

# Aplicación de estimación de pose para la calidad del movimiento en sentadilla búlgara

Juan Jose Nuñez, Juan Jose Castro

*Programa Ingeniería de sistemas  
Universidad de san buenaventura, Cali, Colombia.*

---

## Abstract

---

La sentadilla Búlgara es un ejercicio de piernas unilateral de fuerza ampliamente utilizado por sus beneficios en el fortalecimiento de los músculos de la pierna y la mejora del equilibrio. No obstante, una ejecución incorrecta puede conllevar un riesgo significativo de lesiones en rodillas y espalda. Mantener una postura adecuada con el torso erguido y la rodilla alineada con el pie, resulta fundamental para maximizar los beneficios y reducir la probabilidad de lesiones. Sin embargo, muchos practicantes carecen de retroalimentación experta en tiempo real, debido a la limitada disponibilidad de entrenadores especializados o al aumento del entrenamiento autónomo mediante recursos digitales sin supervisión directa

**Objetivo y propuesta.** Se propone un sistema automático de corrección postural para la sentadilla búlgara basado únicamente en visión por computadora, con el propósito de ofrecer una retroalimentación técnica inmediata que contribuye a la mejora del rendimiento y la prevención de lesiones. El sistema utilizara una cámara RGB para captura la ejecución del ejercicio. Además, usamos un modelo de *pose estimación* (MediaPipe Pose) que extrae 33 puntos clave por fotograma. Con estas secuencias de *landmarks* se entrena un modelo secuencial Bi-GRU que analiza cada repetición completa, detecta desviaciones frecuentes como la inclinación excesiva del tronco, el valgo de rodilla o la profundidad insuficiente y emite retroalimentación correctiva en tiempo real.

**Resultados esperados.** Este trabajo busca tanto como una reducción y clasificación de errores técnicos y retroalimentación inmediata por repetición, con mejora de la técnica, reducción del riesgo de lesiones y aumento de la eficacia del entrenamiento, demostrando el potencial de soluciones asequibles basadas en visión por computadora para asistencia fuera del laboratorio.

**Keywords:** pose estimation; evaluación de movimiento; Bulgarian split squat; retroalimentación en tiempo real; rehabilitación.

## 1. Introduction

La sentadilla búlgara es un ejercicio unilateral clave para fortalecer el tren inferior y mejorar el equilibrio del core. Su correcta ejecución exige un control postural, dado como la inclinación excesiva del torso, la desalineación de la rodilla adelantada (valgo) o la colocación inadecuada del pie posterior pueden reducir la eficacia del ejercicio. Sin embargo, suele ser necesaria la supervisión experta, la cual no siempre es accesible. Este escenario motiva el desarrollo de sistemas automatizados que asistan al usuario durante la práctica del ejercicio.

En los últimos años se han desarrollado ideas tecnológicas que aprovechan la visión por computadora para evaluar posturas en ejercicios. Por ejemplo, se han presentado aplicativos para detección y medición de la postura en el curl de bíceps utilizando visión por computadora [1]. Mientras que en el ámbito deportivo, herramientas de videoanálisis como Kinovea que permiten la medición objetiva de gestos técnicos [2]. Con los avances en inteligencia artificial, han surgido modelos capaces proporcionar retroalimentación en tiempo real basados en estimación de pose [3][4]. La precisión de algoritmos 2D como MediaPipe Pose ha sido estudiada en ejercicios de fisioterapia, permitiendo rastrear

decenas de puntos clave en tiempo real y proveer retroalimentación inmediata [5]. Asimismo, existen métodos de estimación 3D con redes profundas que logran ángulos articulares cercanos a mediciones reales [6].

Asimismo, se han propuesto sistemas integrales de monitoreo del desempeño físico. Mennella et al. presentaron un sistema de aprendizaje profundo para evaluar diversos ejercicios en el hogar [7]. Mientras que entornos clínicos, se han validado algoritmos de pose 2D para pruebas clínicas como el Single Limb Stance [8]. Otras líneas de trabajo exploran arquitecturas espaciotemporales que están basadas en Transformers para valorar la ejecución de ejercicios [9] o técnicas robustas de segmentación con modelos como Swin-UNet en escenarios con oclusiones [10]. En este contexto, las plataformas basadas en OnePose y OnePose permiten evaluar las posturas de yoga o entrena funcionales mediante análisis de angular y segmentación [11][12]

Revisiones recientes sintetizan los marcos computacionales para evaluación de ejercicios y el uso de cámaras RGB/RGB-D en monitoreo musculoesquelético y entrenamiento asistido [13].

No obstante, la mayoría de estas propuestas presentan limitaciones específicas cuando se aplican a ejercicios unilaterales o con alto componente de estabilidad, como la sentadilla bulgara. En particular, este se enfoca en movimientos bilaterales o simétricos, lo que dificultaba la detección de desalineación laterales, también empleando conjuntos de datos generados o con una escasa anotación de errores técnicos específicos, además las pocas investigaciones integran una evaluación multietiquetada, que será capaz de reconocer simultáneamente distintos tipos de fallas posturales y otra limitación que se es que los modelos más precisos requieren una cámara de profundidad o algunas que otra configuración compleja, esto hace que existan muchas más limitaciones a la hora de aplicarlo en entornos reales o de entrenamiento.

Para darle frente a todas estas limitaciones se presenta este trabajo que propone el diseño de un sistema de corrección postural que está basado en inteligencia artificial, que se orientó al análisis

automático de la sentadilla bulgara mediante visión RGB y estimación 2D de pose. Para darle solución el sistema integrará un modelo secuencial (Bi-GRU), que tendrá la capacidad de analizar series temporales de landmarks y características angulares para detectar en tiempo real errores de ejecución, como la inclinación del tronco, el valgo de la rodilla y la profundidad insuficiente que se realiza

En síntesis, las contribuciones de este estudio es el uso de conjuntos de datos específicos de la sentadilla bulgara con anotaciones de errores técnicos, la formulación de un modelo secuencial ligero que permitirá la inferencia en tiempo real, también la incorporación de características geométricas interpretables, que aumenta la precisión y transparencia del sistema

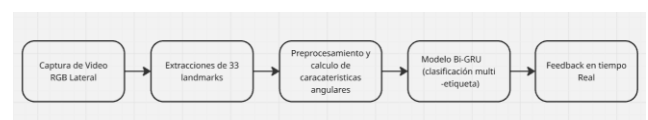
Finalmente, este documento se estructura con la segunda sección que se describe toda la metodología empleada, incluyendo la sección de conjuntos de datos, el preprocesamiento y la arquitectura empleada. La tercera sección se presentan los resultados experimentales y un análisis. En la última sección se discuten las principales implicaciones y limitaciones del sistema

## 2. Metodología

Este apartado presenta un enfoque experimental claro y reproducible, que abarca la selección del conjunto de datos, su contenido y preprocesamiento, la arquitectura propuesta con detalle de capas e hiperparámetros, y el protocolo de entrenamiento y evaluación. La metodología se estructura en varias etapas, como se describe a continuación.

### 2.1 Visión general del sistema.

El sistema consta de cuatro etapas: (i) captura RGB lateral (o ángulo óptimo) del ejercicio; (ii) extracción por fotograma de 33 landmarks corporales con MediaPipe Pose (posición 2D y visibilidad) [6]; (iii) construcción de secuencias por repetición y extracción de características geométricas; y (iv) clasificación de calidad y feedback en tiempo real mediante un modelo Bi-GRU.



## 2.2 Conjunto de datos.

Se empleó el conjunto de datos público **Dataset\_csv\_with\_data\_augmentation** (Kaggle, 2023), específicamente seleccionado por su enfoque en ejercicios de sentadilla búlgara y por la disponibilidad de secuencias temporales etiquetadas que registran la ejecución completa del movimiento.

El dataset incluye 2480 repeticiones correspondientes a 62 sujetos, con un total de 148000 frames. Todas las clases están balanceadas entre la ejecución correcta y los tres tipos de errores técnicos.

Clase	Descripción	Cantidad de repeticiones	Proporción
E1	Tronco inclinado	640	25.8
E2	Valgo de rodilla	610	24.6
E3	Profundidad insuficiente	600	24.2
Correcta	Ejecución correcta	630	25.4

Esta elección se fundamenta en que, a diferencia de otros conjuntos de datos de *pose estimation* más generales orientados a actividades múltiples o gestos cotidianos, este dataset proporciona **anotaciones precisas sobre errores técnicos propios del ejercicio**, lo que resulta esencial para entrenar un modelo de corrección postural especializado.

El conjunto incluye coordenadas de *landmarks* obtenidas mediante **MediaPipe Pose** (Google Research, 2020), normalizadas y complementadas con técnicas de aumento de datos a través de **Albumentations** (Buslaev et al., 2020), lo que amplía la variabilidad visual y mejora la capacidad de generalización del modelo ante diferentes condiciones de captura. Cada muestra contiene las coordenadas ( $x$ ,  $y$ ) de 33 puntos clave por fotograma, etiquetas binarias que indican la correcta o incorrecta ejecución de la repetición, y metadatos asociados al número de frames, video

de origen y tipo de error técnico.

Frame_id	landmark	x	y	visibilidad	Etiquetas_tronco	Etiquetas_valgo	Etiquetas_profundidad
001	Rodilla_derecha	0.455	0.732	0.96	0	1	0

En el ejemplo se muestran las características que hacen que el *Dataset\_csv\_with\_data\_augmentation* sea una base de datos adecuada para tareas de análisis postural en ejercicios unilaterales, garantizando suficiente diversidad, precisión y representatividad para el entrenamiento de modelos secuenciales orientados a la detección automática de desviaciones en tiempo real.

## 2.3. Preprocesamiento y segmentación de repeticiones.

**Filtrado de frames:** Se excluyeron frames de mala calidad donde la visibilidad global de la pose fuera baja (p.ej., aquellos con < 80% de landmarks válidos detectados) o con ruido extremo (jitter) en las coordenadas.

**Normalización:** Las coordenadas de los landmarks se centraron y escalan utilizando la media y desviación estándar calculadas sobre el conjunto de entrenamiento (por cada coordenada), para asegurar invariancia a la escala y posición de la persona en el cuadro.

$$\tilde{x}_i = \frac{x_i - \mu_x}{\sigma_x}, \quad \tilde{y}_i = \frac{y_i - \mu_y}{\sigma_y}$$

donde  $\mu$  y  $\sigma$  *sigmas* se calcularon sobre el conjunto de entrenamiento.

**Delimitación de repeticiones:** Se identificaron automáticamente los inicios y finales de cada repetición de sentadilla. Para ello se analizó la señal de velocidad vertical de puntos clave como cadera y rodilla, aplicando un umbral suave para detectar ciclos de descenso-ascenso.

$$v_t = (y_{cadera,t} - y_{cadera,t-1})/\Delta t$$

Este es la fórmula del ciclo de decenso- ascenso completo que define una repetición cuando la duracion supera 1s y no excede los 3s

Cálculo de características geométricas: Además de los landmarks crudos, se calcularon varias features angulares y de movimiento para cada repetición, las cuales complementan la secuencia de entrada y mejoran la interpretabilidad clínica de la evaluación [6][12][14]. En, se extrajeron las siguientes características por :

Ángulo de rodilla( $\theta_{rod}$ ): Ángulo articulado formado por la cadera, rodilla y tobillo de la pierna adelantada en cada instante (flexión de rodilla)(°).

Ángulo de cadera( $\theta_{hip}$ ): Ángulo formado por el tronco, la articulación de la cadera y la rodilla (indicativo de la flexión de la cadera (°).

Inclinación del tronco( $\theta_{trunk}$ ): Ángulo del tronco respecto a la vertical (calculado a partir de la línea hombro-cadera), que refleja cuánto se inclina el torso hacia adelante(°).

Profundidad de la sentadilla( $dsq$ ): Ángulo mínimo de rodilla alcanzado en cada repetición (un ángulo menor equivale a mayor profundidad de la sentadilla)(°).

Suavidad del movimiento( $Sm$ ): Indicadores de continuidad o temblor, basados en la variabilidad de la velocidad y aceleración angulares (p. ej., desviación de la velocidad angular de la rodilla), para detectar ejecuciones inestables( $\theta_{rod}$ ).

#### 2.4. Etiqueta de errores y esquema de salida.

Se adoptó un esquema de clasificación multi-label por repetición, dado que una sentadilla puede presentar múltiples fallas simultáneamente. Cada repetición  $r$  tiene un vector de etiquetas:

$$yr = [yE1, yE2, yE3, ycorr] \in \{0,1\}$$

Donde cada componente de indica la presencia(1) o la ausencia (0) del error

Las categorías de error definidas (no excluyentes entre sí) se basan en la siguiente consideración:

- Tronco inclinado: el torso se inclina excesivamente hacia adelante (más de  $\sim 10-15^\circ$  fuera de la vertical en la fase inferior del movimiento).
- Valgo de rodilla: la rodilla de la pierna adelantada se desvía medialmente (hacia adentro) respecto al eje pie-cadera durante la fase de bajada o subida.
- Profundidad insuficiente: la sentadilla no alcanza suficiente profundidad, definida típicamente como flexión de rodilla  $> 90^\circ$  en la posición más baja (ángulo obtuso, por encima de la paralelo).

Para la ejecución correcta se asigna etiquetas cuando ninguna de las anteriores condiciones de error se presenta en la repetición. Esta categoría se utiliza junto con las anteriores en un enfoque multi-etiqueta, marcando las repeticiones libres de error.

#### 2.5. Arquitectura y entrenamiento.

Para cada repetición segmentada, la entrada al modelo secuencial consiste en la secuencia temporal de vectores de características por frame. Cada vector por time-step concatena:

1. las coordenadas de los 33 landmarks 2D normalizados (MediaPipe Pose) junto con sus valores de visibilidad
2. las características geométricas derivadas mencionadas (ángulos, profundidad, etc.).

Esto produce un vector de entrada de tamaño fijo por frame (aprox. 99 dimensiones si se consideran 33 puntos con 3 atributos cada uno, ya sea coordenadas (x,y) + visibilidad, o incluyendo coordenada de profundidad z).

Capa	Tipo	Parametro principal
1	Bi-GRU	128 unidades, drop out=0
2	Bi-GRU	64 unidades
3	Activacion(Sigmoid)	4 salidas(E1-E3 + Correcta)

La grafica muestra la arquitectura que se

implementara. La primera capa opera de manera de modo bidireccional para capturar patrones temporales en ambas direcciones, la segunda capa con 64 unidades y finalmente la capa final que es la de activación sigmoideal que produce  $k$  salidas, en este caso son 4 salidas erróneas

De este modo, el modelo emite una predicción binaria para cada tipo de error y para la corrección, indicando la presencia o ausencia de cada condición en la repetición analizada.

Para entrenar la red se utilizó una función de pérdida de binary cross-entropy apropiada para clasificación multi-etiqueta. El optimizador usado fue Adam con una tasa de aprendizaje inicial de  $1 \times 10^{-3}$ . Se incorporaron mecanismos de ajuste automático, funciona de tal forma que si la métrica de validación se estancaba, la tasa de aprendizaje se reducía a la mitad, gracias a ReduceLROnPlateau con factor 0.5, paciencia de 7 epochs para afinar el aprendizaje, y se aplicó early stopping con paciencia de 15 epochs sin mejora para prevenir sobreajuste.

Finalmente, para el entrenamiento con un tamaño de lote de 32 secuencia, usando padding y masking para manejar longitudes variables de cada repetición, esto da más posibilidad de agrupar las secuencia de longitudes similares en buckets con el fin de optimizar todo el relleno

Para el aumento de datos temporal, se introdujeron pequeñas variaciones en las secuencias como una ligera deformación temporal (time warp) en la velocidad de ejecución, también eliminación aleatoria de algún frame intermedio (dropout de frames bajo), se añadió ruido Gaussiano de baja magnitud añadido a las coordenadas de entrada, e inversión horizontal controlada del video (espejado) cuando no altera la semántica de la lateralidad del ejercicio. Todas estas técnicas aumentan la robustez del modelo ante variaciones en la cadencia y posición, sin cambiar la esencia del movimiento. El modelo resultante es capaz de inferir en tiempo real, dado que la GRU es ligera y las secuencias de una repetición son cortas.

## 2.6. Métricas y protocolo.

La evaluación del sistema se realiza a nivel de repetición completa usando un conjunto de pruebas, se hicieron dos métricas, que están son:

### Métricas primarias.

Donde se consideran el F1-score macro y la F1-score por clase para cada una de las capas. Adicionalmente, se reporta la exactitud global (Acc) por repetición que es el porcentaje de repeticiones clasificadas totalmente correctas, esto sirvió como un como indicador complementario del rendimiento general del sistema.

### Métricas secundarias.

En el cual se evaluarán los **tiempos de inferencia**, considerando la latencia promedio tanto por repetición como por frame, con el objetivo de cumplir con un tiempo máximo de  $\leq 100$  ms por repetición en CPU moderna o  $\leq 30$  ms en una GPU de gama media. Finalmente, se examinará la **robustez angular**, es decir, la variación en el desempeño del modelo cuando se modifica ligeramente el ángulo de la cámara o bajo distintas condiciones de iluminación, siempre que el conjunto de datos proporcione dichas variaciones.

Para el protocolo de evaluación se contempla reportar el conjunto de las pruebas finales, el cual no se utilizó en entrenamiento ni en el ajuste de los hiperparámetros. Dado que las particiones se realizaron por video/sujeto, la evaluación es estricta a nivel de repeticiones de videos completamente inéditos para el modelo sin fuga entre frames de un mismo video. Todos los indicadores se calculan agregando los aciertos/errores por repetición completa. Para estimar la confianza de las métricas obtenidas, se calculan intervalos de confianza al 95% mediante bootstrap sobre las repeticiones de prueba. De esta forma, la metodología propuesta asegura una evaluación objetiva y reproducible del sistema de corrección postural, desde la captura del movimiento hasta la generación de feedback multi-etiqueta en tiempo real, siguiendo lineamientos experimentales rigurosos y acordes con trabajos previos en la literatura.

## 3. Results and discussion

Aquí van todos los resultados

- [5] Y. Y. Lee et al., "Validation of a video-based pose estimation algorithm for BESS assessment," 2025
- [6] G. Cai et al., "Application of Swin-UNet and pose estimation-based athlete motion quality detection," 2025
- [7] A. Hernández-Rayas et al., "Postural rehabilitation with machine vision," 2025
- [8] K. Manikandan et al., "AI-based posture correction, real-time exercise tracking and feedback," 2024
- [9] F. B. Pereira et al., "Técnica del vídeo análisis con KINOVEA," 2024
- [10] A. Chander et al., "RGB camera-based rehabilitation exercise assessment using transformer framework," 2025
- [11] S.-C. Yeh et al., "Yoga pose recognition and motion analysis for home-based fitness monitoring," 2025
- [12] W. Zhang et al., "Combined MediaPipe and YOLOv5 range of motion assessment system," 2024
- [13] M. Rajesh et al., "Spatiotemporal neural radiance fields for AI driven motion quality analysis," 2025

#### 4. References

- [1] Y. Liao et al., "A review of computational approaches for evaluation of rehabilitation exercises," 2020
- [2] N. K. Mangal et al., "A review of the evolution of scientific literature on technology-assisted approaches using RGB-D sensors," 2021
- [3] W. Simoes et al., "Accuracy assessment of 2D pose estimation with MediaPipe for physiotherapy exercises," 2024
- [4] C. Mennella et al., "A deep learning system to monitor and assess rehabilitation exercises," 2023









