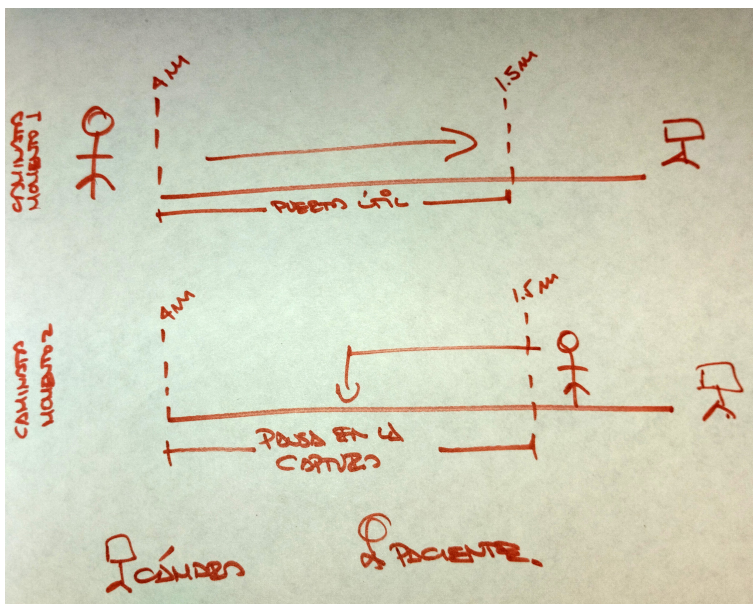


Avances en la aplicación de la ingeniería a la valoración de personas con la enfermedad de Parkinson



Melissa García Peña
Carlos Alfonso Herrán Sánchez
Jonatan Ordoñez Burbano
Christian Camilo Urcuqui
Andrés Navarro Cadavid

**Avances en la aplicación de la ingeniería
a la valoración de personas con la
enfermedad de Parkinson**

**Avances en la aplicación de la ingeniería
a la valoración de personas con la
enfermedad de Parkinson**

**Melissa García Peña
Carlos Alfonso Herrán Sánchez
Jonatan Ordoñez Burbano
Christian Camilo Urcuqui
Andrés Navarro Cadavid**

Avances en la aplicación de la ingeniería a la valoración de personas con la enfermedad de Parkinson. Melissa García Peña, Carlos Alfonso Herrán Sánchez, Jonatan Ordoñez Burbano, Christian Camilo Urcuqui y Andrés Navarro Cadavid.

1 ed. Cali, Colombia. Universidad Icesi, 2023

112 p., 17x24 cm

Incluye referencias bibliográficas

ISBN: 978-628-7630-06-2

<https://doi.org/10.18046/EUI/iys.4.2023>

1. Parkinson disease 2. Early diagnosis 3. Medical informatics application (MeSH) I.Tit 616.28 – dc22

© Universidad Icesi, 2023

Facultad de Ingeniería

Rector: Esteban Piedrahita Uribe

Decana Facultad de Ingeniería: Norha Villegas

Coordinador editorial: Adolfo A. Abadía



Edición, producción y diseño: José Ignacio Claros V.

Imagen de portada: Nicolás Javier Salazar / Jonatan Ordoñez

Publicado en Colombia / *Published in Colombia*

El contenido de esta obra no compromete el pensamiento institucional de la Universidad Icesi ni le genera responsabilidades legales, civiles, penales o de cualquier otra índole frente a terceros.

 Editorial
Universidad
Icesi

Calle 18 #122-135 (Pance), Cali-Colombia

editorial@icesi.edu.co

www.icesi.edu.co/editorial

Teléfono: +57 (602) 555 2334

Este cuarto volumen de la colección Ingeniería & Salud, tal como sus antecesores, reporta trabajos conjuntos realizados por el Grupo de Investigación en Informática y Telecomunicaciones (i2t) y el grupo de investigación Biomédica de la Universidad Icesi, junto con el Grupo de Investigación en Parkinson y Transtornos del Movimiento de la Fundación Valle del Lili.

Melissa García Peña

Máster en Ciencia de Datos e Ingeniera de Sistemas de la Universidad Icesi (Cali, Colombia). Se desempeña como business analyst - product owner en Perficient Latin America. Sus áreas de interés profesional están centradas en la gestión de proyectos y en la búsqueda de soluciones mediante la ciencia de datos. melisa.garcia.pena@gmail.com

Carlos Alfonso Herrán Sánchez

Magister en Ciencia de Datos de la Universidad Icesi (Cali, Colombia), con pregrado en Ingeniería Informática y Especialización en Gerencia Informática. Se desempeña como arquitecto de software en Tecnoquimicas S.A. Sus áreas de interés profesional son: las clean architectures, el domain-driven design y el desarrollo de software a partir de metodologías ágiles. c.herran@outlook.com

Jonatan Ordoñez Burbano

Ingeniero de Sistemas de Software con Maestría en Informática y Telecomunicaciones (énfasis en investigación) de la Universidad Icesi (Cali, Colombia). Se desempeña como desarrollador fullstack en Patagonian, empresa donde tiene a su cargo la implementación de módulos en React, Typescript y Express para el soporte, mantenimiento y creación de una plataforma de lenguaje conversacional en línea. Jonatan-Ordonez@hotmail.com

Christian Camilo Urcuqui

Máster en Informática y Telecomunicaciones e Ingeniero de Sistemas con énfasis en Administración e Informática de la Universidad Icesi (Cali, Colombia), Especialista en Deep Learning y Natural Language Processing; científico de datos en la industria de software; docente investigador, miembro del Grupo de Investigación en Informática y Telecomunicaciones (i2t) y Coordinador del Club de Hacking de la Universidad Icesi. Sus áreas de interés profesional son: ciberseguridad, ciencia de datos y e-salud. ulcamilo@gmail.com

Andrés Navarro Cadavid

Doctor Ingeniero en Telecomunicaciones de la Universidad Politécnica de Valencia (España), Máster en Gestión Tecnológica e Ingeniero Electrónico de la Universidad Pontificia Bolivariana (Medellín, Colombia); profesor de tiempo completo y director del Grupo de Investigación en Informática y Telecomunicaciones (i2t) de la Universidad Icesi (Cali, Colombia); investigador senior (Minciencias); miembro senior del IEEE y Director Regional de IEEE COMSOC LATAM (2022-2024); consultor internacional; y miembro de Grupo de Estudio 1 de la Unión Internacional de Telecomunicaciones. anavarro@icesi.edu.co

Tabla de contenido

Nota del editor	13
Presentación	17
Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson.	19
1. Introducción	20
2. Marco teórico	21
3. Estado del arte	26
4. Método	32
5. La investigación	36
6. Hallazgos, conclusiones y trabajo futuro	50
7. Referencias	52
Índice de tablas	58
Índice de figuras	59
Anexo 1. Exploración del dataset	61
Anexo 2. Diccionario de variables de balanceo de brazos y marcha	69
Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson	73
1. Introducción	74
2. Marco teórico	80
3. Estado del arte	83
4. Metodología	88
5. Resultados	93
6. Discusión	100
7. Conclusiones y trabajo futuro	102
8. Referencias	104
Índice de tablas	109
Índice de figuras	110

Siglas

AIC	Akaike Information Criterion
ASUM-DM	Analytics Solutions Unified Method for Data Mining
CAOBA	Centro de Excelencia y Apropiación en Big Data y Data Analytics
CRISP-DM	Cross Industry Standard Process for Data Mining
CURS	Columbia University Rating Scale
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
ECA	Ensayos Controlados Aleatorios
EEG	ElectroEncephalography
EMEP	Escala de la Marcha para la Enfermedad de Parkinson
FFT	Fast Fourier Transform
FoG	Freezing of Gait
k-NN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
MDS-UPDRS	MDS-Unified Parkinson's Disease Rating Scale
NUDS	Nothwestern University Disability Scale
OFDM	Orthogonal Frequency Division Multiplexing
PCI	Parkinson Cognitive Impairment
RGBD	Red, Green, Blue + Deep
SBT	Split Belt Training
SDLC	Software Development Life Cycle
SIMD	Single Instruction, Multiple Data
SVM	Support Vector Machine
UPDRS	Unified Parkinson's Disease Rating Scale

Nota del editor

Con este volumen se completan cuatro ediciones de Ingeniería & Salud, colección que presenta resultados de investigación realizadas conjuntamente entre los grupos de investigación i2t y Biomédica de la Universidad Icesi, en asocio con la Fundación Valle del Lili desde 2012, y que se ha enfocado en el desarrollo de soluciones tecnológicas de bajo costo que apoyen el diagnóstico y seguimiento de la Enfermedad de Parkinson (EP) y otras enfermedades neurodegenerativas de alta prevalencia.

“Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de Parkinson”, el primer volumen de esta colección, recogió una serie de publicaciones que habían sido presentadas en reconocidos eventos científicos internacionales y agregó dos capítulos inéditos, todo ello con el fin de presentar de manera ordenada los resultados obtenidos hasta la fecha de su edición en sus investigaciones sobre las posibilidades y formas de uso de un dispositivo electrónico de juegos (Kinect™) aplicado como una herramienta de captura de movimientos capaz de complementar de manera acertada, oportuna y relevante a la evaluación clínica que realizan los profesionales de la salud para diagnosticar la EP o monitorear aspectos propios de su evolución, y aportarle elementos cuantitativos. En el curso de esta investigación, no solo se probó la utilidad del dispositivo, sino que se desarrolló una serie de herramientas que hacen posible su aprovechamiento. Por su bajo costo y eficacia, el producto representa un aporte significativo a los sistemas de salud pública de los países en desarrollo, pues pone a su alcance herramientas tecnológicas de baja complejidad operativa, que no requieren de grandes inversiones y son adecuadas para su uso en la consulta clínica.

“Identificación de marcadores clínicos, cognitivos y motores en pacientes con enfermedad de Parkinson en estadios tempranos: evaluación clínica

complementada con un dispositivo de análisis del movimiento”, el segundo libro de la colección presenta un avance en esa misma línea de investigación, pero hace mayor énfasis en los temas médicos, con lo cual facilita el entendimiento de la enfermedad a profesionales ajenos al campo de la salud. A partir del reconocimiento del valor de usar dispositivos tecnológicos dentro del entorno clínico para obtener medidas objetivas, como una herramienta eficaz para el seguimiento, pronóstico y personalización del manejo de la enfermedad, esta investigación se enfocó en la determinación de los marcadores clínicos cognitivos y motores que caracterizan el compromiso de personas con enfermedad de Parkinson en estadios tempranos. El texto reporta los resultados de un estudio desarrollado con veinticinco personas de 45 a 87 años de edad con enfermedad de Parkinson (e igual número de controles pareados por edad) en el cual: se realizaron análisis descriptivos para las variables clínicas; se usó la prueba de Mann-Whitney para comparar las variables continuas entre grupos y los análisis estratificados por edad; y se realizó un análisis de regresión logística para determinar los marcadores clínicos cognitivos y motores que explican el compromiso temprano de pacientes con EP. La investigación mostró cómo la evaluación clínica, junto con la implementación del dispositivo de análisis del movimiento, permiten cuantificar las diferencias motoras sutiles en los pacientes con EP en estadios tempranos, algo de gran valor si se considera la provechosa relación existente entre detección temprana y calidad de vida.

“Enfermedad de Parkinson: análisis motor desde la ingeniería”, el tercer volumen de esta colección, mantiene su enfoque en aportar elementos que reduzcan el nivel de subjetividad en la valoración que hacen los especialistas de salud para el diagnóstico de la enfermedad de Parkinson o el monitoreo de su evolución y tratamiento, mediante el uso de instrumentos no tradicionales, y presenta los resultados de dos proyectos de investigación: el primero, “Sistema de detección de anomalías motoras a partir del análisis del balanceo de los brazos”, que busca establecer un método objetivo de medición de las variables de la marcha a través del desarrollo de un sistema hardware-software capaz de detectar anomalías motoras a partir de la observación del movimiento de las extremidades superiores, utilizando acelerómetros, que tiene como novedad su enfoque en el movimiento de los brazos (no de las piernas, como en las ediciones anteriores); el segundo, “Técnicas de análisis de señales para el uso de e-Motion en el diagnóstico y seguimiento de la enfermedad de Parkinson”, que va un paso más allá y complementa todos los esfuerzos anteriores, reconoce la necesidad de definir técnicas de análisis que permitan transformar la información obtenida con los mecanismos reportados en estas investigaciones en valores objetivos

que puedan ser usados en el contexto clínico, para lo cual se enfoca en proveer técnicas de análisis de señales de marcha y balanceo de los brazos, y en generar algoritmos de procesamiento que pudieran ser incluidos en los análisis de pacientes.

El volumen actual surge al no hallar estudios previos que determinen si existe una relación entre las extremidades inferiores y superiores de las personas que padecen la enfermedad de Parkinson. Como problema de investigación se definió la ausencia de un modelo de marcha completo, que incluya las cuatro extremidades y permita discriminar si un paciente padece o no la enfermedad de Parkinson. En su desarrollo se evaluó un modelo de clasificación estadístico (regresión logística), varios de machine learning (árboles de decisión, naïve Bayes, random forest y support vector machine) e inferencia causal. La investigación permitió evidenciarla relación entre las extremidades superiores e inferiores de las personas cuando presentan enfermedad del Parkinson e identificar un efecto causal entre unas pocas variables de las extremidades inferiores con una única variable de los brazos. La relación causal, así como el uso de técnicas de clasificación y aprendizaje de máquina son una aproximación novedosa al problema, cuyo análisis clínico se publicó en el volumen 16 de la revista *Frontiers in Human Neuroscience* [1].

José Ignacio Claros V.

Presentación

Desde 2010, el Grupo de Investigación en Informática y Telecomunicaciones (i2t) de la Universidad Icesi y el Grupo de Investigación en Parkinson y Trastornos del Movimiento de la Fundación Valle del Lili vienen trabajando en la identificación y desarrollo de alternativas tecnológicas basadas en las ingenierías telemática y de sistemas, que faciliten la detección temprana y la valoración de la evolución de la enfermedad de Parkinson a costos accesibles para el estándar latinoamericano. Sus avances han sido reportados a la comunidad científica en: conferencias, a través de publicaciones en revistas nacionales e internacionales y en tres volúmenes de la colección Ingeniería & Salud editada por la Universidad Icesi.

Al desarrollo de estas investigaciones se han unido, como reflejo de una política institucional, estudiantes de maestría de la Universidad Icesi, quienes se han integrado a i2t y han enfocado sus trabajos de grado en el aporte al desarrollo de soluciones que continúen y complementen los esfuerzos de los dos grupos de investigación. Es el caso de Melissa García Peña, Carlos Alfonso Herrán Sánchez y Jonatan Ordoñez Burbano, cuyos trabajos han servido de base para la preparación de este cuarto volumen de la colección Ingeniería & Salud.

La primera sección de este libro, “Modelo que relaciona los datos de las extremidades de un paciente con posible diagnóstico de enfermedad de Parkinson”, aborda un vacío, la constatación de la existencia o no de una relación entre las extremidades inferiores y las extremidades superiores de las personas que padecen esta enfermedad. Con el apoyo de expertos y de modelos de selección, como forward y backward, se eligieron las variables más representativas de un grupo de características disponibles en una base de datos de pacientes y se evaluaron varios modelos de clasificación; además, se probaron modelos estadísticos de regresión y herramientas de

aprendizaje de máquina (machine learning) para identificar personas con la enfermedad. Esta investigación encontró que random forest, uno de los más populares algoritmos de machine learning, obtuvo los mejores resultados, con un nivel de predicción de 83.6 %, y permite afirmar, ahora sí con evidencia cuantitativa, que existe una relación entre las cuatro extremidades. Si bien en el proyecto únicamente se logra identificar un efecto causal entre unas pocas variables de extremidades inferiores con una única variable de los brazos (coeficiente de asimetría), los resultados abren un camino para continuar la exploración de esta relación que sirve de base para el concepto de inferencia causal.

La segunda sección de este libro, “Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson”, está dirigido más hacia un tema de usabilidad de la investigación realizada durante estos doce años, en la medida en que sus resultados le brindan autonomía al personal de salud en el manejo de la información técnica que arrojan las pruebas de marcha realizadas con e-motion, el sistema desarrollado por los grupos de investigación citados. Durante este proyecto, se desarrolló un sistema de visualización que permite presentar los resultados de la marcha al personal médico sin necesidad de la participación de un ingeniero para la manipulación y preparación de los datos y el despliegue de los resultados. La implementación de este sistema de procesamiento y visualización automatiza el flujo de análisis de la marcha y facilita el acceso fácil y rápido a información segura, útil en el diagnóstico y la valoración de la evolución de la enfermedad. SI bien la investigación publicada en este libro solo aborda los resultados obtenidos con los primeros sensores Kinect v1, al momento de escribir estas líneas ya se ha avanzado en la inclusión de otros sensores disponibles en el mercado.

Andrés Navarro Cadavid
Director del Grupo de Investigación i2t

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

Melissa García Peña, Carlos Alfonso Herrán Sánchez, Christian
Camilo Urcuqui, Andrés Navarro Cadavid

Universidad Icesi Cali, Colombia

1. INTRODUCCIÓN

La enfermedad del Parkinson está catalogada como el segundo trastorno neurodegenerativo más común a nivel mundial [2], [3], se caracteriza por provocar el deterioro neurológico progresivo o la muerte neuronal, su principal trastorno bioquímico es la disminución del contenido de dopamina de la sustancia negra y del núcleo estriado [4]. La EP afecta directamente la calidad de vida de los pacientes y el movimiento de su cuerpo, por lo general, los síntomas comienzan de forma gradual [5] y se distinguen por presentar rigidez, temblor de reposo e inestabilidad en la postura [2], [6].

El diagnóstico de la enfermedad de Parkinson solo puede ser realizado por un médico experto, especializado en esta enfermedad, mediante el uso de escalas o métricas estándar de evaluación del paciente, tales como: la escala de calificación de la enfermedad de Parkinson unificada, de la Sociedad de Trastornos del Movimiento (MDS-UPDRS); y la escala de Hoehn y Yahr (H&Y). Ellas ayudan a medir: las características motoras, tales como la velocidad al caminar, la longitud de la zancada y la cadencia del movimiento; las características mentales; y el estado conductual o psicológico del paciente [4], [5], [6]. Dicho análisis o diagnóstico es subjetivo pues depende de la observación del médico que lo realice [2], [5], [6].

El diagnóstico de esta enfermedad se ha estudiado de manera independiente, no de forma homogénea, ni se ha revisando cada aspecto o cada variable en conjunto, por ejemplo, las extremidades (brazos y piernas) se han analizado por separado, pero no se cuenta con un resultado sobre la relación que pueda existir entre las extremidades superiores e inferiores para el caso de la enfermedad de Parkinson.

En los estudios e investigaciones que se han realizado sobre esta enfermedad, se han utilizado técnicas de machine learning, a partir de herramientas que permiten obtener la estabilidad/balanceo de la persona, y mediante cámaras KinectTM, para obtener datos tales como el tiempo de zancada y la postura [2], [4].

La ausencia de un estudio que permita relacionar los datos provenientes de las extremidades superiores e inferiores de las personas con enfermedad de Parkinson para su diagnóstico es evidente, esta investigación se dirigió a suplir ese vacío buscando la respuesta a la siguiente pregunta de investigación: ¿Es posible encontrar un modelo de marcha completo que incluya tanto los movimientos de las extremidades superiores como los de las inferiores y permita discriminar si un paciente padece o no la enfermedad de Parkinson?

El objetivo general del proyecto fue definido como “formular un modelo que permita validar la relación que podría existir entre los movimientos de las extremidades superiores e inferiores, para determinar si una persona padece de la enfermedad de Parkinson”, y como objetivos específicos:

- identificar los modelos estadísticos y de machine learning que relacionen los movimientos de las extremidades superiores e inferiores, para determinar si un paciente padece de la enfermedad de Parkinson;
- identificar las variables más representativas en los modelos utilizados para determinar si un paciente padece la enfermedad de Parkinson;
- evaluar los modelos identificados considerando las variables más representativas que relacionen los movimientos de las extremidades superiores e inferiores de los pacientes con la enfermedad; y
- validar el resultado del mejor modelo seleccionado mediante el juicio de un conjunto de médicos expertos.

2. MARCO TEÓRICO

2.1. LA ENFERMEDAD DEL PARKINSON

El estudio de la enfermedad del Parkinson es más frecuente en las personas mayores de sesenta años, pues es poco probable que una persona joven padezca esta enfermedad; pero que sea poco probable no quiere decir que no sea posible [7].

La fisiopatología, que hace referencia a los estudios del origen de diferentes enfermedades y permite explorar y explicar sus síntomas, busca un tratamiento para los síntomas motores y no motores de esta enfermedad; con base en ella, se afirma que la enfermedad de Parkinson se da por la pérdida de las neuronas y de la pigmentación en la sustancia negra del cerebro; sin embargo, para los trastornos cognitivos no existe hasta el momento un acuerdo unitario, aunque cabe aclarar que se ha dicho que la dopamina contribuye al deterioro intelectual de las personas con esta enfermedad [8], [9].

Se han detectado dos patrones que contribuyen a que una persona con enfermedad de Parkinson padezca demencia: el primero incluye pérdida del lenguaje, amnesia y deterioro de aprendizaje, entre otros, y por lo general se presenta en la etapa terminal de la enfermedad; el segundo hace alusión a la pérdida del habla, el inicio de olvidar y algunas alteraciones visoespeciales, y afecta directamente el estado del ánimo de la persona, desmotivándola, haciéndola sentir depresión, entre otros síntomas [9].

Una característica esencial de la sintomatología de la enfermedad de Parkinson es la inestabilidad de los pacientes, un síntoma difícil de identificar a simple vista al inicio de la enfermedad, pues por lo general solo se hace evidente a partir de la etapa 3 de la escala Hoehn y Yahr. Sin embargo, según el UK Brain Bank Criteria [10], en el diagnóstico de esta enfermedad es necesario:

- considerar que la identificación de los movimientos está asociada con otro síntoma principal, como es el caso de la rigidez, el temblor o la inestabilidad;
- estudiar y validar la historia clínica del paciente, revisar si tiene antecedentes cerebrovasculares, lesiones múltiples en la cabeza, algún tumor cerebral, antecedentes familiares, características atípicas de la enfermedad o problemas que incluyan el cerebro o las neuronas; y
- contar con criterios de apoyo para la enfermedad, como son la asimetría del cuerpo, la respuesta del precursor metabólico de la dopamina y el avance de la enfermedad.

Como se mencionó, en el diagnóstico de la enfermedad se utilizan escalas, las más usuales [11] se describen a continuación.

ESCALA DE HOEHN Y YAHR

Está basada en una clasificación por estadios, su fácil aplicación y su relación con el deterioro motor y la calidad de vida son sus principales ventajas. El estadio de un paciente se determina a partir de las características de los síntomas, la extensión de la afección y la discapacidad física ocasionada [12]. Sin embargo, es muy limitada debido a su falta de linealidad y el mayor peso que le otorga a la inestabilidad postural frente a las demás manifestaciones motoras [13]. En esta escala, se divide a los pacientes en cinco estadios: inicio del padecimiento (estadio I); afectación bilateral y axial (estadio II); afectación bilateral (estadio III); afectación bilateral con inestabilidad postural importante (estadio IV); y enfermedad completamente desarrollada (estadio V) [11].

UNIFIED PARKINSON'S DISEASE RATING SCALE (UPDRS)

Esta escala multidimensional consta de 45 ítems divididos en cuatro secciones [14], la mayoría de ellos se califica en una escala de 0 (normal) a 4 (afección severa), excepto algunos de la sección 4 en los que se valora su presencia o ausencia. Los dominios evaluados son: estado mental, comportamiento y estado de ánimo (sección 1); actividades de la vida diaria (sección 2); examen motor (sección 3); y complicaciones de la terapia (sección 4) [11].

ESCALA UNIFICADA DE LA ENFERMEDAD DE PARKINSON MODIFICADA POR LA MDS (MDS-UPDRS)

La base de esta escala es la UPDRS, de ella conserva la estructura original de cuatro secciones, pero modificándolas. El principal cambio es la inclusión de una sección que integra elementos no motores de la enfermedad. Sus cuatro secciones son: experiencias no motoras de la vida diaria (sección I); experiencias motoras de la vida diaria (sección II); examen motor (sección III); y complicaciones motoras (sección IV). Todos los ítems se pueden calificar de 0 (normal) a 4 (severo) [12].

ESCALA DE ZIEGLER-BLETON

Tal como ocurre con la de Hoehn y Yahr, esta escala tiene como base la clasificación de los pacientes en estadios dependiendo de su nivel de compromiso, solo que incluye únicamente tres: el primero, con pocas manifestaciones clínicas (correspondiente a los estadios I y II de Hoehn y Yahr); el segundo, con la enfermedad instalada (correspondiente a los estadios III y IV de Hoehn y Yahr); y el tercero, con la enfermedad completamente desarrollada (correspondiente al estadio V de Hoehn y Yahr) [11].

NOTHWESTERN UNIVERSITY DISABILITY SCALE (NUDS)

La base de esta escala es el deterioro de la calidad de la vida diaria de las personas que padecen la enfermedad de Parkinson, por ello, su valoración se enfoca en dos áreas: la primera, vestirse, hablar y caminar, actividades que se califican, dependiendo del nivel de autonomía con una puntuación que va de 0 a 10; la segunda, comer y alimentarse, que se califica con una puntuación de 0 a 5. Su coeficiente de correlación con la escala de Hoehn y Yahr es significativa [11].

COLUMBIA UNIVERSITY RATING SCALE (CURS)

Esta es una de las escalas más utilizadas, su evaluación consta de siete ítems que se puntúan con valores entre 0 y 4. Valora: bradicinesia, marcha, postura, temblor en reposo, temblor postural, discinesias y posturas distónicas. Su mayor desventaja es que no toma en cuenta las actividades de la vida diaria ni el tiempo durante el cual el paciente se encuentra en Off respecto del tiempo total en que está despierto [11].

ESCALA DE LA MARCHA PARA LA ENFERMEDAD DE PARKINSON (EMEP)

Evalúa veintiún ítems agrupados en cuatro categorías: exploración, capacidad funcional, complicaciones a largo plazo y aspectos sociales. Cada

uno de los ítems se califica con valores de 0 (normal) a 3 (incapacidad o limitación excesiva). Aunque incluye cuatro categorías, la mayor parte de los ítems pertenecen a las dos primeras [11] que son las que se refieren a aspectos más funcionales.

2.2. CIENCIA DE DATOS

La ciencia de datos según Oracle [15] es “un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas científicos para extraer valor de los datos”. En consecuencia, “los científicos de datos combinan una variedad de habilidades, entre ellas estadística, informática y conocimiento empresarial, para analizar datos recopilados de la web, de teléfonos inteligentes, de clientes, sensores y otras fuentes” [15]. La ciencia de los datos aplica las capacidades interdisciplinarias mencionadas en los modelos estadísticos, como el análisis de negocios, incorporando la informática, el modelado, las estadísticas, la analítica y las matemáticas, así como el *machine learning*.

Existen cuatro tipos de estudio en la ciencia de datos: exploratorio, descriptivo, correlacional y correlacional causal. En este proyecto, el más usado es el correlacional, un tipo de estudio que determina cuánta relación tienen ciertas variables en un conjunto de datos, es decir, de qué manera los cambios en una variable influyen en el valor de las demás [16].

La selección de variables significativas se realiza utilizando modelos tales como la regresión, el stepwise (paso a paso), los forward, el LDA (*Linear Discriminant Analysis*) y los árboles de decisión.

Como se mencionó, para este proyecto es importante seleccionar las variables más significativas según el dataset que se tiene para la investigación. Por tal motivo, es necesario tener un criterio de selección que permita identificar las variables representativas para un modelo, pues el criterio que se vaya a utilizar depende del propósito de cada modelo. La selección de variables se puede clasificar, según Sutter y Kalivas [17] y Derksen y Keselman [18], primordialmente en selección forward y eliminación backward, aunque existen otros métodos como: *stepwise* y *generalized simulated annealing*.

En selección forward se adiciona una variable tras otra al modelo y él finaliza cuando todas las variables han sido incluidas; cada variable que ha sido incluida no puede ser eliminada después y su adición se hace si disminuye el AIC (*Akaike Information Criterion*). Este método hace uso de 2^p combinaciones, donde p es la cantidad de variables. La eliminación por *backward* inicia con todas las variables y se van eliminando una a una siempre que su ausencia disminuya el AIC.

Los modelos estadísticos, por su parte, se componen de formulas y ecuaciones matemáticas simples de aproximación a la realidad útiles para realizar predicciones; también son un conjunto de distribuciones de probabilidad en el espacio muestral S [19] o una distribución de probabilidad, que es la clave de estos modelos, construida para tomar decisiones a partir de los datos [20].

Los datos se preparan mediante su recolección, compendio, análisis e interpretación. Los modelos estadísticos sirven para predecir una variable dependiente a partir de una o varias variables independientes posiblemente correlacionadas [21]. Esta predicción se realiza por medio de métodos como los de muestreo, simulación, diseño de experimentos, modelamiento estadístico e inferencia.

Los modelos estadísticos se aplican con el fin de aportar en los procesos de toma de decisiones en las ramas de la investigación, la industria y el gobierno [22], [23]. Muchos científicos utilizado métodos tradicionales para el análisis descriptivo, que van desde el muestreo de datos hasta la interpretación de resultados. Estos modelos, si bien son significativos al procesar grandes volúmenes de datos, requieren de equipos computacionalmente robustos [24]. Algunos ejemplos de modelos estadísticos de clasificación son: la regresión lineal [25] y la regresión logística [26].

De otro lado, machine learning es la parte de la inteligencia artificial que se ocupa del diseño y la aplicación de algoritmos de aprendizaje. Es un buen mecanismo para encontrar similitudes entre datos; hacer reconocimiento facial, reconocimiento de imágenes y reconocimiento de voz; pronosticar; clasificar; solucionar problemas en robótica. En medicina su uso es relevante, entre otros, para obtener diagnósticos médicos [27].

Machine learning presenta tres tipos de aprendizaje: supervisado, no supervisado y semi-supervisado. En este proyecto se utilizó aprendizaje supervisado, el cual consiste en “aprender con base en etiquetas o datos bien definidos, es decir con base en una serie de entradas etiquetadas con una clase o valor, llamadas predictores, con los que el algoritmo aprende” [28, p. 56]. El objetivo de este tipo de aprendizaje es predecir una clase o valor a partir de dichos predictores.

En otras palabras, este aprendizaje consta de tener una base de datos etiquetados o clasificados, hacer una partición de ella y entrenar un algoritmo para que luego, con la otra parte del dataset, la parte del test se pueda predecir a qué etiqueta pertenece cada una de las observaciones que se encuentran ahí [29], [30].

Algunos ejemplos de los algoritmos son: los árboles de decisión [31], BayesNet [32], K-Nearest Neighbors (k-NN) [33], Support Vector Machine (SVM) [34], naive Bayes [35] y random forest [36]. En la evaluación de los modelos toma gran relevancia que los algoritmos de clasificación se puedan valorar mediante técnicas de análisis sensitivo de K-Fold Cross Validation [37], utilizando diferentes semillas aleatorias.

2.3. INFERENCIA CAUSAL

La inferencia causal es la afirmación causal sobre la relación entre dos factores en términos de causa y efecto. Entender porqué sucede algo de manera que se puedan modificar comportamientos para mejorar los resultados a futuro. La inferencia causal es una herramienta estadística que permite que los algoritmos de inteligencia artificial y aprendizaje automático razonen de manera similar.

La inferencia causal tiene su estándar de oro en los Ensayos Controlados Aleatorios (ECA) o en las pruebas A/B, de manera que se puedan efectuar controles sobre un grupo de datos y ninguno sobre los restantes y posteriormente medir los resultados de ambos grupos [38]. La inferencia causal interviene con gran frecuencia en dominios de la medicina [39], la educación [40] y el gobierno [41].

Actualmente se cuenta con una biblioteca de Python que aporta en el análisis y pensamiento causal denominada “DoWhy”. Esta biblioteca aporta tres diferencias claves en comparación con los software disponibles, tales como: supuestos de identificación explícitos, que permiten probar automáticamente las suposiciones establecidas utilizando datos observados; separación entre identificación y estimación, que centra el esfuerzo de inferencia causal en la identificación y libera la estimación utilizando cualquier estimador estadístico disponible para un estimado objetivo; y comprobación de robustez automatizada, que facilita la ejecución automática de comprobaciones de sensibilidad y solidez en la estimación obtenida.

Para implementar la inferencia causal con DoWhy se deben llevar a cabo cuatro pasos: modelado del problema causal; identificación de un objetivo estimado; estimación del efecto causal; y refutación de la estimación [42].

3. ESTADO DEL ARTE

Previo al inicio del proyecto, se consultaron los trabajos de investigación que se describen a continuación.

EXPLORING MACHINE LEARNING TO ANALYZE PARKINSON'S DISEASE PATIENTS

En este artículo, Urcuqui et al. [6] indican que el uso de la tecnología ha permitido estudiar la marcha de forma cuantitativa y objetiva. Para describir su trabajo de investigación es necesario revisar cómo se obtuvieron los datos, cuáles fueron las técnicas de aprendizaje automático que utilizaron y qué resultados obtuvieron.

Los datos se obtuvieron mediante la captura de los movimientos utilizando un software que mide la distancia de un cuerpo, llamado e-Motion; además, usando un Kinect™ se tomaron muestras de veinticuatro mujeres, doce de ellas con enfermedad de Parkinson, y treinta y seis hombres, dieciocho de ellos con enfermedad de Parkinson. Cada persona recorrió tres veces una distancia de cuatro metros de largo por uno y medio de ancho.

El estudio se enfocó en la marcha de cada persona y de él se obtuvieron diecisiete características: aceleración, velocidad, tiempo de paso, tiempo de oscilación, tiempo de postura, porcentaje de oscilación, distancia recorrida, edad, duración de cada ciclo de marcha, zancada integrada por dos pasos, número de ciclo, relación entre el tiempo total y la postura, número de pasos dados en un minuto, duración de la prueba, longitud del paso, número de pasos y longitud de la zancada. Luego se adicionaron cinco más: identificador único por cada paciente, número de la prueba, género, articulación y etiqueta con/sin enfermedad de Parkinson.

Para la evaluación de los datos, se utilizó R y se evaluaron cuatro algoritmos: regresión logística, naïve Bayes, random forest y árboles de decisión. Se centraron en los resultados que no contaban con falsos positivos ni falsos negativos, es decir, que no clasificaron de forma incorrecta a las personas con enfermedad de Parkinson.

Para dicho experimento, dividieron los datos en una proporción 90/10, entrenamiento y test respectivamente, e hicieron uso de una validación cruzada con diez particiones. Al ejecutar los algoritmos mencionados, el mejor modelo, random forest, obtuvo un 82 % de precisión y 12 % de falsos positivos y falsos negativos. Finalmente, con algoritmos exploratorios descubrieron que las características edad y velocidad, son las más representativas para esta enfermedad.

USING KINECT TO CLASSIFY PARKINSON'S DISEASE STAGES RELATED TO SEVERITY OF GAIT IMPAIRMENT

La investigación que presentan Dranca et al. [43] se basa en el desarrollo de un sistema utilizando Kinect™ para distinguir las diferentes etapas de la

enfermedad de Parkinson. Este estudio se realizó con una población clínica de treinta pacientes con EP que pertenecían a tres diferentes grupos: pacientes en una etapa temprana, sin deterioro axial; pacientes con mayor deterioro de la marcha, pero sin congelación de la marcha (Freezing of Gait, FoG); y pacientes con enfermedad de Parkinson y FoG avanzados. A todos ellos se les generaron registros por medio de dos dispositivos Kinect™ que obtenían datos mientras caminaban en medio del pasillo de un hospital.

Los conjuntos de datos obtenidos con el Kinect™ fueron pre-procesados, lo que permitió la identificación de ciento quince características. Se aplicaron algunos métodos para la selección de las características relevantes (selección de características basada en correlación, ganancia de información y evaluación de subconjuntos de consistencia) y algunos métodos de clasificación (árboles de decisión, redes bayesianas, redes neuronales y k-NN) con el objetivo de encontrar el método más preciso para la clasificación de la correspondiente etapa de la enfermedad.

El clasificador que proporcionó los mejores resultados es un caso particular de una red bayesiana (similar a un clasificador naïve Bayes) construido a partir de un conjunto de siete características relevantes seleccionadas por el método de selección de variables basado en correlación. La precisión obtenida para ese clasificador utilizando un 10 - fold cross validation fue de 93.40 %. Las características relevantes están relacionadas con los ángulos de la tibia izquierda, los ángulos del húmero izquierdo, las inclinaciones frontales y laterales, los ángulos del antebrazo izquierdo y el número de pasos durante el giro.

Aunque la precisión del 93.40% se considera alta, el estudio también demostró que fue posible obtener una precisión mejor (96.23 %), para un clasificador que solamente utiliza características extraídas en los pasos al caminar ligeramente en desplazamiento y en los pasos al caminar girando.

SPLIT-BELT TREAD MILL WALKING IN PATIENTS WITH PARKINSON'S DISEASE: A SYSTEMATIC REVIEW

Seuthe et al. [44] enfocan su investigación en la marcha y la postura de los pacientes con enfermedad de Parkinson. Consiguen sus datos usando las bases de Web of Science, PubMed y PsychINFO. Las características que se seleccionaron para mejorar la marcha y la postura fueron tomadas de las recomendaciones de los artículos que hacían referencia a las caminatas SBT (*Split Belt Training*).

Examinaron entre diez y veinte personas por estudio; la mayoría de las personas con enfermedad de Parkinson era mayor de 60 años, en el grupo

de control la edad oscilaba entre 60 y 65 años. El 25 % de las personas con Parkinson y el 35 % del grupo de control correspondió a mujeres. Se realizaron cinco estudios cuando el paciente caminaba por el SBT, uno de ellos definió el mejor y el peor lado según la longitud de la pierna que haya tenido un paso más largo durante la prueba.

En uno de los estudios, fue evidente que la reducción de la velocidad del cinturón del lado con el paso más largo conllevaba a una mejora en la coordinación y la simetría bilateral de la marcha después de estar en SBT; mientras que el paso con longitud más corta presentó efectos secundarios.

En cuanto a la prueba y las variables de la marcha se tuvo en cuenta que algunos estudios mostraban la adaptabilidad baja, media y alta; otros promediaban los resultados durante las fases de adaptación y los efectos que surgen luego de esta. Para las pruebas se usó un sistema de captura de movimiento con un intervalo en las muestras de 100 a 240 Hz. Además, se usaron varios métodos para obtener los parámetros de la marcha, lo que hizo difícil la comparación de los resultados.

Dentro de los parámetros se evaluó: la cadencia, el tiempo de la zancada, la duración de la instancia, la oscilación, la variabilidad de la marcha y el deterioro cognitivo (PCI, Parkinson Cognitive Impairment).

Respecto de la adaptación de la marcha en personas con/sin enfermedad de Parkinson en condiciones SBT, no se encontró una diferencia en la adaptabilidad: la coordinación bilateral de la marcha tuvo el mismo grado en ambos grupos y no se encontraron diferencias significativas entre las extremidades superiores e inferiores. Concluyeron que una persona con enfermedad de Parkinson en un estadio bajo o medio, se adapta muy bien en la marcha y en la coordinación bilateral al caminar en un SBT, tal cual como lo hace una persona que no tiene la enfermedad. A otras personas, les ayudó a modular la asimetría de la marcha.

BOTH COORDINATION AND SYMMETRY OF ARM SWING ARE REDUCE IN PARKINSON DISEASE

Koh et al. [45] enfocan su investigación en la asimetría de los brazos y el uso de acelerómetros portátiles para registrar la aceleración del movimiento del brazo en ocho pacientes con enfermedad de Parkinson de la misma edad y género, durante una caminata de ocho minutos por persona (seis de ellos efectivos al recortar el primero y el último para evitar el efecto del inicio y la detención del paciente). Las características que tuvieron en cuenta en su investigación fueron: la asimetría de oscilación del brazo; la correlación cruzada máxima y la fase relativa del balanceo bilateral.

El análisis estadístico se hizo con MATLAB, usando ANOVA unidireccional en regresión lineal para evaluar las dos primeras características mencionadas y probar la última. Se usó una prueba de Kuiper de dos muestras, con $p < 0.05$ y se tuvo en cuenta el R^2 . Sus resultados mostraron una asimetría de oscilación del brazo bastante alta ($p = 0.002$) y una correlación cruzada máxima baja ($p < 0.001$) para personas con enfermedad de Parkinson.

En conclusión, la investigación confirmó que en la asimetría de oscilación de las extremidades superiores para las personas con enfermedad de Parkinson hay una gran reducción en el balanceo de los dos brazos al caminar y que la correlación cruzada máxima es significativamente menor, lo cual indica que hay poca coordinación en ambos brazos y que, por lo tanto, esto contribuye en gran manera a la percepción clínica.

SISTEMA DE DETECCIÓN DE ANOMALÍAS MOTORAS A PARTIR DEL ANÁLISIS DEL BALANCEO DE LOS BRAZOS

Rincón y Navarro [46], a partir de la observación de: la importancia del análisis de marcha en el diagnóstico y seguimiento de la enfermedad de Parkinson; la subjetividad que supone la evaluación a simple vista; y el elevado costo y complejidad de los laboratorios de marcha, buscaron establecer un método objetivo de medición de las variables de la marcha y para ellos desarrollaron un sistema hardware-software, basado en el uso de acelerómetros, capaz de detectar anomalías motoras a partir de la observación del movimiento de las extremidades superiores.

En el componente hardware, se desarrollaron dos brazaletes, uno para cada brazo, “cada uno con un par de acelerómetros de tres ejes conectados a microcontroladores electrónicos que se comunican a través de Bluetooth 4.0 con una aplicación para teléfonos inteligentes Android (también desarrollada por el proyecto), desde donde la información percibida se envía a la nube para que sea almacenada y quede disponible para su posterior análisis” [46, p. 65].

Las pruebas del sistema se realizaron con 51 pacientes con enfermedad de Parkinson y 37 controles pareados por edad, quienes realizaron una caminata de diez metros sobre una superficie plana utilizando los brazaletes. “Con la información de cuatro muestras de cada persona se constituyó una base de datos, cuya información fue analizada utilizando cinco técnicas: análisis de asimetría, para verificar diferencias entre ambos brazos; análisis correlacional, para determinar la regularidad de los pasos; análisis de Fourier, para determinar la magnitud de temblores; análisis Wavelet, para determinar la complejidad de las señales a través de la dimensión fractal; e

integración numérica, para obtener la velocidad a partir de la aceleración” [46, p. 65]. Los resultados obtenidos les permitieron demostrar la capacidad del sistema desarrollado para clasificar de manera inequívoca personas con enfermedad de Parkinson y personas sin problemas motores.

TÉCNICAS DE ANÁLISIS DE SEÑALES PARA EL USO DE E-MOTION EN EL DIAGNÓSTICO Y SEGUIMIENTO DE LA ENFERMEDAD DE PARKINSON

Como base para su investigación, Castaño y Navarro [47] destacan dos aspectos: el primero, la subjetividad que supone la aplicación de las escalas formales que usan los médicos para valorar algunos síntomas reconocidos de la enfermedad de Parkinson —tales como temblor, rigidez en las extremidades y desordenes de la marcha y en la postura corporal—, cuando se aplican sin ayudas técnicas, mediante observación directa del neurólogo; el segundo, las limitaciones derivadas de aspectos como costo, complejidad tecnológica y requerimientos de espacio, para el uso de los laboratorios de marcha, desarrollos tecnológicos complejos que permiten recolectar información sin sesgos y arrojan mediciones objetivas útiles para evitar la variabilidad en el diagnóstico

Asimismo, los investigadores reconocen el valor de las alternativas que se han desarrollado con herramientas de menor complejidad y costo, tales como e-Motion, un sistema que utiliza la cámara RGBD (Red, Green, Blue + Deep) de un Kinect™ para capturar información de las articulaciones de una persona, establecer su posición y obtener datos de tiempo y distancia tan precisos como los que se conseguirían en un laboratorio de marcha.

Indican sin embargo que para que e-Motion sea una herramienta útil, se “requiere de técnicas de análisis que permitan transformar la información obtenida (...) en valores objetivos que puedan ser usados en el contexto clínico” [47, p. 65] y con ello en mente dirigen su proyecto de investigación a proveer técnicas de análisis de señales de marcha y balanceo de los brazos y a generar algoritmos de procesamiento que puedan ser incluidos en los análisis de pacientes usando e-Motion. Sus resultados representan un gran aporte en la medida en que permiten que el sistema “sea una alternativa real, capaz de complementar de manera válida, objetiva y replicable, el criterio del neurólogo durante el diagnóstico de esta enfermedad” [47, p. 65].

AUTOMATIZACIÓN DE PRUEBAS CLÍNICAS PARA EL DIAGNÓSTICO Y MONITOREO DE LA ENFERMEDAD DE PARKINSON

Arango, Ariza y Navarro [48] recogen en este documento los resultados de varios proyectos de investigación realizados entre 2012 y 2018 por los

grupos de investigación i2t y Biomédica de las facultades de Ingeniería y Salud de la Universidad Icesi, en asocio con la Fundación Valle del Lili, dirigidas a comprobar la validez de usar dispositivos de juego, en particular el Kinect™, como “dispositivo de captura de movimiento capaz de aportar información que complementa de manera acertada, oportuna y relevante, a la evaluación clínica de personas con enfermedad de Parkinson, que realizan los profesionales de la salud” [48, p. 17].

En esta compilación se reúne una serie de documentos que los investigadores han presentado a la comunidad científica en eventos internacionales con material inédito, y con ellos construye un documento de referencia completo y continuo sobre el desarrollo de esta línea de investigación. El libro abre con “Enfermedad de Parkinson y análisis de marcha” [49], un capítulo inédito dirigido a suplir la necesidad de entregarle al lector un punto de partida uniforme; el segundo capítulo [50] es una adaptación de “*A reliability assessment software using Kinect to complement the clinical evaluation of Parkinson’s disease*” [51], ponencia presentada en Milán (Italia), durante la 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); el tercero [52] es una adaptación de dos ponencias: “*Spatiotemporal gait variables using wavelets for an objective analysis of Parkinson disease*” [53], presentada en Gjøvik (Noruega) durante la 15th International Conference on Wearable, Micro and Nano Technologies for Personalized Health (pHealth 2018), y “*Automated gait analysis using Kinect and Wavelets*” [54], presentada en Ostrava (República Checa) durante la IEEE International Conference on E-health Networking, Application & Services. El libro finaliza con “Sistema de apoyo a la evaluación clínica de la marcha para personas con EP” [55], un texto inédito que resume y reúne de manera ordenada las lecciones aprendidas en todo este proceso y representa un punto de partida para trabajos posteriores sobre esta materia, como el que se presenta en este documento.

4. MÉTODO

El proyecto se desarrolló utilizando ASUM-DM (*Analytics Solutions Unified Method for Data Mining*) [56], una extensión y refinamiento realizado por IBM de la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) [57] (FIGURA 1).

ASUM-DM combina la gerencia de proyectos tradicional con los principios de la metodología ágil [56] y se ha utilizado con excelentes resultados en proyectos de analítica, tal como lo define la alianza CAOBA (Centro de Excelencia y Apropiación en *Big Data* y *Data Analytics*), iniciativa que tiene

como objetivo generar soluciones en diversos sectores industriales, gubernamentales y académicos a partir de big data y data analytics [59].

El proyecto se gestionó mediante una secuencia de diez fases adaptadas del ASUM-DM original (FIGURA 2), las cuales se describen a continuación.



Figura 1. Metodología CRISP- DM [58]

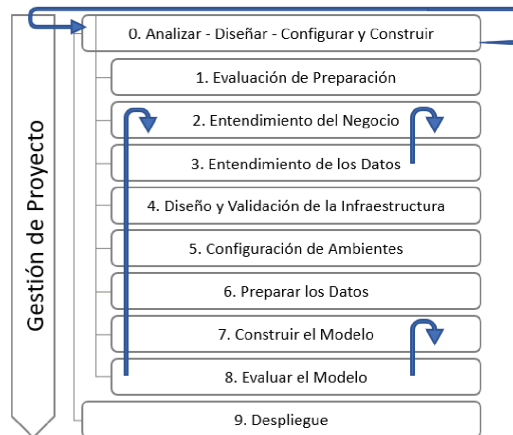


Figura 2. Adaptación de ASUM-DM

Fase 0. Análisis, diseño, configuración y construcción: fase circular iterativa en la cual se comprenden y asimilan los objetivos, las expectativas y los requisitos de la investigación, se preparan los datos y se construyen y evalúan los modelos.

- Fase 1. Evaluación de la preparación: fase en donde se evalúan: los elementos de implementación o administrativos; los elementos de construcción y ubicación; los elementos de tecnología y las fuentes de datos; y los elementos de volúmenes que podrían afectar la implementación del proyecto, y se identifican las soluciones que podrían mitigar el impacto de dichos elementos.
- Fase 2. Entendimiento del negocio: fase en donde se comprende la definición del problema, los objetivos descritos y la contextualización y antecedentes desde la perspectiva de investigación, y tiene como resultado el plan de proyecto.
- Fase 3. Entendimiento de los datos: fase encargada de acceder, observar y explorar los datos disponibles, determinar su calidad y describir los resultados en la documentación del proyecto.
- Fase 4. Diseño y validación de la infraestructura: fase encargada del diseño de la arquitectura de los ambientes y las estrategias de autenticación y autorización.
- Fase 5. Configuración de ambientes: fase a cargo de la configuración de los entornos de desarrollo, el control de la calidad y la producción de acuerdo con el diseño de la arquitectura.
- Fase 6. Preparación de los datos: fase encargada de realizar las siguientes actividades: fusionar conjuntos de datos y/o registros; seleccionar un subconjunto de datos de muestra; agregar registros; obtener nuevos atributos; ordenar los datos para modelar; eliminar o reemplazar espacios en blanco o faltantes valores; y dividir los conjuntos de datos entre entrenamiento y prueba.
- Fase 7. Construcción del modelo: fase encargada de procesar los datos generados en la fase de preparación de datos y de elaborar modelos por medio de la selección de las técnicas de modelado indicadas en el marco teórico; previo a la construcción de cada modelo, en ella se crea un procedimiento para probar la calidad y validez del modelo.
- Fase 8. Evaluación del modelo: fase en donde se evalúa el nivel de logro de los objetivos propuestos. Una vez satisfechas las necesidades del negocio, se revisa todo el proceso para identificar aspectos que se hayan pasado por alto y se decide cómo proceder con los resultados obtenidos.

- Fase 9. Despliegue: fase encargada primordialmente de determinar la estrategia adecuada y definir las acciones de liberación, pero también de la entrega del informe final y de la presentación de los resultados.

El proyecto se desarrolló en un periodo de nueve meses, entre marzo y noviembre de 2020. En la TABLA 1 se presentan las fases a alto nivel y sus actividades generales. Por su carácter circular e iterativo, no se incluye la fase 0. Para un detalle sobre la extensión de cada fase, se recomienda consultar [60].

Tabla 1. Fases y actividades del proyecto

Fases	Actividades
1. Evaluación de la preparación	Plan del proyecto
2. Entendimiento del negocio	<ul style="list-style-type: none"> • Recopilación del contexto del negocio. • Objetivos de negocio. • Criterios de éxito de negocio. • Inventario de recursos, requerimientos, suposiciones, restricciones y contingencias. • Definición de roles
3. Entendimiento de los datos	<ul style="list-style-type: none"> • Recolección de la data inicial. • Descripción de la data. • Exploración de la data. • Verificación de la calidad de la data
4. Diseño y validación de la infraestructura	<ul style="list-style-type: none"> • Diseño de la infraestructura técnica
5. Configuración de ambientes	<ul style="list-style-type: none"> • Configuración de ambientes
6. Preparación de los datos	<ul style="list-style-type: none"> • Selección de los datos. • Limpieza de los datos. • Construcción de los datos. • Integración de los datos. • Formateo de los datos
7. Construcción del modelo	<ul style="list-style-type: none"> • Selección técnica del modelo. • Generación del diseño de las pruebas. • Construcción del modelo. • Evaluación del modelo
8. Evaluación del modelo	<ul style="list-style-type: none"> • Evaluación de resultados. • Revisión del proceso. • Determinación de los próximos pasos
9. Despliegue	<ul style="list-style-type: none"> • Publicación del modelo final. • Elaboración del reporte final. • Cierre del proyecto. • Identificación de las lecciones aprendidas

5. LA INVESTIGACIÓN

A continuación se presenta el proceso de construcción de la solución planteada (FIGURA 3) como respuesta a los objetivos propuestos.

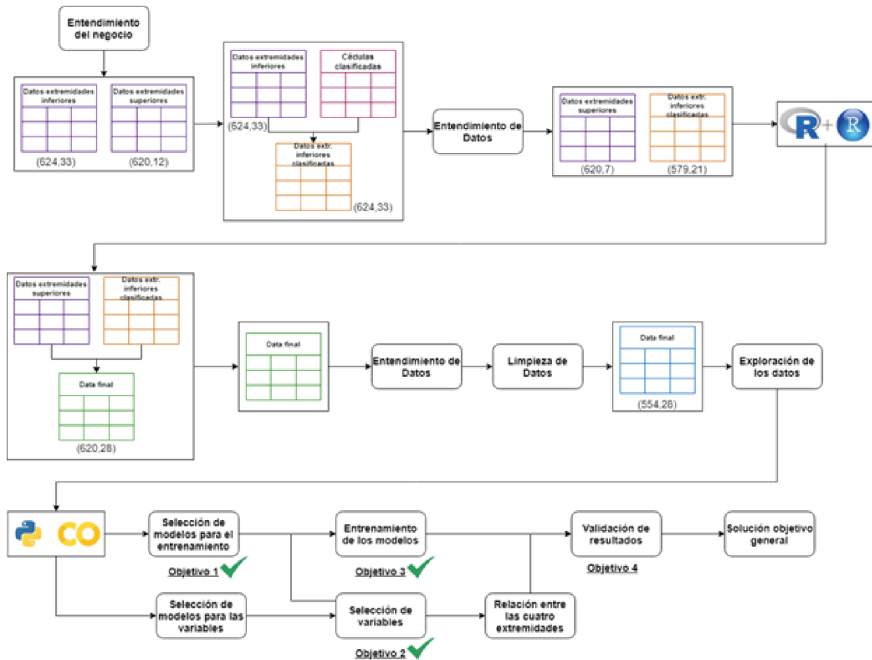


Figura 3. Proceso general de la investigación

5.1. PRIMERA ETAPA

Al inicio del proyecto se realizaron varias sesiones con el fin de obtener el conocimiento de los expertos, en ellas se comprendió su función a nivel profesional y en la investigación y se planteó y entendió el objetivo general del proyecto y sus objetivos específicos. Al finalizar, se recibió de ellos la información base, esto es tres *datasets*: el primero, con información sobre las extremidades superiores; el segundo, con información sobre las extremidades inferiores; y el tercero, con la clasificación de quienes participaron en la prueba, discriminando quiénes padecían la enfermedad de Parkinson y quiénes no. En todos los casos, las personas de este grupo eran voluntarias y no dependían de ayudas para caminar último; no había además alguna particularidad por edad, estatura o género.

Una vez obtenida la data, se procedió a entenderla con ayuda de expertos, hecho esto, se unieron los datasets y se realizaron las transformaciones necesarias para su limpieza. En la TABLA 2 se presenta una descripción de las variables incluidas en el *dataset* final.

Tabla 2. Diccionario de variables del *dataset* final

Nombre	Descripción (unidad de medida)
id_test	Número de Identificación de cada prueba realizada por cada persona
magnitud_left	Desplazamiento en sentido anteroposterior del brazo izquierdo (metros).
time_left	Tiempo que tarda el brazo izquierdo en ir desde el punto más atrás hasta el punto más adelante, en sentido anteroposterior (segundos).
velocity_left	Velocidad del brazo izquierdo calculada como la razón entre <i>magnitud_left</i> y <i>time_left</i> (metros por segundo, m/s).
magnitud_right	Desplazamiento en sentido anteroposterior del brazo derecho (metros).
time_right	Tiempo que tarda el brazo derecho en ir desde el punto más atrás hasta el punto más adelante, en sentido anteroposterior (segundos).
velocity_right	Velocidad del brazo derecho calculada como la razón entre <i>magnitud_right</i> y <i>time_right</i> (m/s).
asim_coef	Coefficiente de asimetría del balanceo de los brazos (valor variable entre 0 y 1).
ret_total_time_test	Tiempo total que toma la persona en realizar la caminata (segundos).
ret_total_distance_test	Distancia total recorrida por la persona al realizar la caminata (metros)
ret_total_distance_test	Distancia total recorrida por la persona al realizar la caminata (metros)
ret_velocidad_total	Velocidad total de la persona al realizar la caminata (m/s)
ret_total_time_left	Tiempo total que el tobillo izquierdo permaneció en el área de captura (segundos).
ret_total_distance_left	Distancia total recorrida por el tobillo izquierdo en el área de captura (metros).
ret_total_time_left_swing	Tiempo total que tarda la fase de balanceo del pie izquierdo (segundos).
ret_total_time_left_support	Tiempo total que tarda la estabilidad del pie izquierdo (segundos).

Tabla 2. Diccionario de variables del dataset final (cont.)

Nombre	Descripción (unidad de medida)
ret_time_left_swing	Tiempo que tarda la fase de balanceo del pie izquierdo (segundos).
ret_time_left_support	Tiempo en estabilidad de la pierna izquierda (segundos).
ret_num_step_left	Número de pasos dados por la pierna izquierda.
ret_length_left_step	Longitud del paso de la pierna izquierda (metros).
ret_total_time_right	Tiempo total que el tobillo derecho permaneció en el área de captura (segundos)..
ret_total_distance_right	Distancia total recorrida por el tobillo derecho en el área de captura (metros).
ret_total_time_right_swing	Tiempo total que tarda la fase de balanceo del pie derecho (segundos).
ret_total_time_right_support	Tiempo total que tarda la estabilidad del pie derecho (segundos).
ret_time_right_swing	Tiempo total que tarda la fase de balanceo del pie derecho (segundos).
ret_time_right_support	Tiempo en estabilidad de la pierna derecha (segundos).
ret_num_step_right	Número de pasos dados por la pierna derecha
ret_length_right_step	Longitud del paso de la pierna derecha (metros)
EP_Classifier	Tipo de participante: con enfermedad de Parkinson (1); sin ella (0)-

Dentro de la sección de conocimiento al dataset final, se encontraron 41 “n.a” para la variable piernas, algo que seguramente se genera por tratarse de pruebas en donde solo se cuenta con la captura de la información de los brazos. Dado que se pretende revisar la relación entre las cuatro extremidades, es necesario tener la información de ambas piernas y ambos brazos por cada id_test. Por esta razón, se tomó la decisión de eliminar las 41 filas que presentan dicha pérdida de valores.

Por otra parte, para la variable de clasificación se encontró un valor inconsistente, aunque el número 2 no se encuentra registrado como posible valor en el diccionario de variables, 24 registros lo tienen. Por ello, se eliminaron esos registros. Hecho lo anterior, el dataset pasó de una dimensión de 620x28 a una de 554x28.

Luego de limpiar la data, se generó la línea base. Se obtuvo un valor de 62,82 % según la cantidad de registros de los tests realizados entre personas con enfermedad de Parkinson (1) y 37.18 % personas sin esta enfermedad

(0). Se considera que se tiene una buen línea base teniendo en cuenta que esta enfermedad puede ser controlada, más aún, cuando no se necesita de ayuda para caminar ni mantenerse de pie.

Finalmente, se generaron dos dataset: uno con los valores sin normalizar, porque algunos algoritmos de clasificación no tienen necesidad de la normalización; y otro con los valores normalizados, con el fin de que los datos puedan ser comparables, pues las medidas están en diferentes escalas, algunos en metros y otros en segundos (ver Anexo 1 en [60]).

5.2. EXPLORACIÓN DE LOS DATOS

Dentro de la exploración de los datos se espera encontrar registros sin simetría, porque como se mencionó, las personas diagnosticadas con EP presentan rigidez, temblor, poca estabilidad y otras características que hacen pensar que los datos van a presentar asimetría y mucha dispersión. Además, los movimientos entre brazos y piernas son diferentes, incluso en personas que no presentan la enfermedad, la zancada es mayor al recorrido que hace el brazo de atrás hasta el cuerpo y del cuerpo hacia adelante (la magnitud del brazo).

Por otro lado, aunque todas las variables presentaron outliers se decidió no eliminarlos, porque los datos con los que se cuenta no son perfectos, es decir, no se puede esperar que personas diagnosticadas con enfermedad de Parkinson tengan coordinación, estabilidad, longitud simétrica, una buena magnitud en los brazos y en general, que las mediciones de las cuatro extremidades sean simétricas o perfectas, como se verá más adelante en las gráficas relevantes de esta sección.

En primera instancia, para tener una referencia sobre lo diferente que son los datos de los pacientes y tener una idea acerca de su caracterización individual, se muestra, a través de la moda para dos variables: magnitud del brazo izquierdo y distancia total por prueba, qué tan agrupados pudiesen estar los datos. Para ambas se presentó una agrupación de 554 registros, esto es, el total de cantidad de datos que se tienen.

Continuando con la exploración, se procedió a visualizar una matriz de correlación, donde se tomó la decisión de reducir la cantidad de variables, sin tener que elegir las manualmente. Dicha selección se realizó para las variables que tuvieran una correlación con otras variables superior a 0,35. Este número se eligió a ensayo y error así: inicialmente se eligieron las correlaciones superiores a 0,9, sin embargo, las variables entre brazos y piernas no mostraron similitud; se bajó entonces el número a 0,5, pero tampoco

co se obtuvo un buen resultado; al bajar a 0,4 se evidenció la similitud entre una variable de brazos y otra de piernas. Por último, se bajó el número de correlación a 0,35, con él, la aproximación entre variables de piernas y brazos fue un poco mayor, aunque no mucho, como se puede observar en la FIGURA 4. Sin embargo, es un comienzo muy coherente porque, como se ha mencionado, las cuatro extremidades presentan asimetría.

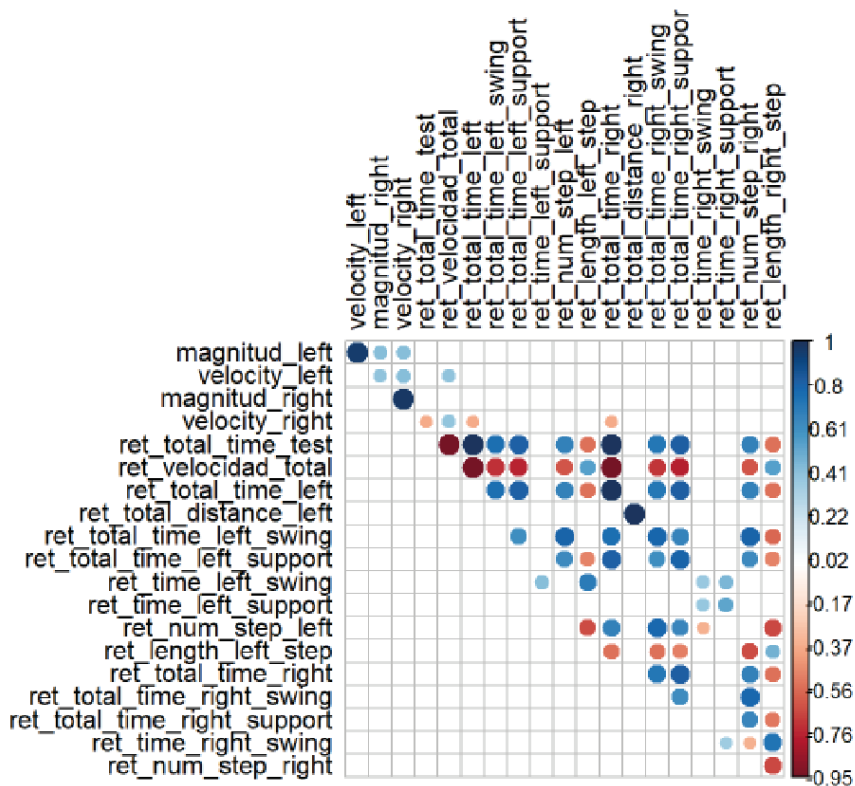


Figura 4. Matriz de correlación

Finalmente, se realizó la exploración multivariada con respecto a la variable objetivo, que para el proyecto es EP_Classifier. A continuación, se muestran las gráficas más representativas para el proyecto.

En la FIGURA 5 se observa una correlación mayor al 0.9; además, los pacientes diagnosticados con EP presentan una menor velocidad en las pruebas sobre las mediciones de la pierna derecha, mientras que quienes no

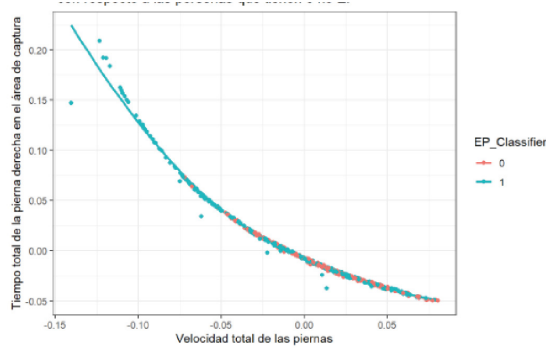


Figura 5. Correlación mayor a 0.9 entre velocidad de las piernas y el tiempo de la pierna derecha en el área de captura, respecto de las personas que padecen o no EP

son diagnosticados con EP presentan una mayor velocidad sobre la pierna derecha.

En la FIGURA 6 se aprecia que los brazos de los pacientes diagnosticados con EP permanecen más cerca al cuerpo que los de las personas que no tienen EP.

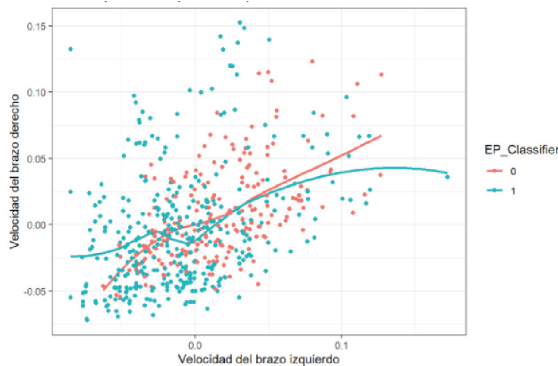


Figura 6. Velocidad del brazo izquierdo frente a la del brazo derecho respecto de las personas que padecen o no EP

En la FIGURA 7 se puede observar cómo: las personas diagnosticadas con EP presentan mayor rigidez y disminución en la velocidad; en la mayoría de los pacientes la rigidez se presenta en mayor proporción sobre el brazo izquierdo; las personas que no presentan EP tienen mayor velocidad en el movimiento de sus brazos y una velocidad muy similar.

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

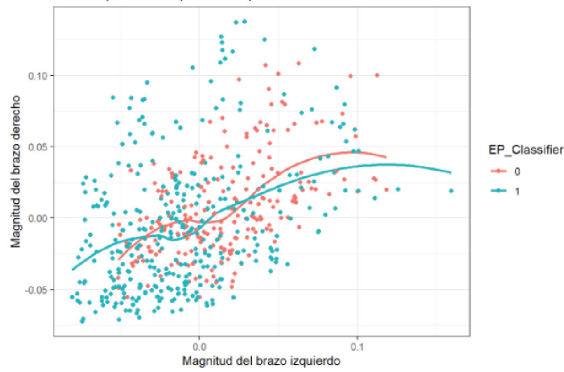


Figura 7. Magnitud de los brazos izquierdo y derecho respecto de las personas que padecen o no EP

En la FIGURA 8, se observa que: el tiempo de la pierna derecha con relación al tiempo del brazo derecho es mayor para las personas que tal vez tienen esta enfermedad más avanzada, mientras que las personas que no tienen EP presentan un rango más centrado a 0, sin mayor dispersión, aunque no podríamos determinar si una nueva medición en el gráfico se clasificaría como EP o no EP (ANEXO 1 en [60]).

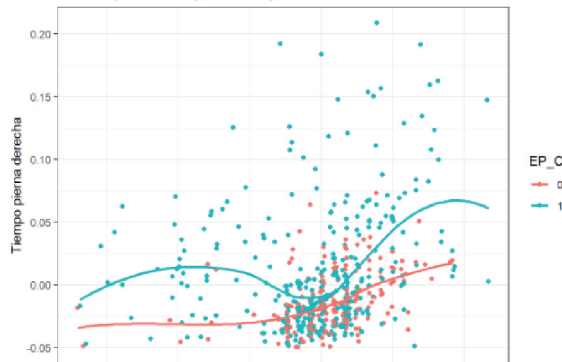


Figura 8. Tiempo del brazo derecho frente al tiempo de la pierna derecha respecto de las personas que padecen o no enfermedad de Parkinson

5.3. SELECCIÓN DE VARIABLES

Para la selección de las variables, se utilizaron los modelos forward y backward, ambos en R y Python. Hecho esto, se consultó a los expertos qué

variables consideraban de mayor relevancia. El objetivo de elegir estas variables fue ejecutar los modelos de clasificación, validar cuál arrojaba el mejor resultado y usar esas variables para correr el algoritmo de inferencia causal (ver Anexo 2 en [60]).

Las variables más relevantes respecto de la variable objetivo (EP_Classifier), según el modelo backward corrido en R, son:

- magnitud_left;
- asim_coef;
- ret_time_left_swing; y
- ret_length_left_step.

Para los expertos, las variables más relevantes respecto de la variable objetivo son:

- magnitud_left;
- magnitud_right;
- asim_coef;
- ret_time_left_swing;
- ret_length_left_step;
- ret_time_right_swing; y
- ret_length_right_step

El universo de variables está formado por todas las que fueron descritas en la TABLA 2, excepto id_test

5.4. EVALUACIÓN DE LOS MODELOS

Hasta esta sección se han cumplido los dos primeros objetivos específicos y se comienza a contar el trabajo que se hizo para cumplir el tercero. A continuación, se presentan los modelos ejecutados y los resultados obtenidos.

Para todos los modelos que se van a mencionar, se usaron métricas de clasificación Hold-Out y k-Fold Cross-Validation. Para el primero se partió la data en entrenamiento (80 %) y pruebas (20 %); Para el segundo, con el fin de garantizar que los resultados sean independientes a la partición entre entrenamiento y pruebas, se manejó un k iteraciones igual a 5.

Los siguientes modelos se ejecutaron en R con sus respectivas descripciones: regresión logística, árbol de decisión sin procesamiento, árbol de decisión pre-poda, árbol de decisión post-poda, naïve Bayes y random forest. Por otra parte el modelo Support Vector Machine se ejecutó en Python.

En la TABLA 3 se presenta el resultado de cada modelo por cada una de las variables y en la TABLA 4 los resultados de su matriz de confusión. En ellas: B corresponde a las variables seleccionadas según el modelo Backward; E, a las seleccionadas por el experto; y T a todas las variables.

Tabla 3. Resultado de las métricas de evaluación de los modelos

Modelo\ Variables	Kappa			Accuracy		
	B	E	T	B	E	T
Regresión logística	0.433	0.415	0.400	0.745	0.736	0.718
Support Vector Machine	0.487	0.572	0.347	0.766	0.802	0.706
Árbol de decisión sin procesamiento	0.404	0.388	0.268	0.727	0.718	0.654
Árbol de decisión pre-poda	0.404	0.471	0.476	0.727	0.745	0.745
Árbol de decisión post-poda	0.449	0.388	0.268	0.681	0.718	0.654
Naïve Bayes	0.466	0.423	0.414	0.745	0.718	0.690
Random forest	0.611	0.650	0.661	0.818	0.836	0.845

Modelo\ Variables	Recall			Precision		
	B	E	T	B	E	T
Regresión logística	0.560	0.560	0.634	0.696	0.676	0.619
Support Vector Machine	0.819	0.819	0.819	0.699	0.796	0.741
Árbol de decisión sin procesamiento	0.585	0.585	0.560	0.648	0.631	0.534
Árbol de decisión pre-poda	0.707	0.731	0.756	0.630	0.638	0.632
Árbol de decisión post-poda	0.585	0.585	0.560	0.571	0.631	0.534
Naïve Bayes	0.707	0.731	0.902	0.644	0.6	0.552
Random forest	0.756	0.780	0.731	0.756	0.780	0.833

Tabla 4. Resultado de la matriz de confusión de los modelos

Modelo\ Variables	Falsos positivos			Falsos negativos		
	B	E	T	B	E	T
Regresión logística	0.696	0.676	0.619	0.766	0.763	0.779
Support Vector Machine	0.699	0.796	0.741	0.568	0.677	0.627
Árbol de decisión sin procesamiento	0.648	0.631	0.534	0.767	0.763	0.731
Árbol de decisión pre-poda	0.630	0.638	0.632	0.812	0.825	0.836
Árbol de decisión post-poda	0.571	0.631	0.534	0.750	0.763	0.731
Naïve Bayes	0.644	0.600	0.552	0.815	0.816	0.906
Random forest	0.756	0.780	0.833	0.750	0.869	0.851

Como se observa en la TABLA 3, los mejores resultados se obtuvieron cuando se entrenaron los modelos con todas las variables. Sin embargo, se eligió seguir trabajando con las variables que seleccionaron los expertos, porque fueron los segundos mejores resultados y cuando se lleguen a tener más registros, el procesamiento es mucho más rápido y será más confiable. Teniendo en cuenta lo anterior, en la misma tabla se evidencia que el modelo que mejor clasificó fue random forest. Por otra parte, cabe resaltar que, aunque el modelo con menos falsos negativos fue SVM (ver TABLA 4), se sigue eligiendo random forest (el segundo con menos falsos negativos) porque es el que con un 0.20 supera la baseline de 62.82 % previamente indicada.

5.5. VALIDACIÓN DE LA RELACIÓN ENTRE LOS MOVIMIENTOS DE LAS EXTREMI- DADES SUPERIORES E INFERIORES

Para finalizar con la solución del objetivo general, se trabajó con inferencia causal para identificar si entre las extremidades superiores e inferiores existía alguna relación. Se usó la librería DoWhy v.0.5 en Python; las variables que se pasaron fueron las elegidas por los expertos, pues como se pudo observar en la evaluación de los modelos, estas arrojaron mejores resultados (ver Anexos 5 y 6 en [60])

La FIGURA 9 corresponde al flujo de trabajo de DoWhy e incluye cuatro pasos para modelar un problema de inferencia causal:

- modelar, en donde se hace explícito cada supuesto causal, por medio de un gráfico;
- identificar, en donde por medio del gráfico obtenido en la modelación, se encuentra la mejor expresión para la estimación causal;
- estimar, que proporciona una estimación causal para la expresión encontrada en el punto 2; y
- refutar, es decir, verificar que la estimación sea robusta o sólida.

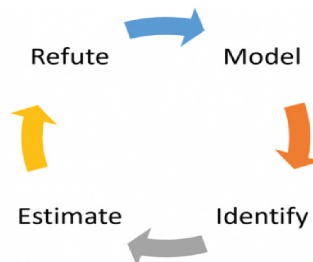


Figura 9. Flujo de trabajo de DoWhy [46]

Se plantearon las siguientes probabilidades causales:

- $P(\text{Magnitud_left} \mid \text{do}(\text{ret_velocidad_total}, \text{ret_time_left_swing}, \text{ret_length_left_step}, \text{ret_time_right_swing}, \text{ret_length_right_step}))$
- $P(\text{Magnitud_righth} \mid \text{do}(\text{ret_velocidad_total}, \text{ret_time_left_swing}, \text{ret_length_left_step}, \text{ret_time_right_swing}, \text{ret_length_right_step}))$
- $P(\text{asim_coef} \mid \text{do}(\text{ret_velocidad_total}, \text{ret_time_left_swing}, \text{ret_length_left_step}, \text{ret_time_right_swing}, \text{ret_length_right_step}))$
- $P(\text{ret_velocidad_total} \mid \text{do}(\text{Magnitud_left}, \text{Magnitud_righth}, \text{asim_coef}))$
- $P(\text{ret_time_left_swing} \mid \text{do}(\text{Magnitud_left}, \text{Magnitud_righth}, \text{asim_coef}))$
- $P(\text{ret_length_left_step} \mid \text{do}(\text{Magnitud_left}, \text{Magnitud_righth}, \text{asim_coef}))$
- $P(\text{ret_time_right_swing} \mid \text{do}(\text{Magnitud_left}, \text{Magnitud_righth}, \text{asim_coef}))$
- $P(\text{ret_length_right_step} \mid \text{do}(\text{Magnitud_left}, \text{Magnitud_righth}, \text{asim_coef}))$

Se decidió abordar la relación de las extremidades superiores e inferiores mediante la selección de todas las variables de un tipo de extremidad, cruzándolas con una sola variable del tipo de extremidad opuesta, de manera que se pudiese encontrar causalidad entre ellas. Esta relación se modeló de tal forma que todas las variables fueron analizadas, de tal manera que fuera posible observar cuál de los modelos presenta el mejor resultado.

Para un mejor entendimiento, se creó un gráfico causal, tal como se muestra en las FIGURAS 10 y 11, en las cuales se presentan los siguientes supuestos:

- el conjunto de la velocidad total que toma el paciente en realizar la caminata “ret_velocidad_total”, el tiempo que tarda la fase de balanceo del pie izquierdo “ret_time_left_swing”, la Longitud del paso de la pierna izquierda “ret_length_left_step”, el tiempo que tarda la fase de balanceo del pie derecho “ret_time_right_swing” y la Longitud del paso de la pierna derecha “ret_length_right_step” afecta o no al coeficiente de asimetría del balanceo de los brazos “asim_coef”, lo que podría dar lugar a que la persona presente la enfermedad del Parkinson; y
- el conjunto del desplazamiento en sentido anteroposterior del brazo izquierdo “magnitud_left”, el desplazamiento en sentido

anteroposterior del brazo derecho “magnitud_right” y el coeficiente de asimetría del balanceo de los brazos “asim_coef” afectan o no la longitud del paso de la pierna derecha “ret_length_right_step”, lo que podría dar lugar a que la persona tenga la enfermedad del Parkinson.

De este planteamiento de probabilidades, se obtuvieron los resultados que se presentan en la TABLA 5.

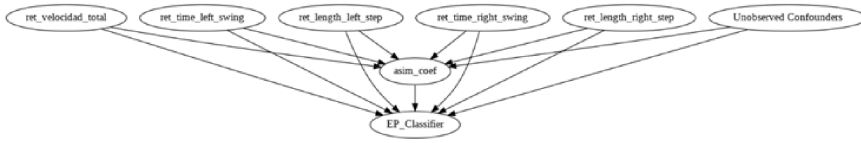


Figura 10. Inferencia causal variable asim_coef

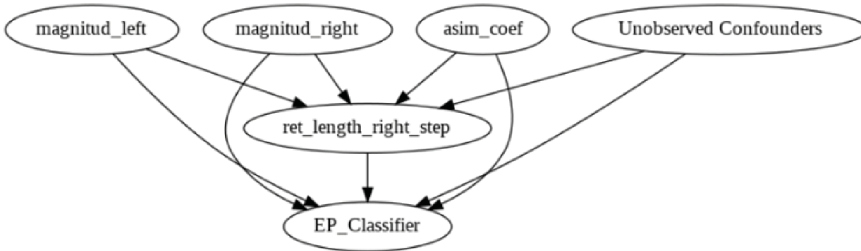


Figura 11. Inferencia causal variable ret_length_right_step

Tabla 5.Resultado de la inferencia causal

Variable	Estimador	Refutación		
		Random common	Placebo treatment	Data subset
magnitud_left	-1.4031477552	-1.3972214254	0.01135989936	1.41537124481
magnitud_rigth	-1.7915436496	-1.7316929403	0.04478764833	-1.8119066182
asim_coef	1.53694724647	1.52082631425	0.00463027302	1.53140757819
ret_velocidad_total	-0.6373051994	-0.6375642949	-0.0206750400	-0.6377562139
ret_time_left_swing	-0.1088931577	-0.1026910427	-0.0078798607	-0.1099028668
ret_length_left_step	-0.4645240085	-0.4645253742	0.01006026372	-0.4639745011
ret_time_right_swing	-0.1200940579	-0.1239362638	-0.0030353285	-0.1115686309
ret_length_right_step	-0.3865592295	-0.3900639846	-0.0150873183	-0.3887911729

Por otro lado, para la ejecución de la estimación se usaron los métodos `linear_regression`, para la refutación:

- `random_common_cause`, el cual intenta probar la robustez de la estimación adicionando una variable aleatoria independiente, si la estimación es sólida, el resultado no cambia mucho, de lo contrario, la estimación es débil;
- `placebo_treatment_refuter`, que sustituye la variable de tratamiento por una variable independiente, se espera que el resultado tienda o llegue a cero; y
- `data_subset_refuter`, en el cual se selecciona un conjunto de datos al azar y el efecto estimado no debería de cambiar significativamente.

De su aplicación se concluye, por los tres métodos, que la estimación por cada prueba de causalidad es robusta.

5.6. VALIDACIÓN: ALGORITMO VS. CLASIFICACIÓN POR LOS EXPERTOS

Para validar los resultados del algoritmo de clasificación elegido, teniendo en cuenta que no fue posible hacer pruebas presenciales, se eligieron aleatoriamente diez pacientes del *dataset*, solo con las características relevantes para este proyecto, es decir, las elegidas por los expertos.

Cada persona y sus características fueron presentados en gráficos, con el fin de que los dos doctores, con base en ellas pudieran clasificarlos (con/sin enfermedad del Parkinson) (ver TABLA 6).

Tabla 6. Clasificación de los pacientes elegidos aleatoriamente y clasificación de los expertos

Paciente	Clasificación	Experto 1	Experto 2
25	Tiene EP	Tiene EP	No tiene EP
58	No tiene EP	Tiene EP	Tiene EP
97	No tiene EP	No tiene EP	No tiene EP
123	Tiene EP	Tiene EP	Tiene EP
133	No tiene EP	Tiene EP	No tiene EP
182	Tiene EP	Tiene EP	Tiene EP
207	Tiene EP	Tiene EP	Tiene EP
393	Tiene EP	Tiene EP	Tiene EP
410	Tiene EP	Tiene EP	Tiene EP
494	Tiene EP	No tiene EP	No tiene EP

Cabe resaltar que este proyecto no tenía como fin enfocarse en la presentación de las características por cada paciente para presentar los datos a doctores o, en general, clínicamente. Por tal motivo, teniendo en cuenta que al tener varias gráficas se presentó una dispersión para lograr entender la marcha de los pacientes, expresado por los doctores, estos obtuvieron una precisión del 60 % de exactitud y, uno de ellos (el más experto) clasificó los pacientes haciendo uso de solo la gráfica que contenía la magnitud del brazo izquierdo y derecho (FIGURA 12).

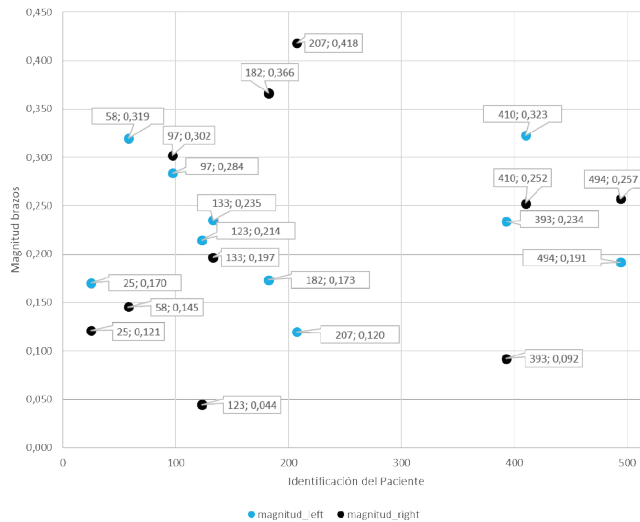


Figura 12. Magnitud de los brazos por cada paciente elegido aleatoriamente

El algoritmo tuvo una exactitud de 90 %, como se presenta en la FIGURA 13; las variables más representativas son las que se presentan en la FIGURA 14, las cuales están de acuerdo con la apreciación que hizo el experto en el párrafo anterior, es decir, el algoritmo no está alejado de la realidad clínica.

Predicción (10 registros)		Real	
		Tiene EP	No tiene EP
Pronóstico	Tiene EP	2	1
	No tiene EP	0	7

Figura 13. Clasificación del algoritmo

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

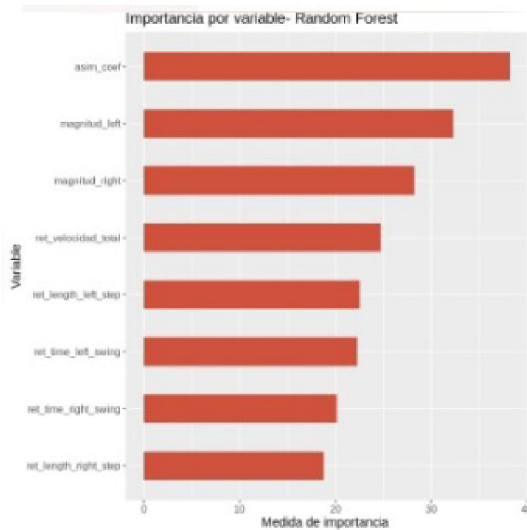


Figura 14. Orden de características más representativas para el algoritmo random forest

RELACIÓN ENTRE LAS EXTREMIDADES SUPERIORES E INFERIORES

Como se mencionó, la variable más representativa a nivel causal, según la librería DoWhy, fue `asim_coef`. Con dicho resultado, se hizo la presentación a los expertos y su apreciación fue muy positiva. Antes de mencionarla, es importante saber que a nivel clínico no se habla de causalidad, sino de incidencia o relación.

Por otra parte, los doctores expresaron que es de bastante interés saber que la variable principal o de tratamiento, como la llama el algoritmo, haya sido `asim_coef`, porque esta incluye la variable de magnitud de los brazos, pues clínicamente hablando, de acuerdo con el experto, esta es la de mayor peso en el cálculo del coeficiente de asimetría de los brazos de un paciente, es decir, a nivel general, los brazos dicen más que las piernas cuando se desea identificar a estos pacientes.

6. HALLAZGOS, CONCLUSIONES Y TRABAJO FUTURO

Random forest clasificó el 90 % de los pacientes correctamente y tuvo un mejor resultado que la clasificación realizada por los doctores, sin alejarse de la realidad clínica. La estimación causal presentó la cantidad de cambio en el valor de resultado cuando interviene y cambia el tratamiento. Es decir, si se aumenta el valor del tratamiento `asim_coef` en “1.0”, el valor

de resultado de EP_Classifier cambia en “1.536”. Como el resultado de la estimación es positivo, el modelo nos indica que es significativo ya que el resultado aumenta con el tratamiento. Las demás variables que se probaron con el modelo entregaron resultados negativos, lo que indica que el resultado disminuyó con el tratamiento.

Se puede entonces afirmar que las variables de las extremidades Inferiores tienen una relación causal con la variable `asim_coef` de la extremidad superior, lo que da como resultado que dicha relación ejerce causalidad sobre el resultado en términos de si una persona tiene o no enfermedad del Parkinson.

En conclusión, se puede hacer uso de técnicas que ayuden a entender la causalidad que tienen algunas características de la enfermedad del Parkinson sobre otras y conocer la fuerza de las suposiciones que se identifican para saber si existe alguna relación entre las extremidades superiores e inferiores. Para este proyecto y, teniendo en cuenta el dataset con que se trabajó, la única variable causal fue `asim_coef`, la cual tuvo efecto sobre las cinco variables de las extremidades inferiores citadas. Esto, de acuerdo con el entendimiento clínico, no es muy alejado de la realidad, porque la variable contiene la magnitud de los brazos, la que para los expertos es una variable que representa muy bien a una persona con enfermedad de Parkinson.

Como trabajo a futuro, se propone realizar: mecanismos de visualización para evaluar a un paciente clínicamente, con el fin de que el experto sea capaz de reconstruir la marcha de la persona evaluada; realizar la parte de inferencia causal con más pacientes, pues con mayor cantidad de información el algoritmo será más preciso; y adicionar variables tales como la edad y el género, con el fin de tener un *dataset* más robusto.

7. REFERENCIAS

- [1] B. Muñoz-Ospina, D. Álvarez-García, H. J. C. Clavijo-Moran, J. A. Valderrama-Chaparro, M. García-Peña, C. A. Herrán, C. C. Urcuqui, A. Navarro-Cadavid, J. Orozco, “Machine learning classifiers to evaluate data from gait analysis with depth cameras in patients with Parkinson’s disease,” *Front Hum Neurosci*, vol. 16; 2022, mayo 19. <https://doi.org/10.3389/fnhum.2022.826376>
- [2] Navarro A., Castano Y., Valderrama J., Muñoz B., & Orozco J. (2019). Objective levodopa response in Parkinson’s disease: A study within the medical consultation using RGB-D camera (Kinect).
- [3] Connolly, B. S., & Lang, A. E. (2014). Pharmacological treatment of Parkinson disease: a review. *Jama*, 311(16), 1670-1683.
- [4] B. M. Ospina, J. A. V. Chaparro, J. D. A. Paredes, Y. J. C. Pino, A. Navarro, & J. L. Orozco, “Objective Arm Swing Analysis in Early-Stage Parkinson’s Disease Using an RGB-D Camera (Kinect®),” *J Parkinsons Dis*, vol. 8, no. 4, pp. 563–570, 2018.
- [5] Kleinholdermann U; Melsbach J; Pedrosa DJ & Der Nervenarzt, ISSN: 1433-0407, 2019 Dec; Vol. 90 (12), pp. 1232-1238; Publisher: Springer Verlag; PMID: 31654235, Base de datos: MEDLINE
- [6] Urcuqui C., Castaño Y., Delgado J., Navarro A., Diaz J., Muñoz B., & Orozco J. (2018). Exploring Machine Learning to analyze Parkinson’s Disease Patients.
- [7] Chana, P., & Albuquerque, T. S. D. (2010). Enfermedad de Parkinson. Chile. Universidad de Santiago de Chile. CETRAM.
- [8] Ostrosky-Solis, F. (2000). Características neuropsicológicas de la enfermedad de Parkinson. *Rev Neurol*, 30(8), 788-796.
- [9] Marín, M., Daniel, S., Carmona, V., Ibarra, Q., & Gámez, C. (2018). Enfermedad de Parkinson: fisiopatología, diagnóstico y tratamiento. *Revista de la Universidad Industrial de Santander. Salud*, 50(1), 79-92.
- [10] Reichmann, H. (2010). Clinical criteria for the diagnosis of Parkinson’s disease. *Neurodegenerative diseases*, 7(5), 284-290.
- [11] de la Cuerda, R. C., Jiménez, A. M., Sánchez, V. C., & Cabezas, M. M. (2004). Escalas de valoración y tratamiento fisioterápico en la enfermedad de Parkinson. *Fisioterapia*, 26(4), 201-210.

- [12] M. Rodríguez-Violante, y A. Cervantes-Arriaga, “La escala unificada de la enfermedad de Parkinson modificada por la Sociedad de Trastornos del Movimiento (MDS-UPDRS): aplicación clínica e investigación,” *Arch Neurocién*, vol. 19, no. 3, pp. 157-163; 2014
- [13] C. G. Goetz, W. Poewe, O. Rascol, et al, “Movement disorder society task force on the Hoehn and Yahr staging scale: status and recommendations,” *Mov Disord*, vol. 19. No. 9, pp. 1020-1028; 2004, sep. <https://doi.org/10.1002/mds.20213>.
- [14] S. Fahn, y R. L. Elton, “Unified Parkinson's Disease Rating Scale (UPDRS),” en *Recent developments in Parkinson's disease*, vol 2, pp. 153-164. Florham Park, NJ: Macmillan Health Care information; 1987.
- [15] Oracle [portal], ¿Qué es la ciencia de datos? Disponible: <https://www.oracle.com/co/data-science/what-is-data-science.html#:~:text=La%20ciencia%20de%20datos%20es,extraer%20valor%20de%20los%20datos>
- [16] Cauas, D. (2015). Definición de las variables, enfoque y tipo de investigación. Bogotá, Colombia: Universidad Nacional de Colombia.
- [17] J.M. Sutter; J.H. Kalivas (1993). Comparison of Forward Selection, Backward Elimination, and Generalized Simulated Annealing for Variable Selection., 47(1-2), 60–66. doi:10.1006/mchj.1993.1012
- [18] Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2), 265–282. doi:10.1111/j.2044-8317.1992.tb00992.x
- [19] McCullagh, P. (2002). What is a statistical model?. *Annals of statistics*, 1225-1267.
- [20] Davison, A. C. (2003). *Statistical models* (Vol. 11). Cambridge University Press.
- [21] Flannelly, Laura & Flannelly, Kevin & Jankowski, Katherine. (2014). Independent, Dependent, and Other Variables in Healthcare and Chaplaincy Research. *Journal of health care chaplaincy*. 20. 161-170. 10.1080/08854726.2014.959374.
- [22] Eroglu, Özgür & Yuksel Serdar. (2019). Statistical Method Selection on Medical Research. 2019/1. 5. 10.26449/sss.1219.

- [23] Steinberg & David. (2016). Industrial statistics: The challenges and the research. *Quality Engineering*, 28. 45-59. 10.1080/08982112.2015.1100453.
- [24] Singh, Rakesh & Tiwari, Neeraj & Prasad, R.C. (2015). Significant Role of Statistics in Computational Sciences. *International Journal of Computer Applications Technology and Research*, 4. 952-955. 10.7753/IJCATR0412.1014.
- [25] G. K. Uyanık and N. Güler, (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 106. 234–240. 10.1016/j.sbspro.2013.12.027.
- [26] Abdallah Bashir, Abdellabashir@uofg.edu.sd. (2014). Logistic Regression Classification for Uncertain Data. *Research Journal of Mathematical and Statistical Sciences*, 2. 1-6.
- [27] Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- [28] C. C. Urcuqui y A. Navarro, Coords., *Ciberseguridad: los datos tienen la respuesta*, Cali, Colombia: Universidad Icesi
- [29] Dietterich, T. G. (1997). Machine-learning research. *AI magazine*, 18(4), 97-97. <https://doi.org/10.1609/aimag.v18i4.1324>
- [30] Mitchell, T. M. (2006). *The discipline of machine learning* (Vol. 9). Pittsburgh: Carnegie Mellon University, School of Computer Science, Machine Learning Department.
- [31] Song, Yan-Yan & Lu, Ying. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27. 130-5. 10.11919/j.issn.1002-0829.215044.
- [32] Liu S., Zhu M. & Yang Y. (2013). A Bayesian Classifier Learning Algorithm Based on Optimization Model. <https://doi.org/10.1155/2013/975953>
- [33] Cunningham, Pdraig & Delany, Sarah. (2007). k-Nearest neighbour classifiers. *MultiClassif Syst*.
- [34] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.
- [35] Kaviani, Pouria & Dhotre, Sunita. (2017). Short Survey on Naive Bayes Algorithm. *International Journal of Advance Research in Computer Science and Management*, 04.

- [36] Ali, Jehad & Khan, Rehanullah & Ahmad, Nasir & Maqsood, Imran. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues (IJCSI)*. 9.
- [37] Rodríguez, Juan & Pérez, Aritz & Lozano, J.A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 32. 569 - 575.
- [38] Susan Athey. 2015. Aprendizaje automático e inferencia causal para la evaluación de políticas. En *Actas de la 21ª Conferencia Internacional ACM SIGKDD sobre Descubrimiento de Conocimiento y Minería de Datos (KDD '15)*. Association for Computing Machinery, Nueva York, NY, EE. UU., 5-6. DOI: <https://doi.org/10.1145/2783258.2785466>
- [39] Pearl, Judea. (2001). Causal Inference in the Health Sciences: A Conceptual Introduction. *Health Services and Outcomes Research Methodology*. 2. 189-220. 10.1023/A:1020315127304.
- [40] Gustafsson, Jan-Eric. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement1. *School Effectiveness and School Improvement*. 24. 10.1080/09243453.2013.806334.
- [41] Finkelstein A., & Hendren N. (2018). Welfare Analysis Meets Causal Inference. <https://doi.org/10.3386/w27640>
- [42] Sharman A., & Kiciman E. (2019). DoWhy | Making causal inference easy. <https://microsoft.github.io/dowhy/readme.html>
- [43] Dranca, L., de Abetxuko Ruiz de Mendarozketa, L., Goñi, A. et al. Using Kinect to classify Parkinson's disease stages related to severity of gait impairment. *BMC Bioinformatics* 19, 471 (2018). <https://doi.org/10.1186/s12859-018-2488-4>
- [44] Seuthe J, D'Cruz N, Ginis P, Weisser B, Berg D, Deuschl G, Nieuwboer A, & Schlenstedt C. (2019). Split-belt treadmill walking in patients with Parkinson's disease: A systematic review.
- [45] Koh, S. B., Park, Y. M., Kim, M. J., & Kim, W. S. (2019). Influences of elbow, shoulder, trunk motion and temporospatial parameters on arm swing asymmetry of Parkinson's disease during walking. *Human movement science*, 68, 102527. <https://doi.org/10.1016/j.humov.2019.102527>

- [46] D. Rincón y A. Navarro, “Sistema de detección de anomalías motoras a partir del análisis del balanceo de los brazos,” en *Enfermedad de Parkinson: análisis motor desde la ingeniería*, Ingeniería & Salud, vol. 3, pp. 19-62, Cali, Colombia: Universidad Icesi
- [47] Y-G. Castaño y A. Navarro, “Técnicas de análisis de señales para el uso de e-Motion en el diagnóstico y seguimiento de la enfermedad de Parkinson,” en *Enfermedad de Parkinson: análisis motor desde la ingeniería*, Ingeniería & Salud, vol. 3, pp. 63-111, Cali, Colombia: Universidad Icesi.
- [48] J. D. Arango, Y. Ariza-Araújo y A. Navarro, Comps., *Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de Parkinson*, Ingeniería & Salud, vol. 1, Cali, Colombia: Universidad Icesi.
- [49] J. D. Arango, Y. Ariza-Araújo, y A. Navarro, “Enfermedad de Parkinson y análisis de marcha,” en *Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de parkinson*, Ingeniería & Salud, vol. 1, pp. 23-39, Cali, Colombia: Universidad Icesi.
- [50] J. D. Arango, B. Muñoz, W. Ágredo, Y. Ariza-Araújo, J. L. Orozco, y A. Navarro, “Confiabilidad del uso de Kinect™ como complemento de la evaluación clínica,” en *Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de parkinson*, Ingeniería & Salud, vol. 1, pp. 41-52, Cali, Colombia: Universidad Icesi.
- [51] J. D. Arango, B. Muñoz, W. Ágredo, Y. Ariza-Araújo, J. L. Orozco, y A. Navarro, “A reliability assessment software using Kinect to complement the clinical evaluation of Parkinson’s disease,” in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6860–6863, 2015.
- [52] Y. J. Castaño, J. D. Arango, A. Navarro, y B. Muñoz, “Análisis automatizado de marcha usando Kinect™ y wavelets,” en *Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de parkinson*, Ingeniería & Salud, vol. 1, pp. 53-66, Cali, Colombia: Universidad Icesi.
- [53] Y. J. Castaño, J. D. Arango, y A. Navarro, “Spatiotemporal gait variables using wavelets for an objective analysis of Parkinson disease,” *Stud Health Technol Inform*, vol. 249, pp. 173-178; 2018.

- [54] B. Muñoz, Y.J. Castaño, J. D. Arango, y A. Navarro, "Automated gait analysis using a Kinect camera and wavelets," 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), 2018, doi: 10.1109/HealthCom.2018.8531161.
- [55] J. D. Arango, Y. Ariza-Araújo, y A. Navarro, "Sistema de apoyo a la evaluación clínica de la marcha para personas con EP," en Automatización de pruebas clínicas para el diagnóstico y monitoreo de la enfermedad de parkinson, Ingeniería & Salud, vol. 1, pp. 67-81, Cali, Colombia: Universidad Icesi.
- [56] J. Haffar, "Have you seen asum-dm?" <ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf>, 2015.
- [57] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," Journal of Data Warehousing, vol. 5, no. 4, 2000.
- [58] https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png
- [59] CAOBA. Perfil Alianza Caoba, Reporte técnico. Universidad de Los Andes. Julio 2017.
- [60] M. García-Peña y C. Herrán, Modelo que relaciona los datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad del Parkinson, tesis de maestría, Universidad Icesi, Cali, Colombia. 2020.

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

Índice de tablas

Tabla 1. Fases y actividades del proyecto	35
Tabla 2. Diccionario de variables del dataset final	37
Tabla 3. Resultado de las métricas de evaluación de los modelos	44
Tabla 4. Resultado de la Matriz de Confusión de los modelos	44
Tabla 5. Resultado de la inferencia causal	47
Tabla 6. Clasificación de los pacientes elegidos aleatoriamente	48

Índice de figuras

Figura 1. Metodología CRISP- DM	33
Figura 2. Adaptación de ASUM-DM	33
Figura 3. Proceso general de la investigación	36
Figura 4. Matriz de correlación	40
Figura 5. Correlación mayor a 0.9 entre velocidad de las piernas y el tiempo de la pierna derecha en el área de captura, respecto de las personas que padecen o no	41
Figura 6. Velocidad del brazo izquierdo frente a la del brazo derecho respecto de las personas que padecen o no	41
Figura 7. Magnitud de los brazos izquierdo y derecho respecto de las personas que padecen o no	42
Figura 8. Tiempo del brazo derecho frente al tiempo de la pierna derecha respecto de las personas que padecen o no	42
Figura 9. Flujo de trabajo de DoWhy	45
Figura 10. Inferencia causal variable asim_coef	47
Figura 11. Inferencia causal variable ret_length_right_step	47
Figura 12. Magnitud de los brazos por cada paciente elegido aleatoriamente	49
Figura 13. Clasificación del algoritmo	49
Figura 14. Orden de características más representativas para el algoritmo random forest	50

Anexo 1

Exploración del dataset

Librerías

```
#install.packages("dplyr")  
library(tibble)
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

¿Qué encontrarás en el documento?

1. Diccionario de variables

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

1/15

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

2. Conociendo el dataset

3. ¿Qué tipo de problemas se encuentran?

4. Limpieza de datos

1. Diccionario de variables

- **id_cc**: Numero de Identificación del Paciente
- **EP_Classifier**: Indica si el paciente tiene EP o no. [1: tiene, 0: No tiene]
- **prom_total_time**: Tiempo promedio total que toma el paciente en realizar la caminata (segundos)
- **prom_total_distance**: Distancia promedio que toma el paciente en realizar la caminata (metros)
- **prom_speed_total**: Velocidad promedio que toma el paciente en realizar la caminata (mtrs/seg)
- **leg_prom_time_ankle_left**: Tiempo promedio que el tobillo izq permanecio en area de captura (seg)
- **leg_prom_distance_ankle_left**: Distacia promedio del total del tobillo izquierdo en el area de captura (metros)
- **leg_prom_num_step_left**: promedio de numero de pasos del pie izquierdo en el Área de captura (Cantidad pasos)
- **leg_prom_length_step_left**: promedio de la longitud de paso del pie izquierdo en el Área de captura. (mtrs)
- **leg_left_Swing_T**: tiempo promedio que tarda la fase de balanceo del pie izquierdo. (seg)
- **leg_left_Stance_T**: tiempo promedio que tarda la fase de apoyo del pie izquierdo. (seg)
- **leg_left_Speed_T**: razón entre prom_time_distance_left y prom_time_ankle_left. (m/s)
- **leg_prom_total_time_swing_stance_left**: suma entre los dos tiempos promedio de la fase de apoyo y de balanceo, es decir, Swing_T_L + Stance_T_L (seg)
- **leg_Swing_Prop_L**: hace referencia al porcentaje o proporción de la fase de balanceo del pie izquierdo, se calcula como la razón entre Swing_T_L y Prom_total_time_swing_stance_left. [0 -1]
- **leg_prom_time_ankle_right**: Tiempo promedio que el tobillo der permanecio en area de captura (s)
- **leg_prom_distance_ankle_right**: Distacia promedio del total del tobillo der en el area de captura (m)
- **leg_prom_num_step_right**: promedio de numero de pasos del pie der en el Área de captura (Cantidad pasos)
- **leg_prom_length_step_right**: promedio de la longitud de paso del pie der en el Área de captura. (mtrs)
- **leg_right_Swing_T**: tiempo promedio que tarda la fase de balanceo del pie der (seg)
- **leg_right_Stance_T**: tiempo promedio que tarda la fase de apoyo del pie der (seg)
- **leg_right_Speed_T**: razón entre prom_time_distance_right y prom_time_ankle_right. (m/s)
- **leg_prom_total_time_swing_stance_right**: suma entre los dos tiempos promedio de la fase de apoyo y de balanceo, es decir, Swing_T_R + Stance_T_R (seg)
- **leg_Swing_Prop_R**: hace referencia al porcentaje o proporción de la fase de balanceo del pie derecho, se calcula como la razón entre Swing_T_R y Prom_total_time_swing_stance_right. [0 -1]
- **coef_asim_sup**: Coeficiente de asimetría del balanceo de brazos o extremidades superiores. Entre 0 y 1
- **arm_prom_mag_left**: hace referencia al desplazamiento en sentido anteroposterior del brazo izquierdo (m)

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

2/15

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

- **arm_prom_time_left**: hace referencia al tiempo que tarda el brazo izquierdo en ir desde el punto más atrás hasta el punto más adelante, en sentido anteroposterior (s)
- **arm_prom_speed_left**: es la velocidad del brazo izquierdo calculada como la razón entre prom_mag_left y prom_time_left (m/s)
- **arm_prom_mag_right**: hace referencia al desplazamiento en sentido anteroposterior del brazo derecho (m)
- **arm_prom_time_right**: hace referencia al tiempo que tarda el brazo derecho en ir desde el punto más atrás hasta el punto más adelante, en sentido anteroposterior (s)
- **arm_prom_speed_right**: es la velocidad del brazo derecho calculada como la razón entre prom_mag_right y prom_time_right (m/s)

Lectura del dataset

```
data <- read.csv(file.choose())
```

2. Conociendo el dataset

```
dim(data_copia)
```

```
## [1] 105 29
```

Inicialmente se tienen 29 variables con 105 observaciones

```
head(data_copia, 5)
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

3/15

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## EP_Classifier prom_total_time prom_total_distance prom_speed_total
## 1 1 3.059333 2.484848 0.8145219
## 2 1 2.343000 2.468224 1.0554940
## 3 1 3.875133 2.475834 0.6505319
## 4 1 2.790167 2.443123 0.8799880
## 5 0 3.115429 2.480864 0.7973345
## leg_prom_time_ankle_left leg_prom_distance_ankle_left leg_prom_num_step_left
## 1 3.059333 2.605027 1.333333
## 2 2.343000 2.523508 1.600000
## 3 3.875133 2.490881 2.933333
## 4 2.790167 2.286961 2.333333
## 5 3.116000 2.453242 2.000000
## leg_prom_length_step_left leg_left_Swing_T leg_left_Stance_T leg_left_Speed_T
## 1 0.9707956 0.7045000 1.006000 1.1743959
## 2 1.0994162 0.7550000 0.727600 0.9284692
## 3 0.6228767 1.2530222 1.660067 1.5557280
## 4 0.5993927 0.7113333 1.311000 1.2200323
## 5 0.7592697 0.8258571 1.372571 1.2701599
## leg_prom_total_time_swing_stance_left leg_Swing_Prom_L
## 1 1.878896 0.3749542
## 2 1.683469 0.4484786
## 3 2.808750 0.4461138
## 4 1.931366 0.3683059
## 5 2.096813 0.3940134
## leg_prom_time_ankle_right leg_prom_distance_ankle_right
## 1 3.059333 2.605027
## 2 2.342300 2.523508
## 3 3.875133 2.490881
## 4 2.790167 2.286961
## 5 3.115143 2.453242
## leg_prom_num_step_right leg_prom_length_step_right leg_right_Swing_T
## 1 1.333333 0.9739963 0.7296667
## 2 1.600000 0.9339181 0.6935500
## 3 2.933333 0.6182212 1.1773333
## 4 2.333333 0.7235279 0.7840000
## 5 2.000000 0.7790259 0.8698571
## leg_right_Stance_T leg_right_Speed_T leg_prom_total_time_swing_stance_right
## 1 0.8406667 1.1743959 1.904063
## 2 0.7338000 0.9281919 1.621742
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

4/15

Avances en la aplicación de la ingeniería a la valoración de personas con la enfermedad de Parkinson

29/11/22, 19:50 Exploracion dataset Enfermedad del Parkinson

```
## 3      1.7218667      1.5557280      2.733061
## 4      1.3528333      1.2200323      2.004032
## 5      1.1238571      1.2698065      2.139664
## leg_Swing_Prop_R coef_asim_sup arm_prom_mag_left arm_prom_time_left
## 1      0.3832157      0.1217286      0.2252723      1.3646667
## 2      0.4276575      0.1078440      0.1622943      1.1940000
## 3      0.4307746      0.1370692      0.2112503      1.1072089
## 4      0.3912113      0.3337022      0.1009451      0.9669167
## 5      0.4065392      0.1147047      0.1372861      1.1846667
## arm_prom_speed_left arm_prom_mag_right arm_prom_time_right
## 1      0.1650750      0.1705084      1.327500
## 2      0.1359249      0.2368073      1.155167
## 3      0.1907953      0.1726007      1.128084
## 4      0.1043990      0.3813142      1.067833
## 5      0.1158859      0.1904321      1.327143
## arm_prom_speed_right
## 1      0.1284433
## 2      0.2049984
## 3      0.1530033
## 4      0.3570915
## 5      0.1434903
```

```
tail(data_copia, 5)
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

5/15

29/11/22, 19:50 Exploracion dataset Enfermedad del Parkinson

```
## EP_Classifier prom_total_time prom_total_distance prom_speed_total
## 101      0      2.572000      2.480143      0.9678004
## 102      0      2.333000      2.476032      1.0625728
## 103      0      2.211333      2.466311      1.1156796
## 104      0      2.043333      2.448736      1.1985066
## 105      1      2.366500      2.459831      1.0414633
## leg_prom_time_ankle_left leg_prom_distance_ankle_left
## 101      2.576829      2.507883
## 102      2.333000      2.429427
## 103      2.211333      2.314663
## 104      2.043333      1.897018
## 105      2.367175      2.555056
## leg_prom_num_step_left leg_prom_length_step_left leg_left_Swing_T
## 101      1.750000      0.7418156      0.5427857
## 102      1.333333      0.9795533      0.5043333
## 103      1.000000      1.1045643      0.4720000
## 104      1.000000      1.2553010      0.4703333
## 105      1.833333      0.9266716      0.6566404
## leg_left_Stance_T leg_left_Speed_T leg_prom_total_time_swing_stance_left
## 101      0.7960893      1.0274920      1.570278
## 102      1.4860000      0.9603087      1.464642
## 103      0.7290000      0.9553585      1.427358
## 104      0.7353333      1.0771290      1.547462
## 105      0.7632281      0.5264670      1.583107
## leg_Swing_Prop_L leg_prom_time_ankle_right leg_prom_distance_ankle_right
## 101      0.3456622      2.574000      2.507883
## 102      0.3443390      2.333000      2.429427
## 103      0.3306808      2.211333      2.314663
## 104      0.3039385      2.043333      1.897018
## 105      0.4147794      2.366500      2.555056
## leg_prom_num_step_right leg_prom_length_step_right leg_right_Swing_T
## 101      1.750000      0.9185081      0.6640000
## 102      1.333333      1.0479873      0.5356667
## 103      1.000000      1.0918217      0.4686667
## 104      1.000000      1.1780874      0.4796667
## 105      1.833333      0.7707596      0.5186667
## leg_right_Stance_T leg_right_Speed_T leg_prom_total_time_swing_stance_right
## 101      1.0095000      1.0263638      1.690364
## 102      0.5983333      0.9603087      1.495975
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

6/15

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

29/11/22, 19:50 Exploracion dataset Enfermedad del Parkinson

```
## 103      0.7316667      0.9553585      1.424025
## 104      0.7170000      1.0771290      1.556796
## 105      0.6770000      0.9262026      1.444869
##      leg_Swing_Prop_R coef_asim_sup arm_prom_mag_left arm_prom_time_left
## 101      0.3928148      0.10298878      0.2757100      1.181750
## 102      0.3580718      0.05656063      0.3532346      1.197333
## 103      0.3291140      0.03235134      0.3058993      1.173000
## 104      0.3081115      0.28615466      0.2160172      1.067750
## 105      0.3589713      0.05018251      0.1289442      1.107500
##      arm_prom_speed_left arm_prom_mag_right arm_prom_time_right
## 101      0.2333065      0.21779341      1.119000
## 102      0.2950178      0.32986165      1.162667
## 103      0.2607837      0.33913721      1.169500
## 104      0.2023106      0.07503496      0.674875
## 105      0.1164281      0.14913263      1.176500
##      arm_prom_speed_right
## 101      0.1946322
## 102      0.2837113
## 103      0.2899848
## 104      0.1111835
## 105      0.1267596
```

str(data_copia)

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

7/15

29/11/22, 19:50 Exploracion dataset Enfermedad del Parkinson

```
## 'data.frame': 105 obs. of 29 variables:
## $ EP_Classifier : int 1 1 1 0 1 1 1 1 1 ...
## $ prom_total_time : num 3.06 2.34 3.88 2.79 3.12 ...
## $ prom_total_distance : num 2.48 2.47 2.48 2.44 2.48 ...
## $ prom_speed_total : num 0.815 1.055 0.651 0.88 0.797 ...
## $ leg_prom_time_ankle_left : num 3.06 2.34 3.88 2.79 3.12 ...
## $ leg_prom_distance_ankle_left : num 2.61 2.52 2.49 2.29 2.45 ...
## $ leg_prom_num_step_left : num 1.33 1.6 2.93 2.33 2 ...
## $ leg_prom_length_step_left : num 0.971 1.099 0.623 0.599 0.759 ...
## $ leg_left_Swing_T : num 0.705 0.755 1.253 0.711 0.826 ...
## $ leg_left_Stance_T : num 1.006 0.728 1.66 1.311 1.373 ...
## $ leg_left_Speed_T : num 1.174 0.928 1.556 1.22 1.27 ...
## $ leg_prom_total_time_swing_stance_left : num 1.88 1.68 2.81 1.93 2.1 ...
## $ leg_Swing_Prop_L : num 0.375 0.448 0.446 0.368 0.394 ...
## $ leg_prom_time_ankle_right : num 3.06 2.34 3.88 2.79 3.12 ...
## $ leg_prom_distance_ankle_right : num 2.61 2.52 2.49 2.29 2.45 ...
## $ leg_prom_num_step_right : num 1.33 1.6 2.93 2.33 2 ...
## $ leg_prom_length_step_right : num 0.974 0.934 0.610 0.724 0.779 ...
## $ leg_right_Swing_T : num 0.73 0.694 1.177 0.784 0.87 ...
## $ leg_right_Stance_T : num 0.841 0.734 1.722 1.353 1.124 ...
## $ leg_right_Speed_T : num 1.174 0.928 1.556 1.22 1.27 ...
## $ leg_prom_total_time_swing_stance_right : num 1.9 1.62 2.73 2 2.14 ...
## $ leg_Swing_Prop_R : num 0.383 0.428 0.431 0.391 0.407 ...
## $ coef_asim_sup : num 0.122 0.108 0.137 0.334 0.115 ...
## $ arm_prom_mag_left : num 0.225 0.162 0.211 0.101 0.137 ...
## $ arm_prom_time_left : num 1.365 1.194 1.107 0.967 1.185 ...
## $ arm_prom_speed_left : num 0.165 0.136 0.191 0.104 0.116 ...
## $ arm_prom_mag_right : num 0.171 0.237 0.173 0.381 0.19 ...
## $ arm_prom_time_right : num 1.33 1.16 1.13 1.07 1.33 ...
## $ arm_prom_speed_right : num 0.128 0.205 0.153 0.357 0.143 ...
```

summary(data_copia)

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

8/15

Avances en la aplicación de la ingeniería a la valoración de personas con la enfermedad de Parkinson

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## EP_Classifier prom_total_time prom_total_distance prom_speed_total
## Min. :0.0000 Min. :2.001 Min. :2.215 Min. :0.3851
## 1st Qu.:0.0000 1st Qu.:2.354 1st Qu.:2.460 1st Qu.:0.8145
## Median :1.0000 Median :2.614 Median :2.469 Median :0.9459
## Mean :0.6667 Mean :2.794 Mean :2.465 Mean :0.9205
## 3rd Qu.:1.0000 3rd Qu.:3.059 3rd Qu.:2.476 3rd Qu.:1.0422
## Max. :2.0000 Max. :5.787 Max. :2.488 Max. :1.2328
##
## leg_prom_time_ankle_left leg_prom_distance_ankle_left leg_prom_num_step_left
## Min. :2.001 Min. :1.897 Min. :1.00
## 1st Qu.:2.354 1st Qu.:2.388 1st Qu.:1.40
## Median :2.614 Median :2.474 Median :2.00
## Mean :2.794 Mean :2.469 Mean :1.86
## 3rd Qu.:3.059 3rd Qu.:2.555 3rd Qu.:2.00
## Max. :5.787 Max. :2.765 Max. :6.00
##
## leg_prom_length_step_left leg_left_Swing_T leg_left_Stance_T leg_left_Speed_T
## Min. :0.2994 Min. :0.3748 Min. :0.4516 Min. :0.7865
## 1st Qu.:0.7174 1st Qu.:0.5526 1st Qu.:0.7461 1st Qu.:0.9554
## Median :0.8540 Median :0.7360 Median :0.8995 Median :1.0767
## Mean :0.8439 Mean :0.7249 Mean :1.0628 Mean :1.1362
## 3rd Qu.:0.9708 3rd Qu.:0.8259 3rd Qu.:1.3187 3rd Qu.:1.2289
## Max. :1.2553 Max. :1.4817 Max. :3.2447 Max. :2.4612
##
## leg_prom_total_time_swing_stance_left leg_Swing_Prop_L
## Min. :1.210 Min. :0.2862
## 1st Qu.:1.522 1st Qu.:0.3611
## Median :1.791 Median :0.3941
## Mean :1.861 Mean :0.3866
## 3rd Qu.:2.022 3rd Qu.:0.4171
## Max. :3.916 Max. :0.4834
##
## leg_prom_time_ankle_right leg_prom_distance_ankle_right
## Min. :2.001 Min. :1.897
## 1st Qu.:2.353 1st Qu.:2.388
## Median :2.614 Median :2.474
## Mean :2.794 Mean :2.469
## 3rd Qu.:3.059 3rd Qu.:2.555
## Max. :5.787 Max. :2.765
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

9/15

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## leg_prom_num_step_right leg_prom_length_step_right leg_right_Swing_T
## Min. :1.00 Min. :0.2891 Min. :0.4930
## 1st Qu.:1.40 1st Qu.:0.7283 1st Qu.:0.5466
## Median :2.00 Median :0.8314 Median :0.6964
## Mean :1.86 Mean :0.8389 Mean :0.7149
## 3rd Qu.:2.00 3rd Qu.:0.9745 3rd Qu.:0.8362
## Max. :6.00 Max. :1.1781 Max. :1.5427
##
## leg_right_Stance_T leg_right_Speed_T leg_prom_total_time_swing_stance_right
## Min. :0.4655 Min. :0.7865 Min. :1.306
## 1st Qu.:0.7466 1st Qu.:0.9554 1st Qu.:1.512
## Median :0.8972 Median :1.0769 Median :1.796
## Mean :1.0489 Mean :1.1362 Mean :1.851
## 3rd Qu.:1.2687 3rd Qu.:1.2289 3rd Qu.:2.004
## Max. :3.2113 Max. :2.4612 Max. :3.840
##
## leg_Swing_Prop_R coef_asim_sup arm_prom_mag_left arm_prom_time_left
## Min. :0.2606 Min. :0.01841 Min. :0.0335 Min. :0.6773
## 1st Qu.:0.3592 1st Qu.:0.07915 1st Qu.:0.1391 1st Qu.:1.0868
## Median :0.3899 Median :0.11518 Median :0.1946 Median :1.1530
## Mean :0.3836 Mean :0.14094 Mean :0.2053 Mean :1.1413
## 3rd Qu.:0.4183 3rd Qu.:0.18929 3rd Qu.:0.2644 3rd Qu.:1.1978
## Max. :0.4920 Max. :0.39847 Max. :0.4712 Max. :1.4049
## NA's :2 NA's :2 NA's :2
## arm_prom_speed_left arm_prom_mag_right arm_prom_time_right
## Min. :0.0279 Min. :0.04591 Min. :0.665
## 1st Qu.:0.1185 1st Qu.:0.13011 1st Qu.:1.000
## Median :0.1710 Median :0.19043 Median :1.138
## Mean :0.1806 Mean :0.20108 Mean :1.126
## 3rd Qu.:0.2355 3rd Qu.:0.25946 3rd Qu.:1.191
## Max. :0.4750 Max. :0.45636 Max. :1.417
## NA's :2 NA's :2 NA's :2
## arm_prom_speed_right
## Min. :0.04448
## 1st Qu.:0.11075
## Median :0.16367
## Mean :0.17811
## 3rd Qu.:0.22075
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

10/15

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## Max. :0.53981
## NA's :2
```

Se revisa la proporción que se tiene con respecto a la variable de interés

```
table(data_copia$EP_Classifier)
```

```
##
##  0  1  2
## 40 60  5
```

3. ¿Qué tipo de problemas se encuentran?

Según lo observado con las funciones `str()` y `summary()`, se obtiene que:

Posibles outliers

- `leg_left_Stance_T`
- `leg_prom_time_ankle_right`
- `leg_prom_num_step_right`
- `leg_right_Swing_T`
- `leg_right_Stance_T`
- `leg_right_Speed_T`
- `leg_prom_total_time_swing_stance_right`
- `coef_asim_sup`
- `arm_prom_time_left`
- `arm_prom_speed_left`
- `arm_prom_mag_right`
- `arm_prom_speed_right`

Valores inconsistentes

- `EP_Classifier`: no tiene 2 como categoría

Valores perdidos

- `coef_asim_sup`
- `arm_prom_mag_left`
- `arm_prom_time_left`

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

11/15

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

- `arm_prom_speed_left`
- `arm_prom_mag_right`
- `arm_prom_time_right`
- `arm_prom_speed_right`

Validar Registros Duplicados

```
## [1] 0
```

- No existen valores duplicados.

Errores de digitaciones

- No existen errores de digitación, ya que las variables son todas cuantitativas excepto la clasificadora.

Valores fuera de Rango

- No existen valores fuera de Rango, ya que las variables presentes no se miden por medio de rangos.

Valores no referenciados en el diccionario de variables

- Todos los valores están referenciados en el diccionario de variables.

Separación de campos

- No se necesita separar ningún campo.

4. Limpieza de datos

Se procede a hacer la limpieza de los datos, según los problemas identificados en el punto 3:

Valores perdidos

Se desea saber si los valores nulos son de distintos registros o de los mismos

```
data_copia[rowSums(is.na(data_copia)) > 0,]
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

12/15

Avances en la aplicación de la ingeniería a la valoración de personas con la enfermedad de Parkinson

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## EP_Classifier prom_total_time prom_total_distance prom_speed_total
## 68 1 2.482889 2.455971 0.9980348
## 74 1 2.256000 2.457344 1.0906754
## leg_prom_time_ankle_left leg_prom_distance_ankle_left leg_prom_num_step_left
## 68 2.484222 2.303553 2.111111
## 74 2.256000 2.435485 1.500000
## leg_prom_length_step_left leg_left_Swing_T leg_left_Stance_T
## 68 0.9210878 0.8094444 0.7581111
## 74 0.6184407 0.3747500 1.0092500
## leg_left_Speed_T leg_prom_total_time_swing_stance_left leg_Swing_Prop_L
## 68 1.078431 1.887875 0.4287595
## 74 0.926304 1.301054 0.2880357
## leg_prom_time_ankle_right leg_prom_distance_ankle_right
## 68 2.485219 2.303553
## 74 2.256000 2.435485
## leg_prom_num_step_right leg_prom_length_step_right leg_right_Swing_T
## 68 2.111111 0.8404894 0.7168218
## 74 1.500000 0.9858084 0.5502500
## leg_right_Stance_T leg_right_Speed_T leg_prom_total_time_swing_stance_right
## 68 0.74096 1.078864 1.795685
## 74 0.71125 0.926304 1.476554
## leg_Swing_Prop_R coef_asim_sup arm_prom_mag_left arm_prom_time_left
## 68 0.3991912 NA NA NA
## 74 0.3726582 NA NA NA
## arm_prom_speed_left arm_prom_mag_right arm_prom_time_right
## 68 NA NA NA
## 74 NA NA NA
## arm_prom_speed_right
## 68 NA
## 74 NA
```

Como los valores nulos son de los mismos registros y, además, de pacientes que solo tienen registros de mediciones promedios en las variables de las extremidades inferiores, se toma la decisión de eliminar las 2 filas que presentan esta pérdida de valores.

```
data_copia <- na.omit(data_copia)
sum(is.na(data_copia))
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

13/15

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

```
## [1] 0
```

Valores inconsistentes

```
table(data_copia$EP_Classifier)
```

```
##
## 0 1
## 40 58
```

Luego de haber quitado los registros con valores inconsistentes, se determina que el baseline es de 59,18% calculando la regla de 3 con respecto a la cantidad de registros de pacientes diagnosticados con Enfermedad de Parkinson. Se considera que se tiene un buen baseline teniendo en cuenta que las personas que tomaron el examen están en la primera fase de esta enfermedad, es decir, que no tienen los síntomas lo suficientemente desarrollados.

Posibles outliers

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

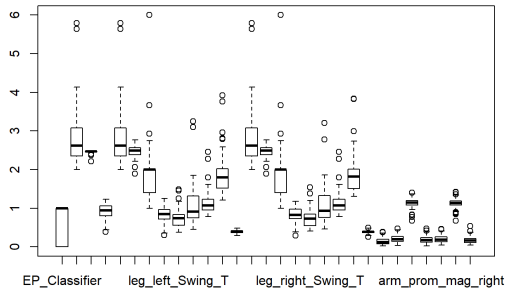
14/15

Modelo que relaciona datos provenientes de las extremidades de un paciente con posible diagnóstico de la enfermedad de Parkinson

29/11/22, 19:50

Exploracion dataset Enfermedad del Parkinson

A continuación, se procede a observar los valores perdidos a nivel general con la siguiente gráfica



```
## [1] 5.6383333 5.7873333 2.4096399 2.3721893 2.2145793 0.3850638 5.6383333
## [8] 5.7873333 2.0647727 1.8970183 2.9333333 6.0000000 3.6666667 0.2993763
## [15] 1.2530222 1.4816667 1.4553333 3.2446667 3.1030000 2.2778542 1.7915532
## [22] 2.4611547 2.8087502 3.7595209 2.9502198 2.7916882 3.9164800 5.6383333
## [29] 5.7873333 2.0647727 1.8970183 2.9333333 6.0000000 3.6666667 0.2891244
## [36] 1.5426667 1.3793333 2.7756667 3.2113333 2.2778542 1.7915532 2.4611547
## [43] 3.8205209 2.9902198 3.8404800 0.2606366 0.4920050 0.3904711 0.3793354
## [50] 0.4712458 0.6733333 0.7736333 1.4048542 0.8276667 0.8337500 0.8202000
## [57] 0.4215078 0.4749709 0.4479014 0.4509654 0.4563569 1.3634000 0.8735417
## [64] 0.8578333 0.8974688 1.4065625 0.8465833 1.4172778 0.8950000 0.8738333
## [71] 0.8454000 0.6748750 0.4102172 0.5398119
```

file:///C:/Users/Lenovo/AppData/Local/Temp/PD_Dataset_Exploracion.html

15/15

Anexo 2

Diccionario de variables de balanceo de brazos y marcha

Cada voluntario hace de tres a cinco pruebas de marcha, en un pasillo de cuatro metros de longitud, las pruebas se normalizan por distancia (1.5m a 4m) y número de pasos (igual número de pasos para la pierna izquierda y la derecha), lo que permite luego promediarlas. Cada prueba de marcha se procesa, de todas ellas se genera un promedio para cada variable.

VARIABLES DE BALANCEO DE BRAZOS

- `prom_mag_left` (promedio de la magnitud de balanceo del brazo izquierdo): se mide en metros, hace referencia al desplazamiento en sentido anteroposterior del brazo izquierdo.
- `prom_time_left` (promedio de tiempo del balanceo del brazo izquierdo): se mide en segundos, hace referencia al tiempo que tarda el brazo izquierdo en ir desde el punto más atrás hasta el punto más adelante, en sentido anteroposterior.
- `prom_speed_left` (promedio de la velocidad de balanceo izquierdo): se mide en metros por segundo (m/s) y corresponde a la velocidad del brazo izquierdo calculada como la razón entre `prom_mag_left` y `prom_time_left`.
- `prom_mag_right` (promedio de la magnitud de balanceo derecha): se mide en centímetros, hace referencia al desplazamiento en sentido anteroposterior del brazo derecho.
- `prom_time_right` (promedio de tiempo del balanceo del brazo izquierdo): se mide en segundos
- `prom_speed_right` (promedio de la velocidad de balanceo izquierdo): se mide en metros sobre segundo (m/s)
- `coef_asim_sup` (coeficiente de asimetría del balanceo de brazos o extremidades superiores): Variable entre 0 y 1.

VARIABLES DE MARCHA (GLOBALES)

- `prom_total_time`: medida en segundos, tiempo total de la prueba de marcha calculado a partir de la articulación virtual `hip_center`, que es la articulación más cercana al centro de cadera físico.
- `prom_total_distance`: medida en metros, se usa el `hip_center` para medir la distancia total recorrida, dado que las pruebas de marcha se normalizan se ajusta a máximo 2,5 metros.

- `prom_speed_total`: medida en m/s, es la razón entre `prom_total_distance` y `prom_total_time`.

VARIABLES PIERNA IZQUIERDA (CADA VARIABLE TIENE UNA LATERALIDAD DEFINIDA)

- `prom_time_ankle_left`: medido en segundos, tiempo total que el tobillo izquierdo permaneció en el área de captura.
- `prom_time_distance_left`: medida en metros, distancia total recorrida por el tobillo izquierdo en el área de captura.
- `Speed_T_L`: medida en m/s, razón entre `prom_time_distance_left` y `prom_time_ankle_left`
- `Swing_T_L`: medida en segundos, tiempo promedio que tarda la fase de balanceo del pie izquierdo.
- `Stance_T_L`: medida en segundos, tiempo promedio que tarda la fase de apoyo del pie izquierdo.
- `Prom_total_time_swing_stance_left`: medida en segundos, suma entre los dos tiempos promedio de la fase de apoyo y balanceo, es decir, `Swing_T_L + Stance_T_L`
- `Swing_Prop_L`: valor entre 0 y 1, hace referencia al porcentaje o proporción de la fase de balanceo del pie izquierdo, se calcula como la razón entre `Swing_T_L` y `Prom_total_time_swing_stance_left`.
- `Stance_Prop_L`: valor entre 0 y 1, hace referencia al porcentaje o proporción de la fase de balanceo del pie izquierdo, se calcula como la razón entre `Stance_T_L` y `Prom_total_time_swing_stance_left`.
- Promedio de `ret_num_step_left`: promedio del número de pasos del pie izquierdo en el área de captura.
- Promedio de `ret_length_left_step`: medida en metros, promedio de la longitud de paso del pie izquierdo en el área de captura.

VARIABLES PIERNA DERECHAS (DEFINICIONES RESUMIDAS)

- Promedio de `ret_total_time_right`: tiempo promedio pierna derecha
- Promedio de `ret_total_distance_right`: distancia promedio de la pierna derecha durante el área de captura.
- $\text{Speed_T_R} = \text{promedio de } \text{ret_total_distance_right} / \text{promedio de } \text{ret_total_time_right}$
- `Swing_T_R`: segundos que dura la fase de balanceo del pie derecho
- `Stance_T_R`: segundos que dura la fase de apoyo del pie derecho
- Promedio de $\text{total_time_swing_stance_right} = \text{Swing_T_R} +$

Stance_T_R

- $\text{Swing_Prop_R} = \text{Swing_T_R} / \text{promedio de total_time_swing_stance_right}$
- $\text{Stance_Prop_R} = \text{Stance_T_R} / \text{promedio de total_time_swing_stance_right}$
- Promedio de `ret_num_step_right`: número de pasos del pie derecho
- Promedio de `ret_length_right_step`: longitud de paso del pie derecho

Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson

Jonatan Ordoñez Burbano, Andrés Navarro Cadavid

Universidad Icesi Cali, Colombia

1. INTRODUCCIÓN

El análisis de la marcha ha ganado particular interés entre los médicos y otros profesionales de la salud, es una herramienta útil para evaluar patrones de marcha relacionados con limitaciones funcionales causadas por enfermedades neurológicas [1], tales como la enfermedad de Parkinson. Esta última puede ser diagnosticada y controlada mediante el análisis de la marcha y de los patrones posturales que pueden indicar un mayor riesgo de caídas [2].

En el contexto clínico, la evaluación de la marcha es realizada por médicos y otros profesionales de la salud, quienes ejecutan un conjunto de pruebas para analizar variables como la velocidad de la marcha, la cadencia, la velocidad de giro y la longitud de la zancada; sin embargo, este tipo de análisis puede ser muy impreciso, pues por lo general, implican evaluaciones “a ojo”, en lugar de mediciones detalladas que permitan la valoración cuantitativa de la salud del paciente [2].

Los sistemas de captura de movimiento, tales como GAITRite o Vicon Motion Systems, son herramientas útiles para registrar patrones de marcha [3], [4], por lo que se utilizan en actividades relacionadas con la rehabilitación física, los deportes, la captura de movimiento y los videojuegos. Son usadas en entornos de laboratorio con el fin de registrar patrones de marcha de forma cuantitativa y precisa, a través de una serie de cámaras, sensores y marcadores [5]. A pesar de su efectividad, presentan dos limitaciones: la primera, su elevado costo, que se constituye en una barrera real para muchos; y sus requerimientos técnicos, pues su configuración es extensa y requiere de laboratorios especializados, preparación del paciente, accesibilidad y experiencia técnica necesaria para su manipulación [2], [6], [7].

Los sistemas de captura de movimiento se utilizan para obtener información fiable y objetiva que fundamente las decisiones clínicas, es útil para cuantificar e incluso predecir los resultados de diferentes intervenciones del personal de salud. El médico tiene que lidiar con una enorme cantidad de datos de los que se derivan varios gráficos, por lo que es necesario dirigir las visualizaciones a las partes relevantes de los datos y vincular partes específicas de la información a las posibles causas subyacentes [8].

El sensor Kinect de Microsoft, por su portabilidad y bajo costo, ha sido identificado como un dispositivo adecuado para realizar mediciones fuera del laboratorio en actividades de evaluación de la marcha, [7], [9]. Este dispositivo permite la medición de variables útiles a bajos costos y con muy poca demanda de tiempo para preparación del paciente. Varios estudios

sobre la marcha han demostrado que su precisión es suficiente para evaluar aspectos relacionados con la enfermedad de Parkinson, si se cuenta con el software apropiado [10], [11].

Se han propuesto diferentes técnicas de procesamiento de señales para diferentes señales biomédicas basadas en la descomposiciones de wavelets, tales como la reducción de ruido, la mejora de imágenes, la detección de microcalcificaciones en mamogramas y la adquisición y reconstrucción de imágenes en tomografías y resonancias magnéticas. Aunque los procesos de transformada wavelet se han utilizado ampliamente, existe consenso en que este tipo de técnicas se deben manejar con precaución [12].

Chau [13] reconoce la capacidad de las principales aplicaciones en la transformada de wavelets en el análisis de marcha para proporcionar información sobre la locomoción humana, pero agrega que esta información no puede obtenerse fácilmente con los métodos existentes [13]. Desde entonces, se han desarrollado herramientas para el diagnóstico de alteraciones de la marcha y el seguimiento del progreso de la rehabilitación, tales como un algoritmo basado en la transformada de wavelet de señales de un giroscopio, que hizo posible identificar los movimientos de despegue y golpe de talón a partir de los datos de velocidad angular del vástago [14].

En medicina, con un enfoque en los desordenes del movimiento, tal como el que se presenta en la enfermedad de Parkinson, se han utilizado transformadas wavelet como herramientas de diagnóstico clínico. Uno de los indicadores de esta enfermedad es la congelación de la marcha, la cual ha sido detectada por un sistema de electroencefalografía (EEG, ElectroEncephaloGraphy) inalámbrico de cuatro canales. Se han obtenido métricas de precisión (accuracy), sensibilidad y especificidad del orden de 75 % [2].

El grupo de investigación i2t de la Universidad Icesi, en colaboración con la Fundación Valle del Lili, una entidad médica de alta complejidad (nivel 3), ha desarrollado la herramienta de software e-Motion Capture System, la misma que utiliza un Kinect como dispositivo principal de captura de movimiento [2]. El sistema calcula variables motoras y espacio-temporales a partir del análisis de la marcha utilizando como dispositivos sensores Kinect y Orbbec Astra. El sistema completo requiere de: una computadora con software e-Motion; un área de captura libre de interferencias; y una persona, con formación en salud, para manejar el sistema [11].

Durante un análisis de marcha, un voluntario camina tres veces por un corredor de 1,5 m de ancho por 4 m de largo en la dirección del dispositivo sensor; la distancia del camino puede llegar incluso hasta seis metros si se

usa el sensor Orbbec Astra. Las pruebas capturadas por el sistema se procesan mediante un algoritmo de procesamiento de señales implementado en MATLAB, que convierte las señales en descomposiciones de *wavelets*. Para evaluar la precisión de las transformaciones wavelet, los resultados se comparan con el juicio clínico de los expertos sobre las fases de la marcha [2], [11].

El proceso de captura de la marcha y obtención de las variables se realiza en tres etapas: la primera consiste en el registro de la marcha mediante los sensores (Kinect u Orbbec Astra); la segunda en tomar las señales obtenidas en el paso anterior y procesarlas mediante el algoritmo implementado en MATLAB, el cual genera un reporte en MS-Excel que se almacena en una computadora local; la tercera, en refinar dicho reporte a través de un conjunto de microprocesos que eliminan las variables innecesarias y agregan otras que son requeridas para el análisis. El informe final se entrega al personal médico, así ellos pueden conocer el estado de salud del paciente.

Una limitación importante de este proceso es que la ejecución del algoritmo de procesamiento de señales y la mejora de los resultados puede tomar entre dos y tres horas por cada dos o tres pacientes. Por este motivo, el registro de la marcha se realiza en un día diferente al del procesamiento de la señal. Asimismo, actualmente un miembro del grupo i2t debe desplazarse hacia el centro médico para el tratamiento de los datos, pues este proceso aún no se ha automatizado; esta misma persona debe además refinar el reporte de las variables, para su entrega al personal médico, su usuario final.

Dados los antecedentes explicados, es evidente la necesidad de crear una mejor solución para el procesamiento y visualización de la información recopilada por el sistema de captura e-Motion (sensor Kinect v1) utilizado para analizar la marcha en pacientes con enfermedad de Parkinson en la Fundación Valle del Lili. Esa solución, de acuerdo con un análisis del problema realizado por un equipo interdisciplinario de las dos instituciones, debería mejorar tanto los procesos asociados al procesamiento de las señales en wavelets como su reporte al personal médico.

Como se mencionó, aunque existen soluciones en el mercado que ofrecen un conjunto de servicios para el análisis de la marcha e incluyen la infraestructura necesaria para operar, en la práctica no están disponibles, pues sus requerimientos —de espacio y técnicos— y su elevado costo no guardan relación con los limitados recursos generalmente disponibles en las entidades de salud de el país.

De acuerdo con el contexto proporcionado acerca de e-Motion, debido a los pasos intermedios que se deben realizar entre la toma de pruebas a los

pacientes y la entrega de los resultados al médico, la plataforma requiere: que el procesamiento de las señales se realice en un día diferente al de la toma de muestras; que una persona técnica visite las instalaciones médicas y acceda al equipo donde se tomaron las capturas; y que el informe que se extrae de este proceso se refine a través de un conjunto de microprocesos que permiten generar el informe final y entregarlo a los médicos.

En este proyecto, el problema del análisis de marcha se aborda desde un enfoque de visualización de datos y procesamiento de grandes volúmenes de datos. Considerando los recursos computacionales limitados, la implementación del algoritmo de procesamiento de señales se enfocó en la ejecución paralela de procesos y en el uso de los recursos de hardware disponibles (donde se ejecutará el algoritmo).

La implementación actual en Matlab no cuenta con este tipo de procesamiento, ya que en su implementación inicial fue pensado para sistemas concurrentes; además, para acceder a este tipo de procesamiento se requiere de una licencia. Los lenguajes de programación modernos usan bibliotecas de procesamiento nativas, un recurso que se puede usar para este trabajo.

La importancia de reducir el tiempo de procesamiento de las pruebas de los pacientes es clara, más aun cuando está previsto, en un futuro cercano, añadir nuevos sensores de captura de movimiento encargados de procesar la información. Estos módulos podrían agregarse como un nodo adicional al sistema de procesamiento.

Por otra parte, la visualización de la información de los pacientes se gestionó usando tecnologías web emergentes que permiten disponer de información proveniente de cualquier dispositivo. Asimismo, es importante eliminar el factor humano en el flujo de procesamiento y entrega de resultados. Por lo tanto, se propuso crear un sistema de visualización que les permita a los médicos conocer los resultados de los pacientes sin necesidad de que una persona trabaje previamente en el procesamiento y refinación de los resultados.

Abordar este problema es importante no solo por su potencial para reducir el tiempo de procesamiento de los datos y producción de reportes, sino porque les permitiría a los médicos acceder a la información de sus pacientes inmediatamente después de haber procesado la prueba, sin necesidad de la intervención de alguien más que refine la información. En trabajos posteriores se podría adaptar la solución a otros dispositivos de captura de movimiento, tales como el Kinect v2 y Orbbec Astra.

Dada la importancia del análisis de la marcha para los profesionales de la salud y su impacto en la evaluación del paciente, el proyecto se enfocó en la explotación de técnicas de procesamiento de big data y algoritmos de procesamiento de señales, para transformar las señales captadas por el sistema e-Motion, con sensor Kinect v1, en descomposiciones wavelet, de tal manera que la información reportada le permita al personal médico reconocer patrones de marcha en pacientes con enfermedad de Parkinson.

Lo anterior presenta los siguientes retos: ¿Cómo aprovechar los recursos de la máquina para procesar las señales captadas por el sistema e-Motion? ¿Cómo evaluar el desempeño de una implementación de procesamiento de señales y la confiabilidad de la información? y ¿Qué tipo de técnicas utilizar para llevar a cabo las tareas de procesamiento en paralelo?

La recopilación de informes de señales capturadas de pacientes con enfermedad de Parkinson en la Fundación Valle del Lili supone una importante restricción, pues este tipo de información requiere de permisos especiales que garanticen su adecuada manipulación, pues se trata de información clínica sensible. Esta restricción se mitigó mediante la suscripción de acuerdos de no divulgación para el uso de los conjuntos de datos disponibles seleccionados para fines de investigación.

Por otra parte, es claro que mientras las bibliotecas de procesamiento de señales de Matlab son suficientemente robustas como para realizar este tipo de tareas, las que han surgido en entornos de tiempo de ejecución para realizar tareas de procesamiento de señales, como Python, requieren de la validación de su confiabilidad para realizar el tipo de tarea específica requerida en este proyecto: la nueva implementación debe ser capaz de realizar dos tareas fundamentales de procesamiento de señales: el análisis de picos y la extrapolación de señales para los valores faltantes o nulos.

Con base en lo anterior, el proyecto definió como objetivo el desarrollo de un sistema para procesar en menor tiempo los resultados de la captura de la marcha realizada con e-Motion (usando como sensor Kinect v1), en pacientes con enfermedad de Parkinson en la Fundación Valle del Lili, de tal manera que el sistema sea capaz de visualizar los resultados para complementar el diagnóstico clínico realizado por los médicos.

Para lograrlo, se definieron como objetivos específicos:

- implementar una solución de software de procesamiento de señales personalizada para el procesamiento de los resultados de la captura de la marcha realizada con e-Motion usando Kinect v1 en pacientes con enfermedad de Parkinson en la Fundación Valle del Lili;

- validar que las señales y los resultados generados por el sistema sean similares a los arrojados por el sistema predecesor, usando métricas de correlación de errores; y que el tiempo de procesamiento se reduzca, mediante pruebas de tiempo de ejecución; e
- implementar un sistema de visualización para los médicos especialistas de la Fundación Valle del Lili, útil para complementar el diagnóstico clínico de los pacientes con enfermedad de Parkinson.

El principal aporte de este proyecto es la implementación de una herramienta de sistema software que ayuda con el procesamiento de las señales captadas por e-Motion, utilizando el sensor Kinect v1 y permite la visualización de los resultados por parte del personal médico, sin intervención de un tercero. Esta aplicación puede servir como base para extender las funcionalidades a otras implementaciones que usaron los sensores Kinect v2 u Orbbec Astra. El enfoque propuesto puede conducir a nuevas opciones de procesamiento de señales, utilizando alternativas de código abierto que ya existen en el mercado para aprovechar al máximo las capacidades computacionales.

Una de las limitaciones del actual procesador de señales de Matlab es que ha sido diseñado para el procesamiento de archivos concurrentes y secuenciales. Si bien Matlab cuenta con una caja de herramientas para realizar este tipo de procesamiento [15], su costo es, en muchos casos, una limitación. Además, el uso de esta herramienta en Matlab requeriría un rediseño del sistema de procesamiento.

La implementación propuesta en este proyecto tiene a Python como su lenguaje principal, se trata de un lenguaje de programación de alto nivel, con una sintaxis sencilla y clara de entender, que admite múltiples paradigmas de programación de forma nativa, incluida la programación paralela (multihilo) [16].

El procesamiento de subprocessos múltiples proporcionado por este lenguaje permite que múltiples subprocessos procesen datos y recursos [16]. La principal ventaja de la implementación propuesta es el rendimiento, ya que se pueden ejecutar múltiples procesos de procesamiento de señales simultáneamente, lo que reduce el tiempo que tarda el sistema en procesar una cantidad N de archivos.

Aunque inicialmente, el sistema está diseñado para procesar señales de tobillo, dado que este es un módulo de procesamiento que se ejecutará en un número N de subprocessos, es posible agregarle en el futuro más módulos de procesamiento, el procesamiento de brazos, por ejemplo.

2. MARCO TEÓRICO

Señal se refiere a una forma de onda eléctrica que cambia su amplitud con el tiempo; las señales se pueden describir completamente en el dominio del tiempo o de la frecuencia; la frecuencia (f) de una señal es el número de ciclos que realiza en cada segundo (el paso de una polaridad a otra y la vuelta a la primera) y se mide en Hertz (H) o ciclos por segundo; el tiempo que tarda una onda en completar un ciclo se denomina período (T) [17] y se calcula como indica la ecuación (1).

$$T = \frac{1}{f} \text{ segundos} \quad (1)$$

La longitud de onda (λ) se define como la distancia que recorre una onda en un tiempo igual a un periodo y se calcula como se indica en la ecuación (2), en donde c es la velocidad de propagación de la onda y f la frecuencia.

$$\lambda = \frac{c}{f} \text{ metros} \quad (2)$$

Una transformada wavelet corresponde a la descomposición de una señal en un conjunto de funciones básicas, mediante operaciones como contracciones, expansiones y traslaciones de una función madre $\Psi(t)$ [18]; esta herramienta matemática se utiliza en el procesamiento de señales, el control de procesos y la detección de anomalías sintomáticas en medicina e ingeniería [19]. La FIGURA 1 corresponde a la descripción de la transformada *wavelet*.

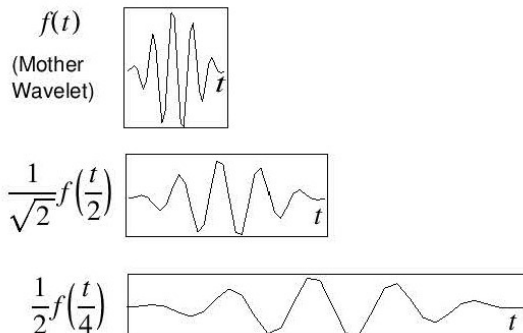


Figura 1. Transformada wavelet [20]

Las *wavelets* madre se clasifican en familias (FIGURA 2) de acuerdo con su utilidad en: Haar, Daubechies, Biorthogonal, Coiflets, Symlets, Morlet, Mexican Hat y Meyer [21]. La familia Daubechies, por ejemplo, es un conjunto de *wavelets* ortonormales que se aplica en el análisis de señales discretas y es útil para la reconstrucción de imágenes.

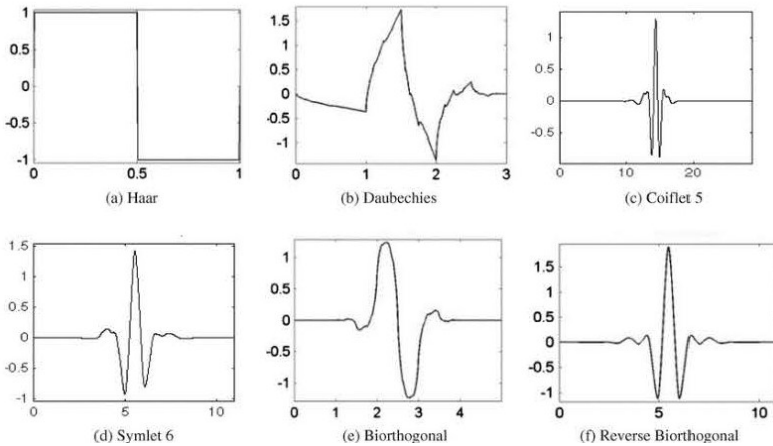


Figura 2. Ejemplos de familias wavelet [22]

La FIGURA 3 es un ejemplo del uso del procesamiento de señales digitales en el diagnóstico y tratamiento de la enfermedad de Parkinson y corresponde a la señal de un tobillo captada en un paciente con dicha enfermedad; en este caso, se aplicó a la señal una transformada de Daubechies con ocho niveles de desvanecimiento. La señal obtenida permitió clasificar las fases de la marcha en fase de apoyo y fase de balanceo. Con los resultados de la clasificación se generaron variables espacio-temporales, como los tiempos de apoyo y de balanceo, la velocidad y la velocidad de la marcha [23].

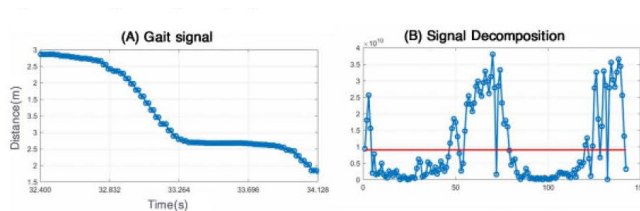


Figura 3. Señal de movimiento del tobillo izquierdo (a); descomposición de wavelet resultante de la señal (b) [23]

El procesamiento paralelo de datos (FIGURA 4) es un método utilizado en informática en el cual una tarea compleja se divide en conjuntos de tareas más pequeñas que se ejecutan simultáneamente en varias CPU (microprocesador), lo que reduce el tiempo de procesamiento [24].

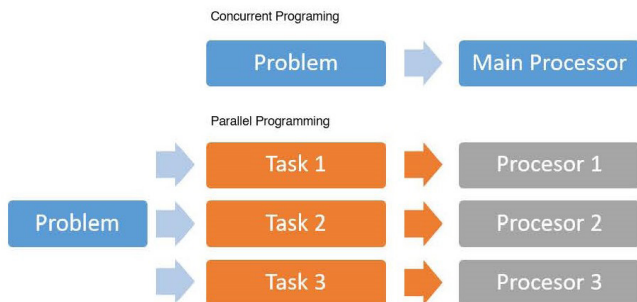


Figura 4. Procesamiento concurrente vs. Procesamiento paralelo [25]

Las tareas de procesamiento digital de señales (DSP, *Digital Signal Processing*) conducen al desarrollo de procesadores especiales que mejoren el rendimiento del sistema, el cual está determinado por el modelo matemático utilizado por él.

Algunos modelos utilizados para la conversión de señales son: la transformada discreta de Fourier (DFT, *Discrete Fourier Transform*), la transformada rápida de Fourier (FFT, *Fast Fourier Transform*) y la tecnología de multiplexación por división de frecuencia ortogonal (OFDM, *Orthogonal Frequency Division Multiplexing*).

Se pueden aplicar técnicas de paralelismo a nivel de operación, creando un procesador especial tipo SIMD (*Single Instruction, Multiple Data*), así el modelo matemático se puede encapsular en un módulo operativo al que se le asignan cadenas de procesamiento [26].

SIMD es una técnica de paralelismo aplicada a un flujo de datos, esta arquitectura aplica la misma operación a un conjunto de datos; una sola unidad de control común despacha la instrucción a diferentes unidades de procesamiento; cada unidad ejecuta la misma instrucción, pero recibe un conjunto de datos distinto; cada unidad de procesamiento se puede replicar para maximizar el uso de los recursos informáticos y reducir el tiempo de procesamiento de datos [24] (ver FIGURA 5).

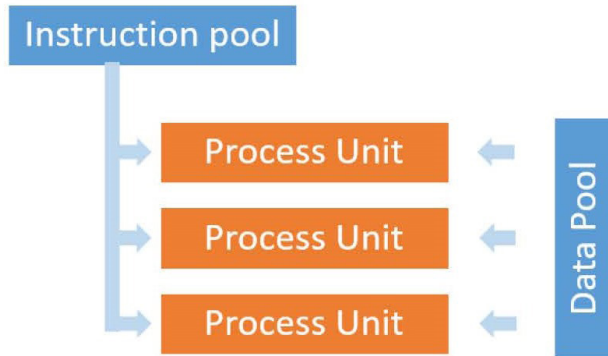


Figura 5. Arquitectura SIMD [27]

3. ESTADO DEL ARTE

La solución propuesta en este proyecto se enfoca en una nueva implementación de procesamiento de señales capaz de transformar las señales captadas por el sistema e-Motion y mostrar los resultados. En este apartado se incluyen implementaciones previas del grupo de investigación i2t de la Universidad Icesi realizadas en colaboración con la Fundación Valle del Lili, que sirven como referencia base.

Los criterios de evaluación se basan en las limitaciones de Matlab, tales como el tiempo de procesamiento, la construcción de informes y la facilidad de uso. La nueva solución, por lo tanto, debe satisfacer las limitaciones propuestas en la descripción del problema, reducir el tiempo de ejecución de la información capturada con los sensores y continuar con el trabajo previo de i2t.

Otras herramientas que solo resuelven parcialmente el problema pero son relevantes para el proyecto, se han incluido también. Para la caracterización de los trabajos relacionados se definieron las siguientes dimensiones:

- código abierto (open source): la herramienta puede ser utilizada e implementada libremente, sin costo;
- visualización, la solución tiene una forma de obtener reportes automáticos o de visualizar los datos;
- técnica de procesamiento, la forma en que la herramienta procesa y transforma las señales: concurrente, paralela o distribuida; y
- dominio del problema, contexto en el que se aplica la herramienta para el procesamiento de señales: análisis de señales generales y análisis de señales clínicas.

En las TABLAS 1, 2 y 3 se presenta el resumen de las soluciones de procesamiento de señales enfocadas en los campos de la marcha y la enfermedad de Parkinson.

Las referencias encontradas se han dividido en tres conjuntos. El primero presenta soluciones completas que permiten el procesamiento y análisis de señales de marcha en entornos controlados (TABLA 1); el segundo, soluciones personalizadas para el análisis de la marcha o el procesamiento de señales digitales centradas en la enfermedad de Parkinson (TABLA 2); y el tercero, herramientas (bibliotecas) que admiten el procesamiento de señales digitales, pero que no representan una solución completa y requieren personalización (TABLA 3). En todos los casos, para efectos de comparación, se incluye una línea con la información correspondiente al proyecto reportado en este documento.

Aunque algunos estudios implementan soluciones de apoyo para el procesamiento de señales o la visualización de los resultados, ninguno parece abordar un flujo de visualización de procesamiento completo. La solución debe estar encaminada a ajustarse a las restricciones y cumplir con parte de los requerimientos del negocio, que es el principal desafío que aborda este proyecto.

Dos de las soluciones investigadas (GAITRIte y Vicon Motion Capture System) se ajustan a las necesidades mencionadas y representan sistemas completos de procesamiento y análisis de señales, sin embargo, las restricciones económicas hacen en muchos casos inviable la compra de estos sistemas. Estas dos soluciones se descartan ya que el objetivo del grupo de investigación i2t es encontrar soluciones a través de tecnologías de acceso abierto. Además, la adquisición de estos sistemas requiere una gran superficie para la distribución de los sistemas de captura, que no está disponible en los centros médicos. El propósito de la solución es que sea portátil como e-Motion.

Otras soluciones se pueden utilizar únicamente para implementar el sistema de procesamiento de señales, ya que proporcionan funcionalidades necesarias para la transformación de las señales en información útil para los médicos. Aunque SciPy ofrece funciones que ayudan en el procesamiento de señales, solo se pueden usar para complementar el procesamiento general de señales. Aunque estas soluciones no cumplen con los requisitos del problema, se pueden utilizar para facilitar el procesamiento de datos.

Algunas implementaciones permiten el procesamiento de diferentes señales de entrada de pacientes con enfermedad de Parkinson. Es el caso del

Tabla 1. Caracterización y trabajos relacionados (soluciones completas)

Referencia	Código abierto	Visualización	Técnica de procesamiento	Dominio del problema	Observaciones
GAITRite [28]	No	Si	No especificada	Análisis general y clínico de la marcha	GAITRite es un sistema de análisis de la marcha que emplea una patalela sensible a la presión para registrar las variaciones de la marcha. La solución incluye una cámara y un software especializado. Debe adquirirse mediante la compra del sistema. [3]
Vicon Motion Capture System [29]	No	Si	No especificada	Análisis general y clínico de la marcha	Vicon es el líder mundial en cámaras, software y sistemas de captura de movimiento. Ofrece sistemas de alta precisión para la captura de movimiento, pero de alto costo. [4]
Este proyecto	Si	Si	Paralela	Análisis de señales clínicas	

Tabla 2. Caracterización y trabajos relacionados (soluciones personalizadas)

Referencia	Código abierto	Visualización	Técnica de procesamiento	Dominio del problema	Observaciones
E-Motion Kinect V1/V2 [11]	No	No	Concurrente	Análisis clínico de la marcha	Esta solución permite el análisis de diferentes partes del cuerpo de los pacientes con enfermedad de Parkinson mediante el uso de Kinect y software especializado. El sistema no tiene un método automático para generar informes y requiere comprar licencias de Matlab para el procesamiento de señales. [11]
Portable Device for Real Time Analysis of Speech of People with Parkinson's Disease [30]	No	Si	Buffer	Análisis clínico del habla	Sistema embebido utilizado para el análisis del habla en pacientes con enfermedad de Parkinson en tiempo real. El sistema implementa un procesador de señal en Python y C++, y muestra la captura de audio en Matlab. [30]
Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect [31]	No	No	Concurrente	Análisis clínico de la marcha	Sistema de análisis de la marcha desarrollado mediante un clasificador bayesiano, MS-Kinect y seguimiento articular 3D; para la captura de imágenes utiliza la caja de herramientas de adquisición de imágenes de Matlab. Utiliza además el entorno computacional de Matlab para el procesamiento de señales digitales. El sistema requiere la compra de una licencia de Matlab para acceder a la caja de herramientas de captura de imágenes. [31]
Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect [32]	No	Si	Concurrente	Análisis clínico de la marcha	Esta aplicación permite estimar las características de la marcha de los pacientes con enfermedad de Parkinson; utiliza Matlab como procesador de señales e implementa una interfaz para configurar la captura de la marcha; no cuenta con un sistema de visualización de los resultados, solo con una interfaz de control. Requiere adquirir una licencia de Matlab para el procesamiento de señales digitales. [32]

Tabla 3. Caracterización y trabajos relacionados (herramientas)

Referencia	Código abierto	Visualización	Técnica de procesamiento	Dominio del problema	Observaciones
PyWavelets [33]	Si	No	Concurrente	Análisis de señales generales	Paquete de Python que implementa varias transformaciones de wavelets discretas n-dimensionales; la herramienta solo incluye una API para el procesamiento de señales y no tiene una extensión para visualizar los informes. Aunque la biblioteca aprovecha un lenguaje de bajo nivel ©, no maximiza las operaciones al aprovechar el procesamiento paralelo o distribuido.
SciPy	Si	No	Concurrente	Análisis de señales generales	La biblioteca es un ecosistema basado en Python de software de código abierto para matemáticas, ciencias e ingeniería y cuenta con un ecosistema especializado para transformadas wavelet. Sin embargo, debe confiar en bibliotecas como matplotlib para la generación de los gráficos y las métricas de los resultados.
Este proyecto	Si	Yes	Paralela	Análisis de señales clínicas	

estudio “Diseño e implementación de un dispositivo portátil para el análisis en tiempo real del habla de personas con enfermedad de Parkinson” [30], el cual propone un sistema de captura de voz y procesamiento de estas señales. Aunque en él se implementa un procesador personalizado usando Python y C++, requiere de Matlab para la visualización de los resultados, por lo que depende de la adquisición de licencias para el uso de la suite Matlab.

En esa misma línea, dos investigaciones proponen el uso de procesadores Matlab para el análisis de la marcha y la extracción de características [31], [32]. Ambos sistemas de procesamiento se desarrollaron en la suite Matlab utilizando cajas de herramientas de Matlab, por lo que están “atados” al equipo físico donde están instalados. Su uso implica los costos adicionales de adquisición de las cajas de herramientas.

El grupo de investigación i2t ha desarrollado procesadores de señal en Matlab, soluciones que requieren del factor humano para el procesamiento y refinamiento de los datos que serán entregados al personal médico. Estos procesadores requieren una computadora y una licencia de Matlab para acceder a los servicios de procesamiento de señales. Además, como se mencionó, el procesamiento manual de los datos requiere personal que sepa cómo ejecutar los algoritmos, lo que crea una indeseable dependencia del médico para tener acceso a los resultados.

El proyecto reportado en este documento combina los atributos mencionados en una única solución orientada al procesamiento de señales y visualización de resultados, basada en un sistema que aprovecha los recursos del sistema, combinado con una arquitectura de procesamiento paralelo y uso de nuevas tecnologías para desarrollar una interfaz para mostrar informes. La herramienta propuesta elimina el factor humano en el procesamiento de datos y facilita a los médicos el acceso directo a los informes.

4. METODOLOGÍA

Este trabajo es un proyecto de desarrollo de software para el que se eligió la metodología del Ciclo de Vida de Desarrollo de Software (SDLC, *Software Development Life Cycle*). Bajo esta metodología, el proceso de creación de software se divide en varias actividades que definen una pequeña parte del proceso general.

En la FIGURA 6 y en la TABLA 4, se presentan y describen, respectivamente, las fases de esta metodología.

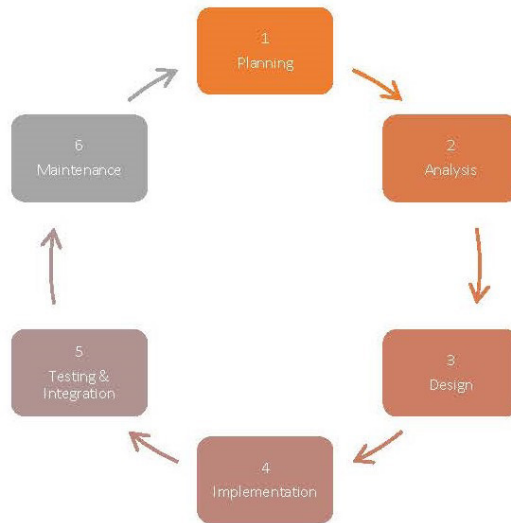


Figura 6. Fases de la metodología SDLC [34]

Tabla 4. Fases de la metodología SDLC

Fase	Descripción
Planificación	Esta fase incluye actividades como la determinación del alcance del proyecto, el estudio de viabilidad, el análisis de riesgos asociados, la estimación de los costos y la asignación de recursos.
Análisis	Esta fase corresponde al proceso de intentar descubrir qué se necesita en el proyecto y de comprender los requisitos del sistema, es decir las características que debe tener.
Diseño	En esta fase se estudian las posibles opciones de implementación del software a construir y se decide su estructura general.
Implementación	En esta fase se deben elegir las herramientas adecuadas, un entorno de desarrollo que facilite el trabajo y un lenguaje de programación adecuado al tipo de software que se va a construir.
Prueba e integración	En esta fase se busca detectar las fallas cometidas en las etapas anteriores para corregirlas. El siguiente paso es poner en funcionamiento el software, por lo que se debe planificar el entorno teniendo en cuenta las dependencias entre sus diferentes componentes.
Mantenimiento	En esta fase se eliminan los defectos detectados durante la vida, se adapta el software a los nuevos requerimientos y se añaden, de ser necesarias, nuevas funcionalidades.

Para la ejecución del proyecto se utilizó un modelo incremental que permite la descomposición del proyecto en una serie de entregas, donde cada uno de ellos ofrece una porción de la funcionalidad total. El proyecto se desarrollará en los siguientes incrementos: el primero enfocado en el desarrollo del sistema de procesamiento de las señales; el segundo, en el desarrollo del sistema de visualización requerido para reportar los resultados del análisis de la marcha a los médicos; y el tercero en la ejecución de un conjunto de pruebas dirigido a validar la operatividad del sistema.

Los datos utilizados para este proyecto serán capturas de movimiento tomadas con el sistema e-Motion con Kinect v1. Estas capturas fueron tomadas a voluntarios de la Fundación Valle del Lili. Algunos sujetos tenían enfermedad de Parkinson en sus primeras etapas y otros estaban sanos y caminaban normalmente. Estas capturas se almacenaron en una computadora operada por el personal médico a quien se le solicitó periódicamente información.

Cada captura realizada usando e-Motion genera un archivo JSON que contiene las posiciones X, Y y Z de una articulación en un momento en el tiempo. En cada período de tiempo t , se capturan las posiciones de las veinticinco articulaciones disponibles por el Kinect [35]. Cuando finaliza el proceso de captura, esta información se condensa en un archivo final en formato JSON, que es procesado por una implementación de procesador de señal hecha en Matlab.

Estos archivos pesan entre 1 y 4 MB, este tamaño depende del tiempo que se tarde en la captura con un paciente. La FIGURA 7 corresponde a un fragmento del archivo generado con e-Motion.

La propiedad `SkeletonLocation` contiene una matriz de objetos, cada uno de ellos contiene la captura de las articulaciones en el tiempo t ; las articulaciones en cada momento t se almacenan en la propiedad `JointLocation`. Para este fragmento se muestra la captura de la base de la columna vertebral y sus coordenadas X, Y y Z.

Para la implementación del nuevo procesador de señales, se limitó a diez el número máximo de archivos que se pueden procesar en cada ejecución. Esta restricción obedece a la necesidad de limitar la cantidad de memoria utilizada para evitar saturar el número de archivos cargados simultáneamente en la memoria.

En la FIGURA 8 se presenta el ejemplo de una señal de profundidad capturada en la articulación del brazo derecho de un paciente (en azul); la señal u onda sinusoidal muestra variaciones a lo largo de la trayectoria o eje de

```
"CaptureTest": {
  "Date": "\\Date(1617653517292-0500)\\",
  "Id": "9c643233-4781-4589-893d-204945f84f12",
  "KinectMachineStatus": [
    {
      "CaptureTestId": "00000000-0000-0000-0000-000000000000",
      "ElevationAngle": null,
      "Id": "aa998ab3-0e5d-450d-8515-0267bfb1fadbf",
      "Location": "Frontal",
      "SkeletonLocation": [
        {
          "Id": "75de3101-014a-46dc-a99c-01ff54f4d02d",
          "JointLocation": [
            {
              "Id": "9736fc50-9a3b-4734-b637",
              "JointPointX": 0.006995074,
              "JointPointY": 0.05861119,
              "JointPointZ": 4.446734,
              "JointTrackingState": "Tracked",
              "JointType": "SpineBase",
            },
            ...
          ]
        },
        ...
      ]
    },
    ...
  ]
}
```

Figura 7. Fragmento del archivo JSON generado por e-Motion

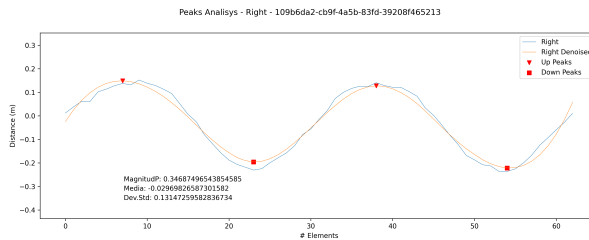


Figura 8. Señal de profundidad del brazo derecho y señal ajustada después del procesamiento

tiempo (# Elementos); también se observan los picos superior e inferior, los cuales corresponden al movimiento del brazo durante la marcha, cuando el brazo está completamente estirado y retraído. Finalmente, después de ejecutar el procesamiento de la señal, la señal (en rojo) se ve más limpia y se corrigen las imperfecciones.

Para validar que los resultados generados por el nuevo procesador de señales sean correctos y similares a los de su antecesor, se utilizaron dos métricas: el coeficiente de correlación de Spearman y la distancia de Hamming, ambas tienen como función medir el error de dos conjuntos de datos. En este caso pertenecen a la descomposición de señales de un grupo de archivos json (capturas de pacientes). A partir de cada archivo procesado, se

genera una descomposición wavelet de los tobillos la cual se compara con el resultado generado por el procesador de señales en Matlab.

El coeficiente de correlación de Spearman, que se describe a continuación, se utilizó en estudios realizados por el grupo de investigación i2t para comparar la eficacia de las cámaras de captura de movimiento, por ello se utilizó en esta tesis para validar los dos procesadores de señal [5]. La distancia de hamming se introducirá como una segunda métrica, con el fin de comparar los resultados; esta métrica se ha utilizado para validar el nivel de precisión en sistemas de descomposición de señales centrados en el procesamiento de imágenes, como el reconocimiento de la textura del iris [36], [37].

Finalmente, para medir el tiempo de ejecución de los procesadores de señal, y la mejora del nuevo procesador respecto de su versión anterior, se utilizaron pruebas de tiempo de ejecución de código, la cual consiste en medir el tiempo que se tarda en ejecutar un fragmento de código con diferentes conjuntos de datos o condiciones [38]. Para ambos procesadores de señales, se mide el tiempo que les toma procesar diez archivos de pacientes. Ambos procesadores se instalaron en la misma computadora para garantizar las mismas condiciones de hardware.

La distancia de Hamming es una medida de similitud natural utilizada para calcular la distancia entre un código binario ideal y su alternativa. La alternativa ideal (código binario ideal) es una alternativa virtual en la que los valores se expresan tan cerca como sea posible de los valores ideales, lo que corresponde a la lógica del pensamiento humano. Finalmente, la distancia de Hamming se calcula como el número de posiciones de bit en donde los dos códigos binarios son diferentes. [39], [40].

El coeficiente de Spearman (ver escala en TABLA 5) es una de las medidas de correlación de uso común en investigación médica, se trata de una prueba no paramétrica que permite medir el rango de asociación entre dos

Tabla 5. Lectura de la correlación de Spearman

Spearman ρ	Correlación
≥ 0.70	Relación muy fuerte
0.40-0.69	Relación fuerte
0.30-0.39	Relación moderada
0.20-0.29	Relación débil
0.01-0.19	Sin relación (o con relación irrelevante)

variables, en este caso, entre dos señales: un valor de correlación tendiente a 1 indica que las dos variables son similares o idénticas, y un valor tendiente a -1, lo contrario.

5. RESULTADOS

5.1. LOS DATOS

Al inicio de la prueba, e-Motion rastrea las articulaciones del cuerpo del paciente mientras este camina en dirección a la cámara. Cada articulación es capturada por Kinect (FIGURA 9) y almacenada en un archivo json con su respectiva unidad de tiempo (época). Aunque el Kinect captura veinticinco articulaciones del cuerpo humano [42], solo se utilizan las señales del tobillo izquierdo, el tobillo derecho y la base de la columna para este análisis de la marcha.

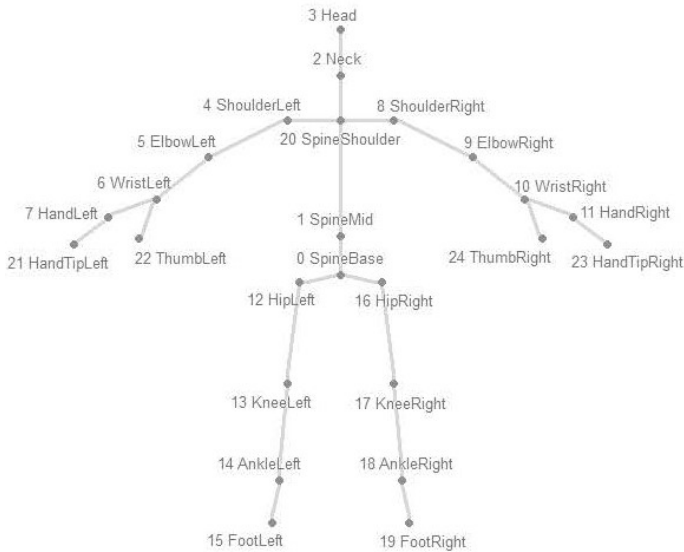


Figura 9. Mapeo de la captura de articulaciones realizada por el sensor Kinect v1 [42]

En el diagrama de la FIGURA 10 se observa el flujo a automatizar, el cual corresponde al proceso posterior a la captura de datos mediante e-Motion. Como se muestra en la imagen, un técnico calificado debe realizar el procesamiento de señales usando Matlab y también debe refinar los resultados obtenidos antes de presentárselos al médico.

Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson

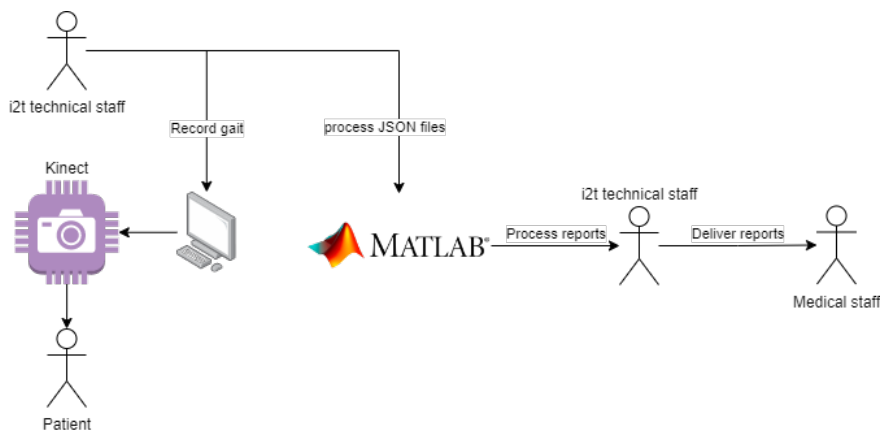


Figura 10. Captura de la señal, procesamiento de wavelet y entrega de resultados

La solución propuesta en este proyecto es capaz de recibir los archivos obtenidos en la captura de la marcha, procesar las señales, refinar —si fuera necesario—, los datos de salida, y presentar su visualización.

5.2. EL SISTEMA

La imagen que presenta la arquitectura del sistema (FIGURA 11) indica que él está formado por dos componentes principales: un servidor de procesamiento de señales y una aplicación web.

APLICACIÓN DE SERVIDOR (PYTHON)

Para aprovechar las bibliotecas de Python (NumPy, SciPy, etc.), se desarrolla un servidor web utilizando Flask para proporcionar una interfaz API

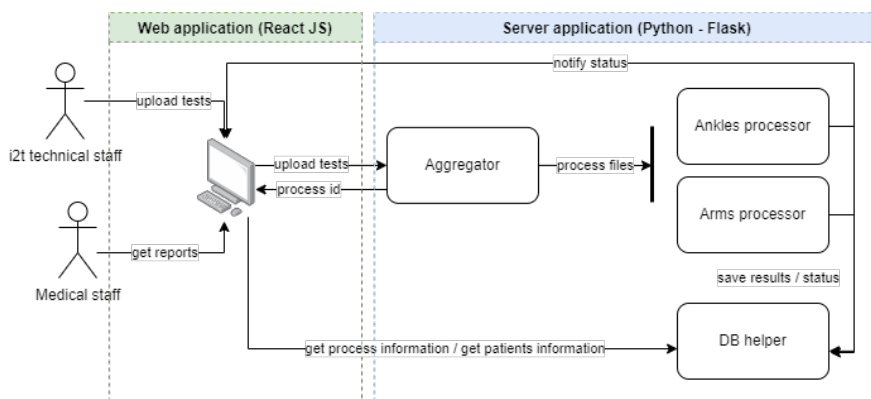


Figura 11. Arquitectura del sistema

REST, que cualquier cliente puede utilizar a través del protocolo HTTP/HTTPS.

La arquitectura del servidor está basada en el patrón de diseño de microservicios Aggregator, el cual consiste en separar una aplicación compleja en servicios pequeños, autónomos, independientes y desplegables. En aplicaciones hechas por microservicios, los módulos grandes se separan en componentes atómicos que pueden funcionar de forma autónoma e independiente.

Estos microservicios están orquestados por el componente Aggregator que recopila y procesa los resultados. En general, Aggregator ejecuta el conjunto de microservicios, procesa la información resultante y devuelve la información al cliente [43].

La aplicación recibe una solicitud con los archivos de prueba del paciente y responde una identificación de proceso asociada con ella; luego, el componente del procesador Signal lanza múltiples subprocesos paralelos con cada uno de los archivos, cada subproceso es responsable del procesamiento de la señal digital. El procesador de señales del componente se encarga de orquestar los subprocesos y notificar el progreso de la tarea a través de un socket. Además, en caso de errores, este componente notifica esta información a través del socket. Finalmente, el componente del procesador Signals sincroniza el proceso en la base de datos para mantener la integridad del proceso.

APLICATIVO WEB (REACT JS)

Se desarrolló una aplicación Web independiente para ser utilizada por médicos y técnicos expertos en i2t, ella se comunica con el servidor a través de una Rest API y permite la carga de archivos y la visualización de informes de pacientes. Esta aplicación se suscribe a un socket de servidor para así mantener una comunicación asíncrona sobre el estado del procesamiento de archivos.

Es importante resaltar que la propuesta de diseño y la validación de la interfaz web se realizaron previamente con el panel de diseño del grupo de investigación y el cuerpo médico de la Fundación Valle del Lili. La implementación de la interfaz es un esfuerzo adicional para complementar el nuevo sistema de procesamiento de señales.

Las FIGURAS 12, 13 y 14 corresponden a capturas de pantalla de la aplicación web. En la FIGURA 15 se presenta un diagrama detallado de la arquitectura del sistema.

Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson

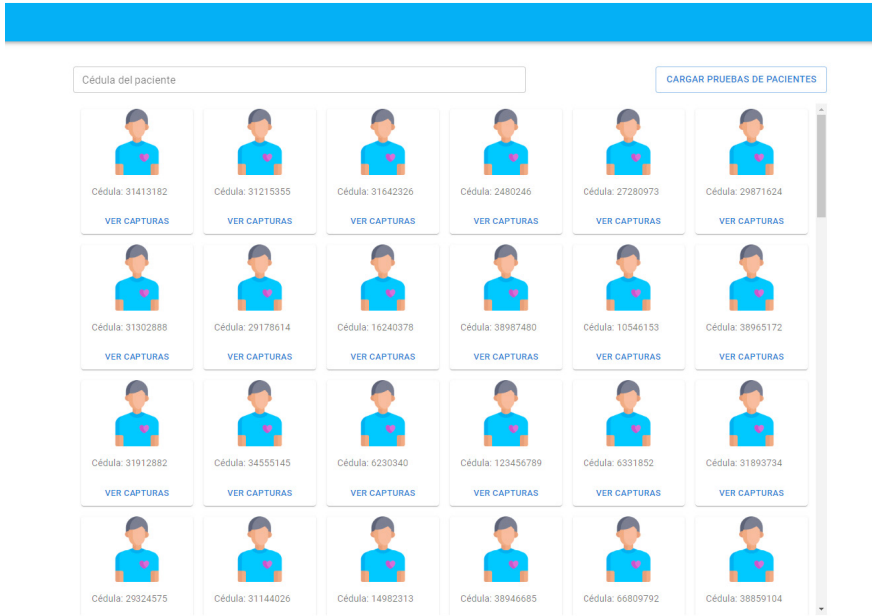


Figura 12. Lista de pacientes

Paciente: 31413182

Resultados análisis tobillos

Imágenes	# de prueba	Initial_time	ret_total_time_test	ret_total_distance_test	ret_velocidad_total	ret_total_time_left	ret_total_distance_left	ret_total_speed_left	ret_time_left_swing	ret_time_left_support
	702052aa-e76c-4506-9f05-9a6725895f1f	25-06-2022 2:57:21 pm	2.749	2.403	0.896	2.752	2.329	0.846	0.469	0.67
	c0ba8724-c0e0-4531-a909-a37789647a27	25-06-2022 2:57:01 pm	3.116	2.488	0.798	3.121	2.829	0.906	0.432	0.53
	1c0840c8-bf47-4769-8952-12159909f93d	25-06-2022 2:56:52 pm	2.456	2.452	0.998	2.456	2.495	1.016	0.434	0.74

DESCARGAR REPORTE

Resultados análisis brazos

Imágenes	# de prueba	Initial_time	magnitud_left	time_left	velocity_left	magnitud_right	time_right	velocity_right	asa
	702052aa-e76c-4506-9f05-9a6725895f1f	25-06-2022 2:57:21 pm	0.081	0.778	0.105	0.186	0.624	0.298	21.336
	f2c64772-1363-4069-996f-781348116c63	25-06-2022 2:57:11 pm	0.219	0.536	0.409	0.045	0.491	0.092	33.324
	c0ba8724-c0e0-4531-a909-a37789647a27	25-06-2022 2:57:01 pm	0.112	0.556	0.201	0.061	0.475	0.127	16.528

DESCARGAR REPORTE

CERRAR

Figura 13. Resultados de las pruebas

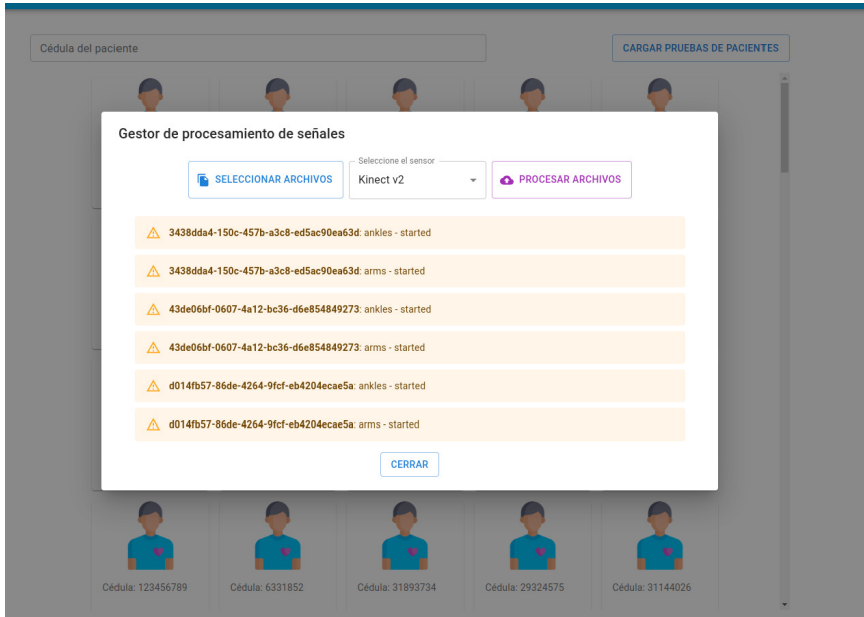


Figura 14. Estado de procesamiento de las pruebas

5.3. HALLAZGOS

Las señales generadas por la nueva herramienta contienen una mayor cantidad de decimales que las generadas por Matlab, por este motivo, sus valores se redondean al mayor número de decimales de las señales originales, esto es cuatro decimales. Los resultados obtenidos para la distancia de Hamming y el coeficiente de Spearman se presentan, respectivamente, en las TABLAS 6 y 7. Por otra parte, en la TABLA 8 se comparan los tiempos de ejecución de los dos procesadores de señal.

Tabla 6. Resultado de la métrica de evaluación de la distancia de Hamming

Paciente	Test id	Tobillo izquierdo	Tobillo derecho
1	3438dda4-150c-457b-a3c8-ed5ac90ea63d	0	0
	109b6da2-cb9f-4a5b-83fd-39208f465213	0	0
	43de06bf-0607-4a12-bc36-d6e854849273	0	0
	9c643233-4781-4589-893d-204945f84f12	0	0
	35855585-9e53-42cc-9d8e-b5b1cf27bceb	0	0

Software automatizado para análisis de marcha que usa Kinect v1 y wavelets como complemento a la evaluación clínica de la enfermedad de Parkinson

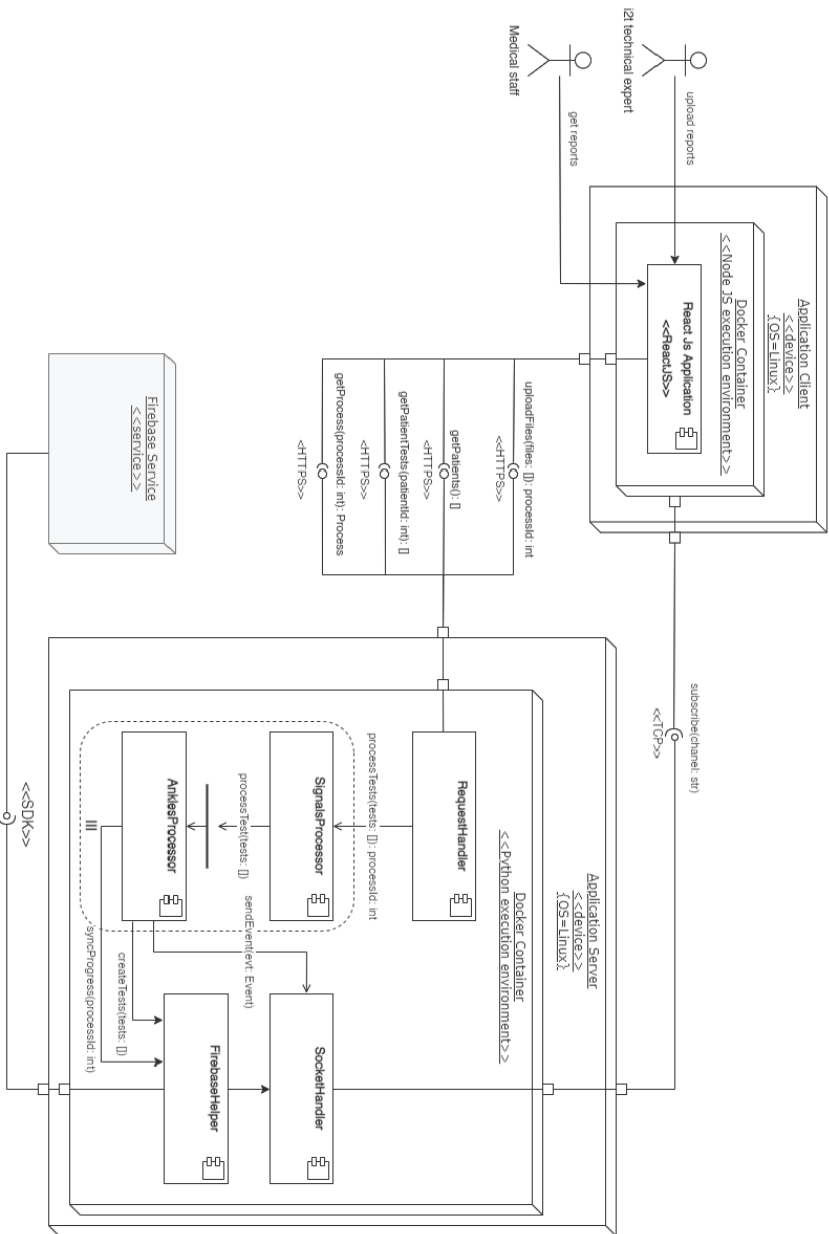


Figura 15. Diagrama de arquitectura del sistema

Tabla 6. Resultado de la métrica de evaluación de la distancia de Hamming (cont.)

Paciente	Test id	Tobillo izquierdo	Tobillo derecho
2	d8b46032-b226-49bb-bc16-99b10c4407cf	0	0
	e9a79b01-58ae-4909-925d-761fe6aaa195	0	0
	817d9690-b1d1-4229-a136-7cc720ed2248	0	0
	5dd15e01-1bb3-492b-8471-ec2083e48b7e	0	0
	d014fb57-86de-4264-9fcf-eb4204ecae5a	0	0

Tabla 7. Resultado de la métrica de evaluación de la correlación de Spearman

Paciente	Test id	Tobillo izquierdo	Tobillo derecho
1	3438dda4-150c-457b-a3c8-ed5ac90ea63d	1	1
	109b6da2-cb9f-4a5b-83fd-39208f465213	1	1
	43de06bf-0607-4a12-bc36-d6e854849273	1	1
	9c643233-4781-4589-893d-204945f84f12	1	1
	35855585-9e53-42cc-9d8e-b5b1cf27bceb	1	0.99
2	d8b46032-b226-49bb-bc16-99b10c4407cf	1	1
	e9a79b01-58ae-4909-925d-761fe6aaa195	1	1
	817d9690-b1d1-4229-a136-7cc720ed2248	0.99	0.99
	5dd15e01-1bb3-492b-8471-ec2083e48b7e	0.99	0.99
	d014fb57-86de-4264-9fcf-eb4204ecae5a	0.99	1

Tabla 8. Comparativo de los tiempos de ejecución de los procesadores de señal

Procesador	Tiempo de ejecución (segundos)
Matlab DSP	48.574
Python DSP	3.431

El primer paso de evaluación para evaluar la confiabilidad del nuevo procesador de señales fue la prueba de distancia de Hamming [34]. Para todos los casos de prueba, obtuvimos un valor de 0 para la distancia de Hamming. Esto implica que la nueva implementación realizada con tecnologías Python puede generar la misma señal que su antecesor implementado en Matlab.

El nuevo procesador de señales se implementó con librerías como Scipy y Numpy, que permiten ejecutar acciones como buscar picos en las señales.

Este tipo de tareas son esenciales cuando se generan nuevas señales para obtener variables de la marcha. Estas nuevas bibliotecas permiten soportar tareas de procesamiento de señales, operando en un entorno empaquetado o incluso en servicios como la nube, sin depender de un conjunto de aplicaciones de escritorio. Además, la nueva implementación se realizó a través de código abierto y no está restringida por el uso de licencias como su antecesor Matlab. El resultado es un servicio independiente que se puede consumir a través de una solicitud HTTP.

6. DISCUSIÓN

El primer paso para evaluar la confiabilidad del nuevo procesador de señales fue la prueba de distancia de Hamming [40]. En todos los casos el resultado obtenido fue 0, lo que implica que la nueva implementación realizada con tecnologías Python puede generar la misma señal que su antecesor implementado en Matlab.

El nuevo procesador de señales se implementó con las librerías Scipy y Numpy, las cuales permiten ejecutar acciones tales como la búsqueda de picos en las señales, tarea esencial cuando se generan nuevas señales para obtener variables de la marcha. Estas nuevas bibliotecas permiten soportar tareas de procesamiento de señales operando en un entorno empaquetado—incluso en la nube—, sin depender de un conjunto de aplicaciones de escritorio.

Dado que la nueva implementación se realizó usando código fuente abierto y no está restringida por el uso de licencias, como su antecesor Matlab, el resultado es un servicio independiente que se puede consumir a través de una solicitud HTTP.

En general, la distancia de Hamming evalúa la diferencia entre dos cadenas de valor, si bien esta métrica se adapta a la evaluación requerida para este proyecto, es posible que se requiera de otras métricas para realizar una evaluación clínica en profundidad.

Por otra parte, como se puede apreciar en la Tabla 7, la prueba de Spearman muestra una alta correlación entre los resultados obtenidos para ambos procesadores de señal, por lo tanto, el procesador de señales implementado en este proyecto proporciona señales iguales o similares a las generadas por Matlab.

Se realizó una tercera prueba para evaluar el tiempo de ejecución de ambos procesadores de señales, sus resultados muestran que el tiempo de procesamiento se puede reducir de aproximadamente 48 segundos a algo

más de 4 segundos, utilizando el nuevo procesador de señales en paralelo, lo que corresponde a una reducción del 92 % con respecto a la implementación de Matlab.

En promedio, procesar una prueba con el procesador implementado en Matlab se tarda 4,8 segundos, que es más tiempo que el que requiere el nuevo procesador en Python en procesar diez pruebas. Sin llegar a utilizar un entorno de procesamiento distribuido, esta solución de procesamiento paralelo es una compensación por los recursos y el tiempo de procesamiento, ya que la solución permite procesar una mayor cantidad de pruebas en menos tiempo y aprovechar los recursos como CPU y RAM de donde se despliega el servicio.

El tipo de arquitectura utilizada presenta desafíos para la orquestación de subprocesos, como la notificación del estado general del proceso. Para esta implementación, el estado de ejecución se conserva en una base de datos, característica que permite conocer el estado general del proceso incluso fuera del proceso. Las arquitecturas basadas en grupos de microservicios [44] utilizan la sincronización de estado a través de una base de datos que refleja el estado general del proceso macro que se está ejecutando. El uso de una base de datos para almacenar el estado del proceso, permite al cliente solicitar el estado de su proceso sin alterar el estado del proceso. El cliente también puede conocer el estado de su proceso mediante técnicas como Long Polling [45] o Websockets. De esta forma, el cliente conoce el estado de la tarea sin tener que depender del inicio o la finalización del proceso.

Otro output del procesador de señales anterior son los gráficos de marcha que surgen del procesamiento de las señales del tobillo. Con la implementación de la aplicación web, el médico tiene acceso a gráficos de marcha y variables espaciotemporales que se condensan en tablas para cada paciente. De esta manera, si tiene acceso a una computadora y conexión a Internet, el médico puede observar los resultados de sus pacientes desde esta plataforma. Con esta herramienta adicional, el médico ya no depende del técnico de i2t para tener acceso a estos informes, ya que puede acceder a la plataforma y conocer esta información de primera mano. Si bien la plataforma web solo admite archivos Kinect v1, se espera que estos resultados conduzcan a la integración de archivos Kinect v2 y Orbbec Astra. Eventualmente, todos los procesadores de señales se integrarán en una sola plataforma capaz de procesar pruebas de pacientes y brindar acceso a estos informes a los médicos.

Desde el punto de vista técnico de la plataforma, la aplicación web se implementa como un servicio separado del servidor, lo que significa que solo

consume los recursos necesarios para representar el contenido y ejecutar solicitudes al servidor. Los recursos de la aplicación también se cargan en el lado del cliente (computadora del usuario), lo que permite trasladar la carga de procesamiento del contenedor a la máquina del cliente [46].

7. CONCLUSIONES Y TRABAJO FUTURO

La prueba de distancia de Hamming y la correlación de Spearman permitieron comprobar que los datos generados por el nuevo procesador de señales son muy similares a los resultados generados por Matlab. A partir de estas nuevas señales procesadas, el sistema obtiene variables espaciotemporales similares a la versión anterior.

La prueba de tiempo de ejecución mostró una mejora del 92 % en el tiempo que le toma al procesador ejecutar un total de diez archivos de prueba y generar los informes. El tiempo que tarda la nueva herramienta en procesar diez pruebas es el mismo tiempo que tarda en procesar una el procesador Matlab. Por lo tanto, es posible desarrollar un procesador de señales personalizado utilizando bibliotecas de código abierto, capaz de generar los mismos resultados que las suites o herramientas de pago y obtener una mejora en el tiempo en el procesamiento de las pruebas de los pacientes, utilizando una arquitectura basada en el procesamiento paralelo.

El uso de Python como lenguaje de la nueva herramienta de procesamiento permite ampliar la capacidad del sistema para ofrecer esta funcionalidad en la nube, incluso convertir el sistema en un servicio independiente que no requiere de una computadora física para su ejecución.

Si bien uno de los desafíos en el desarrollo de la herramienta fue replicar el algoritmo de procesamiento, la ayuda de bibliotecas de Python (Numpy y Scipy) facilitó la manipulación de los datos. Las bibliotecas cuentan con funciones especializadas para trabajar cómodamente con los vectores que componen una señal. Una parte importante del procesamiento fue encontrar los valores faltantes o aquellos que no se capturaron de la articulación del tobillo. Aunque Scipy ofrece funciones para encontrar valores perdidos por métodos de interpolación o extrapolación, fue necesario crear una función personalizada que implementa un modelo de regresión lineal para encontrarlos.

La nueva herramienta de procesamiento de señales, al ser un servicio web basado en la nube, refleja un avance tecnológico en los servicios de procesamiento de señales. La herramienta permite migrar el servicio previamente ejecutado en una computadora Matlab a un entorno que no re-

quiere del factor humano. A diferencia de herramientas como GAITRite y Vicon Motion Capture System, el nuevo procesador se adapta a las necesidades requeridas por el médico para complementar el diagnóstico clínico, sin requerir de un espacio físico adicional en el que adaptar el sistema o de realizar pagos para su uso.

La aplicación web utilizada para cargar pruebas y reportar resultados se desarrolló utilizando una biblioteca javascript (React JS [47]), la cual facilita la creación de nuevos componentes reutilizables en poco tiempo. La interfaz puede aumentar fácilmente las funciones visuales gracias a las bibliotecas de secuencias de comandos que se desarrollan a diario. La aplicación web puede incluso adaptarse a las necesidades del médico por tratarse de un componente independiente del servicio de procesamiento.

Los objetivos propuestos, declarados en la introducción, se lograron a cabalidad: se creó un procesador de señales personalizado; se validó su correcto funcionamiento mediante la distancia de Hamming, la correlación de Spearman y las pruebas de tiempo de ejecución; y se desarrolló un sistema de visualización para informar los resultados al médico.

Gracias a la validación de los resultados con la nueva implementación del servicio de procesamiento en Python, es posible continuar extendiendo las funcionalidades de la herramienta al procesamiento de otro tipo de señales para complementar el diagnóstico médico. Por ejemplo, la versión del procesador para Kinect v2 permite obtener variables espacio-temporales y de marcha a partir del análisis de los brazos. Además, en esta nueva herramienta se podrá incluir una nueva versión del procesador desarrollado para el Orbbec Astra, que permitirá contrastar los resultados del Kinect vs este nuevo sensor.

La proyección de esta nueva herramienta es poder crear una suite de servicios para el procesamiento de señales utilizando los sensores disponibles en el grupo de investigación i2t de la Universidad Icesi; unificar estos servicios complementará el diagnóstico clínico del médico mediante el análisis de diferentes partes del cuerpo y el uso de diferentes sensores para captar la información del paciente.

8. REFERENCIAS

- [1] T. A. L. Wren, C. Lening, S. A. Rethlefsen, and R. M. Kay, “Impact of gait analysis on correction of excessive hip internal rotation in ambulatory children with cerebral palsy: A randomized controlled trial,” *Dev. Med. Child Neurol.*, vol. 55, no. 10, pp. 919–925, 2013, doi: 10.1111/dmcn.12184.
- [2] B. Munoz, Y. J. Castano-Pino, J. David Arango Paredes, and A. Navarro, “Automated gait analysis using a kinect camera and wavelets,” 2018 IEEE 20th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2018, pp. 1–5, 2018, doi: 10.1109/HealthCom.2018.8531161.
- [3] “GAITRite | World Leader in Temporospacial Gait Analysis.” <https://www.gaitrite.com/> (accessed Apr. 18, 2021).
- [4] “Vicon | Award Winning Motion Capture Systems.” <https://www.vicon.com/> (accessed Apr. 18, 2021).
- [5] J. D. Mejia-Trujillo et al., “Kinect™ and Intel RealSense™ D435 comparison: a preliminary study for motion analysis,” 2019 IEEE Int. Conf. E-Health Networking, Appl. Serv. Heal. 2019, no. October, pp. 3–7, 2019, doi: 10.1109/HealthCom46333.2019.9009433.
- [6] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, “Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study,” *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–22, 2020, doi: 10.3390/s20185104.
- [7] A. Muro-de-la-Herran, B. García-Zapirain, and A. Méndez-Zorrilla, “Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications,” *Sensors (Switzerland)*, vol. 14, no. 2, pp. 3362–3394, 2014, doi: 10.3390/s140203362.
- [8] R. A. Noble and R. White, “Visualisation of gait analysis data,” *Proc. Int. Conf. Inf. Vis.*, vol. 2005, pp. 247–252, 2005, doi: 10.1109/IV.2005.139.
- [9] “Kinect - Windows app development.” <https://developer.microsoft.com/en-us/windows/kinect/> (accessed Apr. 18, 2021).
- [10] A. P. Rocha, H. Choupina, J. M. Fernandes, M. J. Rosas, R. Vaz, and J. P. S. Cunha, “Parkinson’s disease assessment based on gait analysis using an innovative RGB-D camera system,” 2014 36th Annu.

- Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014, pp. 3126–3129, 2014, doi: 10.1109/EMBC.2014.6944285.
- [11] J. D. Arango Paredes, B. Muñoz, W. Agredo, Y. Ariza-Araújo, J. L. Orozco, and A. Navarro, “A reliability assessment software using Kinect to complement the clinical evaluation of Parkinson’s disease,” Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2015-Novem, pp. 6860–6863, 2015, doi: 10.1109/EMBC.2015.7319969.
- [12] M. Unser and A. Aldroubi, “A review of wavelets in biomedical applications,” Proc. IEEE, vol. 84, no. 4, pp. 626–638, 1996, doi: 10.1109/5.488704.
- [13] T. Chau, “A review of analytical techniques for gait data. Part 2: Neural network and wavelet methods,” Gait Posture, vol. 13, no. 2, pp. 102–120, 2001, doi: 10.1016/S0966-6362(00)00095-3.
- [14] K. Aminian, B. Najafi, C. Büla, P. F. Leyvraz, and P. Robert, “Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes,” J. Biomech., vol. 35, no. 5, pp. 689–699, 2002, doi: 10.1016/S0021-9290(02)00008-8.
- [15] “Parallel Computing Toolbox - MATLAB.” <https://www.mathworks.com/products/parallel-computing.html> (accessed May 17, 2022).
- [16] A. Marowka, “On parallel software engineering education using python,” Educ. Inf. Technol., vol. 23, no. 1, pp. 357–372, 2018, doi: 10.1007/s10639-017-9607-0.
- [17] B.-C. Wang, “DIGITAL SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS IN RADAR IMAGE PROCESSING.”
- [18] “Wavelet Transforms - an overview | ScienceDirect Topics.” <https://www.sciencedirect.com/topics/mathematics/wavelet-transforms> (accessed May 09, 2021).
- [19] “Algoritmo para detección de vibraciones anormales en maquinarias utilizando la transformada wavelet.” https://www.researchgate.net/publication/37613214_Algoritmo_para_deteccion_de_vibraciones_anormales_en_maquinarias_utilizando_la_transformada_wavelet (accessed May 09, 2021).
- [20] J. O. Smith III, Continuous wavelet transform, in Spectral audio signal processing, W3K Publishing, 2011. https://ccrma.stanford.edu/~jos/sasp/Continuous_Wavelet_Transform.html

- [21] M. J. Lai, "Popular wavelet families and filters and their use," in *Computational Complexity: Theory, Techniques, and Applications*, vol. 9781461418009, Springer New York, 2012, pp. 2168–2225.
- [22] S. Zope-Chaudhari, P. Venkatachalam, and K. M. Buddhiraju, "Assessment of distortion in watermarked geospatial vector data using different wavelets," *Geo-Spatial Inf. Sci.*, vol. 18, no. 2–3, pp. 124–133, Jul. 2015, doi: 10.1080/10095020.2015.1071064.
- [23] Y. J. Castaño, A. Navarro, J. D. Arango, B. E. Muñoz, J. L. Orozco, and J. Valderrama, "Gait and arm swing analysis measurements for patients diagnosed with Parkinson's disease, using digital signal processing and Kinect," *CEUR Workshop Proc.*, vol. 2178, pp. 71–74, 2018.
- [24] S. H. Roosta, "Parallel Processing and Parallel Algorithms," *Parallel Process. Parallel Algorithms*, 2000, doi: 10.1007/978-1-4612-1220-1.
- [25] M. I. Capel, A. J. Tomeu, and A. G. Salguero, "Teaching concurrent and parallel programming by patterns: An interactive ICT approach," *J. Parallel Distrib. Comput.*, vol. 105, pp. 42–52, Jul. 2017, doi: 10.1016/J.JPDC.2017.01.010.
- [26] A. Igor Anatolyevich Kalmykov, Konstantin Aleksandrovich Katkov, Naumenko Daniil Olegovich and A. V. M. Bronislavovich Sarkisov, "Parallel modular technologies in digital signal processing," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.
- [27] C. J. Hughes, "Single-instruction multiple-data execution," *Synth. Lect. Comput. Archit.*, vol. 32, pp. 1–121, 2015, doi: 10.2200/S00647ED1V01Y201505CAC032.
- [28] C. Shih-Lin Chien, L. Shinn-Zong, L. Chung-Chao, H. Yu-Loong, L. Chi-Wei, Ch. Shin-Yuan, "The efficacy of quantitative gait analysis by the GAITRite system in evaluation of parkinsonian bradykinesia," *Parkinson and related Disorders*, Vol. 12, No. 7, pp. 438-442; 2006. <https://doi.org/10.1016/j.parkreldis.2006.04.004>
- [29] A. Pfister, A. M. West, S. Bronner, J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *J Med Eng Technol*, Vol. 38, No. 5, pp. 274, 280; 2014. doi: 10.3109/03091902.2014.909540.
- [30] J. Camilo and V. Correa, "Design and Implementation of a Portable Device for Real Time Analysis of Speech of People with Parkinson's Disease," 2013.

- [31] A. Procházka, O. Vyšata, M. Vališ, O. Tupa, M. Schätz, and V. Marík, “Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect,” *Digit. Signal Process. A Rev. J.*, vol. 47, pp. 169–177, 2015, doi: 10.1016/j.dsp.2015.05.011.
- [32] O. Tupa et al., “Motion tracking and gait feature estimation for recognising Parkinson’s disease using MS Kinect,” *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–20, 2015, doi: 10.1186/s12938-015-0092-7.
- [33] S. van der Walt, J. I. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Goullart, T. Yu, & The Scikit-Image Contributors, “Scikit-image: image processing in Python, *PeerJ* 2:e453; 2014. <https://doi.org/10.7717/peerj.453>
- [34] N. B. Ruparelia, “Software development lifecycle models,” *ACM SIGSOFT Softw. Eng. Notes*, vol. 35, no. 3, pp. 8–13, May 2010, doi: 10.1145/1764810.1764814.
- [35] M. Naveenkumar and S. Domnic, “Vector Quantization based Pairwise Joint Distance Maps (VQ-PJDM) for 3D Action Recognition,” *Procedia Comput. Sci.*, vol. 133, pp. 27–36, 2018, doi: 10.1016/J.PROCS.2018.07.005.
- [36] T. W. Ng, T. L. Tay, and S. W. Khor, “Iris recognition using rapid Haar wavelet decomposition,” *ICSPS 2010 - Proc. 2010 2nd Int. Conf. Signal Process. Syst.*, vol. 1, 2010, doi: 10.1109/ICSPS.2010.5555246.
- [37] N. Ahmadi, N. Ahmadi, and M. Nilashi, “Iris Texture Recognition based on Multilevel 2-D Haar Wavelet Decomposition and Hamming Distance Approach,” *J. Soft Comput. Decis. Support Syst.*, vol. 5, no. 3, pp. 16–20, Jun. 2018, Accessed: May 16, 2022. [Online]. Available: <http://jscdss.com/index.php/files/article/view/168>.
- [38] J. M. Redondo and F. Ortin, “A Comprehensive Evaluation of Common Python Implementations,” *IEEE Softw.*, vol. 32, no. 4, pp. 76–84, Jul. 2015, doi: 10.1109/MS.2014.104.
- [39] R. Md Saad, M. Z. Ahmad, M. S. Abu, and M. S. Jusoh, “Hamming distance method with subjective and objective weights for personnel selection,” *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/865495.
- [40] M. Norouzi, D. J. Fleet, and R. Salakhutdinov, “Hamming distance metric learning,” *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 1061–1069, 2012.

- [41] P. Schober and L. A. Schwarte, “Correlation coefficients: Appropriate use and interpretation,” *Anesth. Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018, doi: 10.1213/ANE.0000000000002864.
- [42] M. Ousmer, J. Vanderdonckt, and S. Buraga, “An ontology for reasoning on body-based gestures,” *Proc. ACM SIGCHI Symp. Eng. Interact. Comput. Syst. EICS 2019*, Jun. 2019, doi: 10.1145/3319499.3328238.
- [43] V. F. Pacheco, “Microservice Patterns and Best Practices : Explore patterns like CQRS and event sourcing to create scalable, maintainable, and testable microservices.,” p. 357, 2018.
- [44] A. J. Fernández-García, L. Iribarne, A. Corral, J. Criado, and J. Z. Wang, “A microservice-based architecture for enhancing the user experience in cross-device distributed mashup UIs with multiple forms of interaction,” *Univers. Access Inf. Soc.*, vol. 18, no. 4, pp. 747–770, Nov. 2019, doi: 10.1007/S10209-017-0606-0.
- [45] E. Stratmann, J. Ousterhout, and S. Madan, “Integrating Long Polling with an MVC Web Framework.”
- [46] T. Fadhilah Iskandar, M. Lubis, T. Fabrianti Kusumasari, and A. Ridho Lubis, “Comparison between client-side and server-side rendering in the web development,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 801, no. 1, 2020, doi: 10.1088/1757-899X/801/1/012136.
- [47] “React – A JavaScript library for building user interfaces.” <https://en.reactjs.org/> (accessed Nov. 25, 2021).

Índice de tablas

Tabla 1. Caracterización y trabajos relacionados (soluciones completas)	85
Tabla 2. Caracterización y trabajos relacionados (soluciones personalizadas)	86
Tabla 3. Caracterización y trabajos relacionados (herramientas)	87
Tabla 4. Fases de la metodología SDLC	89
Tabla 5. Lectura de la correlación de Spearman	92
Tabla 6. Resultado de la métrica de evaluación de la distancia de Hamming	97
Tabla 7. Resultado de la métrica de evaluación de la correlación de Spearman	98
Tabla 8. Comparativo de los tiempos de ejecución de los procesadores de señal	99

Índice de figuras

Figura 1. Transformada wavelet	80
Figura 2. Ejemplos de familias wavelet	81
Figura 3. Señal de movimiento del tobillo izquierdo / descomposición de wavelet resultante de la señal	81
Figura 4. Procesamiento concurrente vs. procesamiento paralelo	82
Figura 5. Arquitectura SIMD	83
Figura 6. Fases de la metodología SDLC	89
Figura 7. Fragmento del archivo JSON generado por e-Motion	91
Figura 8. Señal de profundidad del brazo derecho y señal ajustada después del procesamiento	91
Figura 9. Mapeo de la captura de articulaciones realizada por el sensor Kinect v1	93
Figura 10. Captura de la señal, procesamiento de wavelet y entrega de resultados	94
Figura 11. Arquitectura del sistema	94
Figura 12. Lista de pacientes	96
Figura 13. Resultados de las pruebas	96
Figura 14. Estado de procesamiento de las pruebas	97
Figura 15. Diagrama de arquitectura del sistema	98

Este libro se terminó de imprimir y encuadernar en enero de 2023 en los talleres de Carvajal Soluciones de Comunicación, en la ciudad de Bogotá D.C. En su preparación, realizada desde la Editorial Universidad Icesi, se emplearon los tipos Baskerville MT Std de 8, 9, 10.5, 11 y 16 puntos, Times New Roman de 10 puntos y Cambria Math de 12 puntos. La edición estuvo al cuidado de José Ignacio Claros Vaca.

Los tres volúmenes previos de Ingeniería y Salud han reportado resultados en la investigación de las posibilidades de uso de dispositivos no especializados, más propios de los juegos electrónicos, como herramientas de captura de movimientos capaces de producir datos cuantitativos que faciliten la evaluación clínica que realizan los profesionales de la salud para el diagnóstico y monitoreo de la evolución de la enfermedad de Parkinson. Este cuarto volumen sigue esa línea, en él se reportan dos temas: el primero, la constatación cuantitativa de la relación que existe entre las extremidades —inferiores y superiores— de las personas que padecen Parkinson, con lo que se abre una ruta para la exploración de esta relación que sirve de base para el concepto de inferencia causal; el segundo, el diseño de una herramienta que le permite al personal médico manejar la información técnica que arrojan las pruebas de marcha realizadas con e-motion, el sistema desarrollado por los proyectos previamente reportados, sin necesidad del apoyo de ingenieros, con lo que se reducen tanto los tiempos de entrega de datos útiles como los costos asociados a esta tarea.

