datos está perdida.
5. Calcula los prototipos que dan información sobre la relación entre las variables y la clasificación.
6. Calcula las proximidades entre los pares de casos que pueden usarse en los grupos, localizando valores atípicos o (ascendiendo) dando vistas interesantes de los datos.
7. Ofrece un método experimental para detectar las interacciones de las variables.

#### **Desventajas.**

1. Se ha observado que *Random forest* sobreajusta en ciertos grupos de datos con tareas de clasificación/regresión ruidosas.
2. A diferencia de los árboles de decisión, la clasificación hecha por *Random forest* es difícil de interpretar por el hombre.
3. Para los datos que incluyen variables categóricas con diferente número de niveles, el *Random forest* se parcializa a favor de esos atributos con más niveles. Por consiguiente, la posición que marca la variable no es fiable para este tipo de datos. Métodos como las permutaciones parciales se han usado para resolver el problema.
4. Si los datos contienen grupos de atributos correlacionados con similar relevancia para el rendimiento, entonces los grupos más pequeños están favorecidos sobre los grupos más grandes.

---

**Algorithm 1** SMOTE ( $T, N, k$ )

---

**Require:**

Number of minority class samples  $T$ ;  
 Amount of SMOTE  $N$ ;  
 Number of nearest neighbors  $k$ ;

**Ensure:**

$(N/100) * T$  synthetic minority class samples;      ▷ (If  $N$  is less than 100 %, randomize the minority class samples as only a random percent of them will be SMOTEd.)

```

1: if  $N < 100$  then
2:   Randomize the  $T$  minority class samples
3:    $T = (N/100) * T$ 
4:    $N = 100$ 
5: end if

6: Explore L dimension  $N = (int)(N/100)$     ▷ The amount of SMOTE is
   assumed to be in integral multiples of 100.
7:  $k =$  Number of nearest neighbors
8:  $numattrs =$  Number of attributes
9: Sample: [ ] array for original minority class samples
10: newindex: keeps a count of number of synthetic samples generated,
    initialized to 0
11: Synthetic [ ] : array for synthetic samples.      ▷ Compute k nearest
    neighbors for each minority class sample only.

12: for  $i \leftarrow 1$  to  $T$  do
13:   Compute  $k$  nearest neighbors for  $i$ , and save the indices in the
    nnarray
14:   Populate( $N, i, nnarray$ )
15: end for
16: Populate( $N, i, nnarray$ )    ▷ Function to generate the synthetic samples.
17: while  $N \neq 0$  do
18:   Choose a random number between 1 and  $k$ , call it  $nn$ .    ▷ This step
    chooses one of the k nearest neighbors of  $i$ .
19:   for  $attr \leftarrow 1$  to  $numattrs$  do
20:     Compute:  $dif = Sample[nnarray[nn]][attr] - Sample[i][attr]$ 
21:     Compute:  $gap =$  random number between 0 and 1
22:      $Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$ 
23:   end for
24:    $newindex ++$ 
25:    $N = N - 1$ 
26: end while
return End of populate

```

# Capítulo 4

## Metodología Propuesta

En este capítulo se describe el método propuesto para monitorear el progreso de la enfermedad de Parkinson, obteniendo los requerimientos, la arquitectura y los diagramas que nos ayudan a definir el sistema de una manera clara.

### 4.1. Introducción

Esta investigación tiene la finalidad de entender las variaciones en los síntomas de la enfermedad de Parkinson. Las personas que padecen esta enfermedad neuro-degenerativa pueden tener diferentes síntomas durante el día, afectando la calidad de vida y el desarrollo personal. Este estudio permite monitorear el comportamiento individual del estado de los síntomas de la EP, mediante actividades específicas utilizando los sensores de un dispositivo móvil y cuestionarios a través de una aplicación móvil para iOS.

La metodología propuesta se divide en 4 etapas:

1. Monitoreo y Recolección de Datos
2. Preprocesamiento de los Datos
3. Entrenamiento del Algoritmo de Aprendizaje y Clasificación
4. Evaluación del Modelo.

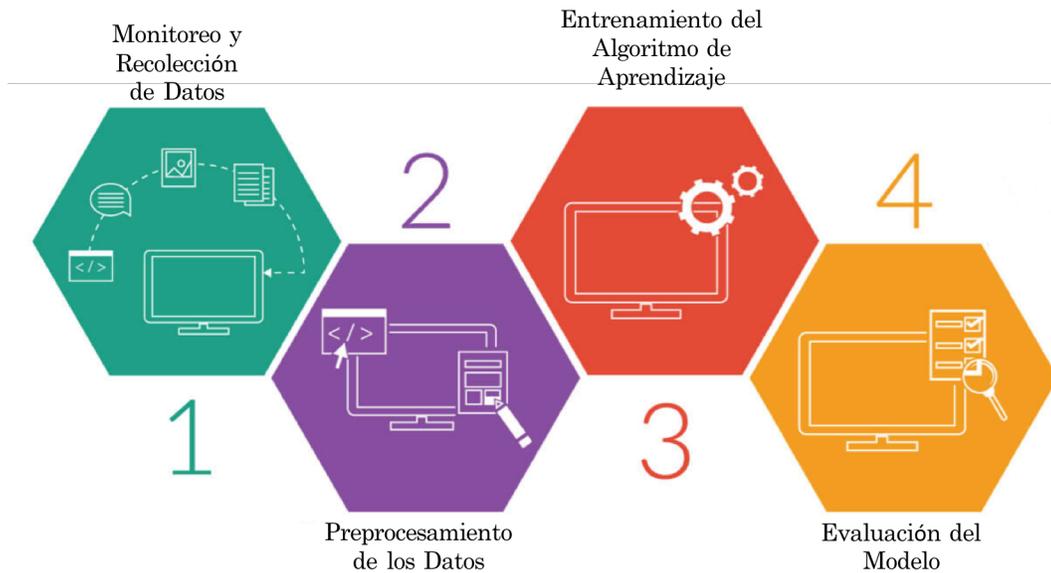


Figura 4.1: Metodología Propuesta

## 4.2. Etapa de Monitoreo

En esta etapa, por medio de una aplicación móvil se solicita a los participantes realizar actividades específicas diseñadas para monitorear el progreso de la enfermedad utilizando los sensores de un dispositivo móvil. La información recolectada se envía a un servidor web denominado Firebase, para su posterior análisis. Adicionalmente, la aplicación móvil brinda el beneficio a los participantes de registrar sus tratamientos en una tarjeta de cuidados y realizar actividades en forma de juegos que le permiten ejercitar su memoria. Las actividades del estudio (monitoreo) incluyen las siguientes tareas:

1. Una prueba de destreza manual.
2. Una prueba de caminata.
3. Una prueba de memoria espacial.
4. Un cuestionario de los síntomas.



Figura 4.2: Etapa de Monitoreo.

### 4.2.1. Destreza Manual

El temblor de las manos es el síntoma más conocido y más frecuente, sobre todo al comienzo de la enfermedad. Suele iniciarse en la extremidad superior, apareciendo en reposo y disminuye con los movimientos, pero aumenta cuando se exige la selectividad de éstos, afectando sobre todo a los miembros superiores [Jankovic (2008)].

El *Finger Tapping Test* [Reitan R. M. (1993)], también llamado Finger Oscillation Test, tiene una larga historia en el campo de la neuropsicología, tanto como parte de la Batería Neuropsicológica Halstead-Reitan (HRNB) como una prueba independiente. Es una de las pruebas más sensibles al daño cerebral y, como resultado, también se incluye en las baterías de detección HRNB para adultos y niños [Ashendorf y cols. (2009)]. Se usa comúnmente en evaluaciones de lesiones cerebrales traumáticas (TBI) y otros trastornos neurológicos como el Parkinson. También se han desarrollado versiones computarizadas alternativas del FTT que se incluyen en algunas pruebas informatizadas orientadas a la investigación clínica [Hubel y cols. (2013)].

Brian, [Bot y cols. (2016)] utiliza una versión computarizada del *Finger Tapping Test* para el diagnóstico de la progresión de la enfermedad de Parkinson. Se seleccionó esta prueba porque permite medir el nivel de temblor de las manos de un paciente. La tarea consiste en presionar dos botones alternadamente durante 20 segundos y se obtienen las siguientes características [Figura 4.3]:

- Coordenadas  $x, y$  del Botón Derecho.

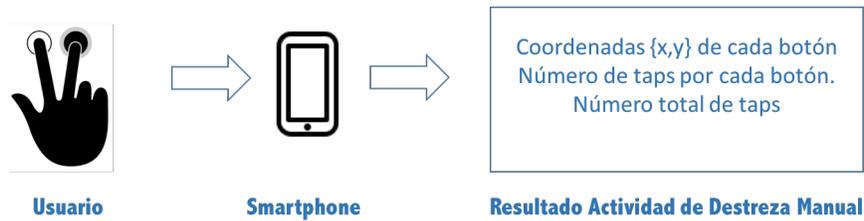


Figura 4.3: Actividad de Destreza Manual

- Coordenadas  $x, y$  del Botón Izquierdo.
- Número de veces que se presiona cada botón.
- Número total de veces que se presionan los botones por actividad.
- Inicio timestamp.
- Tamaño de la pantalla del iphone.
- Información de la medicación.
- Información acerca de la mano dominante y afectada.

### 4.2.2. Caminata

La marcha de los pacientes con la EP se ve alterada por la pérdida de los patrones automáticos de movimiento, apareciendo una marcha patológica con pasos cortos, torpes y rápidos al empezar a caminar en un intento de alcanzar su centro de gravedad y con dificultades para girar [de Fisioterapeutas. AI Macías Jiménez V Crespo Sánchez M Morales Cabezas (2004)]. Esta prueba está diseñada para evaluar el balance y ciclo de marcha mientras las personas realizan una caminata en línea recta durante un tiempo determinado.

El flujo de trabajo más utilizado para evaluar el ciclo de marcha de los participantes es posicionar dispositivos de recopilación de datos en diferentes partes del cuerpo, como extremidades superiores, extremidades inferiores o tronco. De esta forma los investigadores instruyen al paciente con EP a terminar una serie de tareas motoras estándar, que son explotadas para estimar la severidad del paciente con EP en la institución clínica. A partir de esto se extraen características de los síntomas motores y se alimentan estas características a los algoritmos de aprendizaje automático. Finalmente, los

algoritmos extraerán patrones de estas características. Al mismo tiempo, los expertos clínicos observan el rendimiento de los pacientes y proporcionan puntajes para el paciente con EP a través de la encuesta UPRDS. Estas pruebas tienen la característica de llevarse a cabo en un entorno de laboratorio. [Hammerla y cols. (2015)].

Los dispositivos móviles se encuentran equipados con acelerómetros por lo que no se requiere una preparación adicional. [Bot y cols. (2016)] [Arora y cols. (2016)][Ellis y cols. (2015)] [Kim y cols. (2015)] utilizan los resultados proporcionados por dispositivos móviles para evaluar el ciclo de marcha de los pacientes con EP.

El acelerómetro mide la suma de dos vectores de aceleración: la gravedad y la aceleración del usuario. La aceleración del usuario es la aceleración que el usuario imprime al dispositivo. Debido a que Core Motion de los dispositivos iPhone puede rastrear la altitud de un dispositivo usando tanto el giroscopio como el acelerómetro, puede diferenciar entre la gravedad y la aceleración del usuario. De esta forma se puede obtener mediciones en las propiedades de gravedad y de aceleración del usuario.

La información que se obtiene en esta prueba y utilizando el acelerómetro, giroscopio y magnetómetro incorporados en los dispositivos móviles es la siguiente [Figura 4.4]:

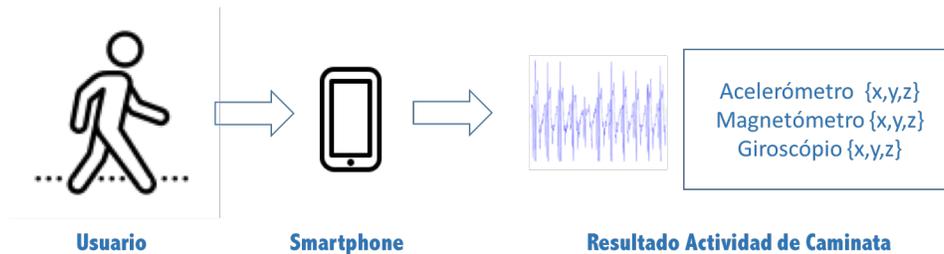


Figura 4.4: Actividad de Caminata

- Información timestamp
- *rotationRate*: coordenadas x, y, z en radianes por segundo
- *userAcceleration*: coordenadas x, y, z en Gs (junto con la gravedad, conforma la lectura del acelerómetro)

- gravedad: coordenadas x, y, z en Gs (junto con la aceleración del usuario, conforma la lectura del acelerómetro)
- *magneticField* (no informativo): y, z, x, exactitud - si la precisión es -1, significa que la estimación del campo magnético no está calibrada (que es el caso para la mayoría de los ejemplos)

### 4.2.3. Memoria Espacial

Los cambios cognitivos están presentes en mayor o menor grado en casi todas las personas con EP y se convierten en una característica prominente a lo largo del curso de la enfermedad. Aunque han recibido menos atención que los síntomas motores, producen efectos evidentes en la vida cotidiana, incluyendo cómo se adaptan las personas a sus síntomas motores. Todavía no se tienen tratamientos definitivos para la disfunción cognitiva en la EP. Sin embargo, es importante reconocer los problemas intelectuales, con el fin de aprovechar los medicamentos y las estrategias de comportamiento actualmente disponibles [WP y cols. (1998)].

El deterioro cognitivo, en forma de déficit visoespacial y memoria así como la demencia, parece ser un aspecto no motor independiente del trastorno que tiene un papel importante en el establecimiento del resultado funcional. La frecuencia y la gravedad del deterioro cognitivo causado por la enfermedad de Parkinson y sus implicaciones para el manejo clínico enfatizan la necesidad de un síntoma que requiera atención separada y tratamiento específico [Kehagia y cols. (s.f.)].

Brian et al.[Bot y cols. (2016)] implementan una actividad de memoria desarrollada por Katherine Possin y Joel Kramer de la Universidad de California en San Francisco con la finalidad de evaluar el deterioro cognitivo de memoria.

Esta actividad está enfocada en la medición del deterioro de la memoria espacial donde se le pide al usuario que observe y recuerde una secuencia de patrones visuales, aumentando el número de patrones por nivel. La longitud de la secuencia de patrones varía automáticamente durante la tarea, aumentando después de completar con éxito una secuencia. La actividad finaliza cuando el usuario ha cometido un error en la secuencia de patrones, a cada nivel se le asigna un *score*, basado en el número de aciertos de la actividad.

#### 4.2.4. Cuestionario Calidad de Vida

El concepto de calidad de vida ha venido cobrando importancia en la población adulta. Existen múltiples instrumentos que se han diseñado para evaluar las dimensiones que integran las mediciones de salud y de calidad de vida que se han diseñado con diversos propósitos. Para conocer y comparar el estado de salud entre poblaciones, un aspecto fundamental para estrategias y programas en políticas de salud y para evaluar el impacto de ciertas intervenciones terapéuticas para modificar los síntomas y función física a través del tiempo. La calidad de vida es un fenómeno que se afecta tanto por la enfermedad, como por el tratamiento (por sus efectos adversos). Los instrumentos para medir calidad de vida deben verse como herramientas adicionales del clínico en la evaluación integral del paciente y en la conducción de ensayos clínicos [Velarde-Jurado y Avila-Figueroa (2002)][Martínez-Jurado Elizabeth (2010)].

Los factores que impactan en la calidad de vida relacionada con la salud en pacientes con EP incluyen la severidad de la enfermedad, duración de la misma, inestabilidad postural y caídas, complicaciones motoras, depresión, ansiedad, dolor, sueño, deterioro cognitivo, alucinaciones y limitaciones de las actividades de la vida diaria.<sup>7</sup> Existen diversos instrumentos clínicos para evaluar tanto la función motora como la no motora [Chaudhuri y cols. (2006)][Goetz y cols. (2008)].

Los instrumentos diseñados para medir calidad de vida, como el Cuestionario de Salud (SF-36), son utilizados frecuentemente; sin embargo, existen instrumentos específicos para la EP que deben usarse preferentemente. Los instrumentos deben ser suficientemente exhaustivos, de forma que contemplen las áreas particularmente afectadas por la EP y su tratamiento. Los cuestionarios deben ser tan breves como sea posible, con un formato de fácil comprensión, deben contener documentación sobre la forma de puntuación, modo de administración y guías de interpretación. Las escalas deben ser sensibles (precisas), reflejando pequeñas diferencias y, además, deben ser sensibles para detectar cambios significativos en el estado del paciente causados por la evolución de la enfermedad o por intervención [Martínez-Jurado Elizabeth (2010)].

Este cuestionario fue diseñado con apoyo del Instituto Nacional de Neurología, con base en el cuestionario EQ-5D [Herdman y cols. (2001)], un

instrumento genérico de medición de la calidad de vida que puede utilizarse tanto en individuos relativamente sanos (población general) como en grupos de pacientes con diferentes patologías y el Cuestionario PDQ39 [Peto y cols. (1998)] (*Parkinson's Disease Quality of Life Questionnaire*) para medir la percepción de calidad de vida en individuos con EP . Contiene 39 cuestiones divididas en ocho dominios: movilidad, actividades de vida diaria, bienestar emocional, estigma, soporte social, cognición, comunicación y desaliento corporal para evaluación de la calidad de vida. Las preguntas que se incluyeron fueron las siguientes:

En una escala del 0 al 10 indique donde 0 significa que el síntoma está ausente y 10 significa que está experimentando la peor angustia posible debido al síntoma.

1. Temblores.
2. Dificultad para hablar.
3. Sentimientos de miedo y ansiedad.
4. Rigidez muscular.
5. Depresión.
6. Dificultad para caminar.
7. Dificultad para realizar actividades de la vida diaria.

Considerando que existe una gran cantidad de factores que influyen en la calidad de vida relacionada con la salud en pacientes con EP, de este cuestionario se espera obtener un instrumento de evaluación en la vida diaria de los participantes que nos permita tener consideración de los síntomas no motores de la enfermedad.

## 4.3. Preprocesamiento de los datos

El flujo de trabajo para la etapa de preprocesamiento [Figura 4.5] incluye la preparación de la información con la finalidad de generar un conjunto de datos más pequeño que el original con información relevante para alimentar el algoritmo de aprendizaje automático y mejorar la eficiencia del proceso de clasificación. En primer lugar se deberán integrar las bases de datos de acuerdo a la información proporcionada por la aplicación móvil y la base de

datos de SageBionetworks, una vez conjuntada la información se procede a una etapa de limpieza y transformación de la información con la finalidad de obtener características esenciales que puedan reflejar los cambios motores en los síntomas de la EP para el modelo de clasificación. Una vez obtenidas las características importantes del modelo de clasificación se aplica un algoritmo para balancear las clases denominado SMOTE. Este algoritmo nos permite realizar crear nuevas muestras sintéticamente utilizando el algoritmo de K-vecinos, que permite reducir la clase mayoritaria cuando la base de datos esta desbalanceada. A través de los pasos anteriores, las características extraídas se pueden utilizar en el algoritmo de aprendizaje automático directamente, porque el conjunto de datos se encuentra etiquetado para extraer los patrones del conjunto de datos.

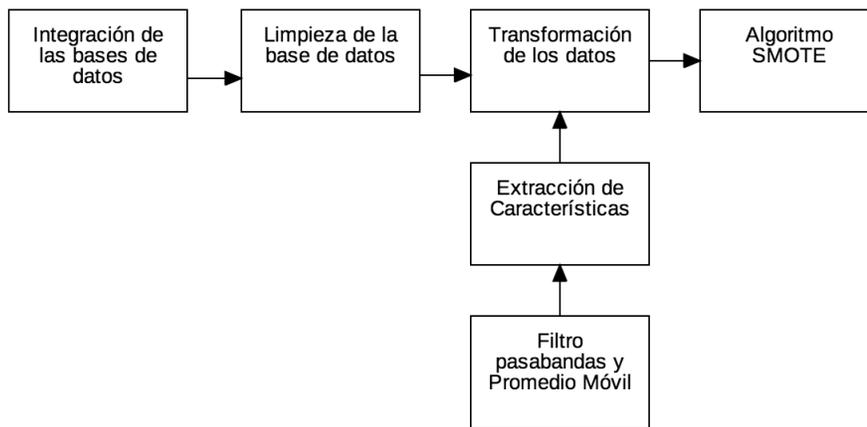


Figura 4.5: Preprocesamiento de la Información.

### 4.3.1. Integración de las bases de datos y limpieza de los datos

Para construir el modelo de clasificación de este trabajo se utilizaron dos fuentes de información, la primera a partir de la aplicación móvil desarrollada en este trabajo y la segunda es una base de datos proporcionada por *SageBionetworks*, una organización de investigación sin fines de lucro que busca desarrollar predicciones de enfermedades y acelerar la investigación en salud a través de la creación de sistemas abiertos, incentivos y estándares

que permite a los investigadores compartir e interpretar datos a gran escala, realizar pruebas colectivas para nuevas hipótesis y contribuir al conocimiento a través de los desafíos de la comunidad. A través del portal *mPower Public Researcher* ponen a disposición de la comunidad científica los resultados del estudio mPower para pacientes con EP. mPower es un estudio basado en aplicaciones móviles que prueba nuevos enfoques para monitorear indicadores clave de la progresión y el diagnóstico de la enfermedad de Parkinson, las mediciones incluyen las actividades de caminata, destreza manual, voz y memoria. [ Bot y cols. (2016)].

La integración de las bases de datos es un paso importante que incluye la homogeneización de la información proporcionada por ambas fuentes especialmente cuando se encuentran en diferentes formatos, la base de datos proporcionada por SageBionetworks se encuentra en un formato SQL y la proporcionada por la aplicación móvil desarrollada en este trabajo se encuentra en un formato JSON.

Pandas es una librería de Python para Data Science que proporciona estructuras de datos denominados *dataframes*. Un *dataframe* es una estructura de datos bidimensional etiquetada con columnas de tipos potencialmente diferentes. Se puede pensar como una hoja de cálculo o una tabla SQL, o un diccionario de objetos Serie. Junto con los datos, puede opcionalmente pasar argumentos de índice (etiquetas de fila) y columnas (etiquetas de columna).

Pandas proporciona herramientas que permiten leer y escribir datos en diferentes formatos: CSV, Microsoft Excel, bases SQL , JSON y formato HDF5, además de seleccionar y filtrar de manera sencilla tablas de datos en función de posición, valor o etiquetas, fusionar y unir datos además de transformar datos aplicando funciones personalizadas.

Esta herramienta resulta de gran utilidad para combinar bases de datos en diferentes formatos, por lo que se procedió a convertir los conjuntos de datos de las diferentes fuentes de información en *dataframes* en Python para su manipulación, de esta manera se almacenan los datos en cuadrículas rectangulares y cada fila de estas cuadrículas corresponde a mediciones o valores de una instancia, mientras que cada columna es un vector que contiene datos para una variable específica, correspondiente a la etapa de extracción de características, se escogió además esta herramienta debido a que las filas

no necesitan contener el mismo tipo de valores.

Una vez convertidas las diferentes fuentes de información en *dataframes* se procede a reestructurar o construir la información; este paso incluye fusionar, unir y concatenar filas y columnas, con la herramienta de pandas es posible realizar estas tareas y generar un único *dataframe* así como utilizar las herramientas de limpieza de datos de la librería pandas que nos permiten eliminar registros duplicados o nulos así como aquellos que contienen información faltante.

### 4.3.2. Transformación de los datos

En esta etapa los datos son transformados o consolidados de forma apropiada para la extracción de información. La transformación de la información incluye una etapa de Filtrado que está compuesta por un Filtro pasabandas y un Filtro promedio móvil, la finalidad de esta etapa es reducir la dimensionalidad de la información y extraer aquellas características que son importantes para el modelo de clasificación. La extracción de características devuelve un subconjunto de las características originales del conjunto original. A continuación se detalla cada paso de esta transformación de los datos.

#### Filtro Pasabanda

Un filtro paso banda es un tipo de filtro electrónico que deja pasar un determinado rango de frecuencias de una señal y atenúa el paso del resto. Bot y cols. (2016) exploran un filtro pasabandas en el proyecto de reconocimiento de patrones con pacientes con EP. En el experimento el umbral del filtro pasabandas va de 1 a 3 Hz. La frecuencia de corte inferior a 1 Hz puede eliminar el efecto del cambio de orientación de los segmentos del cuerpo y la frecuencia de corte más alta puede extraer los componentes de frecuencia de los síntomas motores.

El filtro pasabanda puede filtrar el ruido no deseado en el conjunto de datos y mejorar la confiabilidad y la precisión en el experimento.

#### Filtro Promedio Móvil

El filtro promedio móvil a menudo se explota para eliminar las fluctuaciones extremas del tiempo en las series de datos. Es un cálculo utilizado para analizar un conjunto de datos en modo de puntos para crear series de promedios. Así, las medias móviles son una lista de números, en la cual cada uno es el promedio de un subconjunto de los datos originales.

Por ejemplo, si se tiene un conjunto de 100 datos el primer valor de la serie de medias móviles podría ser el promedio de los primeros 25 términos, luego el promedio de los términos 2 al 26, el tercer elemento de los términos 3 al 27 y así, hasta por último el promedio de los últimos 25 números del 76 al 100.

#### Extracción de características

La extracción de características es uno de los mayores desafíos para la reconstrucción de patrones, pues puede verse influenciado en el desempeño del clasificador. En muestras con alta dimensionalidad, como en nuestro estudio, puede incrementar la complejidad computacional del algoritmo de aprendizaje automático, por lo que extraer las características relevantes que describen los síntomas de la enfermedad de Parkinson más importantes de la aceleración original resulta en información más significativa que puede ayudar a aumentar la velocidad de procesamiento del clasificador así como su desempeño.

De acuerdo con anteriores investigaciones [Bot y cols. (2016)], las siguientes características pueden ser extraídas a partir de los datos de un teléfono móvil :

1. *Media*. Es el valor característico de una serie de datos cuantitativos, objeto de estudio que parte del principio de la esperanza matemática o valor esperado, se obtiene a partir de la suma de todos sus valores dividida entre el número de sumandos.
2. *Valor RMS* (Root Mean Square). Es la raíz cuadrada de la media aritmética de los cuadrados de los valores. El valor RMS está asociado con la amplitud de la aceleración y representa el poder promedio de la aceleración. La media cuadrática para una colección de  $N$  valores  $x_1, x_2, \dots, x_N$  de una variable discreta  $x_1$  viene dada por la fórmula (1):

$$x_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}} \quad (1) \quad (4.1)$$

3. *Desviación Estándar*. Se define como la raíz cuadrada de la varianza de la variable.
4. *Densidad Espectral de Potencia (PSD)*. Para señales continuas en todo momento, como procesos estacionarios, la densidad espectral de potencia o (PSD) describe cómo la potencia de una señal o serie de tiempo se distribuye sobre la frecuencia. En estadística se estudia la varianza de una función a lo largo del tiempo  $x(t)$  (o sobre otra variable independiente), y usando una analogía con señales eléctricas (entre otros procesos físicos), se puede referir al espectro de potencia incluso cuando no hay energía física involucrada. [Stoica y Moses (2005)].

Considerando una ventana finita  $1 \leq n \leq N$  con una señal muestreada en tiempo discreto  $x_n = x(n\Delta t)$  por un periodo de muestreo  $T = N\Delta t$ . Es posible estimar la PSD como sigue:

$$\tilde{S}_{xx}(\omega) = \frac{(\Delta t)^2}{T} \left| \sum_{n=1}^N x_n e^{-i\omega n} \right|^2. \quad (4.2)$$

### 4.3.3. Algoritmo SMOTE para desbalance entre clases

La mayoría de los sistemas que emplean aprendizaje supervisado fueron creados asumiendo que el conjunto de entrenamiento se encuentra bien balanceado. Sin embargo esta suposición frecuentemente no es aplicable en el mundo real.

Un conjunto de datos se encuentra desbalanceado si las clases no están representadas aproximadamente igual, surge con frecuencia en muchas aplicaciones prácticas en el contexto del reconocimiento de patrones médicos y de la minería de datos [Nitesh V. Chawla (2002)]. Existen muchas razones por las que un conjunto de datos puede estar desequilibrado: la categoría a la que se dirige puede ser muy rara en la población o los datos pueden ser difíciles de recopilar.

La mayoría de los enfoques de clasificación existentes están bien desarrollados, asumiendo que el conjunto de entrenamiento está distribuido uniformemente. Sin embargo, se enfrentan a un grave problema de sesgo cuando el conjunto de entrenamiento tiene una distribución altamente desequilibrada (es decir, los datos comprenden dos clases, la clase minoritaria  $C^-$  y la clase mayoritaria  $C^+$ ). El límite de decisión resultante está severamente sesgado a la clase minoritaria y, por lo tanto, conduce a un desempeño pobre según el análisis de la curva característica del operador receptor (ROC). Para este propósito, muchos algoritmos de clasificación han sido investigados intensivamente [Juanjuan Wang (2006)].

En la presente investigación, contamos con un conjunto de datos etiquetado como clase predominante (Clase 0) correspondiente a participantes sin diagnóstico de enfermedad de Parkinson y una clase minoritaria (Clase 1) perteneciente al conjunto de participantes diagnosticados con EP. El porcentaje de los datos etiquetados para la clase predominante es del 69 % mientras que para la clase minoritaria es de 31 %. Por lo que se decidió utilizar una Técnica conocida como SMOTE *Synthetic Minority Over-sampling TEchnique*, para balancear el conjunto de datos.

Para la aplicación de este algoritmo y de acuerdo a la investigación bibliográfica [Nitesh V. Chawla (2002)], es necesario configurar el porcentaje en el cual se generarán nuevos registros de la clase minoritaria así como el número de vecinos cercanos, se probaron diferentes Porcentajes, con la finalidad de encontrar el que mejor se ajuste a nuestro modelo de datos.

El conjunto de datos original donde el Porcentaje de SMOTE es 0 %, representa el conjunto de datos sin agregar nuevos casos minoritarios. En este conjunto de datos, la proporción de clase no ha cambiado.

Para el caso donde el porcentaje es igual a 100 %. El algoritmo SMOTE genera nuevos casos de minorías y agrega el mismo número de casos minoritarios que estaban en el conjunto de datos original. Debido a que SMOTE no aumenta el número de casos pertenecientes a la clase mayoritaria, la proporción de casos de cada clase ahora ha cambiado.

En el tercer caso, donde el porcentaje es igual a 200 %. El algoritmo dobla el porcentaje de casos minoritarios en comparación con el conjunto de

datos original. Esto no resulta en tener el doble de casos minoritarios que antes. Por el contrario, el tamaño del conjunto de datos se incrementa de tal manera que el número de casos de la mayoría se mantiene igual, y el número de casos minoritarios se incrementa hasta que coincida con el valor porcentual deseado. Los resultados de esta configuración pueden visualizarse en el [Cuadro 4.1].

El número de vecinos cercanos determina el tamaño del espacio de funciones que usa el algoritmo SMOTE cuando construye casos nuevos. Un vecino cercano es una fila de datos (un caso) que es muy similar al caso objetivo. La distancia entre dos casos se mide combinando los vectores ponderados de todas las características. Al aumentar el número de vecinos más cercanos se obtienen características de más casos. Manteniendo el número de vecinos cercanos en un número bajo permite que las características que se generen sean más parecidas al conjunto de datos original, en este estudio se seleccionó un número de vecinos = 1.

	Clase 0	Clase 1	Total
Dataset Original	23, 998	10,585	34,583
Porcentaje Smote 0 %	69 %	31 %	
Porcentaje Smote 100 %	23,998	21,170	45,168
	53 %	47 %	
Porcentaje Smote 200 %	23,998	31,755	55,753
	43 %	57 %	

Cuadro 4.1: Configuración del algoritmo SMOTE

#### 4.3.4. Aprendizaje Supervisado

Los algoritmos de aprendizaje supervisado hacen predicciones basadas en un conjunto de ejemplos. Cada ejemplo usado para el entrenamiento se etiqueta con el valor de interés; en este caso, el precio de las acciones. Un algoritmo de aprendizaje supervisado busca patrones en esas etiquetas de valor. Puede usar cualquier información que pueda ser relevante y cada algoritmo busca diferentes tipos de patrones. Una vez que el algoritmo encuentra el mejor patrón posible, lo usa para hacer predicciones de datos de prueba sin etiquetar. El aprendizaje supervisado es un tipo conocido y útil de aprendizaje automático. Cuando los datos se usan para predecir una categoría,

el aprendizaje supervisado también se denomina clasificación. Cuando hay solo dos opciones, se denomina clasificación de dos clases o binomial [Ericson (2017)].

En esta tesis se utiliza un aprendizaje supervisado, ya que se conoce a priori si el participante ha sido diagnosticado con enfermedad de Parkinson o no.

## 4.4. Clasificador

La clasificación trata de asignar las diferentes partes del vector de características a grupos o clases basándose en las características extraídas. En esta etapa se utilizan técnicas de aprendizaje automático. La elección del algoritmo de clasificación depende de las propiedades del conjunto de datos, así como de la longitud y el porcentaje de datos etiquetados. Algunos investigadores han utilizado técnicas de aprendizaje automático supervisado como Máquinas de Soporte Vectorial y clasificadores Bayesianos, para extraer patrones relacionados a ciclos de marcha. [Pedregosa y cols. (2011)][Arora y cols. (2016)].

La entrada del clasificador se define como un vector de características reales  $\vec{x}$ , donde el resultado de salida es

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right), \quad (4.3)$$

donde  $\vec{w}$  es un vector real de pesos y  $f$  es una función que convierte el producto punto a punto de los dos vectores en la salida deseada. El vector de pesos  $\vec{w}$  aprende de un conjunto de muestras de entrenamiento.

La función  $f$  mapea todos los valores por encima de un cierto umbral a la primera clase en este trabajo que pertenezca a la clase de participantes con diagnóstico de EP y el resto a la segunda clase es decir participantes sin diagnóstico. Dependiendo del tipo de clasificador se genera una función  $f$  más compleja que puede dar la probabilidad de que una muestra pertenezca a cierta clase.

El aprendizaje supervisado se basa en la disponibilidad de áreas de entrenamiento, se trata de áreas de las que se conoce a priori la clase a la que pertenecen y que servirán para generar una firma característica de cada una de las clases.

Scikit-learn es una librería en Python para trabajar con técnicas de aprendizaje automático. Incluye modelos de clasificación, extracción de características, regresiones, agrupaciones, reducción de dimensiones, selección de modelos, o preprocesamiento. Posee una API que es consistente en todos los modelos y se integra con el resto de los paquetes científicos que ofrece Python. Esta librería también facilita las tareas de evaluación, diagnóstico.

Microsoft Azure Machine Learning Studio (Azure ML) es una herramienta colaborativa desarrollada por Microsoft para crear, probar y desplegar soluciones de análisis predictivo en sus datos. Machine Learning Studio publica modelos como servicios web que pueden ser consumidos por aplicaciones personalizadas o herramientas de BI como Excel.

En este estudio se propone resolver un problema de clasificación supervisada binomial, donde la clase 0 corresponde a los participantes sin diagnóstico de EP y la clase 1 corresponde a los participantes con diagnóstico de EP. Para este fin se probaron 4 métodos de clasificación basados en aprendizaje supervisado utilizando la librería de Scikit-Learn y Azure ML.

- Maquinas de Soporte Vectorial.
- Redes Neuronales.
- Árboles de decisión.
- Regresión Logística.

#### 4.4.1. Regresión Logística.

La regresión logística asume una distribución logística de los datos, donde la probabilidad de que un ejemplo pertenezca a la clase 1 es la fórmula :

$$p(x; \beta_0, \dots, \beta_D - 1) \tag{4.4}$$

Dónde:

#### 4.4. CLASIFICADOR

---

- $x$  es un vector  $D$ -dimensional que contiene los valores de todas las características de la instancia y que se calcularon anteriormente.
- $p$  es la función de distribución logística.
- $\beta\{0\}, \dots, \beta\{D - 1\}$  son los parámetros desconocidos de la distribución logística.

El algoritmo trata de encontrar los valores óptimos para  $\beta_0, \dots, \beta_{D - 1}$  maximizando la probabilidad de registro de los parámetros dados las entradas [Ericson (2017)].

Para la construcción del Modelo de Regresión Logística se consideraron los siguientes parámetros:

- Tolerancia a la optimización = 0.0000001. Especifica el valor de umbral para optimizar la convergencia del modelo. Si la mejora entre iteraciones es menor que el umbral, el algoritmo se detiene y devuelve el modelo actual
- Peso L1 y L2 de Regularización = 1.0. La Regularización es un método para evitar overfitting, penalizando a los modelos con valores de coeficientea extremos. La regularización funciona mediante la penalización que se asocia con valores de los coeficientes al error de la hipótesis. Un modelo preciso con valores del coeficiente extremas sería penalizado más, pero menos preciso, un modelo con valores más conservadores sería penalizado menos. L1 se puede aplicar a los modelos dispersos, lo cual es útil cuando se trabaja con datos de alta dimensión. Por el contrario, L2 es preferible cuando los datos son escasos.
- Tamaño de la memoria para el optimizador L-BFGS = 20. Determina el tamaño de memoria en (MB) a utilizar para el optimizador L-BFGS.

#### 4.4.2. Boosted Decision Tree.

En la minería de datos, los árboles de decisión se pueden describir también como la combinación de técnicas matemáticas y computacionales para ayudar en la descripción, categorización y generalización de un conjunto dado de datos [Rokach y Maimon (2014)].

Los datos vienen en registros del formulario:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \dots, x_k, Y) \quad (4.5)$$

La variable dependiente,  $Y$ , es la variable objetivo que estamos tratando de comprender, clasificar o generalizar, en este estudio clasificar los participantes con diagnóstico de EP de los que no. El vector  $x$  se compone de las características seleccionadas anteriormente,  $x_1, x_2, x_3$ , etc., que se utilizan para esa tarea.

En los árboles de decisión potenciados se evita el sobreajuste limitando la cantidad de veces que pueden subdividir y la cantidad de puntos de datos permitidos en cada región. El algoritmo construye una secuencia de árboles, cada uno de los cuales aprende a compensar el error dejado por el árbol antes. El resultado es preciso pero tiende a usar mucha memoria.

Para la construcción del Árbol de Decisión se consideraron los siguientes parámetros:

- Número máximo de hojas por árbol = 20. Indica la cantidad máxima de hojas que se pueden crear en cualquier árbol
- Número mínimo de muestras por nodo hoja = 10. Para un número mínimo de muestras por nodo hoja indica el número de casos requeridos para crear cualquier nodo hoja en un árbol
- Tasa de Aprendizaje = 0.2. Este parámetro indica que tan lento o rápido converge a la solución óptima
- Número de árboles construidos = 100. Para el número de árboles construidos indica el número total de árboles de decisión para crear en el conjunto.

### 4.4.3. Máquina de Soporte Vectorial.

Tenemos un conjunto de entrenamiento de  $n$  puntos de la forma

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n) \quad (4.6)$$

donde  $y_i$  es 1 o -1 indicando a la clase donde el punto  $\vec{x}_i$  pertenece. Cada  $\vec{x}_i$  es un vector real  $p$ -dimensional. Queremos encontrar la distancia máxima que divide el grupo de puntos  $\vec{x}_i$  donde  $y_i = 1$  para el grupo de puntos

$y_i = -1$  que es definido de modo que la distancia entre el hiperplano y el punto más cercano  $\vec{x}_i$  de cada grupo es maximizado.

Cualquier hiperplano puede ser descrito como un conjunto de puntos  $\vec{x}$  que satisface :

$$\vec{w} \cdot \vec{x} - b = 0 \quad (4.7)$$

Donde  $\vec{w}$  es el vector normal al hiperplano (no está necesariamente normalizado) . Es parecido a la forma Hesse normal, excepto que  $\vec{w}$  no es necesariamente un vector unitario. El parámetro  $\frac{b}{\|\vec{w}\|}$  determina el offset del hiperplano al origen sobre la forma del vector normal  $\vec{w}$  [D. Gil y Johnson (2009)].

Para la construcción del algoritmo de Máquina de Soporte Vectorial se consideraron los siguientes parámetros:

- Número de iteraciones = 1.
- Lambda = 0.001. Valor de L1 para la regularización. La utilización de valores diferentes a cero evita el sobreajuste del modelo en el conjunto de datos de entrenamiento
- Normalización de características.

#### 4.4.4. Redes neuronales.

Los modelos de redes neuronales se pueden ver como modelos matemáticos que definen una función  $f : X \rightarrow Y$  o una distribución sobre  $X$  o ambos  $X$  y  $Y$ . A veces los modelos están íntimamente asociados con una regla de aprendizaje particular. Un uso común del modelo de Red neuronal es la definición de una clase de tales funciones (donde los miembros de la clase se obtienen variando los parámetros, los pesos de conexión o las características específicas de la arquitectura, como el número de neuronas o su conectividad) [Zell (2014)].

Matemáticamente, la función de red de una neurona  $f(x)$  se define como una composición de otras funciones  $g_i(x)$ , que pueden descomponerse en otras funciones. Esto se puede representar convenientemente como una estructura de red, con flechas que representan las dependencias entre funciones. Un tipo de composición ampliamente utilizado es la suma ponderada no lineal, donde:

$$f(x) = K \left( \sum_i w_i g_i(x) \right), \quad (4.8)$$

donde  $K$  (comúnmente conocida como la función de activación) es alguna función predefinida, como la función tangente hiperbólica o sigmoidea. La característica importante de la función de activación es que proporciona una transición suave a medida que cambian los valores de entrada, es decir, un pequeño cambio en la entrada produce un pequeño cambio en la salida. Lo siguiente se refiere a una colección de funciones  $g_i$  como vector  $g = (g_1, g_2, \dots, g_n)$ .

En la construcción del modelo de red neuronal se consideraron los siguientes parámetros.

- Inicialización de los pesos de entrada = 0.1
- Tasa de aprendizaje = 0.1. Define el tamaño del paso dado en cada iteración, antes de la corrección. Un valor más grande de la tasa de aprendizaje puede hacer que el modelo converja más rápido, pero puede sobrepasar los mínimos locales.
- Momento = 0.0. Especifica un peso aplicable durante el aprendizaje a los nodos de iteraciones anteriores
- Tipo de normalizador = Min-Max
- Número de iteraciones = 100.

## 4.5. Evaluación del Desempeño

La validación cruzada o *cross-validation* es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba.

Lo anterior consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo que se llevará a cabo a la práctica [Mursalin (2013)]. Esta es una técnica muy utilizada en proyectos de Inteligencia Artificial para validar modelos generados.

#### 4.5. EVALUACIÓN DEL DESEMPEÑO

---

Basado en la etapa de preprocesamiento de la Metodología de este trabajo, en el [Cuadro 4.2] se muestra el listado de características utilizadas en el modelo de aprendizaje automático así como su descripción.

Características	Descripción
total_tap	Número total de taps
left_tap	Número de taps Izquierdo
mean_left_tap	Promedio de coordenadas tap Izquierdo
median_left_tap	Media coordenadas tap izquierdo
std_left_tap	Desviación estándar tap izquierdo.
std_right_tap	Desviación estándar tap derecho.
median_right_tap	Media coordenadas tap derecho
mean_right_tap	Promedio de coordenadas tap Derecho
right_tap	Número de taps Derecho
x_mean	Promedio coordenada de aceleración eje x
x_psd	Densidad Espectral de Potencia de aceleración eje x
x_std	Desviación estándar aceleración eje x
x_rms	Valor rms aceleración eje x
y_mean	Promedio coordenada de aceleración eje y
y_psd	Densidad Espectral de Potencia de aceleración eje y
y_std	Desviación estándar aceleración eje y
y_rms	Valor rms aceleración eje y
z_mean	Promedio coordenada de aceleración eje z
z_psd	Densidad Espectral de Potencia de aceleración eje z
z_std	Desviación estándar aceleración eje z
z_rms	Valor rms aceleración eje z
xyz_mean	Promedio coordenada de aceleración eje x,y, z
xyz_psd	Densidad Espectral de Potencia de aceleración eje x,y, z
xyz_std	Desviación estándar aceleración eje x,y,z
xyz_rms	Valor rms aceleración eje x,y,z
professional_diagnosis	Etiqueta correspondiente al valor de la clase

Cuadro 4.2: Características utilizadas en el modelo de aprendizaje automático

Usando el método de validación cruzada, los datos correspondientes a las características del [Cuadro 4.2] se dividieron en un conjunto de entrenamiento compuesto por el 90% de los datos, utilizados para entrenar el clasificador y un conjunto de prueba que consiste en el 10% restante de los datos, utilizados para la evaluación del clasificador). Este método implica entrenar los métodos diez veces usando el conjunto de entrenamiento y evaluación utilizando el conjunto de prueba, asegurando que cada señal de aceleración se produce exactamente una vez en la prueba. Este proceso se repitió 100 veces cada vez que los datos se tomaron aleatoriamente antes del entrenamiento y la prueba.

Para este caso la asociación entre las etiquetas y los datos es mantenido

intacto durante el proceso de tomar aleatoriamente. De esta manera éste método de evaluación intenta evaluar la generalización de la clasificación para otros conjuntos de datos similares desconocidos y por lo tanto ayuda a protegerse contra el sobreajuste u *overfitting*.

---

# Capítulo 5

## Experimentos y Resultados

En esta sección se muestran los resultados obtenidos por la aplicación móvil y las pruebas realizadas con los métodos de aprendizaje automático.

### 5.1. Descripción de la Aplicación Móvil

Para el desarrollo de la aplicación móvil destinada a monitorear los síntomas de la enfermedad de Parkinson se utilizó xCode versión 9.2, que es el entorno de desarrollo para aplicaciones móviles en iOS además del *framework* desarrollado por Apple, denominado *Research Kit* y *CareKit*. La forma en cómo se registra una nueva actividad es a través del Método ORKTask que contiene objetos del tipo ORKStep relacionados a cada instrucción, actividad o cuestionario, dependiendo de la actividad que se desea registrar, estas tareas están conectadas al controlador de la Vista de la Aplicación a través del ORKTaskViewController, finalmente el resultado de cada actividad se refleja en el Método ORKStepResult, como se muestra en la [Figura 5.2].

Con base en el diseño, se solicita a los participantes registrarse en la aplicación móvil y aceptar el formulario de consentimiento informado para participar en el estudio. La aplicación permite recopilar información acerca de los síntomas motores de la EP y de la calidad de vida de los participantes. Estos datos brindan la capacidad de explorar la clasificación de los participantes con la enfermedad de Parkinson y aquellos que se encuentran sanos. Una vez recopilada esta información se utiliza la API de Firebase para la autenticación de los usuarios y el envío de Información al Framework Firebase Database para su posterior análisis como se muestra en la [Figura 5.1].

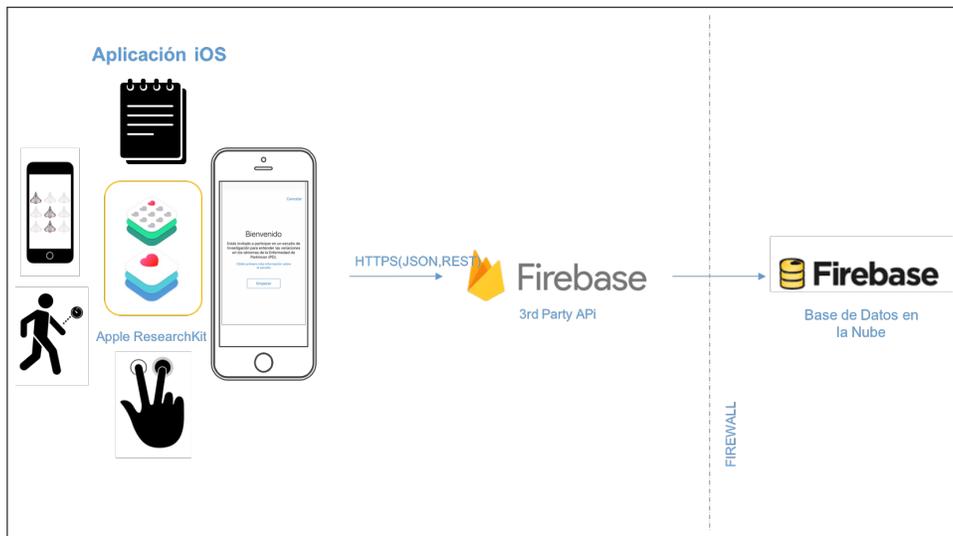


Figura 5.1: Diagrama funcional de la Aplicación Móvil.

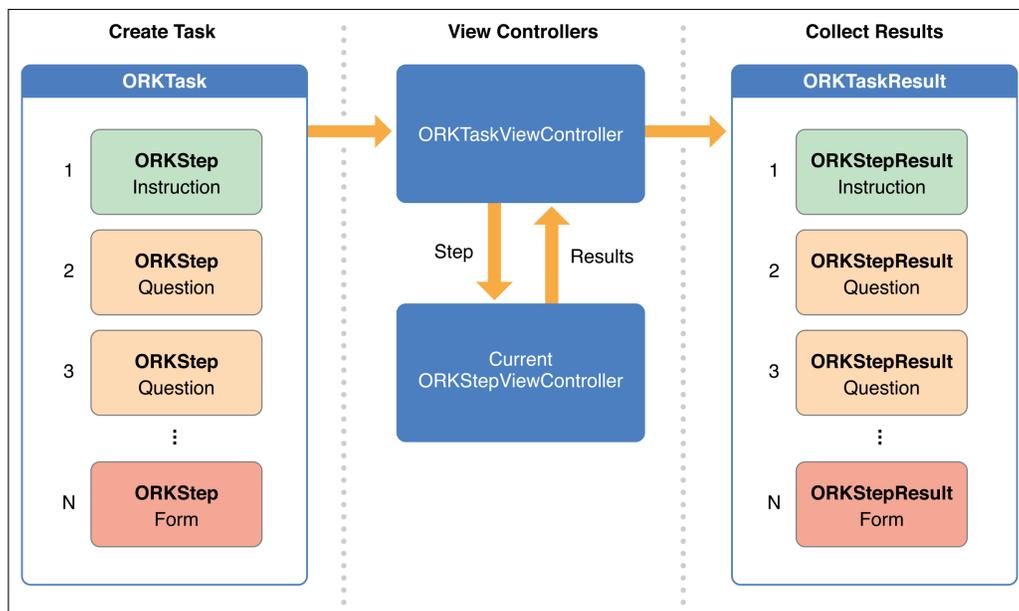


Figura 5.2: Arquitectura del Framework ResearchKit de Apple.

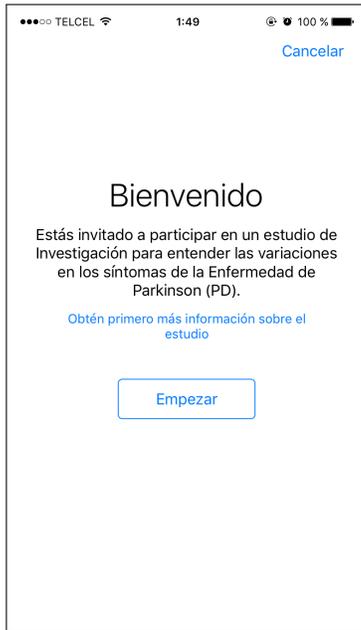
### 5.1.1. Formulario de Consentimiento Informado

El consentimiento informado es el procedimiento mediante el cual se garantiza que el sujeto ha expresado voluntariamente su intención de participar en la investigación, después de haber comprendido la información que se le ha dado, acerca de los objetivos del estudio, los beneficios, las molestias, los posibles riesgos y las alternativas, sus derechos y responsabilidades.

Debido a la naturaleza de los datos y con la finalidad de preservar la confidencialidad de la información proporcionada. Los participantes firmaron y aceptaron un formulario de consentimiento acerca de las condiciones de uso asociadas con este estudio.

Se diseñó un formulario de consentimiento informando sobre este estudio, mediante el cual se garantiza que el participante, es informado y acepta voluntariamente la realización de las evaluaciones médicas después de haber comprendido la información que se le ha dado, acerca de los objetivos del estudio, los beneficios, y las directrices a seguir. Es importante mencionar que el participante deberá aceptar y firmar el consentimiento informado, para formar parte del estudio, de lo contrario no podrá registrarse en la plataforma.

En la [Figura 5.3] se muestran las pantallas del consentimiento informado.



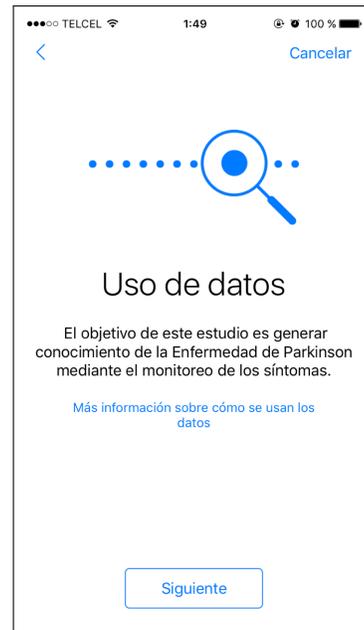
(a) Bienvenida



(b) Recopilación de Datos



(c) Privacidad



(d) Uso de Datos

Figura 5.3: Consentimiento Informado

## 5.1. DESCRIPCIÓN DE LA APLICACIÓN MÓVIL



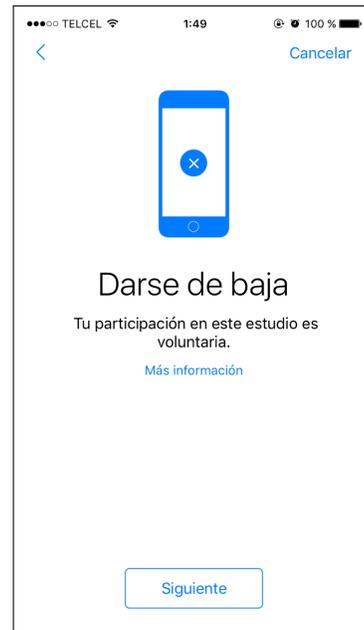
(e) Compromiso de tiempo



(f) Encuesta del estudio



(g) Tareas del estudio

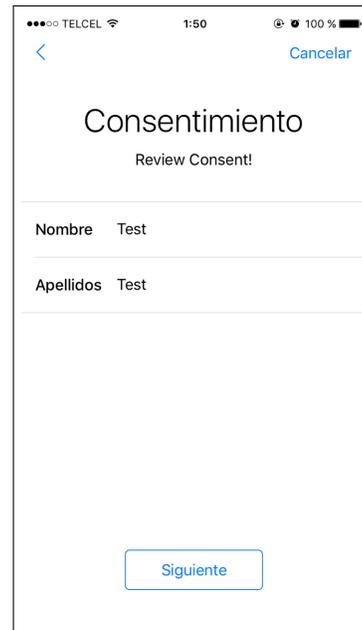


(h) Darse de baja

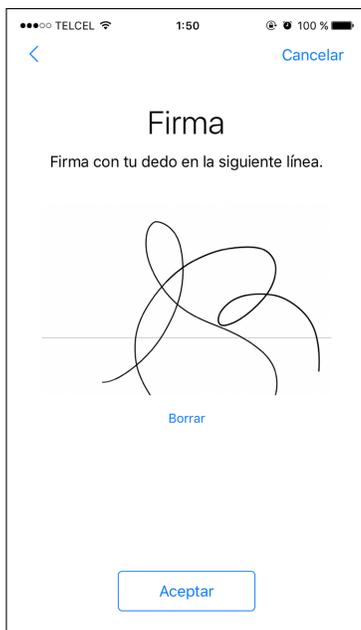
Figura 5.3: Consentimiento Informado



(i) Resumen Consentimiento y aceptación



(j) Datos del participante

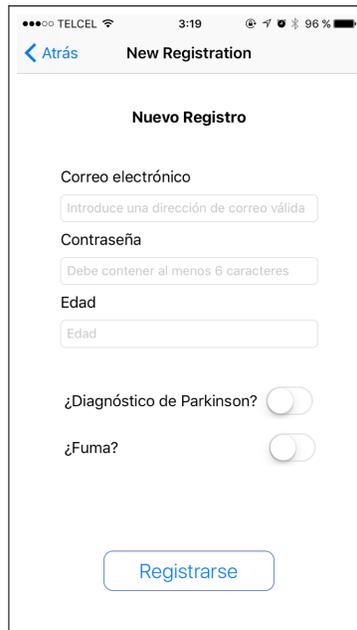


(k) Firma de aceptación

Figura 5.3: Consentimiento Informado

### 5.1.2. Registro

El registro de los participantes incluye una dirección de correo válida en vez de su nombre completo, una contraseña, edad, indicar si es fumador o no y si ha sido diagnosticado con Enfermedad de Parkinson. [ver Figura 5.4]:



The screenshot shows a mobile application interface for a 'New Registration' form. At the top, there is a status bar with 'TELCEL', signal strength, time '3:19', and battery '96%'. Below the status bar is a navigation bar with a blue back arrow labeled 'Atrás' and the title 'New Registration'. The main content area is titled 'Nuevo Registro' and contains the following elements: a text input field for 'Correo electrónico' with the placeholder 'Introduce una dirección de correo válida'; a text input field for 'Contraseña' with the placeholder 'Debe contener al menos 6 caracteres'; a text input field for 'Edad' with the placeholder 'Edad'; two toggle switches, one for '¿Diagnóstico de Parkinson?' and one for '¿Fuma?'; and a blue button labeled 'Registrarse' at the bottom.

Figura 5.4: Registro de la aplicación móvil Parkinson App

## 5.2. Actividades del Estudio

Se solicitó a los participantes realizar las siguientes tareas diariamente:

1. Actividad de Destreza Manual
2. Actividad de Caminata
3. Actividad de Memoria
4. Evaluación de los Síntomas

También se recopiló información demográfica y un cuestionario con los síntomas generales de la enfermedad. Se añadieron tareas opcionales denominadas juegos para incentivar la participación de los individuos y finalmente una *Tarjeta de Cuidados* que les permitiría hacer un seguimiento regular de sus síntomas y monitorear los niveles de dolor.

### 5.2.1. Prueba de Destreza Manual

La prueba de destreza manual mide la destreza y velocidad de la mano. Los participantes deben colocar su teléfono celular sobre una superficie plana y utilizando dos dedos de la misma mano, deberán presionar dos botones de forma alternada durante 20 segundos. Las mediciones de esta prueba incluyen el número de botones presionados, la velocidad y las coordenadas  $[x, y]$  donde se presionó el botón [LeMoyne y cols. (2010)]. En la [Figura 5.5] se muestran las interfaces de la prueba para el dispositivo móvil.



Figura 5.5: Prueba de Destreza Manual

En la [Figura 5.6] se muestra un fragmento de los resultados proporcionados por la prueba de destreza manual, destinada a evaluar los temblores en la mano, como síntoma motor. Asimismo en el [Cuadro 5.1] se observa cada parámetro registrado por la aplicación móvil así como el tipo de dato asociado.

## 5.2. ACTIVIDADES DEL ESTUDIO

Parámetro	Tipo de Dato
TapTimeStamp	double
TappedButtonId	varchar
TapCoordinate	varchar

Cuadro 5.1: Parámetros resultantes de la prueba de destreza manual

ROW_ID	recordId	healthCode	phoneInfo	ButtonRectL	ButtonRectR	timestamp	tappingResults.json	TappingView	medTimepoint
0	95436b71-6f-d597b678-df	iPhone 6	{{0, 156.5}, {{128, 156.5}}	1.426E+12	5397548 {320, 504}	Just after Parkinson medication (at your best)			
1	01077057-fa-9388697d-c0	iPhone 6	{{0, 191.5}, {{128, 191.5}}	1.426E+12	5401650 {375, 603}	I don't take Parkinson medications			
2	c5fbf6f6-bfd-f853506d-d9	iPhone 5c (G	{{0, 156.5}, {{128, 156.5}}	1.426E+12	5504582 {320, 504}	I don't take Parkinson medications			
3	be9ebf5f-d1-49086bf3-5b	iPhone 5c (G	{{0, 156.5}, {{128, 156.5}}	1.4261E+12	5677329 {320, 504}	Another time			
4	d544e6e1-a7-406b537a-73	iPhone 6	{{0, 199}, {1C{{128, 199}},	1.4263E+12	5788732 {375, 603}	Another time			
5	ab739759-bc-56109ba0-7f	iPhone 5c (G	{{0, 156.5}, {{128, 156.5}}	1.4263E+12	6076313 {320, 504}	I don't take Parkinson medications			
6	e259f4ac-d0-52705a16-5c	iPhone 5s (G	{{0, 156.5}, {{128, 156.5}}	1.4265E+12	5885818 {320, 504}	Another time			
7	fc2a19e5-c9-b62c6ea9-57	iPhone 5c (G	{{0, 156.5}, {{128, 156.5}}	1.4265E+12	5920876 {320, 504}	Immediately before Parkinson medication			
8	10ac72da-68-4b3ed13d-a6	iPhone 5s (G	{{0, 149}, {1C{{128, 149}},	1.4266E+12	5930995 {320, 504}	I don't take Parkinson medications			

(a)

TapTimeStamp	TappedButtonId	TapCoordinate
0	TappedButtonRight	{259, 498.5}
0.019051458	TappedButtonNone	{141, 526}
0.085591833	TappedButtonLeft	{141, 526}
0.135695208	TappedButtonNone	{269, 499}
0.228707583	TappedButtonRight	{269, 499}
0.229298958	TappedButtonNone	{122.5, 523}
0.268329083	TappedButtonLeft	{122.5, 523}
0.318285583	TappedButtonNone	{240, 497.5}
0.385182875	TappedButtonNone	{117.5, 517}

(b)

Figura 5.6: a) Fragmento de datos de Prueba de Destreza Manual b) Fragmento de archivo tappingResults.json de la primera tabla.

### 5.2.2. Actividad de Caminata

Esta actividad tiene la finalidad de medir el ciclo de marcha y balance de los participantes mientras caminan. Los participantes deberán caminar en línea recta durante 20 segundos. Las mediciones incluyen datos del acelerómetro y giroscopio del teléfono para cada segundo. En la [Figura 5.7] se muestra la interfaz desarrollada para medir esta característica.

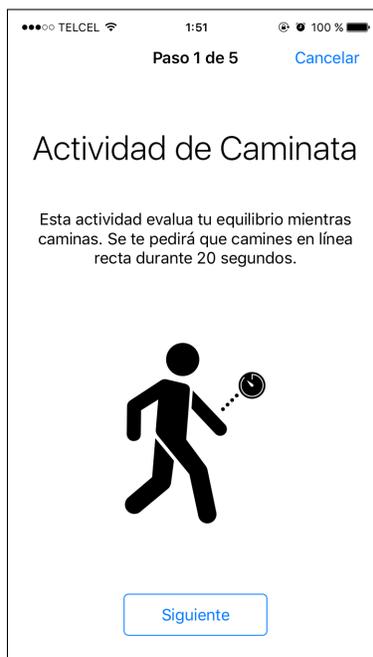


Figura 5.7: Actividad de Caminata

ROW_ID	recordId	healthCode	phoneInfo	deviceMotion.json	medTimepoint
0	0dcb75f3-4b-9a3b504e-62	iPhone 6		5413100	Just after Parkinson medication (at your best)
1	bf5ad14c-7c-8746d33b-bc	iPhone 6 Plus		5709962	I don't take Parkinson medications
2	9a042f44-c4-5bee7cd6-e9	iPhone 6		5795405	I don't take Parkinson medications
3	4ab5ea3e-fb-e26facdc-f8t	iPhone 6		5855050	I don't take Parkinson medications
4	c90f8f8e-cec-e5ce20e1-28	iPhone 5s (G		5955119	Just after Parkinson medication (at your best)
5	1aa5df10-72-45c21ca3-ab	iPhone 5s (G		5998389	Another time
6	9088021e-4a-5b34626f-2f	iPhone 5s (G		6017389	Another time
7	9cef2ef9-c2a-e212f488-3c	iPhone 5s (G		6056879	Another time
8	14d57715-270530d409-a4	iPhone 6		5561981	Just after Parkinson medication (at your best)

Figura 5.8: Fragmento de datos de Actividad de Caminata

En la [Figura 5.9] y [Figura 5.8] se muestra un fragmento de los resultados proporcionados por la prueba de caminata, destinada a evaluar el ciclo de marcha de los participantes. Asimismo en el [Cuadro 5.2] se observa cada parámetro registrado por la aplicación móvil para la prueba motora de caminata, así como el tipo de dato asociado.

## 5.2. ACTIVIDADES DEL ESTUDIO

attitude/y	attitude/w	attitude/z	attitude/x	timestamp	rotationRate/x	rotationRate/y	rotationRate/z	rotationRate/accuracy	userAcceleration/x	userAcceleration/y	userAcceleration/z	gravity/x	gravity/y	gravity/z	magneticField/y	magneticField/z	magneticField/x	magneticField/accuracy
-0.3051289	0.81627514	0.00031729	-0.4905009	19298.8061	-0.2566985	0.15565038	0.34499961	-0.0661836	-0.0303922	0.03782586	-0.497827	0.80096108	-0.3326104	0	0	0	-1	
-0.3034714	0.81578515	0.0008656	-0.4923403	19298.8161	-0.3324519	0.19930542	0.36989641	-0.0838882	-0.0676987	0.02939183	-0.4942825	0.80381316	-0.3310123	0	0	0	-1	
-0.3016474	0.81517325	0.0011848	-0.4944694	19298.8261	-0.4029818	0.2362698	0.35100368	-0.0894447	-0.0802019	0.02758029	-0.4906182	0.80687124	-0.3290177	0	0	0	-1	
-0.2999803	0.81442328	0.00102383	-0.4967147	19298.8361	-0.4662431	0.24394898	0.24554695	0.04368076	0.02753291	0.02492166	-0.4876047	0.8096863	-0.3265727	0	0	0	-1	
-0.2984898	0.81353136	0.00041374	-0.4990695	19298.8461	-0.5208659	0.25510013	0.17849758	0.05484403	0.06967836	0.02384695	-0.4852487	0.81226438	-0.3236669	0	0	0	-1	
-0.2969365	0.81257259	-0.0004084	-0.5015519	19298.8561	-0.5561478	0.28471205	0.168026	0.04115639	0.07175981	0.02268193	-0.4829746	0.81485212	-0.3205487	0	0	0	-1	
-0.2952921	0.81157081	-0.0013461	-0.5041365	19298.8661	-0.5796643	0.3069497	0.16389993	0.0512911	0.05406163	0.02092652	-0.4806582	0.81748986	-0.317298	0	0	0	-1	
-0.2937665	0.81057119	-0.0023142	-0.5066263	19298.8761	-0.5519595	0.27597949	0.13490233	0.08768248	0.03827681	0.01846874	-0.4785821	0.81995538	-0.314062	0	0	0	-1	
-0.292515	0.8096298	-0.0031764	-0.5088462	19298.8861	-0.4900054	0.21961606	0.11319353	0.0921858	0.02583511	0.01861683	-0.4768904	0.82209581	-0.311021	0	0	0	-1	

Figura 5.9: Fragmento de archivo deviceMotion.json de la actividad de Caminata

Parámetro	Tipo de Dato
attitude/y	double
attitude/w	double
attitude/z	double
attitude/x	double
timestamp	double
rotationRate/x	double
rotationRate/y	double
rotationRate/z	double
userAcceleration/x	double
userAcceleration/y	double
userAcceleration/z	double
gravity/x	double
gravity/y	double
gravity/z	double
magneticField/y	boolean
magneticField/z	boolean
magneticField/x	boolean
magneticField/accuracy	boolean

Cuadro 5.2: Parámetros resultantes de la Actividad de Caminata

### 5.2.3. Actividad de Memoria

La Actividad de Memoria mide la capacidad de memoria temporal y espacial de los participantes. En esta prueba se les solicita que memoricen una secuencia de botones y después presionen los botones en el orden en que fueron apareciendo. Si se completa la secuencia de forma exitosa, el nivel aumenta y se agregan más botones a la secuencia. Los resultados de esta

medición incluyen el nivel alcanzado y un *score* que corresponde al número de botones que se memorizaron. En la [Figura 5.10] se muestra la interfaz implementada para esta prueba.

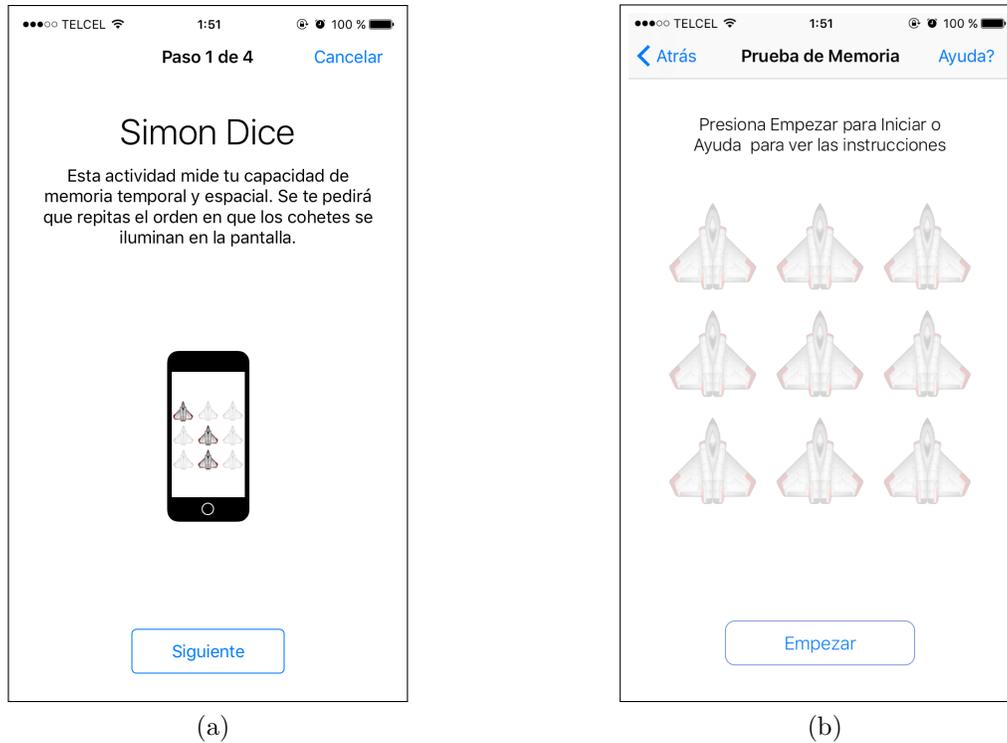


Figura 5.10: Actividad de Memoria

Se propone una actividad de Juego de Memoria con la finalidad de presentar una alternativa para los participantes de recordar diferentes formas a través de un *memorama*. En esta actividad el participante puede activar sus sentidos para memorizar las diferentes cartas del juego. Los resultados incluyen el tiempo que tarda el participante en completar esta actividad. En la Figura 5.11 se presenta la interfaz desarrollada para tal fin.

## 5.2. ACTIVIDADES DEL ESTUDIO



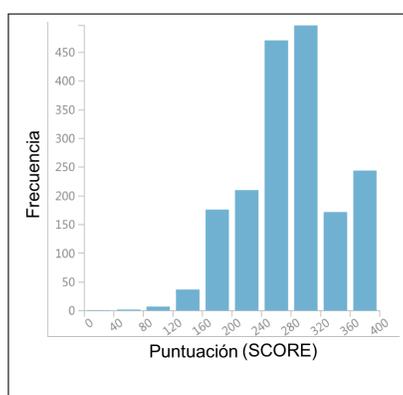
Figura 5.11: Juego de Memoria

Actividad de Memoria EP	
Media	257.51
Mediana	260
Desviación estándar	57.88
Mínimo	0
Máximo	400

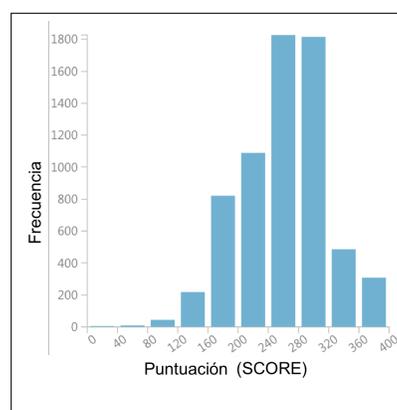
Cuadro 5.3: Resultados por score actividad de Memoria participantes con Diagnóstico de EP.

Actividad de Memoria	
Media	277.62
Mediana	280
Desviación estándar	65.39
Mínimo	0
Máximo	400

Cuadro 5.4: Resultados por score actividad de Memoria participantes sin Diagnóstico de EP.



(a)



(b)

Figura 5.12: Resultados de la Actividad de Memoria a) Participantes sin diagnóstico de EP b) Participantes con diagnóstico de EP.

El [Cuadro 5.4] y [Cuadro 5.3] muestran los valores de media, mediana y desviación estándar para los participantes con y sin diagnóstico de EP, además se observa que obtuvieron mejores puntajes los participantes sin diagnóstico de EP en términos generales.

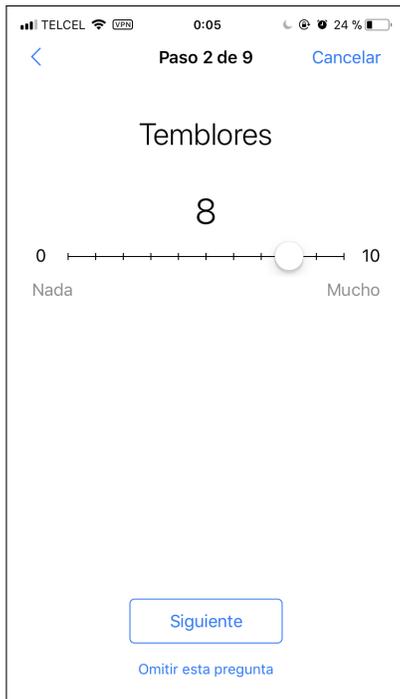
#### 5.2.4. Evaluación de los Síntomas

En general, se conocen más los síntomas motores de la enfermedad de Parkinson, ya que estos son los signos del trastorno que se observan desde el exterior. Estos síntomas principales de la enfermedad incluyen los siguientes: bradicinesia (lentitud en los movimientos): desaceleración o pérdida de los

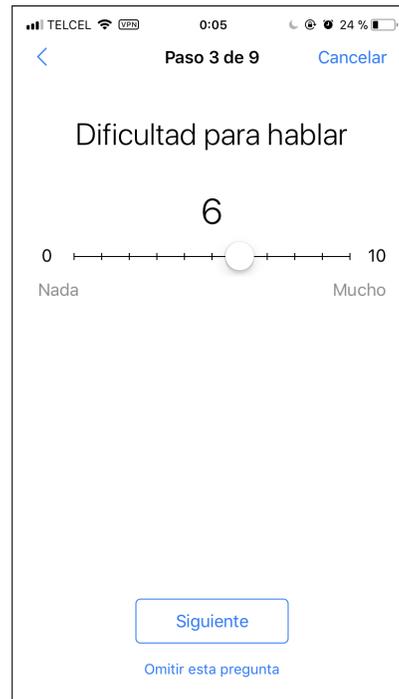
movimientos espontáneos y voluntarios, rigidez, temblor de reposo e inestabilidad postural, es decir problemas al estar de pie o al caminar, o disminución del equilibrio y la coordinación [Xia y Mao (2012)].

Además, pueden aparecer otros síntomas físicos, como dificultades en la marcha y una disminución de las expresiones faciales, debido a la misma perturbación de los movimientos que provocan el conocido temblor y la lentitud. Sin embargo, los médicos reconocen cada vez más la presencia y los efectos de otros síntomas de la enfermedad de Parkinson que a veces se denominan *síntomas no motores o resistentes a la dopamina*. Estos síntomas son comunes y pueden afectar en gran medida a los pacientes con enfermedad de Parkinson. Cabe mencionar los siguientes: Deterioro cognitivo, trastornos del estado de ánimo: depresión y ansiedad, problemas para dormir, presión arterial baja al estar de pie, estreñimiento, dificultades en el habla y para tragar, babeo y pérdida del olfato [Poewe (2008)].

Este cuestionario está diseñado con el objetivo de obtener información acerca de los síntomas presentes en los participantes, así como la intensidad de cada uno de ellos. El cuestionario está basado en el *Cuestionario de Calidad de Vida en la Enfermedad de Parkinson (PDQ-39)* [Jenkinson y cols. (2006)]. En éste, se encuentran relacionadas a las funciones motoras, pero también acerca del estado de ánimo de los participantes. Los participantes deberán contestar en una escala del 0 al 10 si el síntoma se encuentra presente y con qué intensidad se presenta, también se tiene la opción de saltar una pregunta si no se desea contestarla o se sienten incómodos. A continuación se muestra un listado de las preguntas realizadas. En la [Figura 5.13] se muestra el cuestionario implementado en la aplicación móvil.



(a)



(b)

Figura 5.13: Cuestionario Evaluación de los Síntomas

5.2. ACTIVIDADES DEL ESTUDIO

TELCEL 0:05 24%

Paso 4 de 9 Cancelar

Sentimientos de miedo y ansiedad

8

0 Nada 10 Mucho

Siguiente

Omitir esta pregunta

(c)

TELCEL 0:05 24%

Paso 5 de 9 Cancelar

Rigidez Muscular

2

0 Nada 10 Mucho

Siguiente

Omitir esta pregunta

(d)

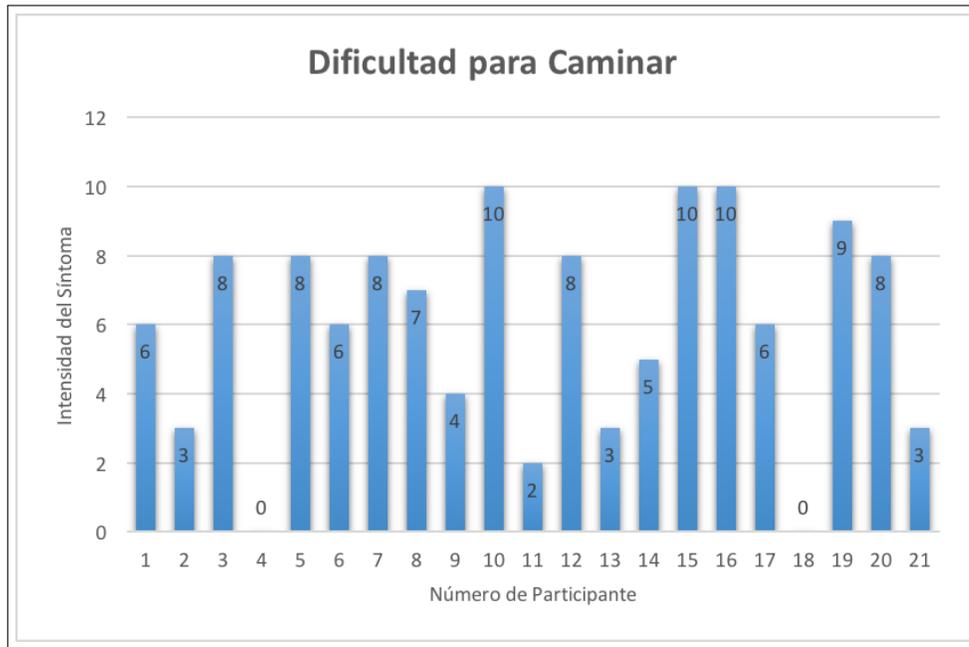
Figura 5.13: Cuestionario Evaluación de los Síntomas



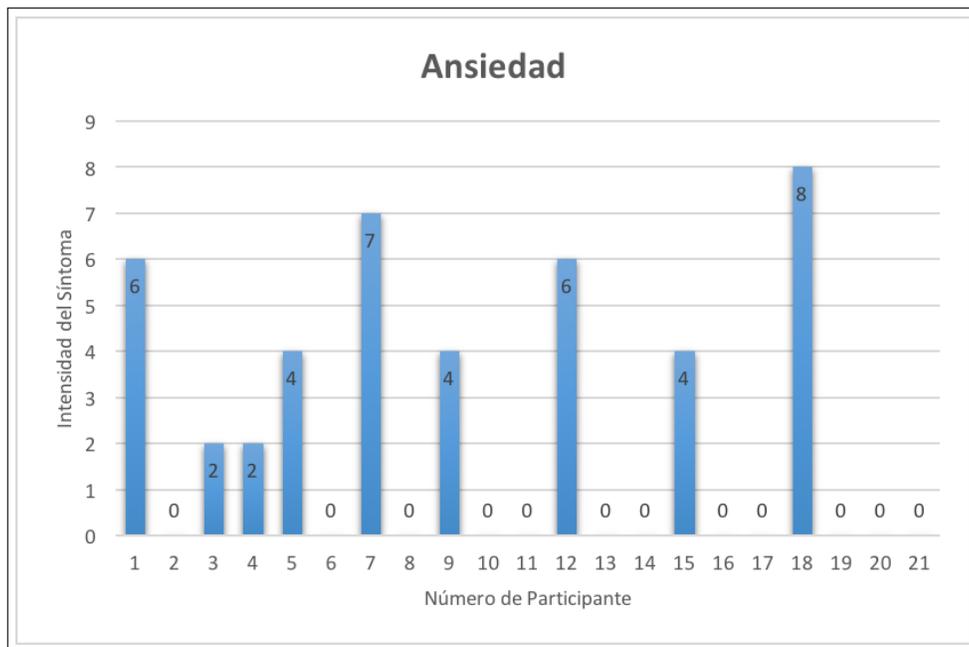
Figura 5.13: Cuestionario Evaluación de los Síntomas

En la [Figura 5.14] se muestran los resultados obtenidos en las pruebas realizadas a 21 participantes con diagnóstico de Enfermedad de Parkinson en el Instituto Nacional de Neurología y Neurocirugía de la Ciudad de México considerando que se encontraban bajo el efecto de los fármacos utilizados para el tratamiento de la enfermedad.

5.2. ACTIVIDADES DEL ESTUDIO

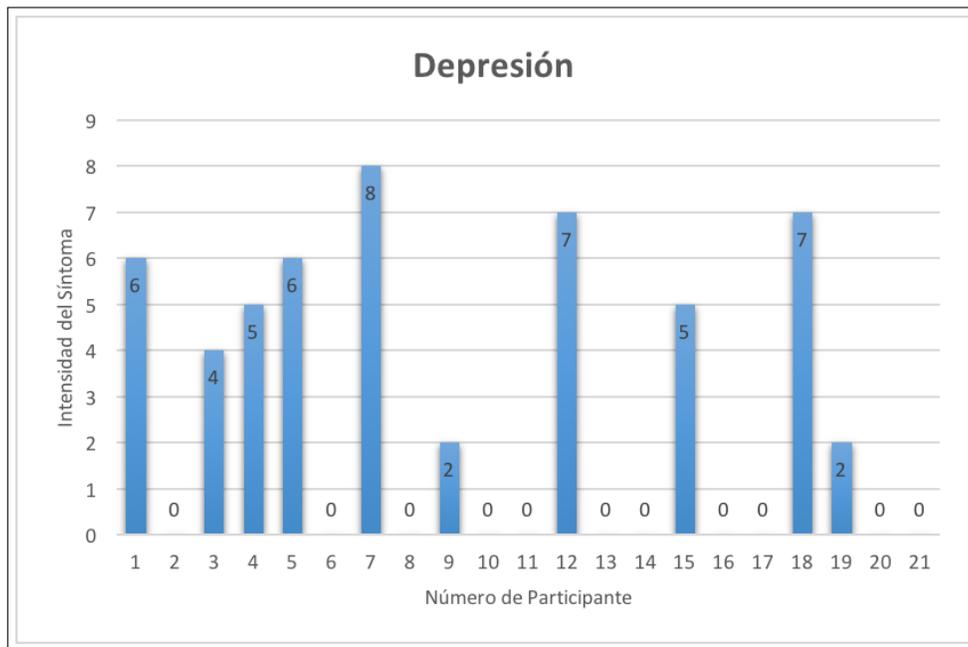


(g)

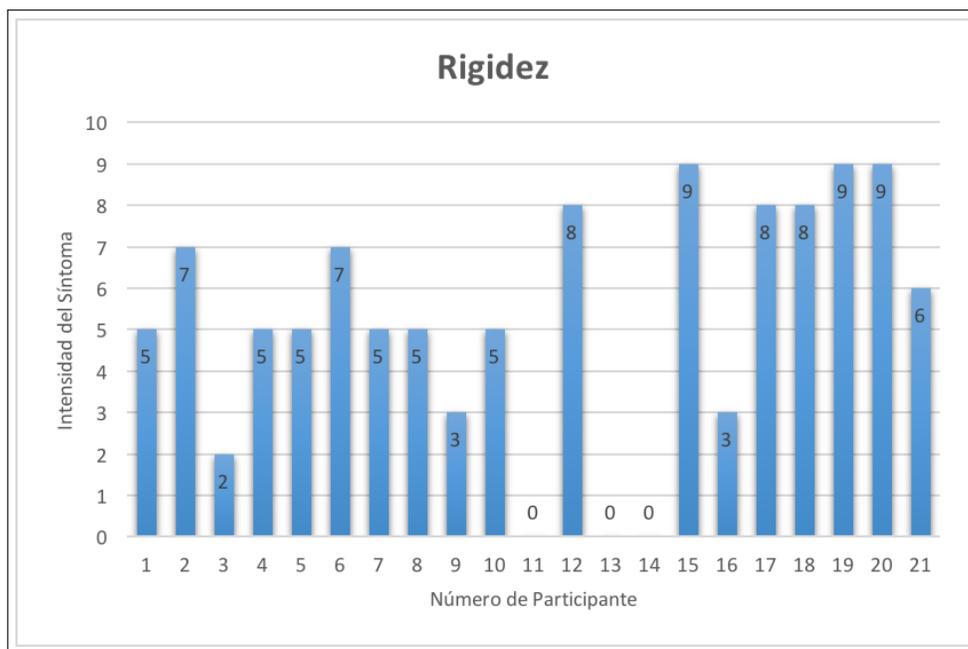


(h)

Figura 5.14: Resultados de las encuestas de acuerdo a la intensidad del síntoma



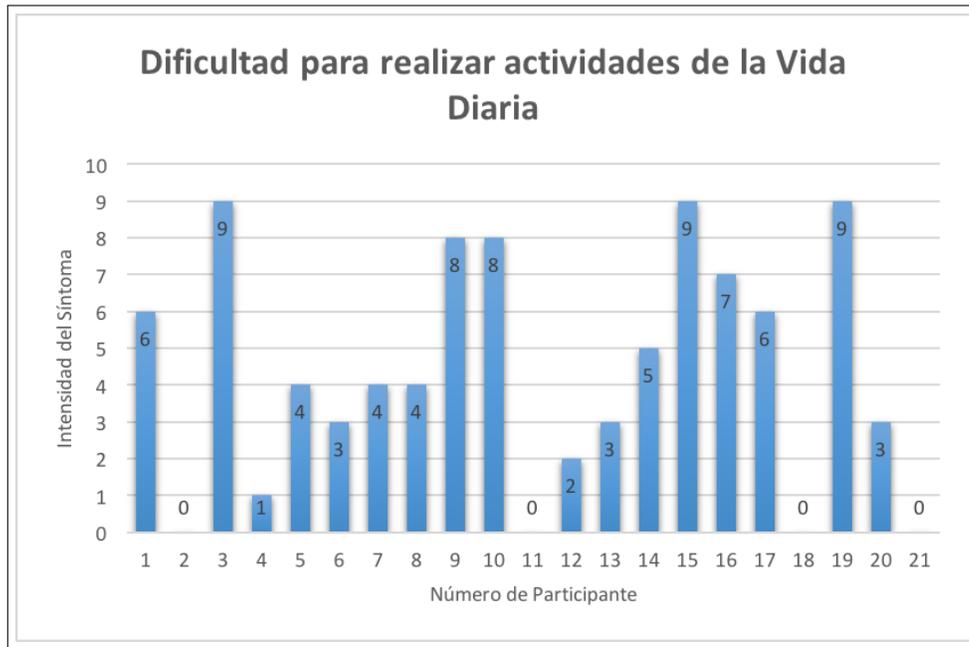
(i)



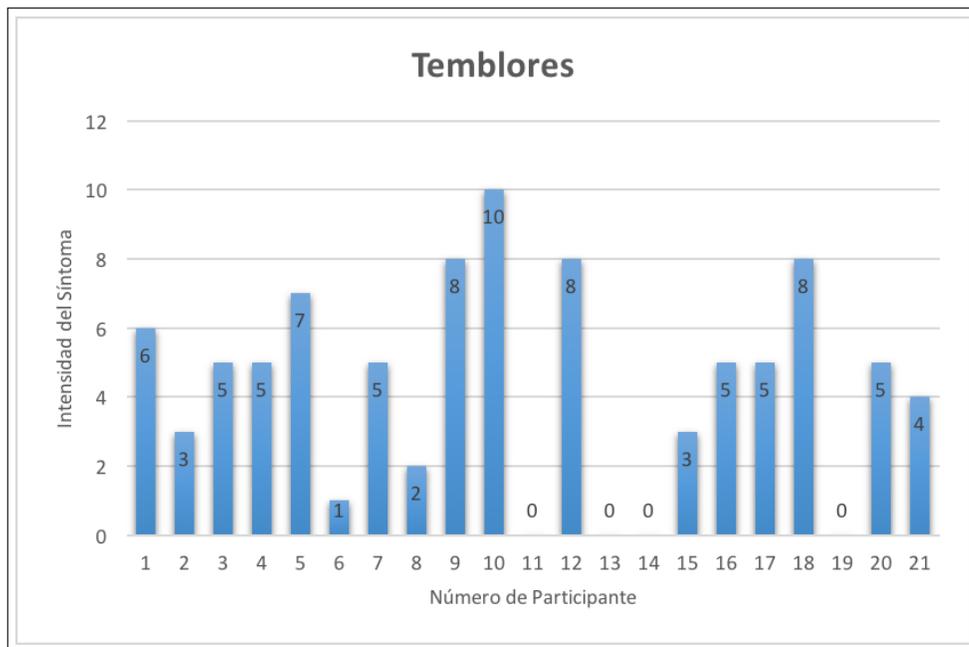
(j)

Figura 5.14: Resultados de las encuestas de acuerdo a la intensidad del síntoma

5.2. ACTIVIDADES DEL ESTUDIO



(k)



(l)

Figura 5.14: Resultados de las encuestas de acuerdo a la intensidad del síntoma

Con base en los resultados obtenidos se obtuvieron los siguientes valores de media y desviación estándar para cada síntoma relacionado con la Enfermedad de Parkinson [Cuadro 5.5].

Síntoma	Media	Desviación Estándar
Ansiedad	2.047	2.765
Dificultad para caminar	5.904	3.144
Dificultad para hablar	4.9523	3.185
Depresión	2.476	2.993
Rigidez	5.190	2.960
Dificultad Actividades Diarias	4.333	3.183
Temblores	4.285	3.002

Cuadro 5.5: Resultados evaluación de los síntomas.

Este cuestionario relacionado con la calidad de vida y los síntomas motores de los pacientes con Enfermedad de Parkinson, demostró que los síntomas son muy variados entre persona y persona, dependiendo del avance de la enfermedad, de la medicación entre otros factores. Es posible que un participante no tenga todos los síntomas relacionados a la enfermedad. Este estudio demostró que una persona puede tener temblores leves pero dificultad para caminar, es decir, los síntomas varían de persona a persona pero también durante el tiempo, pues el mismo participante otro día podría tener temblores importantes pero no problemas con el pensamiento o la memoria. Las pruebas realizadas a pacientes con diagnóstico de Parkinson en el Instituto Nacional de Neurología demostraron que es muy importante tener un control del estado ON/OFF de los participantes, principalmente cuando no se encuentran bajo el efecto de sus medicaciones o bien el tratamiento de Estimulación Cerebral Profunda (DBS) no se encuentra funcionando adecuadamente, los síntomas empeoran y no eran capaces de completar las actividades en el teléfono móvil.

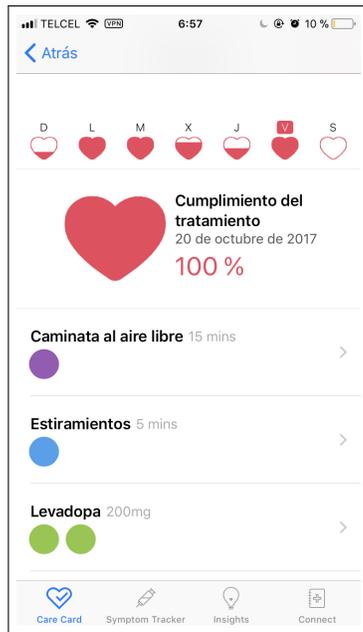
### **5.2.5. Tarjeta de Cuidados**

La finalidad de tener una tarjeta de cuidados personal es que los participantes pueden tener mayor información de su condición médica, pueden hacer un seguimiento regular de sus síntomas y monitorear los niveles de dolor. Es una herramienta visual para los participantes que les permite recordar

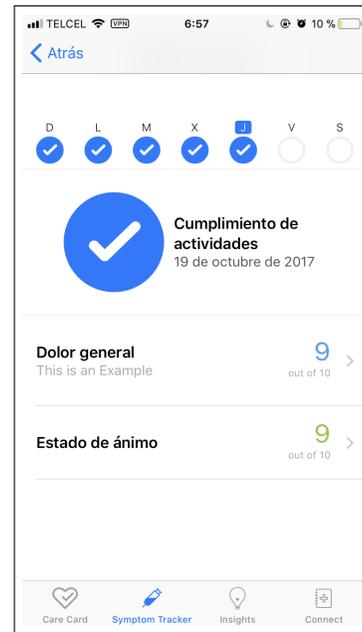
que medicamentos deben tomar durante el día así como dar un seguimiento a actividades motoras que le permitan mejorar su calidad de vida. Cuenta con un reporte a modo de gráfica semanal del cumplimiento de su tratamiento. En la [Figura 5.15] se muestran las interfaces desarrolladas para esta tarjeta de cuidados.

Una vez que el participante ingresa la información acerca de la toma de sus medicamentos prescritos así como algunas actividades como pueden ser Caminata al Aire Libre o Estiramientos, automáticamente se completará el porcentaje de cumplimiento de la actividad hasta llegar al 100 % [Figura 5.15]. Una vez completada esta información el participante puede ingresar manualmente su estado de ánimo así como información si ha presentado algún tipo de dolor general en el cuerpo en una escala del 1 al 10 donde: 0 significa ningún tipo de dolor y 10 significa dolor muy intenso.

Con la información proporcionada por los participantes, la aplicación móvil genera un gráfico semanal que representa la adherencia al medicamento en porcentaje, donde 100 % significa que el participante tomó todas sus medicinas prescritas por el médico así como un comparativo con el dolor presentado en cada día de la semana [Figura 5.15].



(m)



(n)



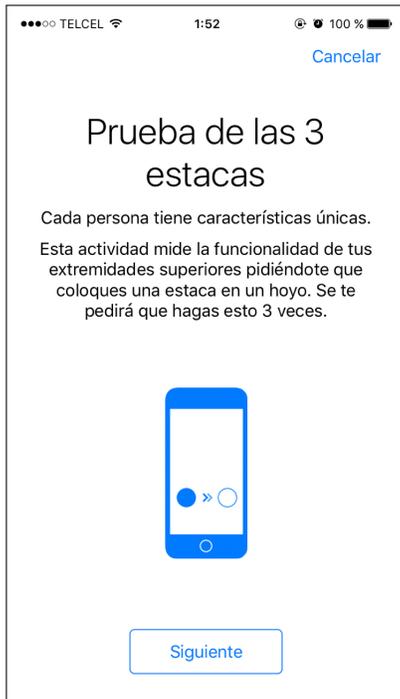
(ñ)

Figura 5.15: Tarjeta de Cuidados

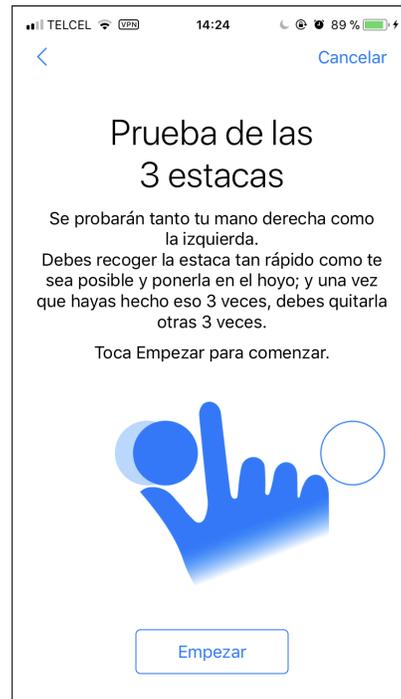
### 5.2.6. Ejercicio Manual. Prueba de las 3 Estacas

La prueba de las estacas es ampliamente utilizada en estudios clínicos para medir el nivel de deterioro manual en las extremidades superiores en varias poblaciones [Haaxma y cols. (2010)]. En individuos con la enfermedad de Pasurkinson la utilidad de esta prueba no ha sido completamente evaluada. Gammon [Earhart y cols. (2011)], encontró que la prueba 9HPT resulta una prueba clínica útil para evaluar la función de la extremidad superior en individuos con EP, teniendo ventajas sobre los métodos utilizados anteriormente debido a la estandarización, valores normativos conocidos por controles sanos, disponibilidad comercial, transportabilidad y facilidad de administración.

Esta actividad marcada como ejercicio permite a los participantes medir la funcionalidad de las extremidades superiores a través de una prueba conocida como Prueba de las 3 estacas. Para ello, deberán recoger una estaca tan rápido como sea posible y colocarla en un hoyo; una vez que se haya colocado 3 veces se deberá quitar otras tres veces. Esta actividad permite interactuar con el dispositivo móvil de manera que se ejerciten los miembros superiores, proporcionando una retroalimentación visual al paciente [Earhart y cols. (2011)]. En la [Figura 5.16] se muestran las interfaces para desarrolladas para esta prueba.



(a)



(b)

Figura 5.16: Prueba de las Estacas

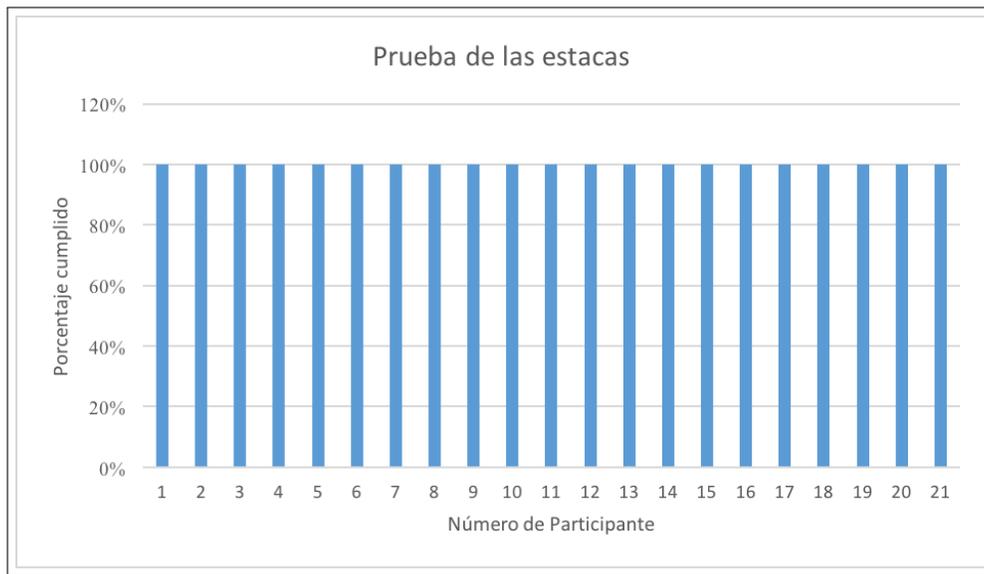


Figura 5.17: Resultados del porcentaje de la Actividad Completada, bajo efecto de medicación.

Los resultados de esta prueba [Figura 5.17] arrojaron que los participantes con diagnóstico de Enfermedad de Parkinson lograron completar la actividad al 100% de las veces cuando se encontraban bajo el efecto de la medicación de Levodopa o bien bajo el efecto del estimulador cerebral profundo conocido como *DBS*, conocido como en estado *ON*. En el caso de que se encontraran en estado *OFF*, es decir sin efecto de la medicación o estimulación cerebral profunda, los participantes eran incapaces de realizar esta actividad, debido al deterioro motor.

### 5.3. Recolección de Datos

Los datos recolectados por la aplicación móvil se guardan en la nube a través de una plataforma denominada Firebase, se contó con apoyo del Instituto Nacional de Neurología en el departamento de Trastorno de Movimiento y Cirugía de Estimulación Cerebral Profunda (*DBS*), para los participantes con diagnóstico de Enfermedad de Parkinson. Los datos generados se guardan en formato *json* para su posterior análisis. El participante debe contar con una conexión a Internet para el envío de la información.

Es importante mencionar que gracias a la colaboración del Instituto Nacional de Neurología, se contó con la participación de 30 pacientes diagnosticados con enfermedad de Parkinson que regularmente asisten a sus citas. El rango de edad de los participantes se encontró desde los 42 años hasta los 80 años. En el caso de los participantes más jóvenes tuvieron un grado de aceptación mayor al uso de teléfonos celulares orientados a monitorear síntomas de la enfermedad de Parkinson. Para los participantes de edad más avanzada quienes no se encuentran acostumbrados al uso de teléfonos móviles en su vida cotidiana, se realizó un entrenamiento para el uso de la pantalla táctil así como una explicación a mayor detalle de las actividades que deberían de realizar. Asimismo todas las actividades se llevaron bajo supervisión para asegurar que se realizaron correctamente. Los participantes mostraron un gran interés así como una motivación por realizar las actividades correspondientes al estudio una vez que leyeron el formulario de consentimiento informado y se mostraban muy entusiasmados por pertenecer a un estudio que tiene la finalidad de generar conocimiento para entender mejor los síntomas relacionados a la enfermedad de Parkinson.

### 5.3.1. Estructura de los Datos

Un total de 9,520 participantes aceptaron unirse al estudio de los cuales 8,003 participantes completaron al menos un cuestionario o tarea después de unirse al estudio. En el [Cuadro 5.6] se muestra el número de participantes por actividad así como el número de muestras que se recogieron para cada instancia. Las características demográficas así como de los participantes auto-registrados con Diagnóstico de EP se observan en el [Cuadro 5.7].

Actividad	No. de Participantes	No. de Muestras
Prueba de Destreza Manual	8003	78887
Prueba de Caminata	3101	35410
Prueba de Memoria Espacial	968	8569

Cuadro 5.6: Número de Participantes por Actividad

Características	
Edad	$\mu = 36,59$
Mujeres	22 %
Hombres	78 %
Participantes Sanos	84 %
Participantes con Diagnostico de Parkinson	16 %
Participantes Fumadores	34 %

Cuadro 5.7: Características de los participantes del estudio

## 5.4. Diseño del Estudio

En esta sección se preparan los datos recolectados por la aplicación móvil y éstos fueron combinados con la base de datos pública de la organización SageBionetworks. La [Figura 5.18] muestra los resultados de la prueba de destreza manual que corresponden a las coordenadas  $(x, y)$  de los botones derecho e izquierdo y la [Figura 5.19] muestra el resultado de las componentes de aceleración triaxial resultado de las mediciones del dispositivo móvil, de un participante sin diagnóstico de la enfermedad de Parkinson.

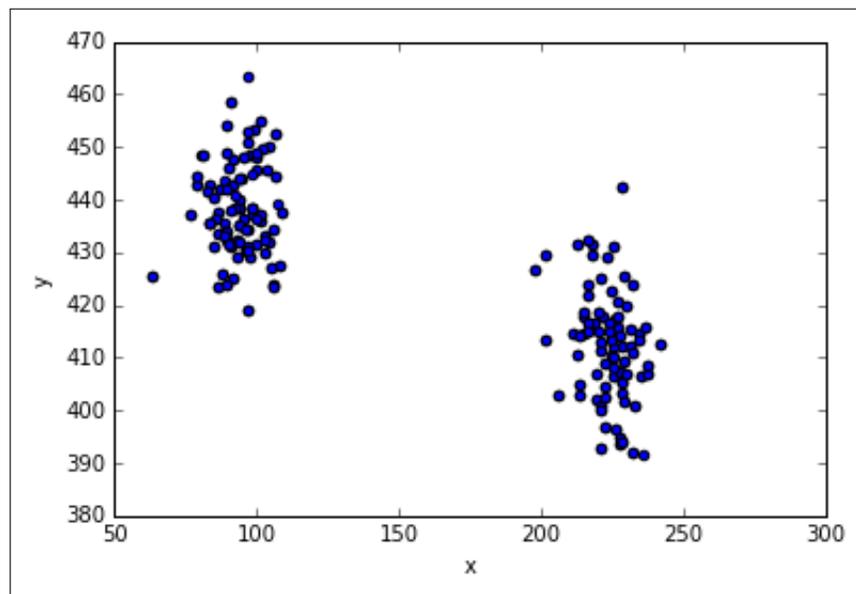
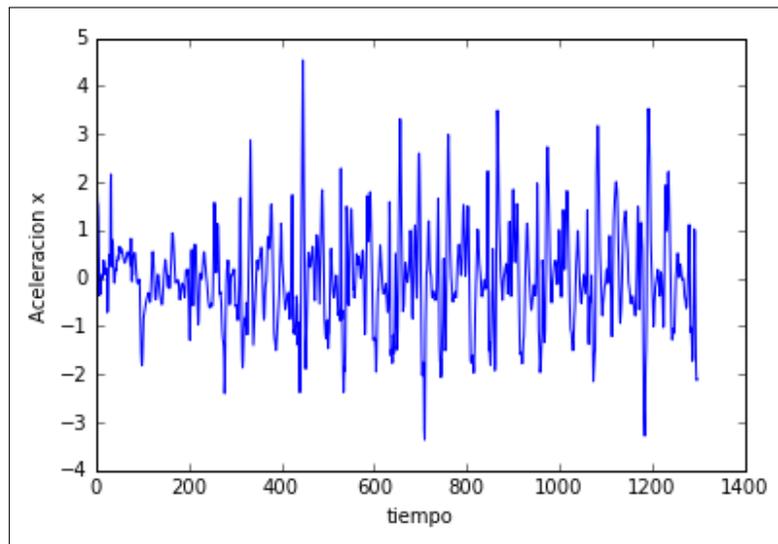
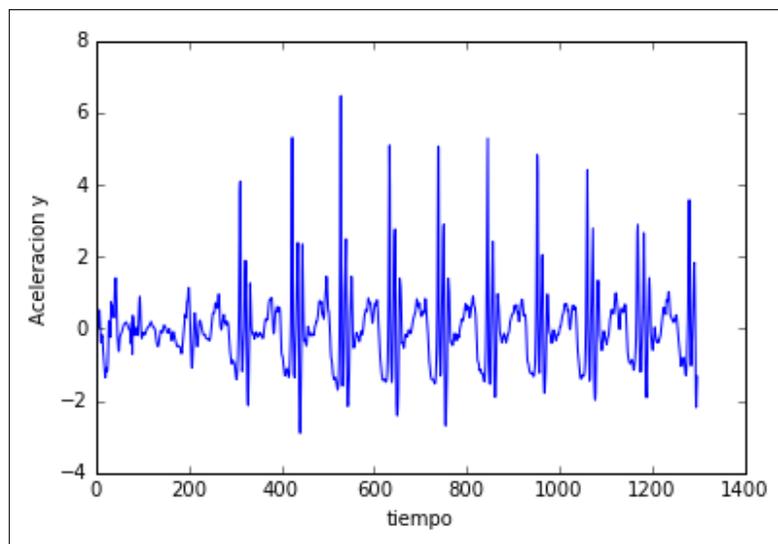


Figura 5.18: Coordenadas Actividad de Destreza Manual



(a) Componente  $x$  de Aceleración



(b) Componente  $y$

Figura 5.19: Componentes de Aceleración de la Prueba de Caminata

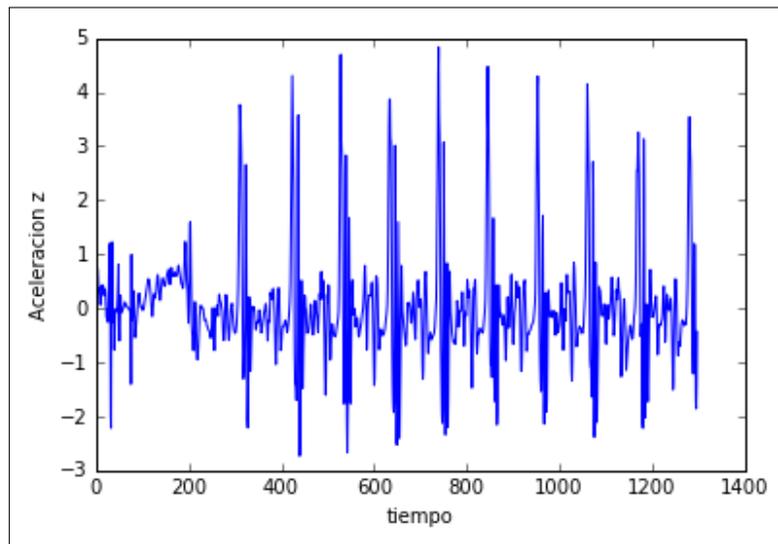
(c) Componente  $z$ 

Figura 5.19: Componentes de Aceleración de la Prueba de Caminata

### 5.4.1. Clasificación de los participantes

De acuerdo con el Capítulo 4 Metodología Propuesta, la etapa de clasificación se realiza una vez que la etapa de procesamiento de los datos se llevó a cabo. Por tanto, se seleccionaron los siguientes algoritmos de clasificación:

1. Regresión Logística (*Logistic Regression*).
2. Red Neuronal (*Neural Network*).
3. Máquina de Soporte Vectorial (*Support Vector Machine*).
4. Árbol de Decisión Impulsado (*Boosted Decision*).

Para validar el desempeño de los clasificadores se utilizó el Método de *k-fold cross validation*, tal como se muestra en la [Figura 5.20]. Los datos de entrenamiento son divididos en  $K$  subconjuntos aleatoriamente, de esos  $K$  subconjuntos se extrae un fragmento que será utilizado como datos de prueba y los demás  $K - 1$  subconjuntos serán utilizados como muestras de entrenamiento. Para este trabajo se utilizó una  $K = 10$  sugerida en trabajos como [Mursalin (2013)].



Figura 5.20: Método de k-fold cross validation con 10 iteraciones

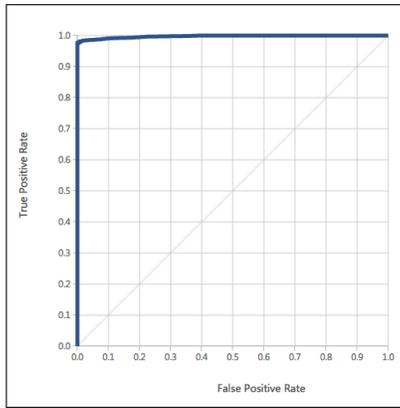
### 5.4.2. Análisis de Resultados Obtenidos

Dentro de los resultados obtenidos se analizaron un total de 34,788 muestras que corresponden a los resultados de la prueba de caminata, después de aplicar las técnicas de preprocesamiento y el algoritmo SMOTE para compensar el desbalance de las clases. Los resultados del desempeño de los clasificadores se muestra en la [Figura 5.8], obteniendo mejores resultados generales para el enfoque basado en Redes Neuronales.

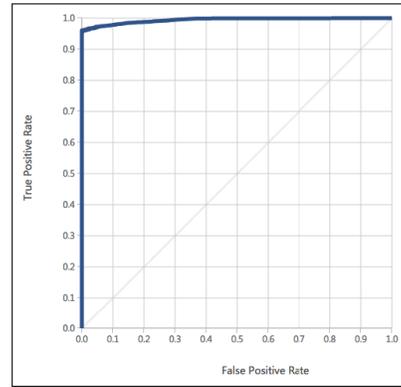
Algoritmo de Clasificación	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0,979	1,0	0,969	0,985
Neural Network	0,995	1,0	0,993	0,997
Support Vector Machine	0,981	0,986	0,986	0,986
<b>Boosted Decision</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>

Cuadro 5.8: Desempeño de los Clasificadores

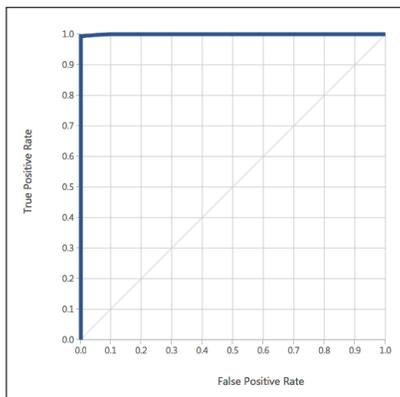
En términos generales el desempeño de los clasificadores se encuentra superior al 90 % y en el caso de *Boosted Decision* se logró identificar en un 100 % el total de las muestras, lo cual muestra que la metodología propuesta es capaz de separar con éxito participantes sanos y aquellos con la enfermedad de Parkinson. Por tanto, se demuestra de esta forma que es posible obtener información objetiva en el diagnóstico de la enfermedad de Parkinson, utilizando aplicaciones móviles y técnicas de aprendizaje automático.



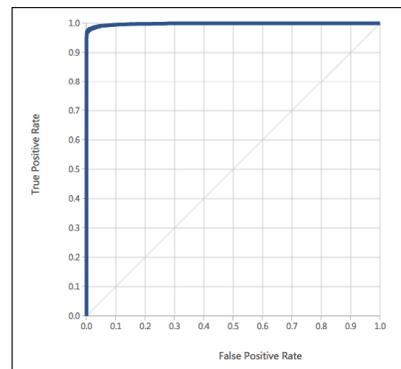
(a) Curva ROC Logistic Regression  
AUC = 0.997



(b) Curva ROC Boosted Decision  
AUC = 1.0



(c) Curva ROC Neural Network  
AUC = 1.0



(d) Curva ROC SVM AUC = 1.0

Figura 5.21: Curvas ROC para los diferentes métodos de clasificación

Las curvas ROC (Receiver Operating Characteristic) presentan la sensibilidad de la prueba que produce resultados continuos, en función de los falsos positivos, lo que es complementario de la especificidad para distintos puntos de corte. En la Figura 5.21 se pueden apreciar estas curvas con sus valores y para cada uno de los métodos de clasificación.

Un espacio ROC se define por FPR y VPR como ejes  $x$  e  $y$  respectivamente y representa los intercambios entre los verdaderos positivos (en

principio, beneficios) y falsos positivos (en principio, costos). Dado que VPR es equivalente a la sensibilidad y FPR es igual a 1-especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a la (1-especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC.

La separación entre los grupos de participantes con Diagnóstico de Enfermedad de Parkinson y sin diagnóstico representa la capacidad discriminativa de una prueba para clasificar sanos como sanos y enfermos como enfermos. Un parámetro para evaluar la bondad de una prueba diagnóstica que produce resultados continuos es el área bajo la curva (AUC). Esta área puede interpretarse como la probabilidad de que ante un par de individuos, uno enfermo y el otro sano, la prueba los clasifique correctamente.

El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada  $(0, 1)$  del espacio ROC, representando un 100 % de sensibilidad (ningún falso negativo) y un 100 % también de especificidad (ningún falso positivo).

A este punto  $(0, 1)$  también se le llama una *clasificación perfecta*. Por tanto, cabe señalar que en los experimentos realizados, en todos los algoritmos de aprendizaje se obtuvieron curvas ROC cercanas a 1, lo cual representa un buen desempeño de los clasificadores.

### 5.4.3. Tiempo de ejecución de los clasificadores.

Las características del sistema de cómputo empleado en el modelo de aprendizaje automático son las siguientes [Cuadro 5.9]:

Sistema Operativo	MacOS Sierra
Procesador	2.4 GHz Intel Core i5
Memoria RAM	8 GB
Disco Duro	256 GB

Cuadro 5.9: Características del sistema de cómputo

En la [Figura 5.22] se muestra de forma gráfica que el modelo que el menor tiempo de ejecución es el *Two Logistic Regression*, mientras que el que

#### 5.4. DISEÑO DEL ESTUDIO

---

tiene un mayor tiempo de cómputo es la *Red Neuronal*. A menudo, el tiempo de entrenamiento depende de la precisión, es decir a mayor precisión un mayor tiempo de entrenamiento, en este caso en particular las redes neuronales tienen un tiempo de entrenamiento mayor con un costo computacional más alto comparado con la Regresión Logística. Para el caso de los árboles de Decisión y Máquinas de Vector soporte también consumen un mayor número de recursos a diferencia de la Regresión Logística. Si el tiempo es limitado, esto puede determinar la elección del algoritmo, especialmente cuando el conjunto de datos es grande.

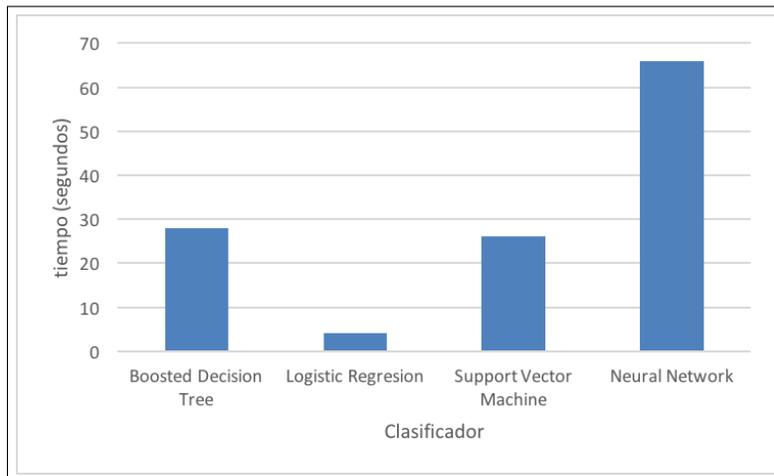


Figura 5.22: Gráfica tiempo de ejecución en segundos.

# Capítulo 6

## Conclusiones y Trabajo a Futuro

En este capítulo se describen las conclusiones de este trabajo, se discuten las limitaciones del mismo y se proponen ideas para una investigación futura.

### 6.1. Conclusiones

La metodología propuesta se centra en el monitoreo no invasivo mediante el uso de un teléfono inteligente, con respecto a los principales síntomas de la enfermedad de Parkinson, como una herramienta objetiva.

En el primer Capítulo se plantearon tres objetivos que se describen a continuación y que se lograron cumplir durante el desarrollo de este trabajo de tesis.

- Se definió una metodología que permite monitorear los síntomas relacionados a la enfermedad de Parkinson, utilizando los sensores inmersos de un dispositivo móvil. A su vez, se realizaron diversos experimentos para lograr que la metodología propuesta cumpla con el objetivo del monitoreo de los síntomas. Por otra parte, la metodología proporciona métricas basadas principalmente en los síntomas motores de la enfermedad, como son el temblor de las manos, problemas al caminar, pero también en el deterioro cognitivo y la calidad de vida de los participantes.

- Se desarrolló una aplicación móvil para iOS que es capaz de monitorear cuantitativamente los síntomas relacionados a la Enfermedad de Parkinson. Esta aplicación móvil fue probada con participantes con Diagnóstico de Parkinson en el Instituto Nacional de Neurología y Neurocirugía obteniendo retroalimentación por parte del departamento clínico en el desarrollo y mejora de la funcionalidad de la aplicación.
- Utilizando algoritmos de aprendizaje automático se clasificaron los datos proporcionados por la aplicación móvil, logrando distinguir participantes con Diagnóstico de Enfermedad de Parkinson y participantes sin diagnóstico hasta con un 100 % de precisión y exactitud. El algoritmo que mejor resultado mostró fue el de *Boosted Decision Tree*.

En la literatura consultada, varios trabajos clasifican participantes con diagnóstico de Enfermedad de Parkinson utilizando los sensores de los dispositivos móviles, sin embargo, los resultados no pueden ser comparados debido a que las condiciones de la prueba difieren en el tiempo tomado para realizar las pruebas y se realizan bajo entornos diferentes.

## 6.2. Limitaciones

- Para realizar las actividades del estudio y tener una mejor comprensión de los síntomas de la Enfermedad de Parkinson, los participantes deberán participar activamente. Sin embargo, para participantes de edad avanzada el uso de teléfonos móviles representa un reto importante pues algunos de ellos no están familiarizados con los dispositivos móviles, así como para los participantes que se encuentran en estado ÓFF' quienes no pueden realizar la prueba.
- Otra limitación es el tamaño de los dispositivos móviles, que en el caso de la prueba de destreza manual, puede representar mayor dificultad para personas de una talla mayor realizar las actividades en una pantalla de menor dimensión.

## 6.3. Trabajo a Futuro

- Desarrollar la aplicación móvil propuesta para sistemas con sistema operativo Android.

- Probar la metodología propuesta con un conjunto mayor de datos de población mexicana, con el objetivo de validar retroalimentar y mejorar el desempeño del sistema.
- Permitir que la aplicación móvil pueda monitorear el registro de voz de los pacientes.
- Añadir actividades en el estudio que permitan monitorear otros síntomas no motores de la enfermedad de Parkinson, tales como el diagnóstico del sueño.
- Utilizando los datos recopilados por la aplicación móvil, es necesario mostrar un seguimiento de la progresión de la enfermedad de Parkinson y obtener métricas objetivas para este propósito.

# Referencias

- Ahmed, F. E. (2005, 06 de Aug). Artificial neural networks for diagnosis and survival prediction in colon cancer. *Molecular Cancer*, 4(1), 29. Descargado de <https://doi.org/10.1186/1476-4598-4-29> doi: 10.1186/1476-4598-4-29
- Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., y Little, M. A. (2016, 2017/10/25). Detecting and monitoring the symptoms of parkinson's disease using smartphones: A pilot study. *Parkinsonism & Related Disorders*, 21(6), 650–653. Descargado de <http://dx.doi.org/10.1016/j.parkreldis.2015.02.026> doi: 10.1016/j.parkreldis.2015.02.026
- Ashendorf, L., Vanderslice-Barr, J. L., y McCaffrey, R. J. (2009). Motor tests and cognition in healthy older adults. *Applied Neuropsychology*, 16(3), 171-176. Descargado de <http://dx.doi.org/10.1080/09084280903098562> (PMID: 20183169) doi: 10.1080/09084280903098562
- Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., ... Trister, A. D. (2016, mar). The mpower study, parkinson disease mobile data collected using researchkit. *Scientific Data*, 3, 160011 EP -.
- Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., y Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics*, 5(3), 229–229. Descargado de <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3959919/> doi: 10.5210/ojphi.v5i3.4814
- Boulos, M. N. K., Wheeler, S., Tavares, C., y Jones, R. (2011, 05 de Apr). How smartphones are changing the face of mobile and participatory

- healthcare: an overview, with example from ecaalyx. *BioMedical Engineering OnLine*, 10(1), 24. Descargado de <https://doi.org/10.1186/1475-925X-10-24> doi: 10.1186/1475-925X-10-24
- Breiman, L. (2017, nov). *Random forest*. Descargado de <https://www.stat.berkeley.edu/users/breiman/randomforest2001.pdf>
- Broen, M. P. G., Marsman, V. A. M., Kuijf, M. L., Van Oostenbrugge, R. J., van Os, J., y Leentjens, A. F. G. (2016, 03). Unraveling the relationship between motor symptoms, affective states and contextual factors in parkinson's disease: A feasibility study of the experience sampling method. *PLOS ONE*, 11(3), 1-11. Descargado de <https://doi.org/10.1371/journal.pone.0151195> doi: 10.1371/journal.pone.0151195
- Burnay, E., Cruz-Correia, R., Jacinto, T., Sousa, A. S., y Fonseca, J. (2013, Jan). Challenges of a mobile application for asthma and allergic rhinitis patient enablement-interface and synchronization. *Telemed J E Health*, 19(1), 13–18. doi: 10.1089/tmj.2012.0020
- Butt, A. H., Rovini, E., Esposito, D., Rossi, G., Maremmanni, C., y Cavallo, F. (2017). Biomechanical parameter assessment for classification of parkinson's disease on clinical scale. *International Journal of Distributed Sensor Networks*, 13(5), 1550147717707417. Descargado de <https://doi.org/10.1177/1550147717707417> doi: 10.1177/1550147717707417
- Chaudhuri, K. R., Martinez-Martin, P., Schapira, A. H., Stocchi, F., Sethi, K., Odin, P., ... Olanow, C. W. (2006). International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for parkinson's disease: The nmsquest study. *Movement Disorders*, 21(7), 916–923. Descargado de <http://dx.doi.org/10.1002/mds.20844> doi: 10.1002/mds.20844
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., y Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. Descargado de <http://dx.doi.org/10.1890/07-0539.1> doi: 10.1890/07-0539.1
- Dang T., O. H., Antolin P. (2001, mar). Fiscal implications of ageing: projections of age-related spending. *Documento de Trabajo del Departamento de Economía de la OCDE*, 3, 160011 EP -.

## REFERENCIAS

---

- de Fisioterapeutas. AI Macías Jiménez V Crespo Sánchez M Morales Cabezas, R. C. A. E. (2004). *Fisioterapia : órgano de la asociación española de fisioterapia*. Editorial Garsi.
- de Lau, L. M., y Breteler, M. M. (2006). Epidemiology of parkinson's disease. *The Lancet Neurology*, 5(6), 525–535.
- del Rey, M. L., Vela-Desojo, L., y de la Cuerda, R. C. (2017). Aplicaciones móviles en la enfermedad de parkinson: una revisión sistemática. *Neurología*. Descargado de <http://www.sciencedirect.com/science/article/pii/S0213485317301639> doi: <https://doi.org/10.1016/j.nrl.2017.03.006>
- DeMaagd, G., y Philip, A. (2015, 08). Parkinson's disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis. *Pharmacy and Therapeutics*, 40(8), 504–532. Descargado de <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4517533/>
- Demidowich, A. P., Lu, K., Tamler, R., y Bloomgarden, Z. (2012). An evaluation of diabetes self-management applications for android smartphones. *Journal of Telemedicine and Telecare*, 18(4), 235-238. Descargado de <https://doi.org/10.1258/jtt.2012.111002> (PMID: 22604278) doi: 10.1258/jtt.2012.111002
- Diener, C. (2017, nov). *Bosques aleatorios (random forests)*. Descargado de <http://bis.ifc.unam.mx/es/ensenanza/metodos-computacionales-para-clasificacion/presentacion-random-forests>
- Earhart, G., T Cavanaugh, J., Ellis, T., P Ford, M., Foreman, K., y Dibble, L. (2011, 12). The 9-hole peg test of upper extremity function: Average values, test-retest reliability, and factors contributing to performance in people with parkinson disease. , 35, 157-63.
- Ellis, R. J., Ng, Y. S., Zhu, S., Tan, D. M., Anderson, B., Schlaug, G., y Wang, Y. (2015, 10). A validated smartphone-based assessment of gait and gait variability in parkinson's disease. *PLOS ONE*, 10(10), 1-22. Descargado de <https://doi.org/10.1371/journal.pone.0141694> doi: 10.1371/journal.pone.0141694

- Ericson, G. (2017, nov). *Cómo elegir algoritmos para aprendizaje automático de microsoft azure*. Descargado de <https://docs.microsoft.com/es-es/azure/machine-learning/studio/algorithm-choice>
- Eskofier, B. M., Lee, S. I., Daneault, J. F., Golabchi, F. N., Ferreira-Carvalho, G., Vergara-Diaz, G., ... Bonato, P. (2016, Aug). Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment. En *2016 38th annual international conference of the ieee engineering in medicine and biology society (embc)* (p. 655-658). doi: 10.1109/EMBC.2016.7590787
- Farley, B., y Clark, W. (1954, September). Simulation of self-organizing systems by digital computer. *Transactions of the IRE Professional Group on Information Theory*, 4(4), 76-84. doi: 10.1109/TIT.1954.1057468
- Figueras, G., Parra, V., Huerta, M., Marzinotto, A., Clotet, R., González, R., ... Sanchez, L. E. (2015). Smartphone application for quantitative measurement of parkinson tremors. En A. Braidot y A. Hadad (Eds.), *Vi latin american congress on biomedical engineering claiB 2014, paraná, argentina 29, 30 & 31 october 2014* (pp. 785-788). Cham: Springer International Publishing. Descargado de [http://dx.doi.org/10.1007/978-3-319-13117-7\\_200](http://dx.doi.org/10.1007/978-3-319-13117-7_200) doi: 10.1007/978-3-319-13117-7\_200
- Forouzanfar, M., Dajani, H. R., Groza, V. Z., Bolic, M., y Rajan, S. (2010, July). Comparison of feed-forward neural network training algorithms for oscillometric blood pressure estimation. En *4th international workshop on soft computing applications* (p. 119-123). doi: 10.1109/SOFA.2010.5565614
- Foundation, P. D. (2016, nov). *What is parkinson's disease?* Descargado de [http://www.pdf.org/en/about\\_pd](http://www.pdf.org/en/about_pd)
- GARCÍA, G. A. J. (2016). *Modelo para el análisis de indicadores biomecánicos bilaterales, en personas con trastornos del movimiento* (Tesis de Master no publicada). Centro de Investigación en Computación.
- Gil, D., y Johnsson, M. (2009). Diagnosing parkinson by using artificial neural networks and support vector machines. , 9(4), 63-71.
- Gil, R. V., y Páez, D. G. (2007). Identificación de sistemas dinámicos utilizando redes neuronales rbf. *Revista Iberoamericana de Automática e*

- Informática Industrial RIAI*, 4(2), 32 - 42. Descargado de <http://www.sciencedirect.com/science/article/pii/S1697791207702078> doi: [https://doi.org/10.1016/S1697-7912\(07\)70207-8](https://doi.org/10.1016/S1697-7912(07)70207-8)
- Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... LaPelle, N. (2008). Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15), 2129–2170. Descargado de <http://dx.doi.org/10.1002/mds.22340> doi: 10.1002/mds.22340
- Gröger, A., Kolb, R., Schäfer, R., y Klose, U. (2014, 01). Dopamine reduction in the substantia nigra of parkinson's disease patients confirmed by in vivo magnetic resonance spectroscopic imaging. *PLOS ONE*, 9(1), 1-6. Descargado de <https://doi.org/10.1371/journal.pone.0084081> doi: 10.1371/journal.pone.0084081
- Güvenir, H. A., Acar, B., Demiroz, G., y Çekin, A. (1997, Sep). A supervised machine learning algorithm for arrhythmia analysis. En *Computers in cardiology 1997* (p. 433-436). doi: 10.1109/CIC.1997.647926
- Haaxma, C. A., Bloem, B. R., Overeem, S., Borm, G. F., y Horstink, M. W. (2010). Timed motor tests can detect subtle motor dysfunction in early parkinson's disease. *Movement Disorders*, 25(9), 1150–1156. Descargado de <http://dx.doi.org/10.1002/mds.23100> doi: 10.1002/mds.23100
- Hammerla, N. Y., Fisher, J. M., Andras, P., Rochester, L., Walker, R., y Plotz, T. (2015). Pd disease state assessment in naturalistic environments using deep learning. En *Proceedings of the twenty-ninth aai conference on artificial intelligence* (pp. 1742–1748). AAAI Press. Descargado de <http://dl.acm.org/citation.cfm?id=2886521.2886562>
- Herdman, M., Badia, X., y Berra, S. (2001). El euroqol-5d: una alternativa sencilla para la medición de la calidad de vida relacionada con la salud en atención primaria. *Atención Primaria*, 28(6), 425 - 429. Descargado de <http://www.sciencedirect.com/science/article/pii/S0212656701704064> doi: [https://doi.org/10.1016/S0212-6567\(01\)70406-4](https://doi.org/10.1016/S0212-6567(01)70406-4)
- Hosmer, S., David W.; Lemeshow. (2000). *Applied logistic regression* (second ed.). Wiley.

- Hubel, K. A., Reed, B., Yund, E. W., Herron, T. J., y Woods, D. L. (2013). Computerized measures of finger tapping: effects of hand dominance, age, and sex. *Perceptual and motor skills*, 116(3), 929–952.
- Inc., A. (2017a, nov). *Develop health and fitness apps that work together*. Descargado de <https://developer.apple.com/healthkit/>
- Inc., A. (2017b, nov). *Researchkit and carekit empowering medical researchers, doctors, and you*. Descargado de <https://www.apple.com/researchkit/>
- Istepanian, R. S. H., Pattichis, C. S., y Laxminarayan, S. (2006). Ubiquitous m-health systems and the convergence towards 4g mobile technologies. En R. S. H. Istepanian, S. Laxminarayan, y C. S. Pattichis (Eds.), *M-health: Emerging mobile health systems* (pp. 3–14). Boston, MA: Springer US. Descargado de [http://dx.doi.org/10.1007/0-387-26559-7\\_1](http://dx.doi.org/10.1007/0-387-26559-7_1) doi: 10.1007/0-387-26559-7\_1
- J, T., y WJ, M. (2016). Logistic regression: Relating patient characteristics to outcomes. *JAMA*, 316(5), 533-534. Descargado de [+http://dx.doi.org/10.1001/jama.2016.7653](http://dx.doi.org/10.1001/jama.2016.7653) doi: 10.1001/jama.2016.7653
- Jankovic, J. (2008). Parkinson’s disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4), 368–376. Descargado de <http://jnnp.bmj.com/content/79/4/368> doi: 10.1136/jnnp.2007.131045
- Jenkinson, C., Heffernan, C., Doll, H., y Fitzpatrick, R. (2006). The parkinson’s disease questionnaire (pdq-39): evidence for a method of imputing missing data. *Age and Ageing*, 35(5), 497-502. Descargado de [+http://dx.doi.org/10.1093/ageing/af1055](http://dx.doi.org/10.1093/ageing/af1055) doi: 10.1093/ageing/af1055
- Juanjuan Wang, H. W. J. Z., Mantao Xu. (2006). Classification of imbalanced data by using the smote algorithm and locally linear embedding. *ICSP2006 Proceedings*.
- Kehagia, A. A., Barker, R. A., y Robbins, T. W. (s.f., 2017/12/03). Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with parkinson’s disease. *The Lancet Neurology*, 9(12), 1200–1213. Descargado de [http://dx.doi.org/10.1016/S1474-4422\(10\)70212-X](http://dx.doi.org/10.1016/S1474-4422(10)70212-X) doi: 10.1016/S1474-4422(10)70212-X

## REFERENCIAS

---

- Kim, H., Lee, H. J., Lee, W., Kwon, S., Kim, S. K., Jeon, H. S., ... Park, K. S. (2015, Aug). Unconstrained detection of freezing of gait in parkinson's disease patients using smartphone. En *2015 37th annual international conference of the ieee engineering in medicine and biology society (embc)* (p. 3751-3754). doi: 10.1109/EMBC.2015.7319209
- LeMoyné, R., Mastroianni, T., Cozza, M., Coroian, C., y Grundfest, W. (2010, Aug). Implementation of an iphone for characterizing parkinson's disease tremor through a wireless accelerometer application. En *2010 annual international conference of the ieee engineering in medicine and biology* (p. 4954-4958). doi: 10.1109/IEMBS.2010.5627240
- Mandal, I., y Sairam, N. (2014). New machine-learning algorithms for prediction of parkinson's disease. *International Journal of Systems Science*, 45(3), 647-666. Descargado de <http://dx.doi.org/10.1080/00207721.2012.724114> doi: 10.1080/00207721.2012.724114
- Martínez-Jurado Elizabeth, R.-V. M., Cervantes-Arriaga Amin. (2010). *Calidad de vida en pacientes con enfermedad de parkinson*. Revista Mexicana de Neurociencia.
- Mursalin, H. M.-. N. M. K., M. (2013). Performance analysis among different classipier including naive bayes, support vector machine and c4. 5 for automatic weeds classipication. *Global Journal of Computer Science and Technology*, 13.
- MV, M., A, S., A, P., y et al. (2017). Feasibility of obtaining measures of lifestyle from a smartphone app: The myheart counts cardiovascular health study. *JAMA Cardiology*, 2(1), 67-76. Descargado de [+http://dx.doi.org/10.1001/jamacardio.2016.4395](http://dx.doi.org/10.1001/jamacardio.2016.4395) doi: 10.1001/jamacardio.2016.4395
- Nitेश V. Chawla, L. O. H. W. P. K., Kevin W. Bowyer. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16.
- N. Sáenz Bajoa, M. Á. B. (2012). Redes neuronales: concepto, aplicaciones y utilidad en medicina. *Atención Primaria*.

- Pan, D., Dhall, R., Lieberman, A., y Petitti, B. D. (2015, 26 de Mar). A mobile cloud-based parkinson's disease assessment system for home-based monitoring. *JMIR mHealth uHealth*, 3(1), e29. Descargado de <http://mhealth.jmir.org/2015/1/e29/> doi: 10.2196/mhealth.3956
- Pan, J. I., y Huang, Y. C. (2015, July). Intelligent fall prevention for parkinson's disease patients based on detecting posture instability and freezing of gait. En *Informatics in control, automation and robotics (icinco), 2015 12th international conference on* (Vol. 01, p. 608-613).
- Parkinson, I., y Society, M. D. (2016, nov). *Mds rating scales*. Descargado de <http://www.movementdisorders.org/MDS/Education/Rating-Scales.htm>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peto, V., Jenkinson, C., y Fitzpatrick, R. (1998, 01 de Apr). Pdq-39: a review of the development, validation and application of a parkinson's disease quality of life questionnaire and its associated measures. *Journal of Neurology*, 245(1), S10–S14. Descargado de <https://doi.org/10.1007/PL00007730> doi: 10.1007/PL00007730
- Pineda, G. F. (2017, nov). *Curso de inteligencia artificial introducción al aprendizaje automático*. Descargado de <http://turing.iimas.unam.mx/~luis/cursos/IA2013-2/slides/s12.aprendizaje.pdf>
- Poewe, W. (2008). Non-motor symptoms in parkinson's disease. *European Journal of Neurology*, 15, 14–20. Descargado de <http://dx.doi.org/10.1111/j.1468-1331.2008.02056.x> doi: 10.1111/j.1468-1331.2008.02056.x
- Powers, D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Printy, B. P., Renken, L. M., Herrmann, J. P., Lee, I., Johnson, B., Knight, E., ... Whitmer, D. (2014, Aug). Smartphone application for classification of motor impairment severity in parkinson's disease. En *2014 36th annual*

## REFERENCIAS

---

- international conference of the ieee engineering in medicine and biology society* (p. 2686-2689). doi: 10.1109/EMBC.2014.6944176
- Provost, F., y Fawcett, T. (2001, marzo). Robust classification for imprecise environments. *Mach. Learn.*, 42(3), 203–231. Descargado de <http://dx.doi.org/10.1023/A:1007601015854> doi: 10.1023/A:1007601015854
- Quinlan, J. (1986). Induction of decision trees. *Kluwer Academic Publishers, Boston*.
- Ramani, D. R. G., y Sivagami, G. (2011, October). Article:parkinson disease classification using data mining algorithms. *International Journal of Computer Applications*, 32(9), 17-22. (Full text available)
- Reitan R. M., W. D. (1993). *The halstead-reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZNeuropsychology Press.
- Resnick, R., Halliday, D., y Krane, K. (2002). *Física* (n.º v. 2). Compañía Editorial Continental. Descargado de <https://books.google.com.mx/books?id=L27MSgAACAAJ>
- Rodríguez, A. G. (2017). *Análisis de señales biomecánicas multiaxiales* (Tesis de Master no publicada). Centro de Investigación en Computación.
- Rokach, L., y Maimon, O. (2008). *Data mining with decision trees*. World Scientific.
- Rokach, L., y Maimon, O. (2014). *Data mining with decision trees: theory and applications*. World scientific.
- Saad, A., Zaarour, I., Zeinedine, A., Ayache, M., Bejjani, P., Guerin, F., y Lefebvre, D. (2013). A preliminary study of the causality of freezing of gait for parkinson’s disease patients: Bayesian belief network approach..
- Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., ... Quattrone, A. (2014). Machine learning on brain mri data for differential diagnosis of parkinson’s disease and progressive supranuclear palsy. *Journal of Neuroscience Methods*, 222(Supplement C), 230 - 237. Descargado de <http://www.sciencedirect.com/science/article/pii/S0165027013003993> doi: <https://doi.org/10.1016/j.jneumeth.2013.11.016>

- Samà, A., Pérez-Lopez, C., Romagosa, J., Rodríguez-Martín, D., Català, A., Cabestany, J., ... Rodríguez-Moliner, A. (2012, Aug). Dyskinesia and motor state detection in parkinson's disease patients with a single movement sensor. En *2012 annual international conference of the ieee engineering in medicine and biology society* (p. 1194-1197). doi: 10.1109/EMBC.2012.6346150
- Schapira, A. H. V. (2005). Present and future drug treatment for parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(11), 1472–1478. Descargado de <http://jnnp.bmj.com/content/76/11/1472> doi: 10.1136/jnnp.2004.035980
- Scholkopf, B., y Smola, A. J. (2001). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA, USA: MIT Press.
- Sposaro, F., Danielson, J., y Tyson, G. (2010, Aug). iwander: An android application for dementia patients. En *2010 annual international conference of the ieee engineering in medicine and biology* (p. 3875-3878). doi: 10.1109/IEMBS.2010.5627669
- Stoica, P., y Moses, R. (2005). *Spectral analysis of signals*. Prentice Hall.
- Trister, A. D., Dorsey, E. R., y Friend, S. H. (2016, 03 de 03). Smartphones as new tools in the management and understanding of parkinson's disease. *Npj Parkinson's Disease*, 2, 16006 EP -. Descargado de <http://dx.doi.org/10.1038/npjparkd.2016.6>
- Velarde-Jurado, E., y Avila-Figueroa, C. (2002, 09). Consideraciones metodológicas para evaluar la calidad de vida. *Salud pública Méx vol.44 no.5, 44*, 448 - 463. Descargado de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0036-36342002000500010&nrm=iso](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342002000500010&nrm=iso)
- Vinuesa, P. (2017, nov). *Regresión lineal simple y polinomial: teoría y práctica*. Descargado de [http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9\\_regresion.pdf](http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9_regresion.pdf)
- WP, G., JD, B., VD, B., S, S., y JC, M. (1998). Cognitive and motor functioning in parkinson disease: Subjects with and without questionable dementia. *Archives of Neurology*, 55(5), 674-680. Descargado de [+http://](http://)

- [dx.doi.org/10.1001/archneur.55.5.674](https://doi.org/10.1001/archneur.55.5.674) doi: 10.1001/archneur.55.5.674
- Xia, R., y Mao, Z.-H. (2012, 02). Progression of motor symptoms in parkinson's disease. *Neuroscience Bulletin*, 28(1), 39–48. Descargado de <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5560285/> doi: 10.1007/s12264-012-1050-z
- Yan, H., Huo, H., Xu, Y., y Gidlund, M. (2010, November). Wireless sensor network based e-health system ?? implementation and experimental results. *IEEE Transactions on Consumer Electronics*, 56(4), 2288-2295. doi: 10.1109/TCE.2010.5681102
- Ying Wang, Y. F. (2009). High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. *Neuro-Image Elsevier*.
- Zell, A. (2014). *Simulation neuronaler netze [simulation of neural networks]*. Addison-Wesley.

# Anexos

## 6.4. Requerimientos Funcionales

RF-01	Registro en el sistema por parte del usuario	Un paciente podrá registrarse directamente en el sistema introduciendo su correo electrónico, contraseña, edad y si ha sido diagnosticado con la enfermedad de Parkinson. El sistema no validará la existencia real del correo ni la autenticidad del diagnóstico
RF-02	Inicio de Sesión	Inicio de sesión de cualquier usuario registrado.
RF-03	Registro de Actividad de Destreza	El paciente podrá realizar una prueba de destreza manual desde la aplicación iOS que permite medir el nivel de temblor de las manos de un paciente
RF-04	Registro de Actividad de Caminata	El paciente podrá realizar una prueba de caminata, con la finalidad de medir el ciclo de marcha, mediante el uso de acelerómetros desde la aplicación iOS
RF-05	Registro de Actividad de Memoria 1	El paciente podrá completar un juego de memoria llamado 'Simón Dice' desde la aplicación iOS que permite medir el nivel de deterioro de la memoria, falta de criterio y confusión.
RF-06	Registro de Actividad de Memoria 2	El paciente podrá completar un juego de memoria llamado 'Memorama' desde la aplicación iOS que permite medir el nivel de deterioro de la memoria, falta de criterio y confusión.
RF-07	Almacenamiento de las Pruebas en la Nube.	Las pruebas registradas por los pacientes o usuarios se enviarán a la Plataforma 'Firebase' para su posterior análisis
RF-08	Procesamiento y Clasificación de los resultados obtenidos.	A través de los resultados obtenidos por la plataforma 'Firebase' y un algoritmo de clasificación, el sistema podrá identificar a los pacientes con enfermedad de Parkinson de aquellos que no.
RF-09	Cerrar Sesión	Cierre de sesión de cualquier usuario registrado.

Cuadro 6.1: Requerimientos Funcionales

## 6.5. Requerimientos No Funcionales

RNF-01	Interfaz Gráfica Intuitiva	La interfaz debe ser intuitiva para que los pacientes con enfermedad de Parkinson puedan acceder sin dificultad y seguir los procedimientos que se indican
RNF-02	Multiusuario	Se requiere que el sistema soporte la interacción con más de un usuario a la vez
RNF-03	Registro de pruebas organizado	Que cada registro de actividad que realicen los usuarios sea almacenado en las categorías pertinentes.
RNF-04	Acceso Vía Internet	Se requiere que todo usuario (registrado o no) tenga acceso al sistema si está conectado a internet.
RNF-04	La aplicación funcionará sobre el SO iOS	Para que los usuarios puedan hacer uso de la aplicación deberán contar con Sistema Operativo iOS en sus celulares.

Cuadro 6.2: Requerimientos No Funcionales

## 6.6. Arquitectura de la Aplicación

La aplicación del usuario, está basada en un Modelo Vista-Controlador. La descripción de cada uno de los tres componentes (modelo, vista y controlador) se describe a continuación:

1. *Modelo*: En esta capa se encontrará la base de datos y la forma en cómo se maneja la información obtenida del usuario y la aplicación propia. Otra de las funcionalidades que se encuentra en la capa es el proceso al cual únicamente tiene acceso el usuario para realizar las actividades de monitoreo de la Enfermedad de Parkinson.
2. *Vista*: Es la capa en la que se encuentra la interfaz de usuario y los elementos con los cuales interactúa el usuario de la aplicación.
3. *Controlador*: Es la capa en la que se transfiere la información a los modelos a partir de las peticiones del usuario, también en esta capa se encuentran los controladores para el funcionamiento de la aplicación, el marco de referencia para la base de datos, autenticación y recursos para el uso de los diferentes sensores con que cuenta el dispositivo móvil.

## 6.7. Análisis de Casos de Uso

A continuación se describe el actor que interviene en la funcionalidad del sistema, así como las operaciones mediante las cuales interactúan con este. Se identifica el actor **usuario** y su operación del sistema en el cuadro 6.3.

Nombre	Usuario
Descripción	Es la entidad principal del sistema, cualquier persona podrá registrarse como este tipo de usuario, sin importar si tienen la enfermedad de Parkinson o no. La cuenta de usuario tendrá los privilegios de realizar cuatro pruebas desde la aplicación móvil, como son una prueba de destreza manual, de caminata y dos juegos de memoria.
Comentarios	El usuario deberá estar registrado en el sistema para realizar las pruebas de progreso de la Enfermedad de Parkinson.

Cuadro 6.3: Descripción del actor Usuario

A continuación se muestra el diagrama de Caso de Uso del Actor Usuario en la Figura 6.1.

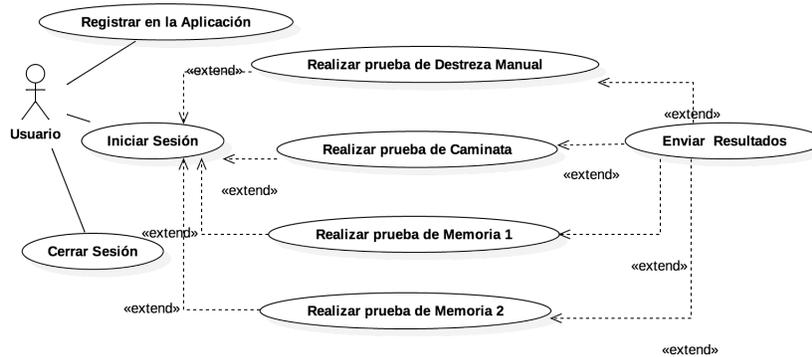


Figura 6.1: Diagrama Casos de Uso Usuario

En el diagrama de la Figura 6.1 se observa un nivel de abstracción general del módulo con el cual interactúa directamente el usuario. El *Actor* adquiere una serie de posibilidades de interactuar con la aplicación. A continuación se realiza una descripción de los casos de uso a un nivel más detallado de acuerdo con lo especificado en el diagrama previo.

## 6.8. Descripción Detallada de los Casos de Uso

En los cuadros siguientes se realiza la descripción detallada de los casos de uso descritos en la Figura 6.1.

Nombre Caso de Uso	Iniciar Sesión
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Inicio de Sesión, el usuario puede estar o no registrado en el sistema.
Flujo de Eventos	2. El usuario ingresa su correo electrónico y contraseña en los campos marcados 3. El sistema autentifica que el formato introducido es correcto. 4. El sistema autentifica que es un usuario registrado.
Condiciones de Salida	5. La aplicación iOS muestra un mensaje de inicio de Sesión Exitoso o que los datos ingresados no se encuentran en la base de datos o el formato es erróneo
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.4: Caso de Uso Iniciar Sesión

Nombre Caso de Uso	Registrar Nuevo Usuario
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Inicio. El usuario no está registrado en el sistema
Flujo de Eventos	2. El usuario presiona el botón Registrar en el Sistema 3. El usuario introduce sus datos como son correo electrónico, contraseña, edad, ubicación geográfica, si es fumador y si cuenta con diagnóstico de Enfermedad de Parkinson 4. La aplicación autentifica que el formato de los datos introducidos es correcto. 5. La aplicación autentifica que el usuario no se encuentre registrado anteriormente. 6. La aplicación guarda los datos en la Base de Datos Firebase
Condiciones de Salida	7. La aplicación iOS muestra un mensaje de Registro Exitoso y regresa a la pantalla de Inicio de Sesión.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.5: Caso de Uso Registrar Nuevo Usuario

Nombre Caso de Uso	Cerrar Sesión
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Actividades, el usuario debe estar registrado en el sistema.
Flujo de Eventos	2. El usuario presiona el botón cerrar sesión 3. El sistema detiene el envío de información y regresa al Menú Inicio
Condiciones de Salida	4. La aplicación iOS muestra la pantalla de Inicio de Sesión con los campos usuario y contraseña en blanco para iniciar sesión o registrarse.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.6: Caso de Uso Cerrar Sesión

Nombre Caso de Uso	Actividad de Destreza Manual
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Actividades. El usuario está registrado en el sistema
Flujo de Eventos	2. El usuario presiona el botón Actividad de Destreza Manual 3. El sistema muestra un mensaje con las instrucciones a realizar 4. El usuario realiza la actividad de presionar dos botones alternadamente durante un período de 20 segundos 5. La aplicación registra el número de veces que se presionaron los botones y las coordenadas geométricas de donde se presionaron los botones. 6. La aplicación guarda los datos de la Actividad en la Base de Datos Firebase
Condiciones de Salida	7. La aplicación iOS muestra un mensaje de que se completó la actividad y el usuario puede realizar una nueva Actividad o Regresar al Menú Actividades.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.7: Caso de Uso Actividad de Destreza Manual

Nombre Caso de Uso	Actividad de Caminata
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Actividades. El usuario está registrado en el sistema
Flujo de Eventos	2. El usuario presiona el botón Actividad de Caminata 3. La aplicación muestra un mensaje con las instrucciones a realizar 4. El usuario realiza la actividad de caminata durante un periodo de 20 segundos. 5. La aplicación registra los valores de lectura del acelerómetro, giroscopio, tiempo de muestreo y magnetómetro. 6. La aplicación guarda los datos de la Actividad en la Base de Datos Firebase
Condiciones de Salida	7. La aplicación iOS muestra un mensaje de que se completó la actividad y el usuario puede realizar una nueva Actividad o Regresar al Menú Actividades.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.8: Caso de Uso Actividad de Caminata

## 6.8. DESCRIPCIÓN DETALLADA DE LOS CASOS DE USO

Nombre Caso de Uso	Actividad de Memoria 1
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Actividades. El usuario está registrado en el sistema
Flujo de Eventos	2. El usuario presiona el botón Actividad de Memoria 1. 3. La aplicación muestra la secuencia a seguir de un juego de memoria llamado 'Simón Dice', 4. El usuario presiona los elementos de acuerdo con la secuencia del juego. 5. El sistema verifica que la secuencia introducida por el usuario coincide con la establecida. 6. Si la secuencia coincide, se incrementa el nivel y se repite desde el paso 4. Si el usuario introduce una secuencia incorrecta el sistema repite el nivel en el que se encuentra. 7. El sistema muestra la puntuación obtenida y el tiempo en que se realizó. 8. La aplicación guarda los datos de la Actividad en la Base de Datos Firebase y envía un mensaje
Condiciones de Salida	9. El usuario presiona terminar juego. La aplicación iOS regresa al Menú Actividades.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.9: Caso de Uso Actividad de Memoria 1

Nombre Caso de Uso	Actividad de Memoria 2
Actores Participantes	Usuario
Condiciones Iniciales	1. Pantalla de Menú Actividades. El usuario está registrado en el sistema
Flujo de Eventos	2. El usuario presiona el botón Actividad de Memoria 2. 3. La aplicación muestra el tablero de juego de 'Memorama' 4. El usuario selecciona y descubre una pareja de imágenes aleatoriamente. 5. El sistema verifica que las imágenes descubiertas sean iguales. 5. Si las imágenes son iguales, las imágenes quedan iluminadas y se procede a encontrar otra pareja de imágenes iguales. 6. Si el par de imágenes descubiertas no son iguales, vuelven a mostrarse ocultas. 7. Cuando se descubren todas las parejas iguales, el sistema muestra la puntuación obtenida y el tiempo en que se realizó. 8. La aplicación guarda los datos de la Actividad en la Base de Datos Firebase
Condiciones de Salida	9. El usuario presiona terminar juego y la aplicación iOS regresa al Menú Actividades.
Requerimientos Especiales	El usuario deberá contar con una conexión a Internet

Cuadro 6.10: Caso de Uso Actividad de Memoria 2

## 6.9. Diagrama de Actividades

En el diagrama de actividades de la Figura 6.2 se muestra el funcionamiento general del sistema.

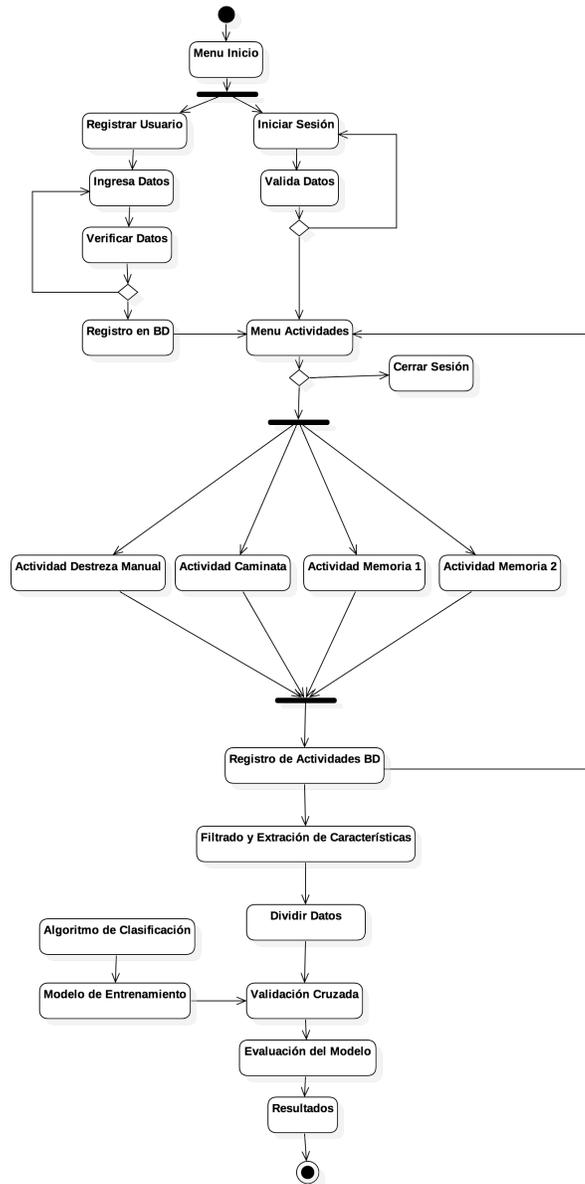


Figura 6.2: Diagrama de Actividades del Sistema

## 6.10. Diagrama de Clases

La definición de los diagramas de clases permite definir la estructura de la aplicación, mostrando sus clases, atributos y las relaciones entre ellos. Los diagramas de clases son utilizados durante el proceso de análisis y diseño, donde se crea el diseño conceptual de la información que se manejará en la aplicación y los componentes que se encargarán del funcionamiento. En la Figura 6.3 se muestra el Diagrama de Clases del Sistema.

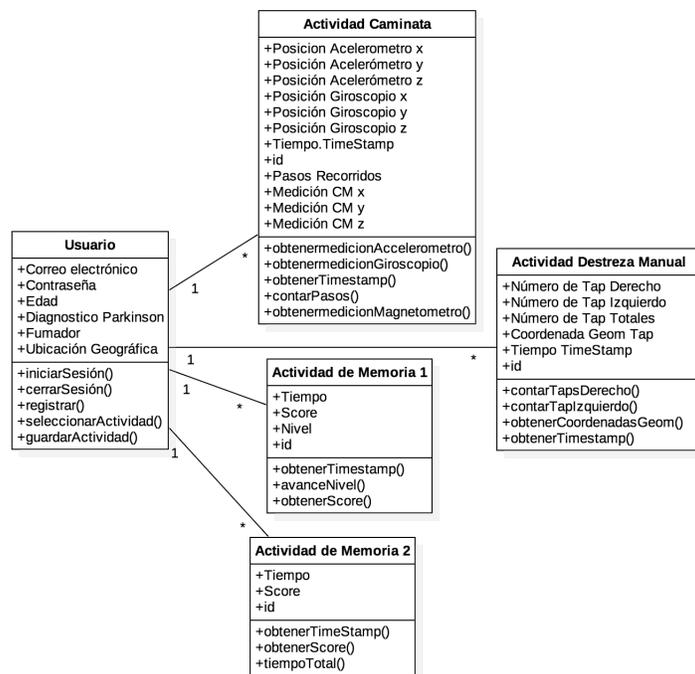


Figura 6.3: Diagrama de Clases del Sistema

En el diagrama se presentan las diferentes clases con sus respectivas relaciones, así como los atributos y operaciones de cada una de ellas. La definición de las clases permitirá definir cómo estará estructurada, diseñada y cuáles serán las relaciones más importantes de la aplicación.

## 6.11. Diagrama de Secuencia

En esta sección se muestran los diagramas de secuencia correspondientes a los casos de uso ya mencionados en la sección anterior; aquí se puede observar el flujo de los datos entre las diferentes clases, dependiendo de las trayectorias principales y alternas de cada uno de los casos de uso.

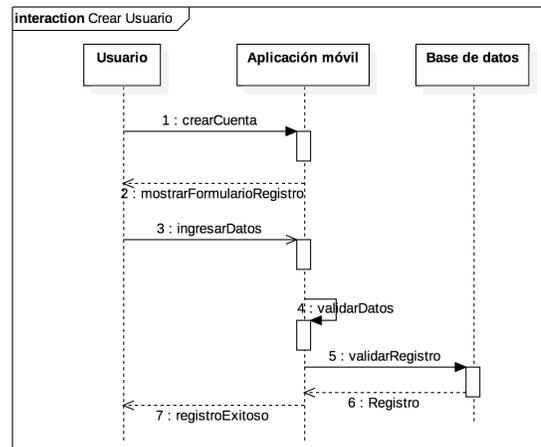


Figura 6.4: Diagrama de Secuencia - Registrar Usuario.

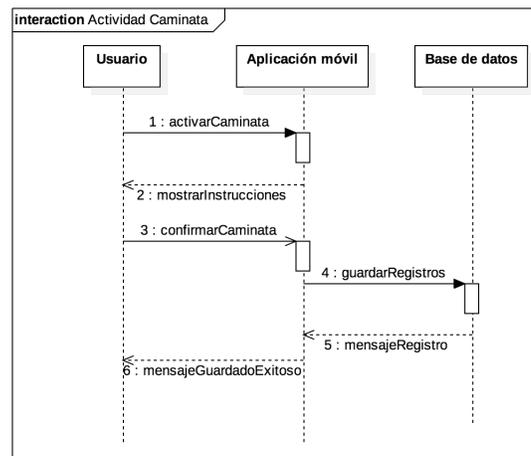


Figura 6.5: Diagrama de Secuencia - Actividad Caminata

## 6.11. DIAGRAMA DE SECUENCIA

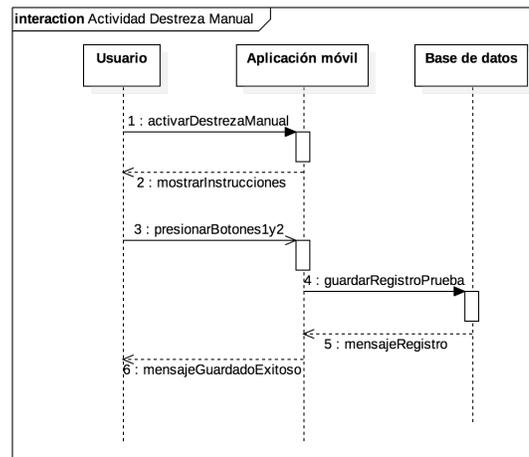


Figura 6.6: Diagrama de Secuencia - Actividad de Destreza Manual

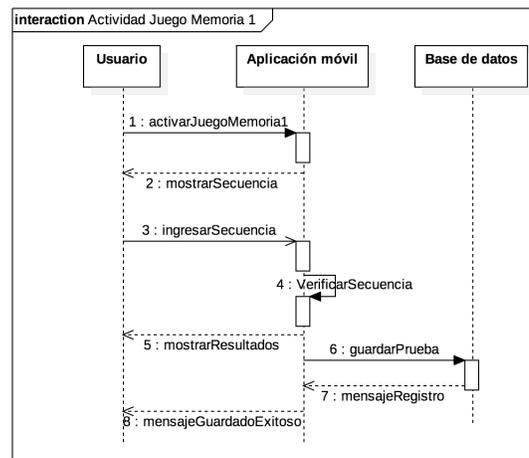


Figura 6.7: Diagrama de Secuencia - Actividad Juego de Memoria 1

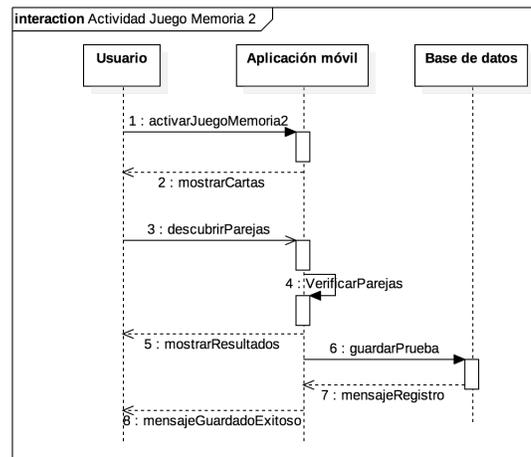


Figura 6.8: Diagrama de Secuencia - Actividad Juego de Memoria 2