Predicción de riesgo de caída en adultos mayores a través de sensores: un enfoque de Machine Learning

Informe de Memoria de Título

Autor:	**Emilio Ramos Montesino**
Profesor Patrocinante:	**Ricardo Flores Huenchullanca**
Profesor Co-Patrocinante:	**Pedro Pablo Pinacho Davidson**

Chile, Región del Biobío, Concepción

22 de agosto de 2025

Agradecimientos

Deseo expresar mis profundos agradecimientos a todas las personas que me acompañaron en este camino que concluye con la presente memoria de título. A mi familia y a mi pareja por apoyarme siempre y estar presente en todo momento. A las amistades que forjé durante la carrera, que me acompañaron día tras día e hicieron más ameno este viaje. A mis profesores patrocinantes, Ricardo Flores y Pedro Pinacho, por darme la oportunidad de trabajar con ellos y tomarse el tiempo de escucharme y brindarme soluciones en los momentos que más perdido me sentía. Al equipo del proyecto FONDEF por integrarme al grupo y permitirme aprender de otras áreas diferentes a la informática. A todas esas personas que aportaron su granito de arena en mi formación como profesional.

A todos ustedes, muchas gracias.

Resumen

Las caídas son actualmente uno de los principales peligros de salud en los adultos mayores, teniendo una alta prevalencia en Chile. Para medir el riesgo de caída existen diversos tests como el Timed Up and Go (TUG), el cual es ampliamente utilizado debido a su sencillez y fácil implementación. A pesar de esto, el TUG solo evalúa la duración total de la prueba, limitando la detección de patrones de marcha y estrategias compensatorias durante el test. Este trabajo propone un enfoque de aprendizaje automático, para predecir el riesgo de caída, utilizando datos de un sensor inercial colocado en la zona lumbar en adultos mayores chilenos al momento de realizar el test TUG. Se entrenaron varios modelos de Machine Learning Clásico, tal que el mejor resultado lo alcanzó el XGBoost con un F1-Score de 91.17 % y un AUC de 0.96. Además, el análisis de interpretabilidad con valores SHAP reveló que las fases de Caminata de Ida y Caminata de Vuelta fueron las más determinantes en la predicción. Finalmente, se desarrolló un dashboard interactivo en Streamlit para visualizar individualmente las predicciones, constituyendo una herramienta exploratoria de apoyo al análisis de la predicción.

Índice general

1. Introducción **5**

1.1. Contexto General 5

1.2. Presentación del problema 6

1.3. Objetivos 7

1.3.1. Objetivo General 7

1.3.2. Objetivos Específicos 7

1.3.3. Restricciones 8

1.3.4. Análisis de Riesgo 8

1.3.5. Operación 9

2. Marco teórico **10**

2.1. Estado del arte 10

2.2. Evaluación de alternativas 13

3. Metodología **15**

3.1. Resumen de metodología 15

3.2. Materiales 16

3.3. Métodos 17

3.3.1. Segmentación del Timed Up and Go 17

3.3.2. Etiquetado de pruebas 18

3.3.3. Extracción de características 21

3.3.4. Modelamiento 22

3.3.5. Características más importantes 23

4. Resultados **25**

4.1. Configuración de los modelos 25

4.2. Resultados generales 27

4.3. Análisis del mejor modelo 28
4.4. Importancia de fases y características 29
4.5. Dashboard 33

5. Conclusiones **38**

Capítulo 1

Introducción

1.1. Contexto General

El envejecimiento de la población es un fenómeno que ha ido incrementando año tras año [1]. Las principales razones de este cambio demográfico son debido al aumento de la expectativa de vida y la baja en la natalidad. Lo anterior conlleva a desafíos en el área de la salud debido al deterioro de las capacidades físicas y mentales de los adultos mayores, una mayor dependencia y riesgos de discapacidad. Es por ello, que una detección temprana junto con tratamientos de rehabilitación y medidas preventivas pueden evitar la disminución de la calidad de vida de este grupo de personas.

Una de las consecuencias asociadas al envejecimiento es el riesgo de caídas. Este evento ha sido definido como una circunstancia accidental donde la persona cae luego de perder el control de su centro de gravedad, sin hacer ningún esfuerzo físico por restaurarlo o siendo este insuficiente. Las caídas son actualmente uno de los principales peligros de salud en los adultos mayores con una prevalencia reportada en Chile de un 35 %. En este país, los adultos mayores superan los 2.8 millones de personas y se espera un aumento del 20 % de esta población, llegando a tener una expectativa de vida de 89 años en promedio [2].

En el área de la salud, los profesionales de la rehabilitación analizan al paciente por medio de la observación y la aplicación de escalas e instrumentos de evaluación, los cuales entregan cierto nivel de objetividad. En la actualidad existen diversos test clínicos para evaluar y medir el balance estático, dinámico y la marcha en personas sanas o con algún problema motor. El Timed Up and Go (TUG) [3] es uno de los métodos más utilizados a nivel mundial por su fácil implementación, bajo costo y que además permite al experto evaluar la movilidad funcional en adultos mayores y población neurológica.

El TUG es un test simple, el cual consiste en que el paciente, partiendo desde una posición

sentada, se levanta de una silla, camina tres metros, gira alrededor de un obstáculo, regresa al punto de partida, gira nuevamente y se sienta. De esta manera, la variedad de cada una de las fases permite analizar distintas capacidades motoras: en la fase de pararse, se evalúa la fuerza y control del tronco inferior para la transición desde la sedestación; en la caminata de ida, se analiza la estabilidad dinámica, la velocidad y la coordinación al caminar; durante el giro en el obstáculo, se observa la capacidad de cambiar de dirección y mantener el equilibrio; en la caminata de vuelta, se vuelve a evaluar la marcha pero considerando la fatiga o adaptación; en el giro para sentarse, se mide el control postural y planificación motora para realizar un giro preciso y en la fase de sentarse se valora el control del descenso y la estabilidad al retomar la posición inicial.

1.2. Presentación del problema

A pesar de su sencillez y su fácil implementación, el TUG mide principalmente el tiempo total como único indicador de riesgo de caída y no evalúa patrones de marcha y posibles estrategias compensatorias que pueden no ser evidentes para el evaluador. Por lo que este test podría verse mejorado al ser combinado con herramientas tecnológicas de análisis de movimiento, como el uso de sensores inerciales IMU (Inertial Measurement Unit) [4].

En Chile, el uso de tecnologías para el análisis de movimiento humano en la práctica clínica no ha tenido un desarrollo acorde a la tendencia mundial. Debido a que aún se opta por utilizar procedimientos clínicos asociados al uso de herramientas con alto costo y que necesitan profesionales capacitados para su operación. Además existen zonas rurales donde el acceso a la salud y a un control continuo de padecimientos de índole motor es complejo, evitando la masificación de sistemas de análisis de movimiento humano.

En este contexto el Proyecto FONDEF ID24I10422 QMUV plantea entre sus objetivos el desarrollo de algoritmos de Inteligencia Artificial (IA) que permitan evaluar el riesgo de caída y la capacidad de marcha en la población chilena utilizando el sensor inercial QMUV propuesto por Ortegas-Bastidas y colaboradores [5].

La presente memoria de título busca contribuir al proyecto analizando las señales de los sensores durante las distintas etapas de test Timed Up and Go instrumentalizado (iTUG) [6]. Con el propósito de obtener las fases del TUG, las señales y las características que más aportan a la detección del riesgo de caída se pretende aplicar métodos de Machine Learning que puedan diferenciar la marcha de personas con y sin riesgo de caída. Brindando de esta manera, un modelo que pueda predecir el riesgo de caída y un análisis exploratorio de los determinantes en el riesgo de caída con un enfoque desde el Machine Learning.

Los modelos serán entrenados y validados a partir de pruebas realizadas del iTUG, las cuales se dividen en dos conjuntos de datos, uno destinado para el entrenamiento de los modelos y el otro para comprobar la capacidad de generalización. La efectividad en la predicción de los métodos será evaluada según métricas de clasificación como F1-Score y Accuracy.

1.3. Objetivos

1.3.1. Objetivo General

Entrenar un modelo de Machine Learning que permita predecir y explicar la presencia de riesgo de caída en las distintas subetapas del test Timed Up and Go (TUG) en adultos mayores usando señales de un sensor inercial.

1.3.2. Objetivos Específicos

1. Aplicar métodos de extracción de características a las señales en cada subetapa del TUG.
2. Implementar y evaluar modelos de Machine Learning que predigan la presencia de riesgo de caída considerando cada subetapa del TUG.
3. Detectar características y señales más importantes al momento de predecir el riesgo de caída.
4. Crear una herramienta de apoyo visual con los resultados obtenidos que evidencie cuáles son las características y señales más relevantes al momento de identificar el riesgo de caída.

La evaluación de los objetivos específicos se realiza mediante un enfoque cuantitativo basado en métricas de desempeño de los modelos y análisis estadísticos. Para la extracción de características, se buscan las más relevantes mediante pruebas estadísticas como el *d-value* de Cohen. Por otro lado, los modelos de predicción son validados según métricas como el F1-Score para asegurar robustez si el dataset está desbalanceado. Además se aplican técnicas de Feature Importance, como el cálculo de los valores SHAP, en el mejor modelo de Machine Learning para identificar cuáles características y señales permite identificar patrones relevantes en cada etapa del TUG. Finalmente, la efectividad del dashboard será medida a través de su capacidad para representar de manera clara y comprensible los resultados obtenidos, asegurando que la información sea interpretable.

1.3.3. Restricciones

La presente propuesta forma parte de un proyecto FONDEF, por lo que se espera presentar resultados dentro de un plazo limitado. El uso de modelos de Machine Learning, por otra parte, puede requerir hardware con altos recursos computacionales, lo que debe ser considerado en la implementación del proyecto. Además, la información de cada paciente debe ser anonimizada para garantizar la privacidad de los datos, los cuales tienen un propósito exclusivamente investigativo. Por otro lado, los datos disponibles son limitados en comparación con los grandes volúmenes de información que suelen manejarse en la Ciencia de Datos, lo que implica un desafío adicional para el desarrollo de modelos precisos y efectivos.

1.3.4. Análisis de Riesgo

Los riesgos técnicos del proyecto están relacionados con la calidad y disponibilidad de los datos obtenidos a partir de los dispositivos IMU. La precisión y calibración de estos sensores pueden afectar la fiabilidad de los parámetros extraídos, lo que influye directamente en la detección de anomalías y el análisis de riesgo de caída. Además, la posibilidad de pérdida de paquetes durante la conexión al grabar las pruebas del IMU puede comprometer la integridad de los datos. Otro desafío importante es la capacidad de los algoritmos para generalizar correctamente, ya que el tamaño reducido del dataset puede limitar su desempeño y afectar la robustez de los modelos de Machine Learning.

Por otro lado, en términos de riesgos de gestión, existen riesgos asociados a la recolección, almacenamiento y segmentación de los datos. Uno de los principales inconvenientes es el posible retraso en la entrega de las pruebas del TUG, lo que podría impactar en el cronograma del proyecto. Además, la existencia de múltiples fuentes de información manejadas por distintas personas incrementa el riesgo de pérdida de datos o inconsistencias en el procesamiento. Finalmente, la segmentación manual de los test TUG basada en el criterio del ojo clínico podría introducir errores, afectando la correcta identificación de las etapas y en consecuencia, la precisión de los análisis posteriores.

Con el objetivo de mitigar los riesgos se pueden tomar medidas como establecer un cronograma realista y flexible que contemple posibles retrasos para tener tiempo de reacción ante posibles inconvenientes. Además se puede considerar la segmentación con algoritmos como una alternativa para eliminar el sesgo humano en la identificación de las etapas del test o en su defecto, utilizar una segmentación sea más flexible. Por otro lado para mejorar la capacidad de generalización de los modelos de Machine Learning, se pueden utilizar estrategias de aumento de datos, así como

explorar la posibilidad de integrar datos de otras fuentes compatibles u otros estudios similares.

1.3.5. Operación

La solución propuesta en este informe de memoria de título tiene un alcance exploratorio, aunque es parte de un proyecto mayor el cual pretende a futuro implementar en los hospitales un dispositivo que permita evaluar el riesgo de caída. Es por esto que el dispositivo debe cumplir ciertas normativas legales para poder ser utilizado en centros médicos.

La norma NCh2893/11, por ejemplo, indica los requisitos de seguridad necesarios para proporcionar protección al paciente (usuario) y al operador del equipo electromédico. Detallando algunos requerimientos como cubiertas de protección, aislación eléctrica, corrientes de fuga, etc. Por otra parte, en materia de datos y privacidad de la información, el ISP ha establecido las 'Políticas de Privacidad' y que se basan en los requerimientos de confidencialidad de los datos de los usuarios establecidos en la Ley N°19.628 en su artículo 19 N°4. Para ello, el equipo de desarrollo de este proyecto considera altos niveles de seguridad y confidencialidad en el manejo de la información confidencial de los usuarios.

Con respecto a la implantación de la solución propuesta en centros médicos se debe tener al menos un computador o celular para conectarse vía Bluetooth al sensor inercial. Además, los profesionales de la rehabilitación serán capacitados para uso del software que contendrá el algoritmo de Machine Learning entrenado. Por último, periódicamente se le realizará mantenimiento a la plataforma software y al modelo entrenado para ir actualizándolos a versiones más avanzadas.

Finalmente, la posible implementación de la solución planteada puede medirse a partir del impacto en la salud pública a través de una serie de criterios. Estos principios pueden verse reflejados en la potencial detección de riesgo de caídas, en una mayor cobertura de zonas rurales del test, en la reducción de barreras económicas para aplicar tratamientos de diagnóstico de riesgo de caída e incluso en la generación de nuevos datos para futuras investigaciones. De esta manera, los criterios anteriores se ven directamente relacionados con el cambio de la calidad de vida de los pacientes.

Capítulo 2

Marco teórico

2.1. Estado del arte

En la actualidad existen diversos tests y escalas para evaluar y medir el balance estático, dinámico y la marcha en personas sanas o con algún problema motor. Uno de los más utilizados a nivel mundial es el Timed Up and Go (TUG) [3], el cual mide el balance dinámico y la movilidad funcional en adultos mayores. El TUG es un test simple, el cual consiste en que el sujeto debe ponerse de pie desde una silla, caminar 3 metros, girar sobre un obstáculo, regresar a la silla caminando y sentarse nuevamente. La variable a controlar en la prueba es el tiempo total en segundos que el sujeto se toma al realizar el test, el cual luego se relaciona con el riesgo de caída [7].

Algunas ventajas del TUG es que puede ser aplicado en cualquier lugar y requiere poco equipamiento, sumando además la corta duración del test. Sin embargo, a pesar de que el test entrega información sobre el riesgo de caída, solo evalúa el equilibrio y balance de forma general según Barry et. al. [8]. Por lo que podría mejorarse al ser combinado con herramientas tecnológicas como los sensores inerciales IMU [4].

Hoy en día se han desarrollado métodos basados en sensores inerciales, debido a la reducción en tamaño y precio de estos dispositivos, que han logrado una medición más confiable y precisa en el análisis del movimiento humano. De esta manera, múltiples propuestas han sido publicadas enriqueciendo el TUG con el uso principalmente de sensores inerciales, a esta variante se le conoce como iTUG [6].

En los últimos años se han ido implementado de forma progresiva las tecnologías IoT (*Internet of Things*) en los contextos de salud. Concretamente en el área de la salud, la conexión remota de dispositivos médicos para el monitoreo del paciente se le denomina *Internet of Medical Things* (IoMT). El-Saleh et al. [9] asegura que IoMT ofrece una gran oportunidad para revolucionar la

industria de la salud a través del monitoreo remoto, mejora en los diagnósticos y procesos en línea. Es por ello que la implementación de variantes como iTUG puede tener un gran impacto en el área de la rehabilitación.

Con respecto al uso de algoritmos de Machine Learning para el análisis de información captada en el test iTUG diversos estudios han sido publicados a lo largo de los años. Friedrich y colaboradores en 2021 [10] realizaron un estudio con 20 adultos mayores, a los cuales se les solicitó llevar un sensor inercial IMU en la cadera 12 horas al día durante 10 meses. Transcurrido el tiempo solicitado, se entrenó una red neuronal (LSTM + Conv1D) con los datos recopilados. Luego los participantes realizaron el test Timed Up and Go usando el sensor IMU. El modelo entrenado predijo con un 95 % de accuracy el riesgo de caída en el test TUG. Sin embargo, al intentar predecir participantes nunca vistos bajó a un 26 % el accuracy.

Por otro lado, Roshdibenam y su equipo en 2021 [11] aplicaron el TUG a adultos mayores de 65 años con 3 sensores IMU ubicados en el pie derecho, pie izquierdo y cuello. Cabe destacar que los participantes fueron diagnosticados como *faller* o *non-faller* por un geriatra. Finalmente, el modelo entrenado obtiene un AUC = 75 % con una sensibilidad de 86 % utilizando una Conv1D.

En 2022, Ko y colaboradores [12] propusieron un modelo de regresión con aprendizaje automático, que predecía la edad de los adultos mayores. En el experimento participaron adultos mayores sanos, los cuales realizaron 3 veces el TUG y una vez el '6 Minutes Walk Test' (6MWT) [13] con un sensor inercial IMU en la espalda baja. Además, las fases del TUG fueron definidas a partir de umbrales en la señal yaw y pitch. El mejor resultado lo obtuvo el modelo XGBoost con un MAE de 3.5 años, utilizando los datos de ambas pruebas.

Años después en 2024, el grupo de investigación de Maiora y colaboradores [14] trabajaron con un dispositivo IMU en el análisis del test TUG segmentado por fases. Los autores extrajeron las características más importantes de cada fase calculando el Test Univariante Chi-Square [15]. Además, los pacientes fueron etiquetados como *fallers* o *non-fallers* según la presencia de caídas en los últimos 6 meses. De esta manera, se entrenó una LSTM Bidireccional (BLSTM), la cual obtuvo un 87 % de Accuracy y un 78 % de AUC.

En el mismo año, Wang y colaboradores [16] proponen un análisis del iTUG usando técnicas de Machine Learning a partir de un sensor IMU ubicado en la zona lumbar y de una cámara de profundidad. Los participantes fueron adultos mayores, a los cuales se les aplicó la Escala de Balance de Berg [17] para determinar si tienen riesgo de caída o no. De esta forma, una vez que segmentan en fases el TUG extraen las características cinemáticas y temporales de cada etapa. El modelo usado para la predicción fue una Regresión Logística, la cual alcanzó un AUC = 88.2 % utilizando la información tanto del sensor como de la cámara. Además, se calcularon cuáles fueron las características que más influyeron en la predicción para identificar parámetros claves asociados

al riesgo de caída.

En Chile son escasas las investigaciones encontradas que analizan la marcha con iTUG. Ortegas-Bastidas y colaboradores en el 2019 [5] propusieron un dispositivo IMU, llamado QMUV, el cual contiene acelerómetro, giroscopio y magnetómetro. Además, desarrollaron algoritmos para segmentar las etapas del TUG y extraer características de la marcha como la cantidad de pasos por etapa, a partir de reglas matemáticas aplicadas sobre la señal. Sin embargo, estas reglas funcionan para personas sanas pero en pacientes con defectos en la marcha tienden a perder eficacia.

Por último, la Tabla 2.1 contiene un resumen de las investigaciones presentadas.

Tabla 2.1: Resumen de investigaciones en iTUG utilizando Machine Learning.

Referencia del paper	Herramientas en iTUG	Criterio de etiquetado de riesgo de caída	Resultados principales
Friedrich et al., Healthcare, 2021 [10]	Sensor inercial en cadera	Evaluación clínica de profesionales	Logra clasificar con un accuracy de un 95 % sujetos que estuvieron 10 meses con el sensor, utilizando una LSTM+Conv1D.
Roshdibenam V., et al., Sensors, 2021 [11]	3 sensores inerciales: pie derecho, pie izquierdo, cuello	Evaluación clínica de profesionales	Usando un modelo Conv1D logra un AUC de 75 % a partir de la información de varios sensores.
Ko et al., Sensors, 2022 [12]	Sensor inercial en lumbar	-	Estima la edad de adultos mayores con un MAE de 3.5 años combinando datos de TUG y 6MWT.
Maiora J., et al., Bioengineering, 2024 [14]	Sensor inercial en lumbar	Presencia de caída en los últimos 6 meses	Clasifica con un AUC de 78 % pacientes con riesgo de caída usando una BLSTM.
Wang X., et al., BMC Geriatrics, 2024 [16]	Sensor inercial en lumbar y cámara de profundidad	Escala de Balance de Berg	Combina cámaras y sensores inerciales para obtener un AUC de 88.2 % con una Regresión Logística, además calcula caraterísticas más relevantes en predicción.

2.2. Evaluación de alternativas

Las investigaciones presentadas en la sección anterior dan cuenta de distintos enfoques relacionados con la identificación del riesgo de caídas en adultos mayores en el Timed Up and Go. Algunos de estos métodos incluyen la recopilación de datos en más de un sensor con el objetivo de obtener información del movimiento humano en distintas partes del cuerpo. Aunque por lo general el uso más común es el de un solo sensor inercial en la espalda baja. Cabe destacar, la combinación de otros dispositivos más avanzados como cámaras de profundidad para tener una fuente de características alternativa para describir mejor la prueba.

Por otro lado, se han aplicado otros tests además del TUG para un análisis más completo de la caminata del paciente, tal es el caso del 6MWT. Sin embargo, si bien esta prueba complementaria es útil para caracterizar de mejor forma la marcha del sujeto, el TUG ya evalúa el balanceo dinámico. Una solución alternativa podría verse en la versión del Timed Up and Go de 6 metros [18], en la cual la caminata de ida y vuelta es mayor, por lo que el esfuerzo físico requerido también lo es. Otros métodos de captación de datos también fueron implementados como el uso de un sensor inercial en la vida cotidiana, pero esto conllevó a un sobreajuste del modelo y por lo tanto una poca capacidad de generalización sobre nuevos pacientes.

Con respecto al modelamiento la mayoría de las investigaciones trataron el problema como una clasificación donde se buscaba predecir si la persona tiene o no riesgo de caída. De forma agregada se etiquetó bajo distintos criterios a los participantes como la Escala de Balance de Berg, el historial de caídas o por diagnóstico de un profesional. Sin embargo, no se etiquetan a las personas por su puntaje en el propio test del TUG, puntuación la cual es el *gold standard* para los profesionales al momento de diagnosticar un paciente con riesgo de caída.

Al momento de elegir un modelo para realizar la predicción algunos autores utilizan algoritmos de Deep Learning, los cuales se adaptan bien a patrones complejos, aunque debido a la poca cantidad de datos existentes en estos estudios se tiende a sobreajustar fácilmente. Por otro lado, otras investigaciones han optado por métodos de Machine Learning Clásico, los cuales se adaptan mucho mejor a datasets más pequeños. Además estos métodos requieren un mayor procesado de los datos por lo que se aplican metodologías como la extracción de características por fases del TUG, el preprocesado de las señales y la selección de las características más representativas.

Un punto clave a destacar es el poco análisis realizado por los estudios sobre la importancia de las características al momento de predecir. Información la cual puede ser relevante para identificar los puntos críticos que el modelo encuentra para diferenciar entre personas con o sin riesgo de caída. Más aún se puede analizar cuáles fueron las señales o etapas del TUG que mayor contribución tuvieron a la predicción. Dando de esta manera, una explicación al riesgo de caída a través de un

enfoque de Machine Learning.

Para finalizar, son pocas las investigaciones encontradas del iTUG evaluando el riesgo de caída en población chilena. Es por ello que cobra gran relevancia el estudio del riesgo de caída en adultos mayores en este país. Además, la aplicación de métodos de Machine Learning pueden llevar a una mejor identificación de indicios de riesgo de caída, dado su gran adaptación a patrones complejos como lo es la marcha.

Capítulo 3

Metodología

3.1. Resumen de metodología

En vistas de cumplir con los objetivos planteados se consideró una etapa de reconocimiento del problema donde a través de reuniones con integrantes del Laboratorio de Biomédica de la Facultad de Ingeniería de la Universidad de Concepción se acordó la problemática sobre la cual trabajar y una propuesta de solución a abordar. Además se crearon varios modelos prototipos utilizando Machine Learning con datos ya existentes para evaluar la viabilidad de la solución planteada. En consecuencia, se realizaron varias recomendaciones para el refinamiento del proceso de toma de datos en el test Timed Up and Go. Este trabajo previo sirvió de ayuda para converger al desarrollo descrito en el proyecto. De este modo, en la Figura 3.1 se diagrama un resumen del flujo de trabajo llevado a cabo en el desarrollo del proyecto, el cual será especificado más en detalle en las siguientes secciones de este capítulo.

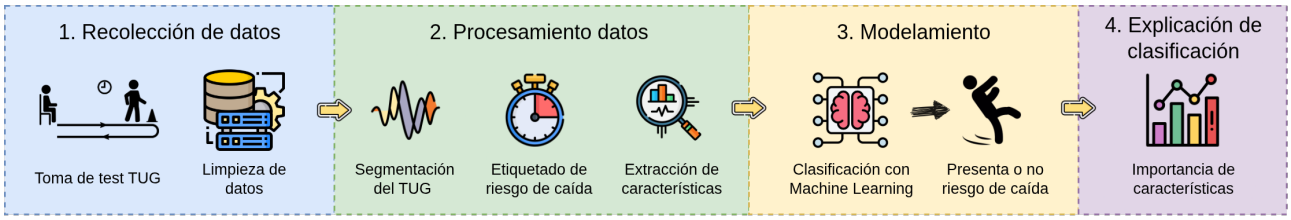
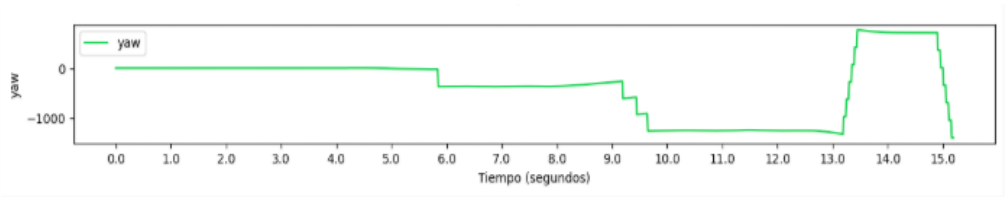


Figura 3.1: Flujo de trabajo ejecutado durante el desarrollo del proyecto.

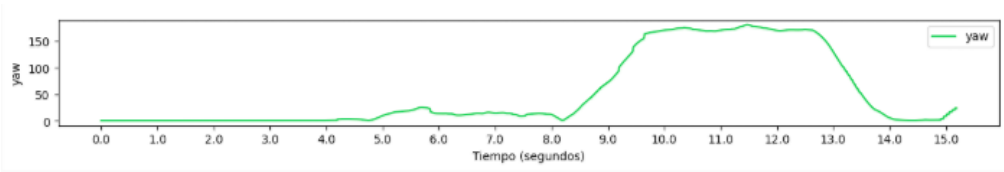
3.2. Materiales

Los datos con los cuales se trabajaron en este proyecto fueron facilitados por el Laboratorio de Biomédica de la Facultad de Ingeniería de la Universidad de Concepción. Estos datos constan de pruebas del Timed Up and Go aplicados a participantes voluntarios de Concepción (Chile), todos con 60 o más años, utilizando un sensor inercial IMU colocado en la espalda baja con una frecuencia de muestreo de $100Hz$. Por otro lado también se registró en una planilla Excel datos personales del paciente como: *edad*, *altura*, *peso*, *sexo al nacer*, *historial de caídas*, *historial médico*, etc.

El test TUG fue aplicado a un total de 62 adultos mayores (16 hombres, 44 mujeres y 2 no especificados), donde cada uno realizó el test 3 veces y algunos volvieron a realizar 3 intentos más del test una semana después. De esta manera el total de pruebas recolectadas fue de 273, de las cuales 2 tests fueron descartados por ser malas mediciones. El sensor inercial grabó en cada prueba las señales de aceleración y velocidad en los 3 ejes, así como las rotaciones de ángulos en cuaterniones. Luego gracias a un *script* en Matlab proporcionado por miembros del Laboratorio de Biomédica se pasaron los ángulos en cuaterniones a ángulos de navegación (yaw, pitch y roll). Además el *script* corrige discontinuidades presentadas en yaw utilizando un umbral adaptativo para identificar saltos anómalos de la señal, el resultado de estas correcciones se puede ver en la Figura 3.2



(a) Señal yaw con discontinuidad.



(b) Señal yaw corregida.

Figura 3.2: Resultado de corrección de una señal yaw mediante umbral adaptativo.

Con el objetivo de segmentar los test del TUG, profesionales kinesiólogos analizaron videos captados por cámaras de profundidad en el software Mokka [19] para dividir por fases cada prueba. Por lo cual se generó una planilla excel con los tiempos de duración de las fases en cada uno de los test. Cabe destacar que en general la frecuencia de muestreo de las cámaras fue de $100Hz$ pero

algunas pruebas las tomó con una frecuencia de $120Hz$.

3.3. Métodos

3.3.1. Segmentación del Timed Up and Go

Las fases del TUG conllevan tareas muy diversas, es por ello que es crucial dividir la prueba en sus diferentes segmentos para obtener una mejor caracterización de las señales. De este modo, se tomó en consideración la segmentación en fases calculada por el ojo clínico y se sincronizó el inicio del test en las grabaciones captadas por la cámara de profundidad y el sensor IMU. Sin embargo el resultado no fue positivo, pues en muchas de las pruebas no calzaba los tiempos descritos por los kinesiólogos con la duración del test en el sensor, un ejemplo de ello se muestra en la Figura 3.3.

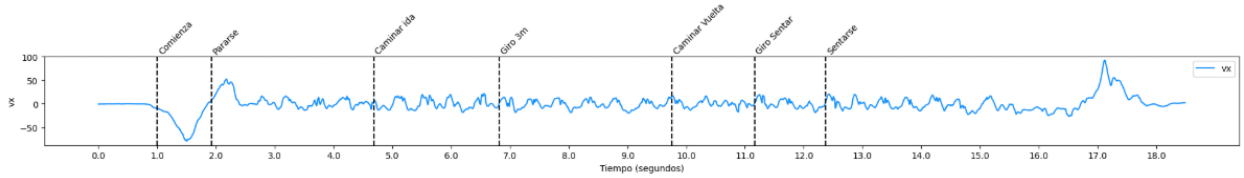
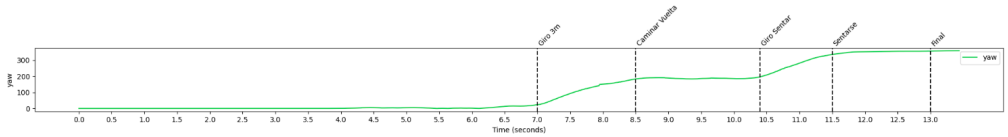
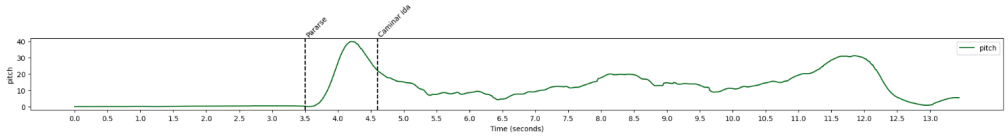


Figura 3.3: Señal del sensor inercial segmentada en fases del TUG según los tiempos descritos por los kinesiólogos.

Como método alternativo de segmentación se utilizaron las señales de pitch y yaw para detectar manualmente donde se encuentran las diferentes fases, metodología la cual está basada en el trabajo de Ortegas-Bastidas y colaboradores [5]. El procedimiento seguido se basó en centrarse en los cambios de valores en la señal yaw, la cual solo cambia cuando una persona está girando, como se ilustra en la Figura 3.4a. De esta forma, se logra identificar las fases: *Giro a los 3m*, *Caminata vuelta*, *Giro para sentarse*, *Sentarse*. Finalmente las fases de *Pararse* y *Caminata ida* fueron delimitadas debido a que la señal pitch genera un *peak* al momento de pararse, esto debido a que de manera natural las personas al pasar de una posición sedante a una posición bípeda inclinan el tronco hacia delante para desplazar el centro de gravedad. Este fenómeno permite identificar con facilidad el cambio de fase entre pararse y caminar como se muestra en la Figura 3.4b.



(a) Segmentación sobre señal yaw.



(b) Segmentación sobre señal pitch.

Figura 3.4: Segmentación en fases del test Timed Up and Go a partir de las señales yaw y pitch.

Si bien la metodología utilizada tiene la desventaja de ser una segmentación manual, por lo que está expuesta al error humano por parte de una persona no experta, se logró obtener una mayor cantidad de pruebas las cuales no habían sido procesadas por los kinesiólogos debido a que no fueron grabadas con las cámaras de profundidad.

3.3.2. Etiquetado de pruebas

De acuerdo al Instructivo de Aplicación del Control de Salud Anual del Adulto Mayor [7] propuesto por el Ministerio de Salud de Chile el 2015, si el test del Timed Up and Go dura 11 segundos o más la persona presenta riesgo de caída. Concretamente, si la duración de la prueba está entre los 11 y 20 segundos el riesgo de caída es leve, mientras que un tiempo mayor a 20 segundos indica un alto riesgo de caída. Como se ve en la Figura 3.5 existen pocos test con un tiempo mayor a 20 segundos. Por lo tanto se decidió etiquetar las pruebas en dos clases: presencia de riesgo de caída y ausencia de riesgo de caída; tomando como umbral definitorio los 11 segundos de duración propuestos por el MINSAL [7].

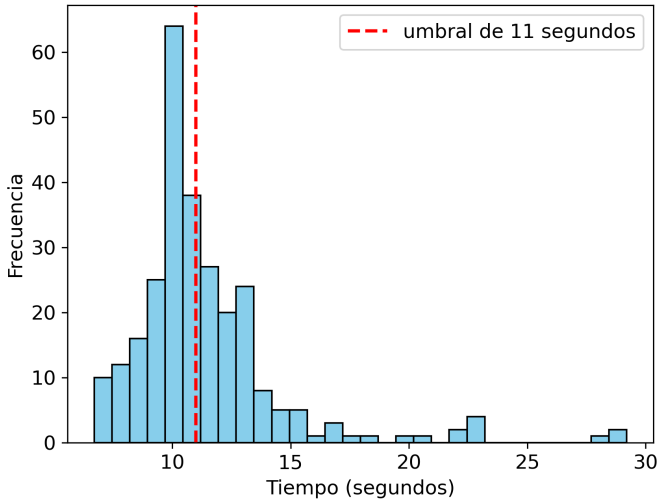


Figura 3.5: Histograma de frecuencia de la duración de las pruebas TUG tomadas. Además se marca el umbral de tiempo utilizado en el etiquetado de las pruebas.

De esta forma, como se puede apreciar en la Figura 3.6, 155 pruebas son etiquetadas con ausencia de riesgo de caída y 116 con presencia de riesgo de caída. Cabe destacar que si bien existe un desbalance entre las clases, este no es tan drástico.

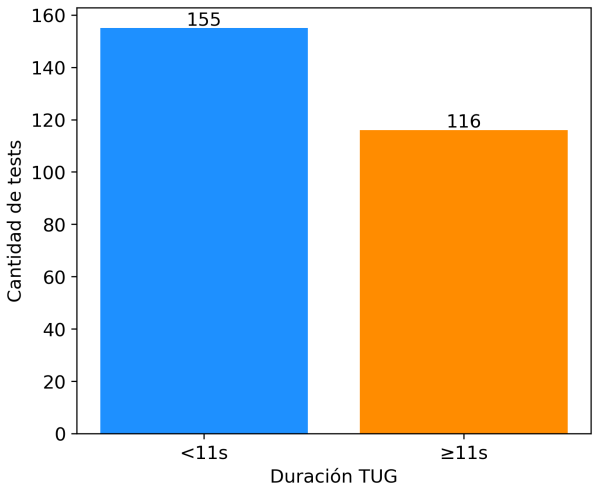


Figura 3.6: Cantidad de pruebas etiquetadas como presencia o ausencia de riesgo de caída.

Aunque el umbral de tiempo definido es un criterio validado para etiquetar los datos, existen otros factores que pueden ser indicios de riesgo de caída, como por ejemplo el historial de caída. Para un adulto mayor la presencia de caídas realizando tareas cotidianas implica el deterioro de sus

capacidades físicas. Lo anterior puede verse reflejado en el gráfico Figura 3.7, donde casi la mitad de las personas que superan los 11 segundos en el TUG se han caído alguna vez. Por otra parte, la mayoría de las personas que realizaron el test bajo el umbral nunca han tenido caídas.

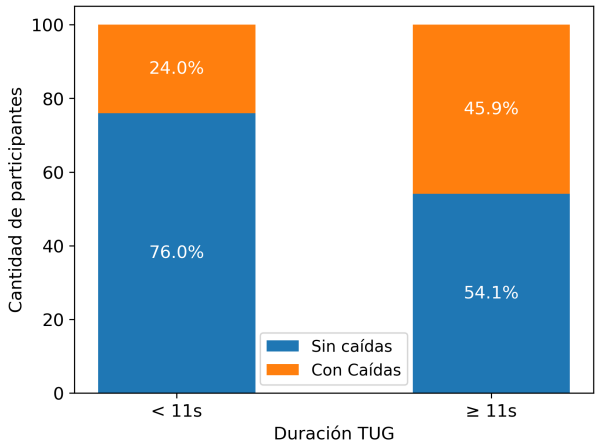


Figura 3.7: Participantes con y sin historial de caídas de acuerdo al umbral de tiempo de 11s.

Por otro lado, la edad también es un factor evidente de la pérdida de funcionalidades, es así como en la Figura 3.8 se puede notar como los distintos grupos etarios van aumentando el tiempo de duración de sus test a medida que la edad es mayor y posicionándose cada vez más por encima del umbral de 11 segundos. De esta manera, los factores presentados respaldan el umbral definido como un decidor representativo de riesgo de caída para la muestra de pruebas recolectada, mas no son incluidos en la fase de modelamiento pues pudieran terminar sesgando al modelo en la predicción.

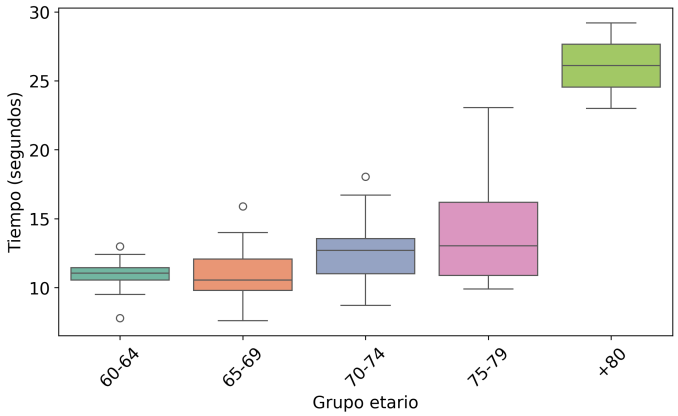


Figura 3.8: Distribución de tiempos de duración en test TUG por grupos etarios.

3.3.3. Extracción de características

Una vez realizada la segmentación de las pruebas por fases del TUG se le extrajo la información más relevante a las señales. Para ello se calcularon las medidas estadísticas: *media*, *mediana*, *desviación estándar*, *valor mínimo*, *valor máximo*, *rango*, *kurtosis*, *skewness*, *entropía* y *root mean square (RMS)* de cada señal en las distintas fases del TUG. Es importante mencionar que estas medidas estadísticas permiten obtener características que no contienen un componente temporal. Por lo que evita que el modelo sea sesgado al momento de la predicción de una etiqueta temporal como lo es el umbral de tiempo.

La Figura 3.9 muestra todo el proceso de extracción de características para cada señal. De este modo, las pruebas están caracterizadas por un total de: *num_señales* $\times$ *num_fases* $\times$ *num_características* = $9 \times 6 \times 10 = 540$ características, las cuales servirán de input para los modelos de Machine Learning descritos en la siguiente sección.

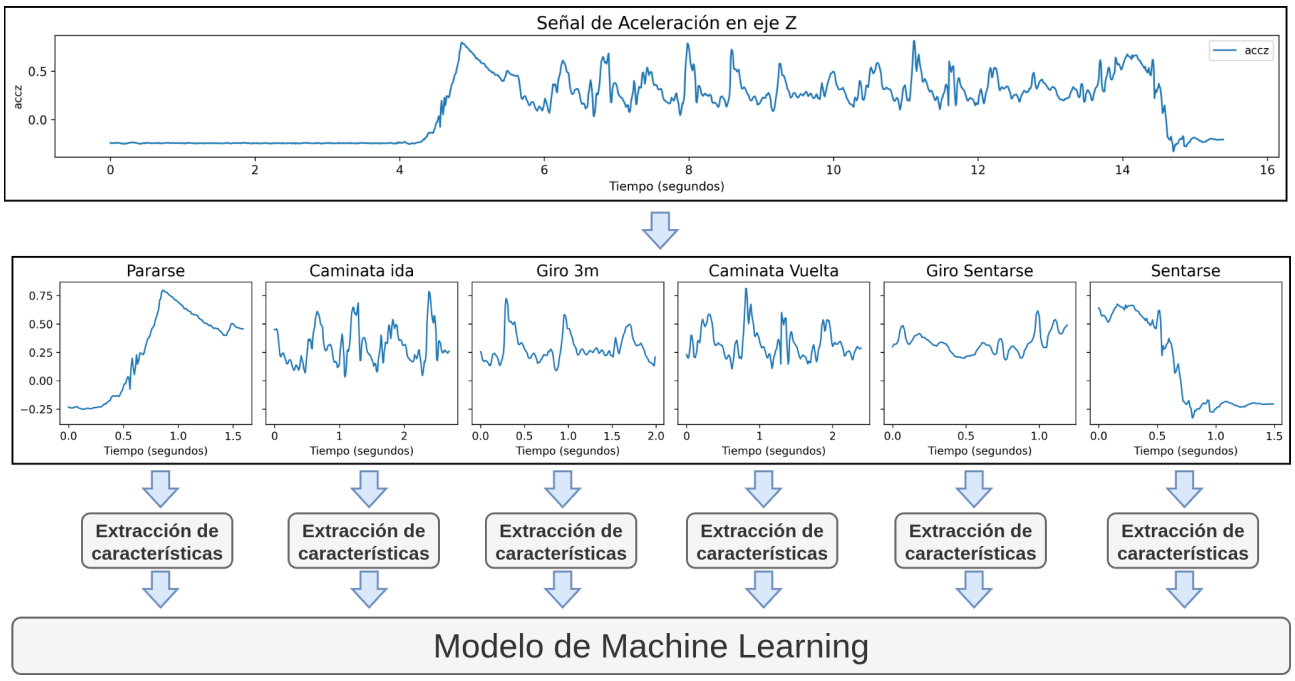


Figura 3.9: Proceso de extracción de características para cada señal.

Para un mejor seguimiento de qué representa cada característica, se nombró cada una con el formato: *medidaEstadística_señal_fase*; el cual se lee de la forma: 'La *medida estadística* de la *señal* en la *fase*'.

3.3.4. Modelamiento

Para clasificar la presencia de riesgo de caída, se dividieron los datos en entrenamiento y testeo con una proporción de 80-20 respectivamente. Dado que los participantes realizaron más de una vez el Timed Up and Go se agruparon los datos por participante antes de particionar el dataset. Dejando las pruebas de 51 personas para entrenamiento y las pruebas de los otros 11 participantes para testeo. De esta manera, se evita que el modelo se aprenda los patrones de una persona en entrenamiento y luego le sea muy fácil en testeo predecir las pruebas de esa misma persona, causando un sobre rendimiento al momento de calcular las métricas.

Los datos fueron estandarizados, tanto entrenamiento como testeo, con una normalización z-score basada en la media y la desviación estándar del conjunto de entrenamiento. Por otro lado, con el objetivo de atacar el desbalance de las clases se aplicó al conjunto de entrenamiento la técnica de oversampling SMOTE, la cual genera nuevos datos sintéticos interpolando puntos entre una instancia de la clase minoritaria y sus vecinos más cercanos. De esta manera el conjunto de entrenamiento, inicialmente con una distribución de 120 tests sin riesgo y 100 tests con riesgo, pasó a tener 120 pruebas para ambas clases. Creando de esta forma, 20 tests sintéticos a partir de los originales con el objetivo de prevenir que el modelo sea sesgado por la clase mayoritaria. Finalmente el dataset consta de 291 pruebas: 240 para entrenar y 51 para testeo.

Dado que el set de datos es relativamente pequeño, se eligieron modelos de Machine Learning Clásico para la predicción. En específico se eligieron los algoritmos basados en árboles de decisión Random Forest y XGBoost. El primero ha demostrado tener buenos resultados para problemas simples y dataset pequeños; y el segundo es una implementación más compleja, la cual utiliza la técnica del descenso del gradiente para optimizar la construcción de los árboles de decisión. Además, XGBoost es usado generalmente para volúmenes de datos medianos y grandes, y tiene una gran adaptación a problemas más complejos. Por último, se probará un enfoque distinto con Support Vector Machine, un algoritmo basado en kernels que se puede ajustar a distintos tipos de datos usando métodos lineales y no lineales.

Las métricas con las cuales se van a evaluar el desempeño de los modelos de Machine Learning son:

1. Accuracy: Proporción de predicciones correctas (tanto positivas como negativas) sobre el total de instancias evaluadas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

2. Precision: Proporción de instancias predichas como positivas que realmente lo son.

$$Precision \ = \ \frac{TP}{TP + FP} \tag{3.2}$$

3. Recall: Proporción de instancias positivas correctamente identificadas.

$$Recall \ = \ \frac{TP}{TP + FN} \tag{3.3}$$

4. F1-Score: Media armónica entre precisión y recall, que equilibra ambas métricas.

$$F1\text{-}Score \ = \ 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.4}$$

Donde:

- TP (True Positives): casos positivos correctamente clasificados.
- FP (False Positives): casos negativos clasificados erróneamente como positivos.
- TN (True Negatives): casos negativos correctamente clasificados.
- FN (False Negatives): casos positivos clasificados erróneamente como negativos.

Cabe destacar que en el modelamiento la clase positiva hace referencia a la presencia de riesgo de caída, mientras que la negativa es la ausencia de riesgo de caída. De forma agregada, dado el desbalance del dataset se tiene un mayor enfoque en la métrica *f1-score*, pues valora de forma equilibrada la tasa de detección de falsos positivos y falsos negativos, evitando de esta manera ser sesgado por modelos que solo aprenden la clase mayoritaria.

3.3.5. Características más importantes

Para la selección de las características que más influyeron en la predicción se aplica el cálculo de los valores Shap [20] a las características del modelo que mejor resultado haya logrado en la etapa de modelamiento. Dado que es un problema de clasificación binaria a cada característica se le atribuirá un valor según su contribución a la predicción individual de un test. Además los valores asignados por contribución serán positivos (si la característica contribuye a predecir la clase con riesgo de caída) o negativos (si contribuye a predecir la clase sin riesgo de caída).

Para una mejor interpretación de los valores Shap se calcula el valor absoluto de la contribución obtenida, de esta forma se puede ver cuáles fueron las características que más ayudaron a diferenciar entre una clase o la otra.

Una vez obtenido los valores Shap absolutos, se puede agrupar y sumar las contribuciones absolutas de las características por fase, debido al formato especificado en la Sección 3.3.3, para obtener de esta manera la contribución de cada fase del Timed Up and Go. Esto permite identificar cuales fueron las fases más críticas al momento de predecir el riesgo de caída. Finalmente para trabajar con un formato más interpretable se calculan los porcentajes de importancia de cada fase.

Análogamente, sumando las contribuciones absolutas de las características por señal, se obtiene la contribución de cada señal del sensor inercial en la predicción, dando cuenta de cuál es la que mayor información entrega para la diferenciación de las personas con y sin riesgo de caída. Por último se lleva los valores obtenidos a porcentajes para facilitar la lectura de los resultados.

Repitiendo el proceso anterior se puede obtener para cada fase las señales que mayor importancia tuvieron, agrupando primero por fase y luego sumando las contribuciones de cada señal en la fase, para finalmente pasarlo a porcentaje.

Los análisis anteriormente comentados son graficados y llevados a un dashboard en la Sección 4.4, la cual busca visualizar los resultados obtenidos.

Capítulo 4

Resultados

4.1. Configuración de los modelos

Los modelos propuestos en el capítulo anterior fueron entrenados utilizando la técnica de *Cross Validation*, la cual busca en todas las configuraciones posibles de hiperparámetros la que mejor resultado obtenga de acuerdo a una métrica. En este caso, al momento de ajustar el modelo a los datos se buscó mejorar la métrica *f1-score* y además se utilizó un K-fold de 5 durante el entrenamiento para asegurar la robustez de los resultados. La Tabla 4.1 muestra todos los hiperparámetros y sus valores que se aplicaron al momento de realizar la validación cruzada para cada uno de los modelos en la etapa de entrenamiento.

Tabla 4.1: Hiperparámetros y sus valores para cada uno de los modelos entrenados.

Modelo	Hiperparámetro	Valores
Random Forest	n_estimators	100, 200, 300, 500
	max_depth	None, 10, 20, 30
	min_samples_split	2, 5, 10
	min_samples_leaf	1, 2, 4
SVM	C	0.1, 10, 100, 200
	kernel	linear, rbf
	gamma	scale, auto
XGBoost	n_estimators	100, 200, 300, 400
	learning_rate	0.1, 0.4, 0.5, 0.6
	gamma	0, 0.01, 0.1

Comenzando por el modelo de Random Forest, el *n_estimators* especifica cuántos árboles se entrenan. Por lo general, más árboles suelen reducir la varianza y mejorar la robustez, pero incrementan el costo computacional. Por otro lado, *max_depth* limita la profundidad máxima de cada árbol, un valor bajo produce árboles poco complejos que mitigan el sobreajuste, mientras que un valor alto permite capturar patrones más detallados a costa de mayor riesgo de *overfitting*. El *min_samples_split*, por su parte, define el número mínimo de muestras requerido para intentar dividir un nodo, valores más altos obligan a que los nodos contengan más observaciones antes de ramificarse, generando árboles más simples y generalizables. Por último, *min_samples_leaf* fija cuántas muestras debe haber como mínimo en cada hoja terminal, de modo que un umbral mayor suaviza las predicciones al evitar hojas demasiado específicas y ayuda a reducir el ruido en los datos.

En el modelo SVM, el hiperparámetro C controla el equilibrio entre maximizar el margen y minimizar los errores de clasificación. Es así como un valor alto castiga con fuerza cada ejemplo mal clasificado, sin embargo un valor bajo tolera más errores para lograr una mayor capacidad de generalización. El hiperparámetro *kernel* determina la función que transforma los datos al espacio donde se calculará el hiperplano. Por lo que con un kernel lineal la separación se realiza directamente en el espacio original, lo que resulta eficiente y adecuado cuando la relación entre las variables es aproximadamente lineal, mientras que con el kernel rbf (*radial basis function*) los datos se proyectan a un espacio de dimensión (potencialmente) infinita, permitiendo capturar fronteras altamente no lineales mediante la métrica de distancia euclidiana. Finalmente, *gamma* influye en la amplitud de la campana del kernel rbf (y de otros kernels basados en distancia): un gamma elevado hace que cada punto de entrenamiento influya solo en su vecindad inmediata, y un gamma pequeño expande dicho alcance, produciendo fronteras más suaves que tienden a generalizar mejor.

Por último en XGBoost, el hiperparámetro *n_estimators* indica cuántos árboles secuenciales se entrenan, a mayor cantidad el modelo puede capturar patrones más complejos, pero con mayor riesgo de sobreajuste y mayor costo computacional. El hiperparámetro *learning_rate* multiplica la contribución de cada árbol antes de agregarla al ensamble, así los valores pequeños ralentizan el aprendizaje, exigen más árboles y suelen mejorar la capacidad de generalización, mientras que valores altos aceleran la convergencia pero pueden provocar un ajuste demasiado específico. Por su parte, *gamma* define la reducción mínima en la función de pérdida que debe lograrse para permitir una división adicional de un nodo; un gamma alto impone un umbral exigente que poda el crecimiento de los árboles y añade regularización, ayudando a controlar el sobreajuste, y un gamma cercano a cero permite divisiones fáciles y genera árboles más profundos, con el potencial de capturar interacciones complejas.

4.2. Resultados generales

Como se puede notar en la Tabla 4.2 los modelos obtienen un buen rendimiento en general. Cabe destacar que los algoritmos basados en árboles de decisión (Random Forest y XGBoost) obtuvieron mejores métricas que el modelo basado en kernel (SVM). En todas las métricas el mejor resultado fue obtenido por XGBoost el cual logró un *f1-score* de 91.17 % y un recall de 92.58 % con los hiperparámetros *gamma*: 0, *learning_rate*: 0.5 y *n_estimators*: 100. Considerando la importancia en el área de la salud de diagnosticar correctamente a las personas que si presentan cierta patología, el *recall* cobra una mayor relevancia en los resultados obtenidos.

Tabla 4.2: Métricas obtenidas con mejores hiperparámetros por modelos en clasificación de riesgo de caída.

Modelo	F1-Score	Accuracy	Recall	Precision
Random Forest	88.80 %	90.19 %	89.46 %	88.23 %
SVM	81.11 %	84.31 %	80.08 %	82.52 %
XGBoost	**91.17 %**	**92.15 %**	**92.58 %**	**90.15 %**

Buscando mejorar los resultados obtenidos anteriormente se consideró utilizar el método de selección de características *Cohen's d-value*. Esta técnica compara los promedios de cada una de las características en las dos clases y los expresa en unidades de desviación estándar. Un *d-value* de 0 indica que casi no hay diferencia, 0.3 se considera pequeña, 0.5 mediana y 0.8 grande. Así se puede saber cuáles características tienen un mayor efecto de diferenciación entre ambas clases. Por lo tanto, filtrar las características según un umbral de *d-value* puede ayudar a eliminar valores que aporten poco a la predicción.

La Tabla 4.3 muestra los *f1-score* obtenidos para distintos umbrales del *d-value*. De esta forma, se puede notar que una umbralización más fuerte no conllevó a mejorar el rendimiento en los modelos de SVM y XGBoost. Sin embargo se destaca que con un *d-value* = 0.3, XGBoost obtuvo el mismo resultado que con una mayor cantidad de características. Por otro lado, Random Forest si alcanzó una mejora en su rendimiento con una umbralización de 0.4 aunque no supera al 91.17 % de *f1-score* alcanzado por XGBoost.

Tabla 4.3: F1-Score obtenido en clasificación por modelos utilizando Cohen's d-value para la selección de características.

Modelo	d-value		
	0.3	0.4	0.5
Random Forest	87.12 %	89.12 %	84.77 %
SVM	90.89 %	86.76 %	84.32 %
XGBoost	91.17 %	89.12 %	89.12 %

4.3. Análisis del mejor modelo

Teniendo en cuenta los resultados de la sección anterior, un análisis más detallado del rendimiento del modelo XGBoost entrenado (sin selección de características) se muestra en la Figura 4.1, donde se presenta la matriz de confusión generada a partir de las predicciones del conjunto de prueba retenido (N = 51). Cabe destacar que la distribución equilibrada de errores presentada en la matriz, confirma la eficacia del modelo para clasificar correctamente ambas clases, lo cual es relevante dada la desproporción original entre ellas. Además, la detección exitosa de la clase minoritaria 'riesgo de caída' destaca la efectividad de las técnicas de *oversampling* aplicadas durante el entrenamiento.

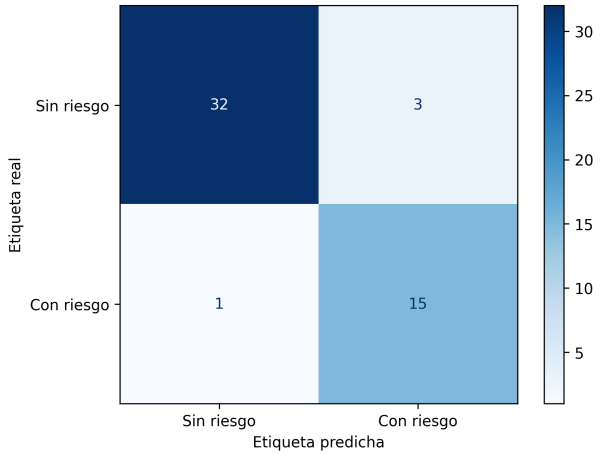


Figura 4.1: Matriz de confusión de predicción de XGBoost sobre set de testeo.

Por otro lado, la solidez de estos hallazgos se refuerza con la curva ROC ilustrada en la Figura 4.2, que muestra un área bajo la curva (AUC) de 0.96, valor que se sitúa dentro del rango de desempeño diagnóstico excelente. Además, la pronunciada ascensión de la curva ROC cerca del

origen indica la gran capacidad del modelo para maximizar la tasa de verdaderos positivos con incrementos mínimos en los falsos positivos, una característica muy deseable dado que el mayor costo clínico está asociado a no detectar casos de riesgo de caída en comparación con emitir falsas alarmas ocasionales.

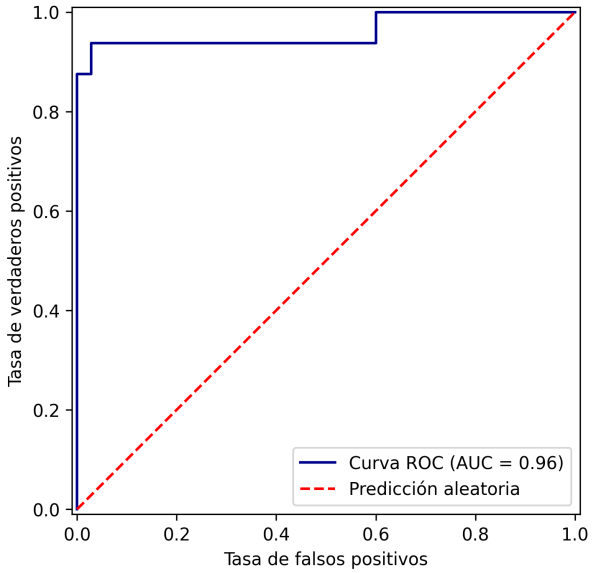


Figura 4.2: Curva ROC sobre la clase positiva (presencia de riesgo de caída) en predicción de XGBoost sobre set de testeo.

4.4. Importancia de fases y características

Una vez obtenido el mejor modelo se calcularon los valores Shap para el XGBoost entrenado, tal como se describe en la Sección 3.3.5. De esta forma, se obtiene una lista de valores Shap absolutos, uno por característica, los cuales representan la contribución a la predicción de las características en cada uno de los test en el set de testeo. Para una mejor interpretación de estos valores se calculó para cada característica su valor Shap promedio, obteniendo una lista de 540 valores de contribución. De esta manera, estos datos sirvieron para realizar una serie de gráficos con el objetivo de entender de forma general cuales fueron las fases, señales y características más importantes en la predicción.

Es así como la Figura 4.3 se obtuvo a partir de agrupar las contribuciones de las características por fases. En esta se puede observar como las fases de marcha, Caminata Ida y Caminata Vuelta, son las más relevantes al momento de la predicción. Sin embargo, el Giro a los 3m destaca por su baja contribución, aún siendo una etapa donde la persona está caminando. Por otro lado, en las

etapas donde la marcha no está involucrada el Giro para Sentarse tiene una notoria importancia por sobre las demás.

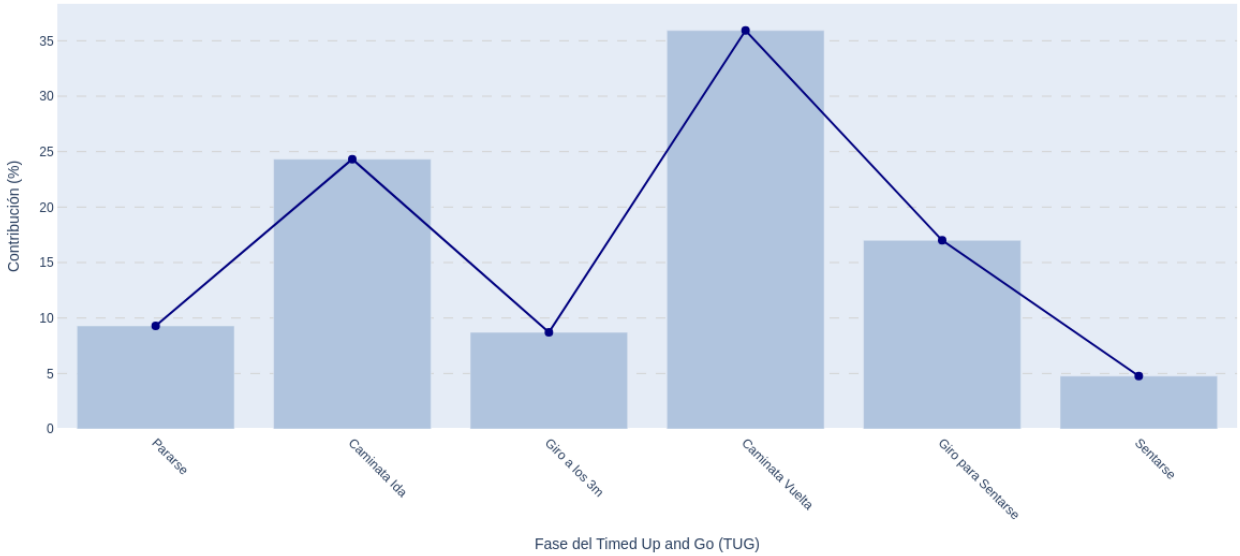


Figura 4.3: Contribución porcentual de las características en cada fase del TUG.

Tomando un enfoque distinto y agrupando las contribuciones por señal se puede obtener la Figura 4.4, en la cual se denota la importancia de las señales captadas por el sensor inercial IMU. Entre ellas las que más destacan son la aceleración en los ejes Y y Z, además de la velocidad en los ejes X y Y. Cabe destacar que a partir de las señales en el eje Y se logra obtener información de las pisadas de los pacientes, por lo cual las caminatas más inestables e inseguras tienden a tener aceleraciones más variadas [21]. Por otra parte es de notar la baja importancia en la aceleración en el eje X, dado que mide las cargas de orientación en ambos pies al pisar. Estos patrones son más notorios en personas con patologías como ACV (Accidente Cerebrovascular) [22], en cambio los participantes del experimento son adultos mayores sanos.

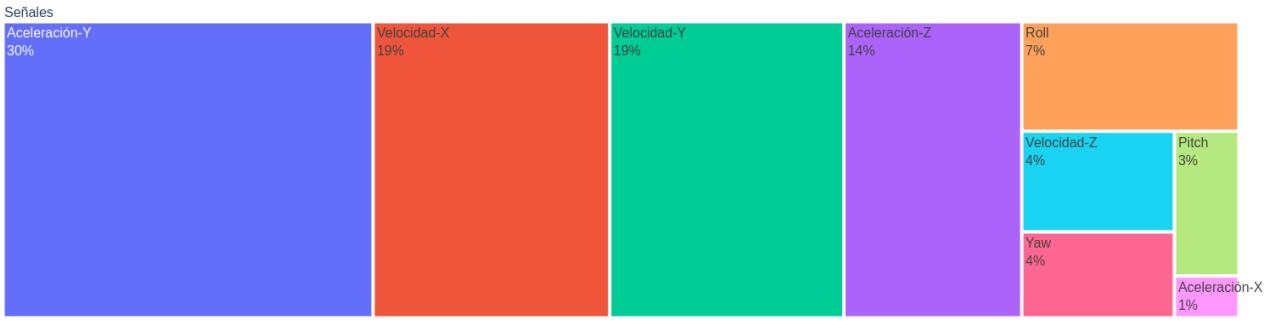


Figura 4.4: Contribución porcentual de las características por señal en predicción.

Haciendo un análisis más a fondo se agrupa por fase y luego por señal para graficar la Figura 4.5. Esta figura consiste en 6 gráficos de barra, uno por etapa del TUG, los cuales muestran las 5 señales con mayor contribución porcentual en sus respectivas etapas. Vale la pena señalar que las señales de aceleración y velocidad están abreviadas de la forma *accx-accy-accz* y *vx-vy-vz* respectivamente. Por otra parte, acorde con el gráfico anterior de tipo treemap, la aceleración en Y, además de las velocidades en X y Y destacan por estar presentes en casi todas las fases. Cabe destacar la fase de Caminata Vuelta, donde la contribución de la aceleración en Y es mucho mayor a la de las demás. A diferencia de las etapas de Pararse y Sentarse donde la distribución de las contribuciones es más homogénea.

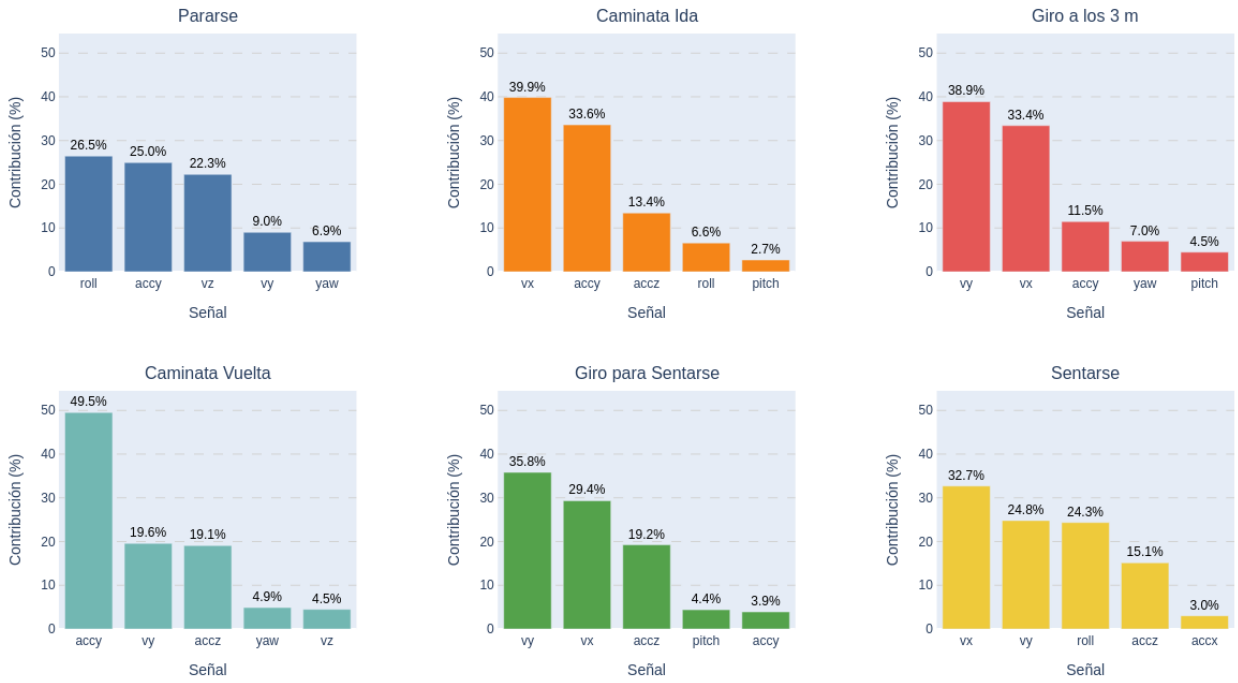


Figura 4.5: Top 5 de señales que mayor contribución porcentual tuvieron en cada fase del TUG.

Finalmente, la Figura 4.6 muestra las 10 características con mayor contribución en la predicción, donde las características están ordenadas en orden descendente, en sentido horario, partiendo desde las 12. De forma agregada, los colores están asociados a la fase a la cual pertenecen en la Figura 4.5. Es importante mencionar que para evitar la saturación de colores en el gráfico, si hay más de una característica de una misma fase el color se irá degradando a medida que la característica tenga menos relevancia. De esta forma, acorde con el gráfico anterior se puede notar que la etapa que más predomina es la Caminata Vuelta, seguido de la Caminata Ida y el Giro para Sentarse. Con respecto a las medidas estadísticas más relevantes, la entropía brinda una gran contribución en varias de las señales pertenecientes a la etapa que mayor aporte tiene, siendo la medida con mayor porcentaje de contribución.

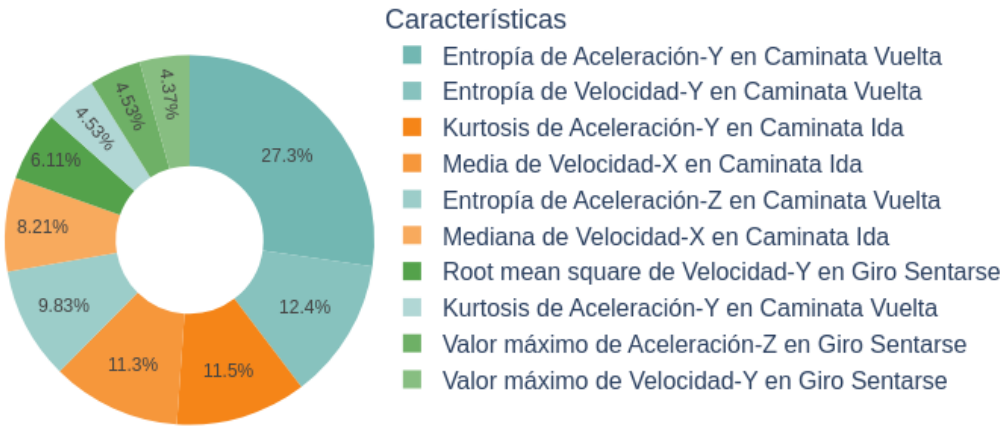


Figura 4.6: Top 10 características con mayor contribución en la predicción.

4.5. Dashboard

Con el objetivo de crear una herramienta visual más interactiva donde poder ver los resultados obtenidos, se creó un dashboard. Esta herramienta fue creada con la librería Streamlit [23] y los gráficos interactivos fueron creados con la librería Plotly [24] en python.

El objetivo del dashboard es poder analizar individualmente los valores Shap obtenidos en cada una de las pruebas a través de los gráficos presentados en la Sección 4.4. De esta manera, se puede examinar diferentes pruebas del Timed Up and Go, mediante valores Shap y la contribución de las características desde distintos puntos de vista, ya sea por señal o por fase del TUG. Dándole así una explicación a la predicción del riesgo de caída a través de un enfoque de Machine Learning. A continuación se explica en orden cada una de las vistas del dashboard, las cuales están contiguas (una abajo de la otra), y los gráficos obtenidos (calculados según el procedimiento descrito en la Sección 3.3.5).

En la Figura 4.7 se puede ver el dataframe que sirvió de conjunto de testeo para el modelamiento en la Sección 3.3.4 y a su derecha un párrafo explicando el contenido del mismo. Además se encuentra un selector abajo del texto que permite configurar el par de valores únicos 'ID paciente - Intento'. Una vez que se hace click en el botón 'Predecir test' el modelo predice la prueba con el 'ID paciente - Intento' seleccionado. Acto seguido, un mensaje aparecerá indicando si en el test el paciente tiene o no indicios de riesgo de caída, como también se muestra en la Figura 4.8.

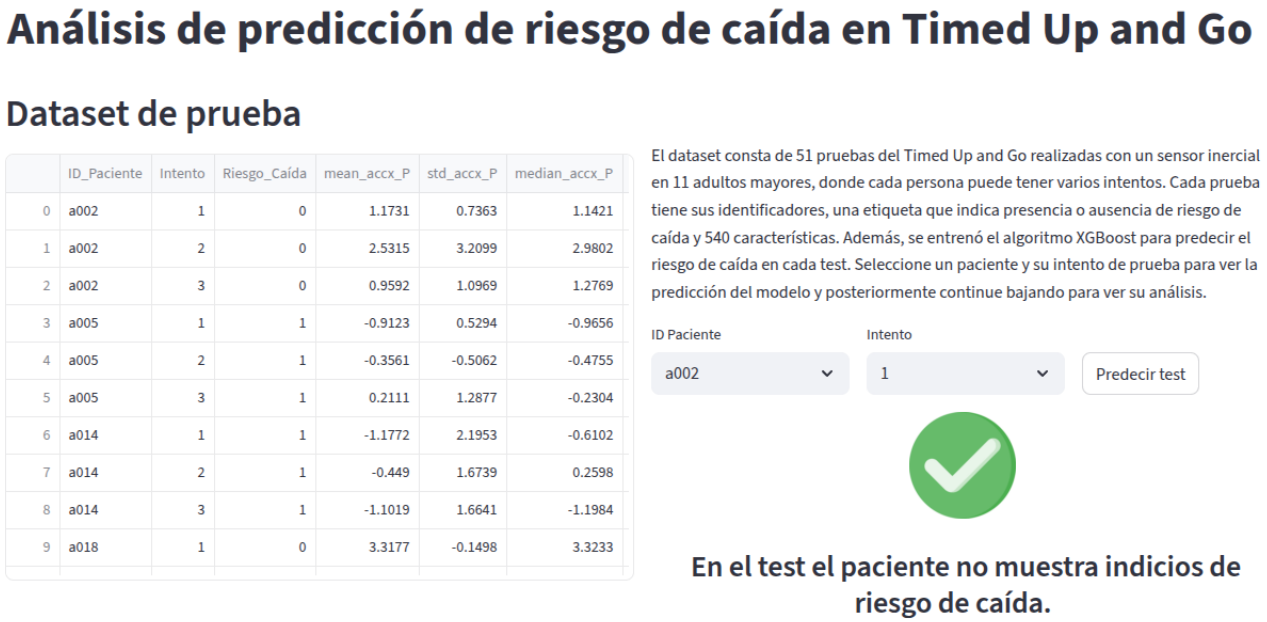


Figura 4.7: Vista del dashboard donde se muestra y explica el dataset, además de predecir un test sin indicios de caídas.

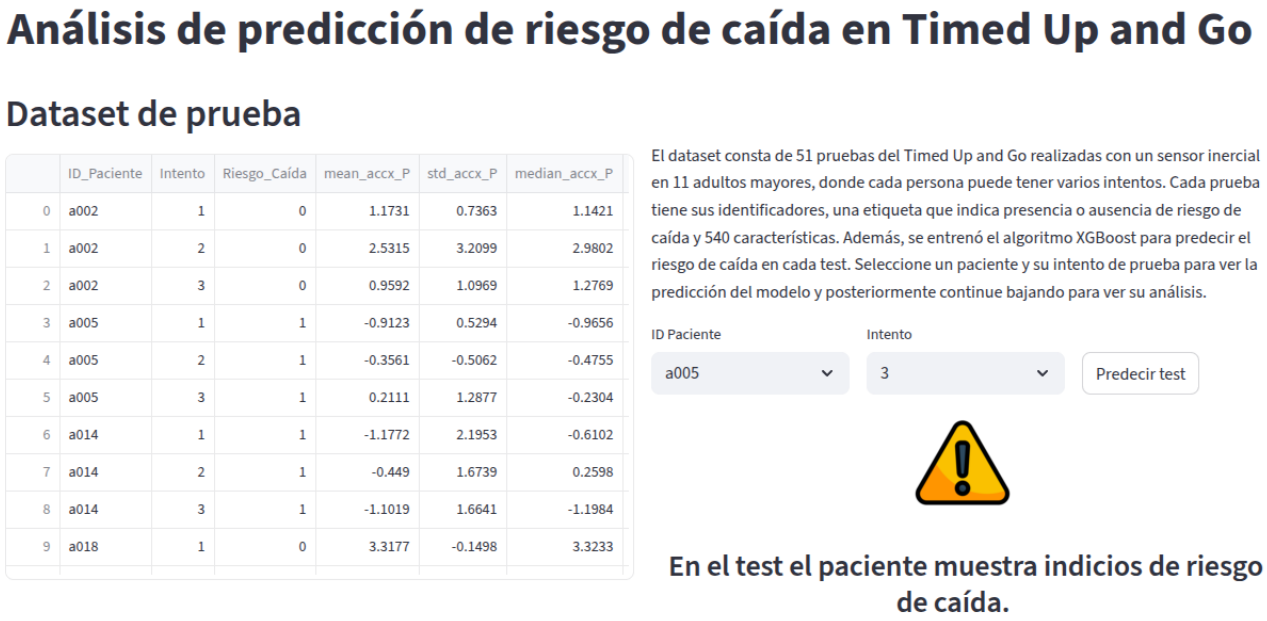


Figura 4.8: Vista del dashboard donde se muestra y explica el dataset, además de predecir un test con indicios de caídas.

Como se muestra en la Figura 4.9 un título con el encabezado ‘Explicación de predicción en ID Paciente - Intento’ da a conocer que la siguiente sección abarcará el análisis de la predicción.

El gráfico de barra, por otra parte, exhibe la contribución de cada una de las fases del Timed Up and Go en la predicción. Adicionalmente un texto es agregado al lado del gráfico como ayuda para interpretarlo. Por otra parte, el botón 'Más Información' abre una ventana emergente que contiene una explicación sobre la aplicación de valores Shap en XGBoost para la obtención de los gráficos, como se puede ver en Figura 4.10.

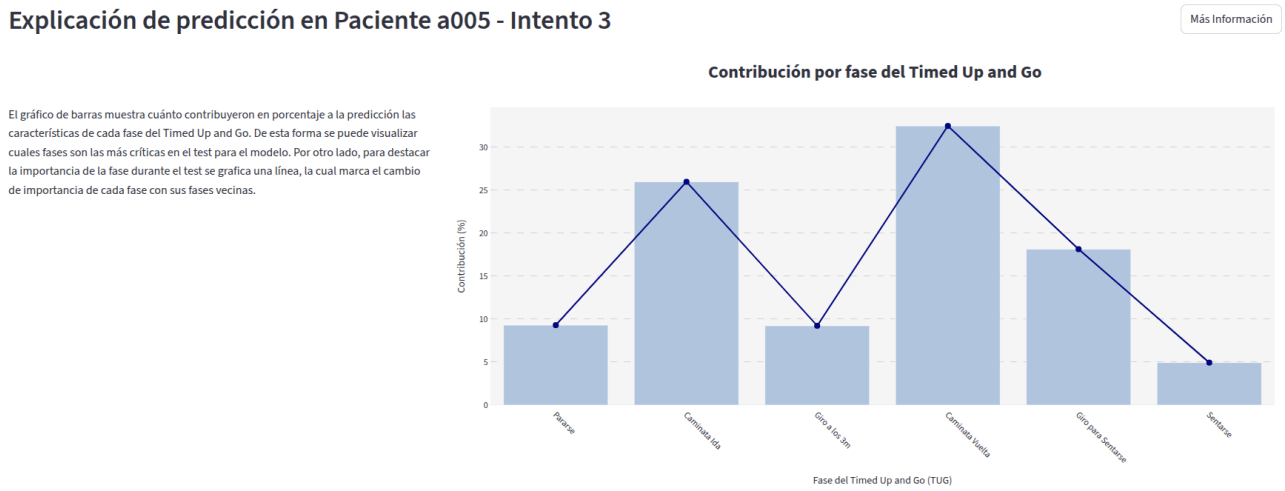


Figura 4.9: Vista donde se muestra un gráfico de barra con la contribución de cada fase en la predicción.

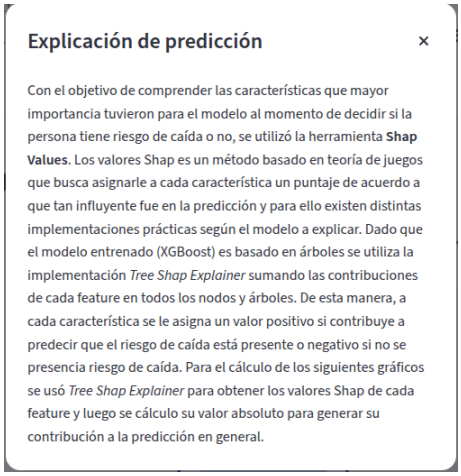


Figura 4.10: Ventana emergente que aparece al presionar botón 'Más Información'.

En la siguiente vista mostrada en la Figura 4.11 se busca visualizar un gráfico de tipo *treemap*, donde se hace énfasis en la contribución de las señales del sensor inercial utilizado para medir

las pruebas del Timed Up and Go. Además, en correspondencia con la vista anterior, un texto explicativo del gráfico se encuentra a su izquierda para facilitar la lectura del mismo.

Figura 4.11: Vista donde se muestra un gráfico treemap con la contribución de cada señal en la predicción

La Figura 4.12 muestra varios gráficos de barra que dan cuenta de un análisis más específico de las señales. Visualizando de esta manera cuáles fueron las 5 señales en cada fase que mayor importancia tuvieron, además de mostrar cómo se reparten porcentualmente la contribución en la predicción. Cabe destacar que cada fase tiene un color único y que las señales de aceleración y velocidad están abreviadas de la forma *accx-accy-accz* y *vx-vy-vz* respectivamente. Finalmente un texto de ayuda es añadido al lado de los subgráficos como complemento a su interpretación.

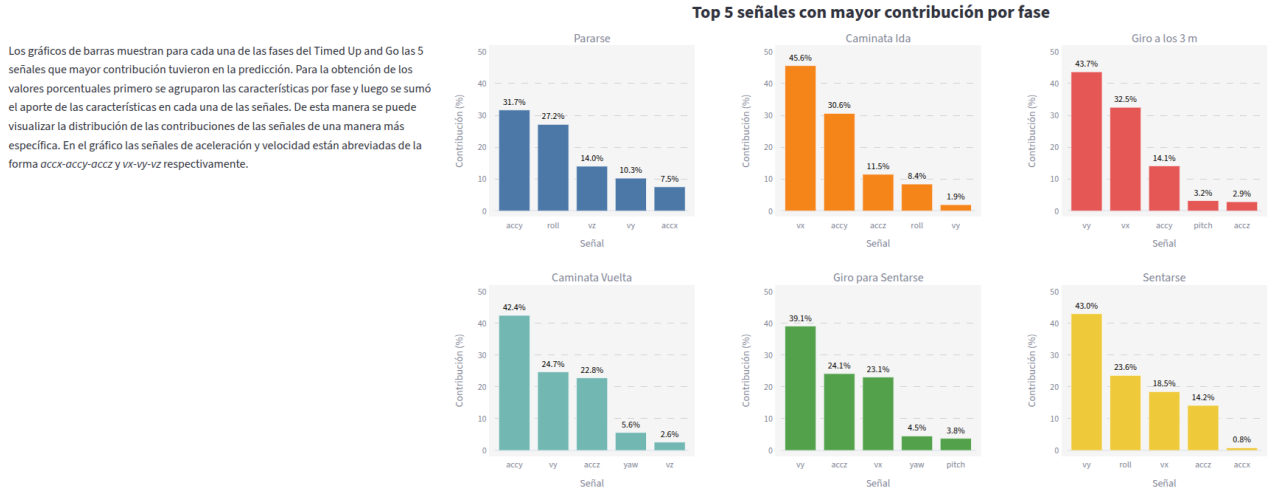


Figura 4.12: Vista donde se muestra gráficos de barra con las señales que mayor contribución tuvieron en cada fase en la predicción.

Por último, la Figura 4.13 presenta un gráfico de torta con las 10 características con mayor contribución en la predicción. Tal como lo especifica el texto de ayuda a la izquierda del gráfico,

las características están ordenadas en orden descendente, en sentido horario, partiendo desde las 12. Además los colores van asociados a la fase a la cual pertenecen en la Figura 4.12. Para evitar la saturación de colores en el gráfico, si hay más de una característica de una misma fase el color se irá degradando a medida que la característica tenga menos relevancia.

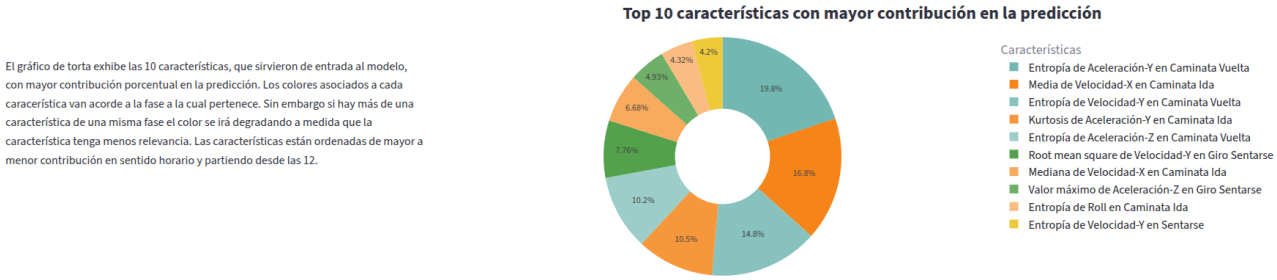


Figura 4.13: Vista donde se muestra un gráfico de torta con las 10 características que mayor contribución tuvieron en la predicción.

Capítulo 5

Conclusiones

El presente trabajo ha dado cuenta del cumplimiento de los objetivos planteados al inicio del informe. Satisfaciendo exitosamente el primer Objetivo Específico al extraer medidas estadísticas de cada una de las fases para las 9 señales. De este modo cada prueba es caracterizada por valores que no dependen del tiempo. Éste método tuvo resultados positivos al momento de la modelación, logrando además caracterizar de una manera más completa las pruebas del Timed Up and Go.

El segundo Objetivo Específico también es efectuado de manera correcta, dado que los resultados obtenidos en la implementación de la solución propuesta dan cuenta de la gran capacidad de los algoritmos de Machine Learning para poder detectar el riesgo de caída en el test iTUG. Esto se ve reflejado tanto en el estado del arte como en el alto rendimiento de los modelos entrenados como XGBoost, alcanzando métricas sobre el 90 % en *f1-score*. Dando cuenta de la robustez de la predicción, aún siendo un dataset desbalanceado. Sin embargo, el uso de técnicas de selección de características no tuvo una mejora en los métricas, igualando o empeorando los resultados. De todas maneras vale la pena señalar que con una menor cantidad de características el modelo pudo obtener iguales métricas que con el total de ellas.

De igual manera, seleccionando el modelo de XGBoost sin aplicar *Cohen's d-value* y calculando los valores Shap al momento de predecir el conjunto de testeo, se analizó las contribuciones de todas las características, satisfaciendo en su totalidad el tercer Objetivo Específico. De esta forma se logró visualizar que las fases del Timed Up and Go que mayor relevancia tuvieron fueron las fases que involucran la marcha. Concretamente la fase de Caminata de Vuelta resultó ser la más informativa, esto debido a que se presenta una mayor fatiga que en la ida y por lo tanto la marcha se ve más afectada. Por otro lado, etapas menos dinámicas como Pararse y Sentarse no entregaron mucha información al momento de predecir, por lo que se presume que estas acciones no se ven tan afectadas en adultos mayores sanos con bajo riesgo de caída.

Por parte de la contribución por señal, vale la pena señalar que la aceleración en Y y las velocidades en X y Y fueron las más relevantes al momento de la predicción. Estando presente entre las 5 señales que mayor contribución tuvieron en casi todas las fases. Cabe destacar que la aceleración en el eje Y tuvo una contribución de casi un 50 % en la fase de Caminata de Vuelta, la cual fue la etapa más determinante en la predicción. Es de resaltar que las rotaciones yaw, pitch y roll no tuvieron una mayor contribución en la predicción en promedio, aún cuando en fases como Pararse y Sentarse es notorio un *peak* en la señal pitch. De igual forma en las etapas que involucran giros el yaw no fue un gran protagonista como si se veía en la señal cruda.

Con respecto a cuales fueron las características más importantes en la predicción, la entropía de la aceleración en Y de la fase Caminata de Vuelta fue la que mayor contribución tuvo con un 27.3 %, resultado el cual va acorde a los anteriores análisis. Es digno de mencionar que la entropía se repitió en 3 de las 10 características más importantes, siendo la medida estadística con mayor contribución en el top 10. Lo anterior nos da entender que el ruido o desorden en las señales es un factor clave que puede diferenciar entre marcha con problema y una sana.

El cuarto Objetivo Específico fue abordado en su totalidad con el desarrollo de un dashboard que permite analizar individualmente cada una de las pruebas. Brindando de esta manera un análisis más específico de la predicción y permitiendo visualizar la variación de la influencia de las características entre pruebas con y sin riesgo de caída. Sin embargo, el análisis hecho tiene un alcance limitado pues es un enfoque desde el Machine Learning, por lo que carece de sustento clínico para que pueda ser una ayuda para los tomadores de decisiones en el área de la rehabilitación.

A pesar de que los resultados obtenidos en este trabajo dan cuenta del gran potencial de predicción de riesgo de caída con modelos de Machine Learning en adultos mayores, queda pendiente como trabajo a futuro validar con profesionales de la rehabilitación el análisis de los determinantes de riesgo de caída realizado en esta memoria de título. Además de considerar un estudio con pacientes que tengan patologías que afecten la marcha, como Parkinson o ACV.

Bibliografía

[1] C. Albala, “El envejecimiento de la poblaciÓn chilena y los desafÍos para la salud y el bienestar de las personas mayores,” *Revista Médica Clínica Las Condes*, vol. 31, no. 1, pp. 7–12, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0716864019301191

[2] A. M. Leiva, C. Troncoso-Pantoja, M. A. Martínez-Sanguinetti, F. Petermann-Rocha, F. Poblete-Valderrama, I. Cigarroa-Cuevas, and C. Celis-Morales, “Factores asociados a caídas en adultos mayores chilenos: evidencia de la encuesta nacional de salud 2009-2010,” *Revista médica de Chile*, vol. 147, no. 7, pp. 877–886, 2019.

[3] D. Podsiadlo and S. Richardson, “The timed üp & go": A test of basic functional mobility for frail elderly persons,” *J. Am. Geriatr. Soc.*, vol. 39, pp. 142–148, 1991.

[4] V. Bonnet, S. Ramdani, C. Azevedo-Coste, P. Fraisse, C. Mazzà, and A. Cappozzo, “Integration of human walking gyroscopic data using empirical mode decomposition,” *Sensors*, vol. 14, no. 1, pp. 370–381, 2014.

[5] P. Ortega-Bastidas, P. Aqueveque, B. Gómez, F. Saavedra, and R. Cano-de-la Cuerda, “Use of a single wireless imu for the segmentation and automatic analysis of activities performed in the 3-m timed up & go test,” *Sensors (Basel, Switzerland)*, vol. 19, no. 7, p. 1647, 2019.

[6] P. Ortega-Bastidas, B. Gómez, P. Aqueveque, S. Luarte-Martínez, and R. Cano-de-la Cuerda, “Instrumented timed up and go test (itug)—more than assessing time to predict falls: A systematic review,” *Sensors*, vol. 23, no. 7, p. 3426, 2023. [Online]. Available: https://doi.org/10.3390/s23073426

[7] Ministerio de Salud de Chile, “Instructivo de aplicación del control de salud anual del adulto mayor,” Departamento de Adulto Mayor, División de Prevención y Control de Enfermedades, Ministerio de Salud de Chile, Instructivo, 2015, disponible en línea.

[Online]. Available: https://diprece.minsal.cl/wrdprss_minsal/wp-content/uploads/2015/05/instructivo-de-control-de-salud-empam.pdf

[8] E. Barry, R. Galvin, C. Keogh, F. Horgan, and T. Fahey, "Is the timed up and go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis," *BMC Geriatrics*, vol. 14, p. 14, 2014. [Online]. Available: https://bmcgeriatr.biomedcentral.com/articles/10.1186/1471-2318-14-14

[9] A. A. El-Saleh, A. M. Sheikh, M. A. M. Albreem, and M. S. Honnurvali, "The internet of medical things (iomt): opportunities and challenges," *Wireless Networks*, vol. 31, pp. 327–344, 2025. [Online]. Available: https://doi.org/10.1007/s11276-024-03764-8

[10] B. Friedrich, S. Lau, L. Elgert, J. M. Bauer, and A. Hein, "A deep learning approach for tug and sppb score prediction of (pre-) frail older adults on real-life imu data," *Healthcare*, vol. 9, no. 2, p. 149, 2021. [Online]. Available: https://doi.org/10.3390/healthcare9020149

[11] V. Roshdibenam, G. J. Jogerst, N. R. Butler, and S.-D. Baek, "Machine learning prediction of fall risk in older adults using timed up and go test kinematics," *Sensors*, vol. 21, no. 10, p. 3481, 2021. [Online]. Available: https://doi.org/10.3390/s21103481

[12] J. B. Ko, J. S. Hong, Y. S. Shin, and K. B. Kim, "Machine learning-based predicted age of the elderly on the instrumented timed up and go test and six-minute walk test," *Sensors*, vol. 22, no. 16, p. 5957, 2022. [Online]. Available: https://doi.org/10.3390/s22165957

[13] R. Intaruk, J. Saengsuwan, S. Amatachaya, and T. Thaweewannakij, "Cut-off score of the 6-minute walk test for determining risk of fall in community-dwelling elderly," *Archives of Allied Health Sciences*, vol. 32, no. 1, pp. 61–70, 2020. [Online]. Available: https://he01.tci-thaijo.org/index.php/ams/article/view/241855

[14] J. Maiora, C. Rezola-Pardo, G. García, B. Sanz, and M. Graña, "Older adult fall risk prediction with deep learning and timed up and go (tug) test data," *Bioengineering*, vol. 11, no. 10, p. 1000, 2024.

[15] M. L. McHugh, "The chi-square test of independence," *Biochemia medica*, vol. 23, no. 2, pp. 143–149, 2013.

[16] X. Wang, J. Cao, Q. Zhao, M. Chen, J. Luo, H. Wang, L. Yu, K. L. Tsui, and Y. Zhao, "Identifying sensors-based parameters associated with fall risk in community-dwelling older adults:

an investigation and interpretation of discriminatory parameters," *BMC Geriatrics*, vol. 24, no. 1, p. 125, 2024.

[17] L. D. Bogle Thorbahn and R. A. Newton, "Use of the berg balance test to predict falls in elderly persons," *Physical Therapy*, vol. 76, no. 6, pp. 576–583, 1996. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8650273/

[18] S.-P. Lee, J. S. Dufek, R. A. Hickman, and S. E. Schuerman, "Influence of procedural factors on the reliability and performance of the timed up and go test in older adults," *International Journal of Gerontology*, vol. 10, no. 1, pp. 37–42, 2016. [Online]. Available: https://doi.org/10.1016/j.ijge.2015.10.003

[19] A. Barré and S. Armand, "Biomechanical toolkit: Open-source framework to visualize and process biomechanical data," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 1, pp. 80–87, Apr. 2014. [Online]. Available: https://doi.org/10.1016/j.cmpb.2014.01.012

[20] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *arXiv preprint*, vol. arXiv:1705.07874, 2017.

[21] S. Mo and D. H. Chow, "Accuracy of three methods in gait event detection during overground running," *Gait Posture*, vol. 59, pp. 93–98, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636217309657

[22] W.-C. Hsu, T. Sugiarto, Y.-Y. Liao, Y.-J. Lin, F.-C. Yang, D.-Y. Hueng, C.-T. Sun, and K.-N. Chou, "Can trunk acceleration differentiate stroke patient gait patterns using time- and frequency-domain features?" *Applied Sciences*, vol. 11, no. 4, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/4/1541

[23] Snowflake Inc. & Streamlit Inc., "Streamlit – a faster way to build and share data apps," Jun. 2025, versión 1.46.1. [Online]. Available: https://github.com/streamlit/streamlit

[24] P. T. Inc. (2015) Collaborative data science. Montréal, QC. [Online]. Available: https://plot.ly