VOLUMEN 45
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Clasificación de patrones corporales en la prueba sit-to-stand mediante trayectorias corporales 2D

Classification of Body Patterns in Sit-to-Stand Repetitions Using 2D Body Trajectories

Jacob Sierra Lemus*[1], Elsa Fernanda Breceda Villagran[1], Paola del Rocío Reyes González[1], Angélica Hernández Rayas[1], Rafael Guzmán[2]

[1]Universidad de Guanajuato, División de ciencias e Ingenierías, Campus León, Cda. Loma del Bosque #103, Lomas del Campestre, 37150 León de los Aldama, Gto.
[2] Carretera Salamanca–Valle de Santiago km 3.5 + 1.8, Comunidad de Palo Blanco, Salamanca, Guanajuato, México.

*j.sierralemus@ugto.mx

Resumen

Sentarse y levantarse de una silla requiere coordinación, equilibrio y control del movimiento. Sin embargo, las evaluaciones habituales suelen resumir esta tarea mediante el tiempo empleado o el número de repeticiones, sin describir cómo se ejecuta cada ciclo. El objetivo de este trabajo fue desarrollar una prueba de concepto capaz de reconocer, a partir de video RGB, cinco patrones simulados de ejecución sit-to-stand. CIGPose convirtió los videos en trayectorias bidimensionales de 14 puntos corporales. Se segmentaron y revisaron 625 repeticiones de cinco participantes; posteriormente, los ciclos se normalizaron, se alinearon en descenso, transición y ascenso, y se utilizaron para entrenar modelos de clasificación. Para evaluar el cambio entre personas, cada participante fue excluido por turno del entrenamiento. La selección más estricta alcanzó 61.3 % de exactitud balanceada media por participante, mientras que el mejor modelo individual obtuvo 65.4 %. Un ensamble ajustado de manera exploratoria llegó a 75.5 % dentro del conjunto disponible. Los resultados indican que las trayectorias corporales contienen información suficiente para diferenciar los patrones estudiados y mostrar una predicción sobre el video. Si bien la prueba de concepto puede considerarse exitosa, la variación entre participantes fue la principal dificultad, por lo que el desempeño deberá confirmarse con una muestra más amplia e independiente.

Palabras clave: sit-to-stand; visión por computadora; estimación de pose; series temporales; clasificación.

Introducción

Poder levantarse de una silla sin ayuda es necesario para realizar actividades tan básicas como caminar, ir al baño, subir escaleras o cambiar de lugar dentro de casa. Aunque suele ejecutarse sin pensar, este movimiento exige coordinar varias acciones en pocos segundos. Al inicio, parte del peso del cuerpo descansa sobre la silla; después, el tronco se inclina hacia delante, el peso se transfiere progresivamente hacia los pies y las piernas impulsan el cuerpo hasta alcanzar la posición de pie. Para sentarse, la secuencia se invierte y el cuerpo debe controlar el descenso hasta recuperar el contacto con el asiento. Por ello, la transición de sentado a de pie, conocida como sit-to-stand o STS, se considera una tarea importante para estudiar la movilidad y la independencia funcional (Janssen et al., 2002; van Lummel et al., 2016).

Este movimiento tan cotidiano aparece en distintas evaluaciones empleadas en rehabilitación, fisioterapia y geriatría. En la prueba de cinco levantamientos se registra cuánto tarda una persona en completar cinco ciclos; en la prueba de 30 segundos se cuenta cuántas veces logra levantarse durante ese periodo. Estas mediciones son sencillas, económicas y fáciles de repetir. Por ejemplo, el programa STEADI de los Centros para el Control y la Prevención de Enfermedades utiliza la prueba de 30 segundos para observar la fuerza y la resistencia de las piernas como parte de la valoración del riesgo de caídas (Centers for Disease Control and Prevention, 2017).

Sin embargo, contar repeticiones o medir el tiempo total responde principalmente a cuánto hizo la persona o cuánto tardó, pero no explica cómo realizó el movimiento. Dos personas pueden completar una prueba en un tiempo parecido y utilizar estrategias diferentes: una puede inclinar más el tronco, otra puede desplazarse hacia un lado, acercar visualmente las rodillas o perder el control durante el descenso. Incluso una misma persona puede modificar su estrategia entre una repetición y la siguiente. Estas diferencias pueden quedar ocultas cuando todo el movimiento se resume en un solo número.

La instrumentación de la prueba permite estudiar esas diferencias. En lugar de registrar solamente el inicio y el final, pueden analizarse por separado el descenso, el momento cercano al asiento y el ascenso, así como la trayectoria y la velocidad de distintas partes del cuerpo. En adultos mayores, van Lummel et al. (2016) encontraron que las duraciones obtenidas mediante una prueba STS instrumentada mostraron asociaciones más claras con el estado funcional, el estado de salud y la actividad física que el tiempo registrado manualmente. También observaron que las fases dinámicas, cuando la persona realmente sube o baja, resultaban más informativas que los periodos en los que permanecía sentada o de pie. Esto no vuelve inútil al cronómetro; muestra que el análisis detallado puede aportar información que una medición global no conserva.

Tradicionalmente, un análisis detallado del movimiento requiere laboratorios con varias cámaras, marcadores reflectantes colocados sobre el cuerpo, plataformas de fuerza y personal especializado. Estos sistemas pueden alcanzar gran precisión, pero su costo, tiempo de preparación y necesidad de un espacio controlado dificultan su uso fuera del laboratorio. La estimación de pose humana ofrece una alternativa más accesible. Esta tecnología utiliza visión por computadora para localizar automáticamente puntos del cuerpo en una fotografía o en cada cuadro de un video. Así, una cámara convencional puede registrar el movimiento sin adherir sensores o marcadores a la persona. Las revisiones recientes señalan que estas herramientas tienen potencial para ampliar el análisis del movimiento, aunque su precisión depende de la cámara, la iluminación, las obstrucciones y el ángulo desde el que se observa a la persona (Avogaro et al., 2023).

Una forma de ampliar el análisis de la prueba STS consiste en convertir el video en una representación numérica del movimiento. En este proyecto se utilizó CIGPose, un modelo de estimación de pose desarrollado por Li et al. (2026). Este tipo de modelo examina cada cuadro del video y localiza puntos corporales como hombros, caderas, rodillas, tobillos y pies. Al ordenar sus posiciones a lo largo del tiempo se obtienen trayectorias que describen cómo se desplazó el cuerpo durante cada repetición.

Esta representación permite concentrar el análisis en el movimiento. En lugar de recibir directamente el rostro, la ropa y el fondo de la grabación, el clasificador trabaja con el recorrido de los puntos corporales. A partir de esta idea surgió la pregunta central del proyecto: ¿las trayectorias bidimensionales obtenidas con una cámara convencional contienen información suficiente para distinguir diferentes formas de ejecutar una repetición STS?

Para responderla se desarrolló una prueba de concepto orientada a reconocer cinco patrones visuales simulados: una ejecución sin alteración conocida, un patrón medial proyectado de rodilla, un desplazamiento hacia la izquierda, un desplazamiento hacia la derecha y un descenso abrupto. Un reto central fue comprobar si el sistema podía reconocerlos en una persona distinta de aquellas con las que había aprendido, pues cada individuo posee una velocidad, una postura y una forma particular de moverse. Con este planteamiento, el objetivo del proyecto fue desarrollar y evaluar un clasificador de patrones simulados de ejecución STS basado en trayectorias corporales bidimensionales extraídas de video RGB.

Materiales y métodos

Captura y organización de los datos

Las sesiones se registraron con una Orbbec Astra Pro utilizada como cámara RGB, con resolución de 1280 × 720 píxeles y 30 cuadros por segundo. La cámara se colocó en una vista diagonal respecto de la silla. Los participantes realizaron ciclos continuos de sentarse y levantarse con los brazos cruzados. El procesamiento se hizo después de la grabación, lo que permitió revisar con calma cada repetición sin comprometer la captura.

El conjunto final contiene 625 repeticiones de cinco participantes, identificados como P01 a P05. Se conservaron 170 repeticiones sin alteración conocida, 128 con patrón medial proyectado de rodilla, 88 con desplazamiento izquierdo, 110 con desplazamiento derecho y 129 con descenso abrupto. Todos los participantes aportaron ejemplos de las cinco categorías. La distribución no es idéntica, pero ninguna clase quedó limitada a una sola persona.

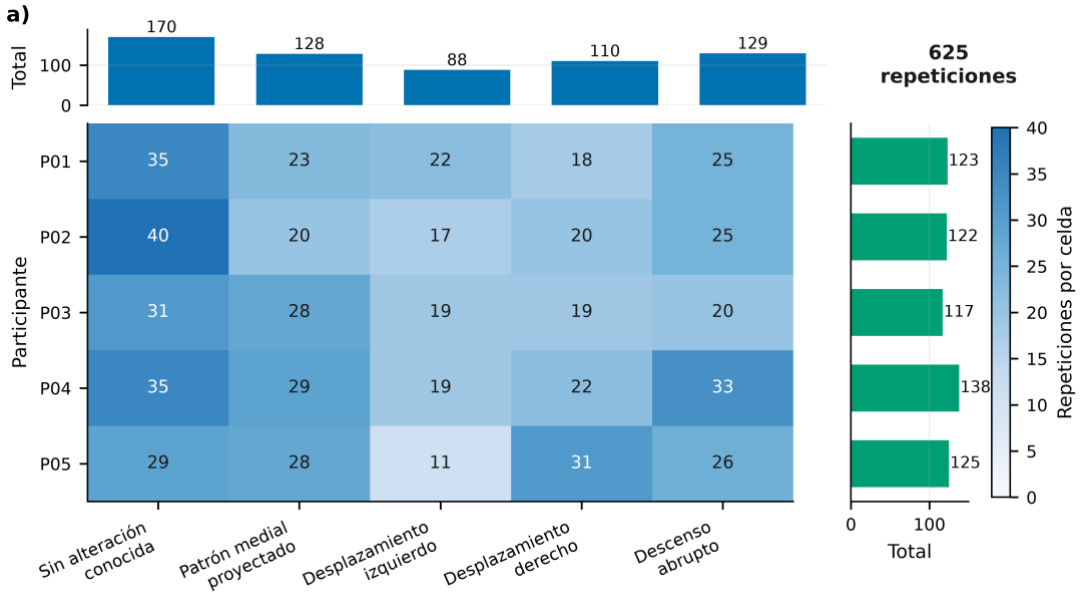


Figura 2. Composición del conjunto de datos por participante y categoría. Las celdas muestran el número de repeticiones revisadas de cada participante en las cinco categorías; las barras superiores y laterales indican los totales por categoría y por participante, respectivamente. El conjunto final incluyó 625 repeticiones de cinco participantes.

Los videos contenían varios ciclos consecutivos de sentarse y levantarse. Para utilizar cada repetición como una muestra independiente, se realizó una segmentación inicial basada en la trayectoria vertical de la pelvis. Para cada ciclo propuesto se registraron los cuadros correspondientes al inicio del movimiento, la llegada al asiento y el final de la repetición. Las repeticiones propuestas se revisaron mediante una interfaz local que mostraba simultáneamente el video RGB, los puntos corporales detectados y los límites temporales asignados. El revisor podía corregir los cuadros de inicio y final y guardar los cambios de manera acumulativa. Solo se conservaron las repeticiones con límites completos y una ejecución considerada válida durante la revisión.

La condición indicada en el nombre de cada video se utilizó únicamente como etiqueta inicial propuesta. La categoría definitiva se asignó de manera individual después de observar cada repetición en la interfaz. Se utilizaron cinco categorías mutuamente excluyentes: sin alteración conocida, patrón medial proyectado de rodilla, desplazamiento hacia la izquierda, desplazamiento hacia la derecha y descenso abrupto. Al concluir la revisión se conservaron 625 repeticiones con límites temporales y categorías definitivas.

Extracción y control de calidad de los puntos corporales.

CIGPose, propuesto por Li et al. (2026), es un modelo de estimación bidimensional de pose corporal completa que analizó cada cuadro de los videos para localizar puntos anatómicos. Se eligió porque identifica detecciones inciertas y emplea las relaciones del esqueleto para reducir errores causados por oclusiones o elementos del fondo, mostrando mayor precisión que modelos como RTMPose y DWPose. Para este proyecto se conservaron 14 puntos correspondientes a hombros, caderas, rodillas, tobillos y pies. Cada punto se representó mediante sus coordenadas horizontal y vertical y un puntaje de confianza de la detección.

Las detecciones con un puntaje inferior a 0.5 se consideraron poco confiables y se trataron como datos faltantes. Cuando el detector perdía un punto durante un máximo de cinco cuadros consecutivos en medio de una trayectoria, su posición se reconstruía mediante una interpolación lineal. Esto significa que se trazaba una transición gradual entre la última posición conocida antes del hueco y la primera posición conocida después de este. Los huecos más largos y los situados al principio o al final de una trayectoria no se rellenaron, porque no existía información suficiente a ambos lados para estimar el movimiento de manera razonable.

Normalización de la posición y el tamaño corporal

Las coordenadas originales dependían de la posición de la persona dentro de la imagen y de su tamaño aparente frente a la cámara. Para reducir estas diferencias, se utilizó como origen el punto medio entre ambos tobillos. Después, las coordenadas se dividieron entre la distancia vertical aproximada que separaba la pelvis de los tobillos. Este procedimiento trasladó todas las trayectorias a un sistema de referencia común. Así, una persona situada ligeramente más a la izquierda o más cerca de la cámara no generaba valores completamente distintos solo por su ubicación en la imagen. La normalización no eliminó las diferencias reales entre personas, pero redujo las variaciones sencillas de posición y escala.

Además de las posiciones normalizadas, se calcularon los cambios de cada punto respecto a la postura inicial de la repetición. También se obtuvieron velocidades y aceleraciones para representar no solo dónde se encontraba cada parte del cuerpo, sino cómo y con qué rapidez cambiaba su posición. Estos cálculos consideraron la duración real de cada fase del movimiento.

Normalización temporal de las repeticiones

Las repeticiones no duraban exactamente lo mismo. Una persona podía realizar el descenso lentamente y levantarse con rapidez, mientras que otra podía distribuir el tiempo de manera diferente. Comparar directamente los cuadros habría provocado que momentos distintos del movimiento ocuparan la misma posición dentro de la secuencia.

Para evitarlo, cada repetición se dividió en tres fases: descenso, transición cercana al asiento y ascenso. Posteriormente, cada fase se remuestreó por separado. El descenso se representó mediante 24 instantes, la transición mediante 12 y el ascenso mediante otros 24. De esta forma, todas las repeticiones quedaron expresadas como secuencias de 60 pasos, en las que los momentos equivalentes del movimiento aparecían en posiciones semejantes.

Filtrado y representación final

Antes de calcular la representación definitiva se aplicó un filtro Butterworth de cuarto orden, fase cero y frecuencia de corte de 4 Hz. Su función fue reducir pequeños saltos producidos por variaciones del detector entre cuadros consecutivos. La aplicación en fase cero evitó desplazar temporalmente los acontecimientos del movimiento; es decir, el suavizado no modificó el instante en que ocurrían el descenso, la llegada al asiento o el ascenso.

La representación final reunió las trayectorias normalizadas de la pelvis, el tronco, las rodillas y los tobillos. También incorporó los cambios respecto a la postura inicial, las velocidades y las aceleraciones. Cada repetición quedó así convertida en una secuencia numérica de la misma longitud, preparada para ser utilizada como entrada de los modelos de clasificación.

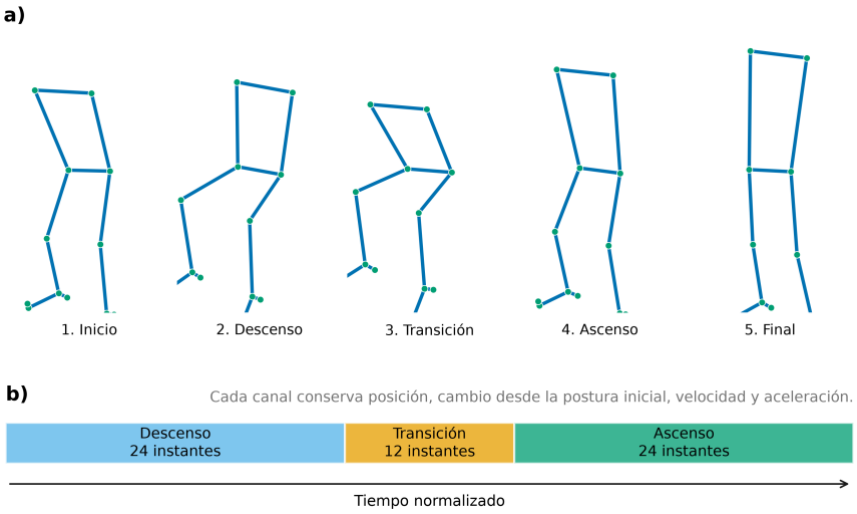


Figura 3. Normalización temporal de las repeticiones sit-to-stand. a) Instantes representativos del ciclo, desde la postura inicial hasta la final. b) Cada repetición se dividió y remuestreó en 24 instantes de descenso, 12 de transición y 24 de ascenso, para obtener secuencias comparables de 60 pasos temporales..

Tarea de clasificación y modelos evaluados

Se evaluaron dos formulaciones del problema. Como referencia inicial se construyó una clasificación binaria que separaba las ejecuciones sin alteración conocida de las cuatro condiciones simuladas agrupadas. Posteriormente, la tarea principal se formuló como una clasificación de cinco categorías: sin alteración conocida, patrón medial proyectado de rodilla, desplazamiento hacia la izquierda, desplazamiento hacia la derecha y descenso abrupto.

El primer modelo utilizado como referencia para la clasificación multiclase fue MiniROCKET. Este método aplica numerosos filtros cortos sobre una serie temporal para localizar fragmentos que ayuden a distinguir las categorías. De manera simplificada, puede imaginarse como un conjunto de pequeñas plantillas que se desplazan sobre las trayectorias y registran en qué lugares aparecen cambios potencialmente útiles (Dempster et al., 2021).

Se compararon modelos de regresión logística, máquinas de soporte vectorial, bosques aleatorios, métodos de boosting, centroides, MiniROCKET y MultiROCKET. También se evaluó una red convolucional temporal pequeña, conocida como TCN. Los modelos recibieron distintas representaciones del movimiento, entre ellas posiciones normalizadas, cambios respecto a la postura inicial, velocidades, aceleraciones y proporciones entre segmentos corporales.

MultiROCKET amplía el procedimiento de MiniROCKET al analizar tanto la trayectoria original como sus cambios entre instantes consecutivos. Además, utiliza diferentes formas de resumir la respuesta de cada filtro (Tan et al., 2022). En las configuraciones evaluadas se utilizaron hasta 5000 kernels. Un kernel es un patrón numérico corto que se desplaza sobre la señal para detectar formas locales. El uso de miles de kernels permite examinar una gran variedad de cambios sin entrenar una red neuronal profunda de gran tamaño.

Aumento de datos

Debido al reducido número de participantes, se probaron transformaciones destinadas a aumentar la variedad de las trayectorias disponibles durante el entrenamiento. El espejado invirtió horizontalmente el movimiento, intercambió los puntos corporales izquierdos y derechos y modificó la etiqueta cuando correspondía a un desplazamiento lateral. La deformación temporal aceleró o ralentizó ligeramente distintas partes de la repetición sin alterar el orden de las fases.

También se utilizó una mezcla de trayectorias. Esta transformación combinó dos repeticiones de la misma categoría procedentes de participantes diferentes para crear una secuencia intermedia. Todas las transformaciones se aplicaron exclusivamente a las repeticiones utilizadas para entrenar el modelo en cada ronda. Las muestras reservadas para la evaluación no se modificaron ni se utilizaron para generar trayectorias sintéticas.

Evaluación por participante

Las 625 repeticiones no se dividieron aleatoriamente. Una separación al azar habría colocado movimientos de una misma persona tanto en el entrenamiento como en la evaluación, lo que permitiría que el modelo aprovechara características personales como la velocidad habitual, la estatura aparente o la forma de inclinarse. Para evitarlo se utilizó una evaluación leave-one-participant-out, abreviada LOPO. El procedimiento se organizó en cinco rondas. En cada una se apartaron todas las repeticiones de un participante, se entrenó el modelo con las otras cuatro personas y después se evaluó con el bloque reservado. De esta manera, cada repetición fue clasificada por un modelo que no había observado durante el entrenamiento a la persona que la ejecutó.

Cuando se realizó selección anidada, la representación, el modelo y sus parámetros se eligieron utilizando únicamente a los participantes disponibles para el entrenamiento de cada ronda. El participante apartado no intervino en esas decisiones.

Métricas de evaluación

La medida principal fue la exactitud balanceada. Esta métrica calcula por separado qué proporción de cada categoría fue reconocida correctamente y después promedia los cinco resultados. Así, todas las categorías tienen el mismo peso, aunque no contengan exactamente el mismo número de repeticiones.

También se registraron el porcentaje general de aciertos, la puntuación F1 macro, la sensibilidad de cada categoría, los resultados de cada participante apartado y las matrices de confusión. Estas últimas muestran qué categorías fueron confundidas entre sí.

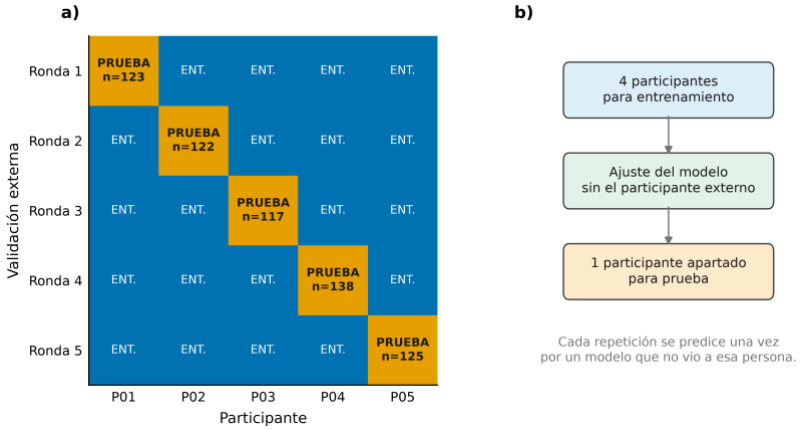


Figura 4. Evaluación leave-one-participant-out (LOPO). a) En cada una de las cinco rondas, todas las repeticiones de un participante se reservaron para prueba y las de los cuatro restantes se utilizaron para entrenamiento. b) El modelo se ajustó sin información del participante apartado, de modo que cada una de las 625 repeticiones fue evaluada una vez por un modelo que no había observado a su ejecutante. ENT.: entrenamiento.

Se usaron dos niveles de comparación. En el nivel exploratorio se revisaron muchas configuraciones usando las cinco rondas LOPO. Esto permite mejorar el prototipo, pero introduce sesgo porque las mismas cinco personas guían la elección. Para obtener una referencia más prudente se aplicó selección anidada: dentro de cada ronda externa, la combinación se decidió solo con las cuatro personas de entrenamiento, sin consultar a la persona reservada.

Resultados

El MiniROCKET multiclase inicial obtuvo 57.1 % de aciertos directos y 55.9 % de exactitud balanceada media por participante. La exactitud balanceada varió entre participantes: alcanzó 76.1 % para P03 y disminuyó hasta 31.1 % para P05. El mejor modelo individual fue MultiROCKET con la representación completa y mezcla de repeticiones entre participantes de la misma clase. Alcanzó 66.4 % de aciertos, 64.6 % de exactitud balanceada global y 65.4 % de exactitud balanceada media por participante.

Evaluación estricta y ensamble exploratorio

La selección anidada alcanzó 61.3 % de exactitud balanceada media por participante. En este procedimiento, el participante apartado no intervino en la selección del modelo ni en el ajuste de sus parámetros. Después se construyó un ensamble, es decir, una combinación de modelos que observan el mismo movimiento desde perspectivas diferentes. Sus cinco componentes fueron el MiniROCKET inicial, el mejor MultiROCKET, un MiniROCKET compacto con espejado y mezcla, cinco detectores independientes y la TCN.

Los pesos del ensamble final se ajustaron después de observar las predicciones LOPO de las cinco personas. En ese conjunto cerrado obtuvo 72.5 % de aciertos y 75.5 % de exactitud balanceada media por participante.

Sistema	Aciertos	Exactitud balanceada global	Promedio balanceado por participante
MiniROCKET inicial	57.1 %	54.8 %	55.9 %
Mejor modelo individual	66.4 %	64.6 %	65.4 %
Selección anidada estricta	59.4 %	60.8 %	61.3 %
Ensamble final exploratorio	72.5 %	74.1 %	75.5 %

Tabla 1. Comparación de los cuatro resultados principales.

La exactitud balanceada da el mismo peso a las cinco categorías. La última fila es exploratoria debido a que los pesos del ensamble se ajustaron después de observar las cinco rondas externas.

VOLUMEN 45
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

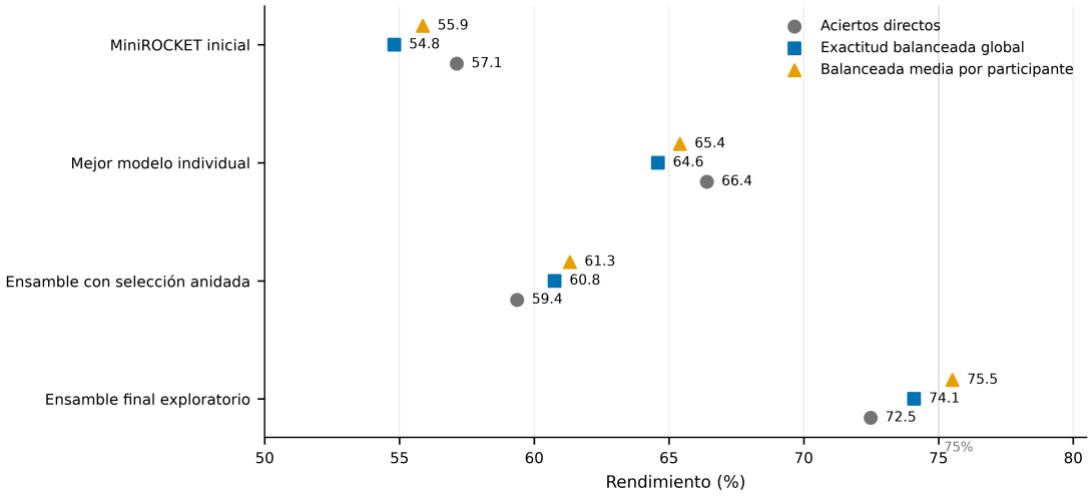


Figura 5. Desempeño de los sistemas evaluados mediante validación LOPO. Se presentan el porcentaje de aciertos, la exactitud balanceada calculada sobre las 625 predicciones agrupadas y el promedio de la exactitud balanceada de los cinco participantes. La selección anidada proporciona la estimación más estricta, mientras que el resultado del ensamble final es exploratorio porque sus pesos se ajustaron después de observar las cinco rondas externas.

La matriz de confusión permite ver algo que un solo porcentaje oculta. Cada fila corresponde a la categoría real y cada columna a la predicción. Una diagonal intensa significa que el modelo acertó; los valores fuera de la diagonal muestran hacia dónde se desviaron los errores. En el ensamble exploratorio, el patrón medial proyectado de rodilla alcanzó 93.0 % de sensibilidad y el descenso abrupto 85.3 %. El desplazamiento izquierdo obtuvo 76.1 % y el derecho 67.3 %.

La categoría más difícil fue «sin alteración conocida», con 48.8 % de sensibilidad. El sistema tendió a asignar alguna de las cuatro alteraciones cuando la repetición no tenía ninguna de las etiquetas conocidas. Esto revela una compensación importante: el ensamble mejoró el reconocimiento de los patrones provocados, pero redujo la sensibilidad de la clase de referencia, que en el MiniROCKET inicial había sido de 71.2 %.

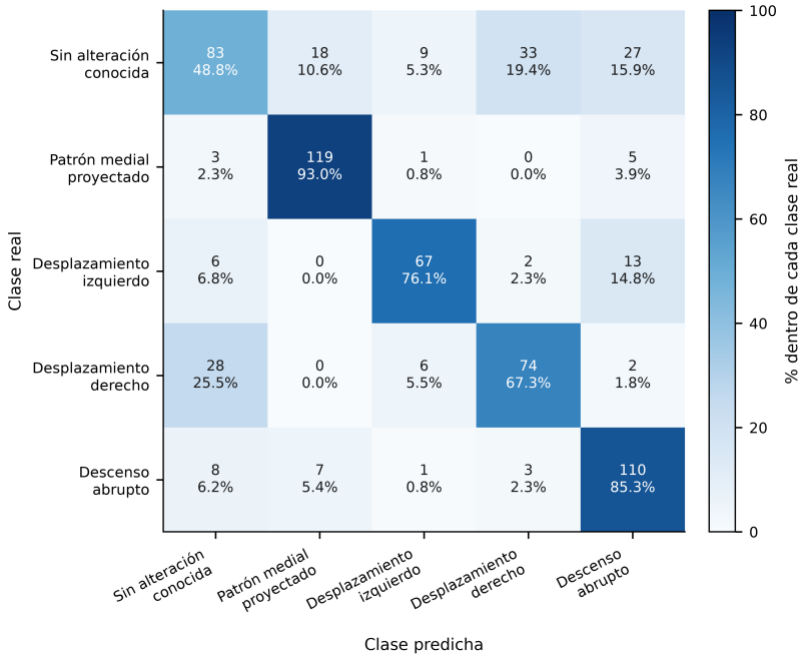


Figura 6. Matriz de confusión del ensamble final exploratorio en las 625 predicciones LOPO. Las filas representan las categorías reales y las columnas las categorías predichas. Cada celda muestra el número de repeticiones y su porcentaje respecto del total de la categoría real; la diagonal corresponde a las clasificaciones correctas. Este resultado describe el desempeño interno del ensamble y no una evaluación independiente..

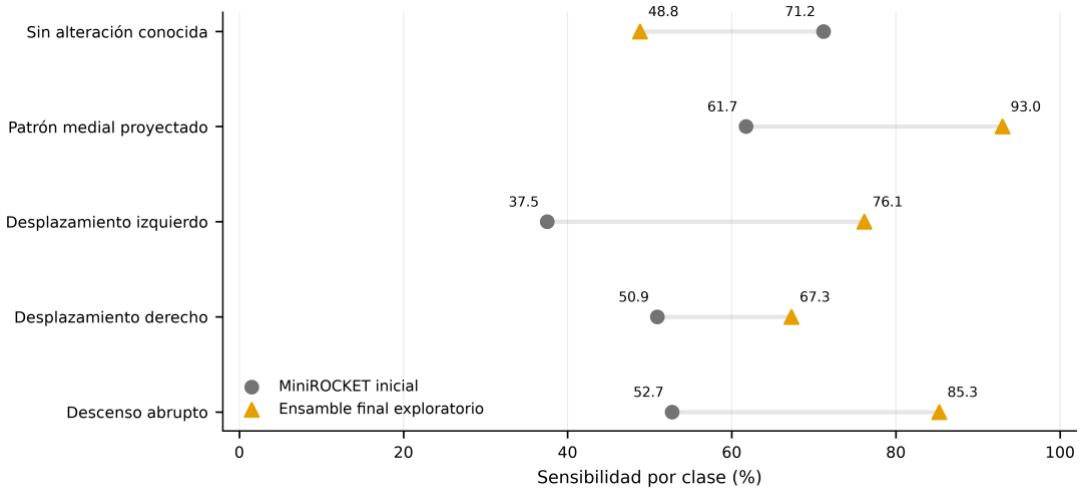


Figura 7. Sensibilidad por categoría del MiniROCKET inicial y del ensamble final exploratorio. La sensibilidad corresponde a la proporción de repeticiones correctamente reconocidas dentro de cada categoría. Los círculos representan el modelo inicial y los triángulos el ensamble; las líneas únicamente facilitan la comparación y no representan intervalos de incertidumbre. Los tamaños de clase fueron 170, 128, 88, 110 y 129 repeticiones, respectivamente.

La variación entre personas

El participante P03 presentó la exactitud balanceada más alta: 80.1 % con el mejor modelo individual y 85.7% con el ensamble exploratorio. P05 presentó los valores más bajos, con 45.5 % y 66.2 %, respectivamente.

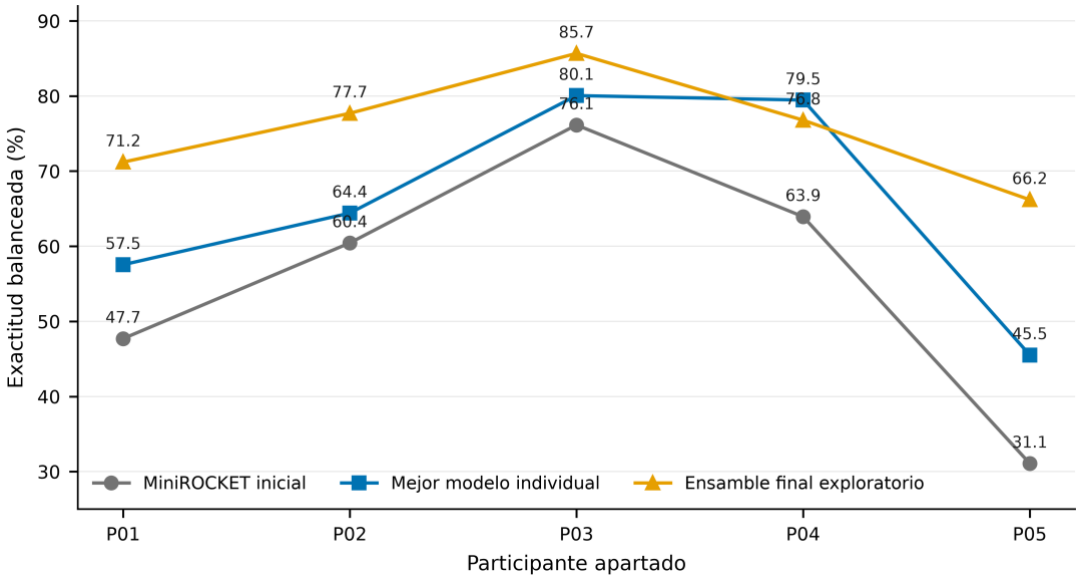


Figura 8. Variación de la exactitud balanceada entre participantes no observados durante el entrenamiento. Cada punto corresponde al resultado de una ronda LOPO para P01, P02, P03, P04 o P05, con 123, 122, 117, 138 y 125 repeticiones, respectivamente. Las líneas facilitan la comparación entre modelos y no representan una secuencia temporal ni intervalos de incertidumbre; el ensamble final conserva su carácter exploratorio.

DISCUSIÓN

La mejora más clara no provino de hacer el modelo más grande, sino de representar mejor el problema. Separar cada repetición en descenso, transición y ascenso permitió comparar momentos equivalentes entre movimientos de distinta duración. Añadir los cambios respecto a la postura inicial, las velocidades y las aceleraciones permitió describir no solo dónde se encontraban los puntos corporales, sino cómo llegaban a cada posición. Esto es relevante porque los estudios de STS instrumentado han encontrado que las fases dinámicas pueden conservar información que se pierde al resumir toda la prueba mediante un solo tiempo (van Lummel et al., 2016).

La representación completa produjo el mejor modelo individual cuando se combinó con la mezcla de repeticiones de una misma categoría entre participantes. Este resultado sugiere que conservar información sobre posición y cambio temporal fue más útil que aumentar únicamente la capacidad del algoritmo. La mezcla también obligó al modelo a trabajar con trayectorias procedentes de personas diferentes, en lugar de depender exclusivamente de la forma de movimiento de un solo participante.

También fue acertado mantener la evaluación por persona. Un reparto aleatorio habría colocado movimientos de un mismo individuo tanto en entrenamiento como en evaluación y habría permitido que el modelo aprovechara características personales. La evaluación LOPO mostró desde el inicio que el desempeño cambiaba considerablemente entre participantes. La exactitud balanceada, por su parte, hizo visible que una mejora general podía ocultar pérdidas en una categoría específica.

La revisión manual de la segmentación y del etiquetado fue otro elemento importante. La segmentación automática basada en el movimiento vertical de la pelvis permitió proponer los ciclos, pero algunos límites finales quedaban adelantados. Durante la revisión fue frecuente añadir alrededor de diez cuadros al final para conservar el ascenso completo. Esta experiencia mostró que la automatización podía reducir el trabajo inicial, pero todavía necesitaba una corrección humana para evitar que una repetición incompleta llegara al entrenamiento.

La revisión también permitió corregir la forma de asignar las categorías. Al inicio se utilizó como propuesta la condición indicada en el nombre del video. Sin embargo, una instrucción dada para todo un bloque no garantizaba que cada repetición se hubiera ejecutado de la misma manera. La decisión final fue considerar cada repetición revisada como la unidad de análisis. De esta forma, los nombres de los archivos se utilizaron para orientar la revisión, pero no sustituyeron la observación del movimiento.

Decisiones que no funcionaron como se esperaba

El primer error de enfoque fue reducir el problema a una comparación entre ejecuciones “sin alteración” y “alteradas”. Aunque esta formulación parecía más sencilla, reunía cuatro movimientos diferentes dentro de una misma categoría y eliminaba precisamente la información que se esperaba mostrar al usuario. El resultado cercano a 53.7 % de exactitud balanceada confirmó que continuar entrenando el mismo clasificador binario no resolvería el problema. La corrección consistió en reformular la tarea para reconocer cinco categorías y revisar las etiquetas de cada repetición.

Otra lección fue que aumentar la complejidad no produce automáticamente una mejor capacidad para reconocer a una persona distinta. Incrementar el número de kernels, repetir el entrenamiento con más semillas y añadir proporciones corporales generó nuevas configuraciones, pero no eliminó la variación entre participantes. Los árboles, las máquinas de soporte vectorial y los modelos basados únicamente en variables resumidas tampoco superaron al mejor modelo de series temporales.

La red convolucional temporal o TCN tampoco destacó como modelo independiente. Sin embargo, sus errores no coincidieron por completo con los de MiniROCKET y MultiROCKET. Esta diferencia permitió utilizarla como un componente complementario del ensamble: un modelo puede aportar información al conjunto incluso cuando su rendimiento individual no es el más alto. El intento de ocultar la identidad de los participantes mediante aprendizaje adversarial produjo un resultado todavía menor. La caída sugiere que, al intentar retirar las características personales, también se eliminó parte de la información necesaria para reconocer los patrones.

El aumento de datos ayudó solamente cuando conservó una estructura plausible del movimiento. Copiar repeticiones no añadía variedad y deformarlas demasiado podía crear velocidades o posiciones poco realistas. La mezcla entre participantes de una misma categoría fue la transformación más útil para el mejor modelo individual. El espejado y la deformación temporal no fueron suficientes por sí solos, aunque contribuyeron a crear componentes con errores diferentes dentro del ensamble.

Por qué el resultado más alto no es la única cifra importante

Los resultados de 75.5 % y 61.3 % no se contradicen, porque proceden de procesos de selección diferentes. El ensamble exploratorio alcanzó 75.5 % después de ajustar sus pesos al observar las predicciones LOPO de los cinco participantes. Cada modelo base había dejado fuera a la persona evaluada, pero la combinación final sí se eligió después de conocer el comportamiento de las cinco rondas.

En la selección anidada, en cambio, la persona apartada tampoco intervino en la elección de los modelos ni en el ajuste de sus pesos. Este procedimiento obtuvo 61.3 % de exactitud balanceada media por participante. La cifra es menor, pero representa mejor lo que podría esperarse si se tuviera que seleccionar el sistema antes de observar a una persona nueva.

El mejor modelo individual alcanzó 65.4 %, aunque también fue identificado después de comparar varias configuraciones sobre las cinco rondas disponibles. Por ello, los tres valores cumplen funciones diferentes. El 65.4 % muestra hasta dónde llegó un solo modelo; el 61.3 % representa la evaluación más estricta; y el 75.5 % describe el mejor ensamble exploratorio construido para el conjunto cerrado.

Esta diferencia es especialmente importante cuando existen pocos participantes. Si se prueban muchas configuraciones con las mismas personas, alguna puede adaptarse particularmente bien a sus movimientos. El resultado no deja de ser útil para construir el prototipo, pero no debe confundirse con una medición independiente de lo que ocurriría con personas futuras.

Qué nos dicen los errores por categoría

El ensamble exploratorio mejoró la sensibilidad de cuatro patrones respecto al MiniROCKET inicial. El patrón medial proyectado de rodilla aumentó de 61.7 % a 93.0 %; el desplazamiento izquierdo, de 37.5 % a 76.1 %; el desplazamiento derecho, de 50.9 % a 67.3 %; y el descenso abrupto, de 52.7 % a 85.3 %. En contraste, la sensibilidad de la categoría sin alteración conocida disminuyó de 71.2 % a 48.8 %.

Esta compensación ayuda a explicar el aumento del promedio general. El ensamble reconoció mejor los cuatro patrones provocados, pero comenzó a asignarlos con mayor frecuencia a repeticiones de referencia. La matriz de confusión mostró especialmente una tendencia a clasificar ejecuciones sin alteración conocida como desplazamientos hacia la derecha o descensos abruptos.

El patrón medial y el descenso abrupto produjeron trayectorias relativamente características. El primero modifica la relación proyectada entre las rodillas y los tobillos, mientras que el segundo cambia la velocidad vertical cerca del asiento. Los desplazamientos laterales fueron más variables y su apariencia dependió en mayor medida de la posición de la persona frente a la cámara.

La categoría sin alteración conocida presenta una dificultad diferente. No describe un único movimiento, sino todas las ejecuciones en las que no se observó ninguno de los cuatro patrones etiquetados. Por ello puede reunir varias maneras de inclinarse, colocar los pies, controlar el descenso o iniciar el ascenso. Es una categoría definida por la ausencia de los otros patrones y, en consecuencia, puede ser más heterogénea que las demás.

La variación entre personas

El desempeño por participante confirmó que el promedio no describe por completo el comportamiento del sistema. P03 obtuvo los resultados más altos durante las principales etapas del desarrollo, mientras que P05 presentó los valores más bajos. El mejor modelo individual alcanzó 80.1 % para P03 y 45.5 % para P05. Con el ensamble exploratorio, estos valores cambiaron a 85.7 % y 66.2 %, respectivamente.

P05 también había requerido mayor atención durante el etiquetado. Sin embargo, su menor desempeño no puede atribuirse únicamente a ese proceso. Sus repeticiones pudieron diferir en velocidad, postura, amplitud o forma de representar las categorías respecto a las otras cuatro personas. Con los datos disponibles no es posible separar cuánto de la dificultad provino del etiquetado y cuánto de un estilo de movimiento menos representado durante el entrenamiento.

El ensamble tampoco mejoró de igual manera a todos los participantes. Para P04, la exactitud balanceada disminuyó de 79.5 % con el mejor modelo individual a 76.8 % con el ensamble exploratorio. Esto muestra que una combinación optimizada para mejorar el comportamiento global puede beneficiar algunos bloques y perjudicar ligeramente otros.

La diferencia entre participantes es una de las conclusiones principales del experimento. Con cinco personas, cada participante representa una quinta parte de las unidades independientes disponibles. Una postura o una estrategia de movimiento poco común puede modificar considerablemente el promedio. Por ello, la gráfica por participante resulta tan importante como el resultado agregado.

Influencia de la cámara

La vista diagonal permitió observar simultáneamente parte del movimiento frontal y lateral, pero también introdujo ambigüedad. Un desplazamiento puede parecer reducido o exagerado según la profundidad y la posición de la persona respecto a la cámara. La normalización disminuyó parte de este efecto, pero una imagen bidimensional no conserva toda la geometría tridimensional del movimiento.

Esta limitación coincide con lo reportado para los sistemas de estimación de pose sin marcadores, cuyo desempeño depende del ángulo de observación y de la configuración de captura (Avogaro et al., 2023; Yeung et al., 2021). El clasificador aprendió los patrones tal como se proyectaron en el montaje utilizado. Cambiar la orientación de la cámara, la distancia o la silla requeriría comprobar nuevamente el procesamiento y el modelo.

El producto final conserva la idea que motivó el trabajo: mostrar el video analizado y colocar en una esquina una etiqueta pequeña con la categoría seleccionada y el puntaje del modelo. El video permite observar la repetición, pero sus píxeles no entran directamente al clasificador. Internamente, el sistema utiliza los puntos corporales, aplica la misma normalización y alineación temporal del entrenamiento y consulta los componentes del ensamble.

El puntaje mostrado expresa qué tan dominante fue una categoría frente a las demás dentro del modelo. Su utilidad principal es hacer visible la decisión del clasificador y permitir que el usuario la compare con el movimiento reproducido. La salida reúne así las dos partes del proyecto: la representación numérica utilizada por el modelo y el video que permite interpretar visualmente la predicción.

Alcance y siguientes pasos

El principal límite del proyecto ya no parece ser la falta de arquitecturas por probar, sino la diversidad de los datos. El conjunto contiene cinco participantes, una configuración de cámara y patrones ejecutados de manera intencional. Estas condiciones permitieron construir una demostración controlada, pero no muestran qué ocurriría con una variedad mayor de cuerpos, ropa, velocidades, sillas y ángulos de observación.

Si fuera posible repetir el montaje, el siguiente avance más informativo sería registrar más participantes y una segunda sesión de cada persona. Esto permitiría distinguir entre reconocer nuevas repeticiones de una sesión conocida, reconocer otra sesión del mismo individuo y reconocer a una persona completamente nueva. También sería conveniente ampliar las ejecuciones de referencia, pues esta fue la categoría más heterogénea y la que perdió mayor sensibilidad con el ensamble.

Conclusiones

Las trayectorias corporales 2D obtenidas de video contienen información suficiente para distinguir los cinco patrones simulados de sit-to-stand en una persona que no participó en el entrenamiento. La evaluación más estricta alcanzó 61.3 % de exactitud balanceada; el mejor modelo individual obtuvo 65.4 %, y el ensamble exploratorio llegó a 75.5 %. Este último resultado muestra el potencial de combinar modelos complementarios, aunque deberá confirmarse con nuevos participantes.

La principal aportación del proyecto fue integrar la segmentación revisada de las repeticiones, la representación temporal del movimiento y la clasificación en un sistema que muestra su predicción sobre el video. Los resultados indican que organizar adecuadamente las trayectorias fue más importante que aumentar por sí sola la complejidad de los modelos. Ampliar la diversidad de participantes y realizar una evaluación externa permitirá determinar hasta dónde puede generalizarse esta prueba de concepto.

Bibliografía/Referencias

Avogaro, A., Cunico, F., Rosenhahn, B., & Setti, F. (2023). Markerless human pose estimation for biomedical applications: A survey. *Frontiers in Computer Science, 5*, Article 1153160. https://doi.org/10.3389/fcomp.2023.1153160

Centers for Disease Control and Prevention. (2017). *30-second chair stand*. https://www.cdc.gov/steadi/media/pdfs/STEADI-Assessment-30Sec-508.pdf

Dempster, A., Schmidt, D. F., & Webb, G. I. (2021). MiniRocket: A very fast (almost) deterministic transform for time series classification. In J. Tang & T. Derr (Eds.), *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 248–257). Association for Computing Machinery. https://doi.org/10.1145/3447548.3467231

Janssen, W. G. M., Bussmann, H. B. J., & Stam, H. J. (2002). Determinants of the sit-to-stand movement: A review. *Physical Therapy, 82*(9), 866–879. https://doi.org/10.1093/ptj/82.9.866

Leightley, D., & Yap, M. H. (2018). Digital analysis of sit-to-stand in masters athletes, healthy old people, and young adults using a depth sensor. *Healthcare, 6*(1), Article 21. https://doi.org/10.3390/healthcare6010021

Li, B., Cao, Z., Li, H., & Guo, Y. (2026). CIGPose: Causal intervention graph neural network for whole-body pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 23484–23494). IEEE. https://openaccess.thecvf.com/content/CVPR2026/html/Li_CIGPose_Causal_Intervention_Graph_Neural_Network_for_Whole-Body_Pose_Estimation_CVPR_2026_paper.html

Li, B., Cao, Z., Li, H., & Guo, Y. (2026). *CIGPose: Causal Intervention Graph Neural Network for Whole-Body Pose Estimation*. arXiv:2603.09418. https://doi.org/10.48550/arXiv.2603.09418

Röhling, H. M., Otte, K., Rekers, S., Finke, C., Rust, R., Dorsch, E.-M., Behnia, B., Paul, F., & Schmitz-Hübsch, T. (2022). RGB-depth camera-based assessment of motor capacity: Normative data for six standardized motor tasks. *International Journal of Environmental Research and Public Health, 19*(24), Article 16989. https://doi.org/10.3390/ijerph192416989

Tan, C. W., Dempster, A., Bergmeir, C., & Webb, G. I. (2022). MultiRocket: Multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery, 36*, 1623–1646. https://doi.org/10.1007/s10618-022-00844-1

van Lummel, R. C., Walgaard, S., Maier, A. B., Ainsworth, E., Beek, P. J., & van Dieën, J. H. (2016). The instrumented sit-to-stand test (iSTS) has greater clinical relevance than the manually recorded sit-to-stand test in older adults. *PLOS ONE, 11*(7), Article e0157968. https://doi.org/10.1371/journal.pone.0157968

Yeung, L.-F., Yang, Z., Cheng, K. C.-C., Du, D., & Tong, R. K.-Y. (2021). Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2. *Gait & Posture, 87*, 19–26. https://doi.org/10.1016/j.gaitpost.2021.04.005