Sistema de Clasificación de Movimiento Humano en Tiempo Real

Juan David Colonia Aldana
Miguel Ángel Gonzalez Arango
Pablo Fernando Pineda Patiño

Resumen

Este informe presenta el desarrollo de un sistema capaz de clasificar en tiempo real cinco tipos de movimientos corporales humanos utilizando video a tiempo real a través de una cámara convencional. La solución integra la detección de poses con MediaPipe y un modelo de aprendizaje automático basado en Random Forest, empleando técnicas de ingeniería de características para mejorar la precisión.

El sistema se implementa como una aplicación web interactiva, permitiendo la identificación automática de movimientos como: caminar, girar, sentarse y ponerse de pie. Con los resultados obtenidos en términos de precisión y velocidad, se afirma que este trabajo demuestra la viabilidad de construir clasificadores de movimiento accesibles y eficientes, con potencial de aplicación en salud, deporte y sistemas interactivos.

I. INTRODUCCIÓN

La clasificación automática de movimientos humanos es un área de gran relevancia dentro de la inteligencia artificial y la visión por computador. Su propósito es analizar y reconocer los patrones de movimiento del cuerpo a partir de secuencias de video, lo que permite interpretar el comportamiento corporal en tiempo real.

Esta tarea implica varios retos: la alta variabilidad en la forma en que distintas personas realizan un mismo movimiento; la influencia de factores externos como el entorno, la iluminación o la vestimenta; y la dificultad de distinguir movimientos que pueden ser similares en ciertos momentos. Además, lograr que los sistemas sean precisos, rápidos y funcionen con tecnologías accesibles, como cámaras convencionales, sin necesidad de sensores especializados, sigue siendo un desafío importante que impulsa el avance en este campo.

II. MARCO TEORICO

A. Detección de poses con MediaPipe

MediaPipe es una biblioteca de visión por computador desarrollada por Google que permite la detección eficiente de poses humanas en imágenes y video. Utiliza modelos de deep learning para identificar 33 puntos clave (landmarks) del cuerpo, proporcionando una representación estructurada de la postura y el movimiento humano. Esta información es la base para el análisis automático de movimientos.

B. Feature Engineering

A partir de los landmarks detectados, se calculan características geométricas relevantes, como ángulos articulares (por ejemplo, en rodillas y caderas), distancias entre puntos clave y proporciones corporales. Estas características permiten capturar patrones distintivos de cada movimiento, facilitando la tarea de clasificación y mejorando la interpretabilidad del modelo.

C. *Modelos supervisados de clasificación*

Para la tarea de clasificación de movimientos, se emplean algoritmos supervisados como XGBoost, Random Forest, SVM y K-NN. Estos modelos aprenden a distinguir entre diferentes tipos de movimiento a partir de ejemplos etiquetados, y se selecciona el de mejor desempeño según métricas de validación.

D. *Normalización de características*

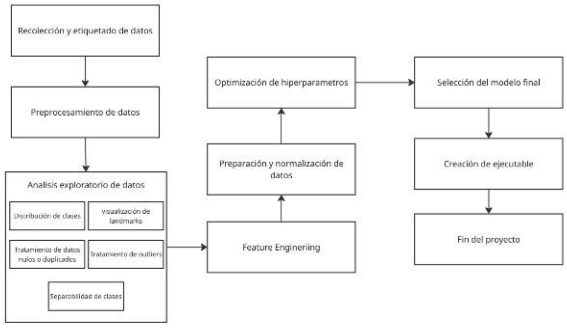
Antes del entrenamiento, todas las características extraídas se normalizan utilizando técnicas como el escalado estándar (StandardScaler), que ajusta los datos para que tengan media cero y varianza unitaria. Esto es fundamental para evitar que variables con diferentes escalas numéricas dominen el proceso de aprendizaje y para mejorar la estabilidad y el rendimiento de los modelos.

E. *Optimización de hiperparámetros*

El ajuste de hiperparámetros se realiza mediante búsqueda en malla (GridSearchCV) y validación cruzada estratificada. Este proceso permite encontrar la configuración óptima de cada modelo, maximizando la precisión y la capacidad de generalización, y evitando tanto el sobreajuste como el sub-ajuste.

III. METODOLOGIA

A continuación, se presenta el flujo de trabajo seguido durante el proyecto:



A. *Recolección y etiquetado de datos*

Se grabaron varios videos de cada uno de los integrantes del grupo realizando los cinco tipos de movimientos: caminar hacia adelante, caminar hacia atrás, girar, sentarse y ponerse de pie. Además, se compartieron y recopilaron videos de otros grupos, enriqueciendo la diversidad del dataset.

El proceso de etiquetado fue manual, de forma que cada video fue asignado a una clase específica según el movimiento realizado, asegurando la correcta correspondencia entre los datos y las etiquetas.

B. *Preprocesamiento*

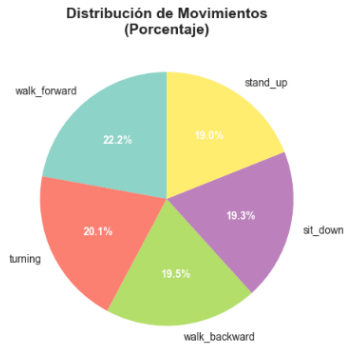
A partir de los videos, se utilizaron los modelos de MediaPipe para extraer los landmarks corporales cuadro a cuadro (un promedio de 30 frames por video). Posteriormente, se realizó un filtrado riguroso: aquellos videos en los que la detección de landmarks presentaba muchas alteraciones, inconsistencias o pérdida de puntos clave fueron descartados, priorizando la calidad y la estabilidad de los datos.

Tras el proceso de preprocesamiento, el conjunto de datos quedó conformado por 38 videos del movimiento de sentarse, 38 de ponerse de pie, 31 de girar, 25 de caminar hacia atrás y 28 de caminar hacia adelante, un total de 160 videos. Así mismo, el dataset contiene aproximadamente 16,000 frames, cada uno con las coordenadas y visibilidad de los landmarks, así como la etiqueta correspondiente al movimiento.

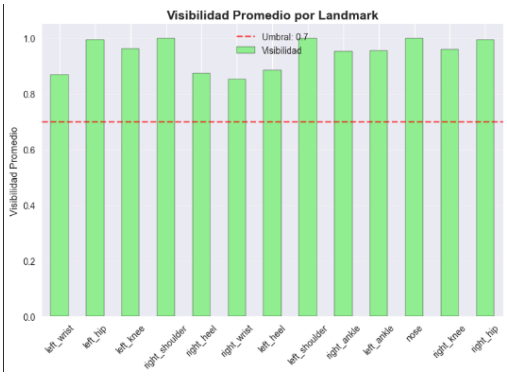
C. *Análisis exploratorio de datos*

Se realizó un análisis exploratorio exhaustivo para comprender la calidad y distribución de los datos. Entre los aspectos analizados se encuentran:

- **Distribución de clases:** Se evaluó el balance de las clases de movimiento, encontrando una distribución relativamente equilibrada con una ratio de balance de 0.856



- **Visualización de landmarks:** El análisis de los landmarks y su visibilidad mostró que todos los puntos clave superan el umbral de 0.7 en visibilidad promedio, lo que indica una buena calidad de captura. Se graficaron la visibilidad promedio por landmark, la distribución general de visibilidad, la visibilidad promedio por tipo de movimiento y el porcentaje de landmarks completamente faltantes, sin detectar landmarks problemáticos.



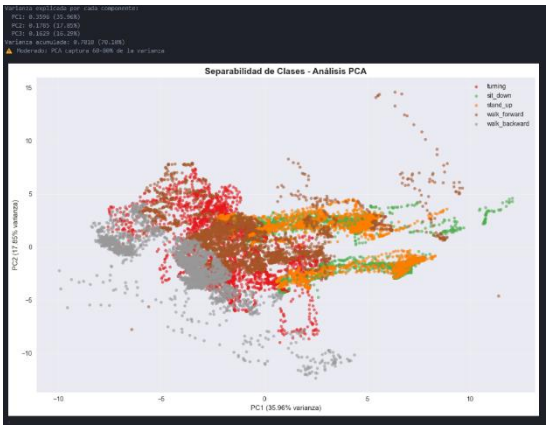
- **Tratamiento de datos nulos y duplicados:** Para el tratamiento de datos nulos y duplicados, se eliminaron frames con más del 70% de landmarks faltantes, se aplicó interpolación temporal para valores faltantes esporádicos (aunque no fue necesario en este caso) y se rellenaron valores de visibilidad faltantes con un valor conservador (0.5). Estas estrategias permitieron maximizar la cantidad de datos útiles sin sacrificar calidad.

- **Tratamiento de outliers:** En cuanto a los outliers, se detectaron y trataron en las coordenadas utilizando el método IQR (factor 2.5), se trataron un total de 651 outliers, menos del 1% de los datos, lo que contribuyó a mejorar la robustez del sistema.

```
Outliers tratados por columna:
  nose_x: 14 outliers (0.09%) - Limitados
  nose_y: 7 outliers (0.04%) - Limitados
  nose_z: 13 outliers (0.08%) - Limitados
  left_shoulder_x: 77 outliers (0.48%) - Limitados
  left_shoulder_y: 32 outliers (0.20%) - Limitados
  left_shoulder_z: 159 outliers (0.99%) - Limitados
  right_shoulder_x: 36 outliers (0.23%) - Limitados
  right_shoulder_y: 25 outliers (0.16%) - Limitados
  right_shoulder_z: 152 outliers (0.95%) - Limitados
  left_wrist_x: 136 outliers (0.85%) - Limitados

Total de outliers tratados: 651
```

- **Separabilidad de clases:** Se aplicaron técnicas de reducción de dimensionalidad (como PCA) para visualizar la separabilidad entre las clases de movimiento.



D. Feature Engineriing

Para mejorar la capacidad de clasificación, se transformaron los landmarks crudos en variables biomecánicas más representativas. Se calcularon ángulos articulares clave (rodillas e inclinación del tronco), distancias y proporciones corporales (como el ancho de hombros, caderas y altura del torso), así como velocidades y aceleraciones para

nueve articulaciones principales. Además, se generaron características específicas para cada tipo de movimiento, como diferencias de velocidad lateral (para giros), cambios de altura de caderas (para sentarse o levantarse), coordinación de piernas (para caminar) y medidas de simetría corporal.

También se incluyeron variables de visibilidad por regiones, lo que permitió robustecer el modelo ante posibles pérdidas de landmarks. Tras la generación, se aplicó selección de características y normalización estándar para asegurar que todas las variables aportaran de manera equilibrada. Como resultado, el conjunto final incluyó 52 features originales y 52 derivadas, sumando un total de 104 características para el entrenamiento del modelo.

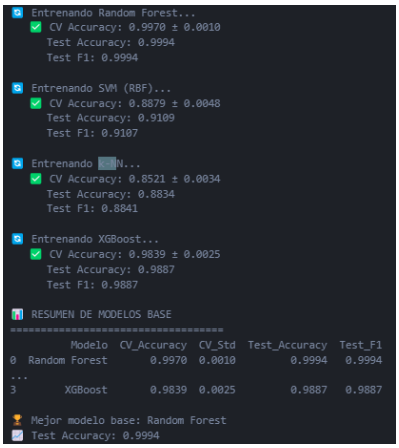
E. Preparación y normalización de datos

Tras la generación y selección de características, el conjunto de datos fue dividido en dos subconjuntos mediante partición estratificada: el 80% de los datos (12,796 muestras) se destinó al entrenamiento y el 20% restante (3,199 muestras) a la prueba, asegurando la representatividad de todas las clases en ambos conjuntos.

Para evitar sesgos debidos a diferencias de escala entre variables, todas las características fueron normalizadas utilizando un escalado estándar (StandardScaler), ajustando los datos para que cada variable tuviera media cero y varianza unitaria. Este paso es fundamental para garantizar un aprendizaje eficiente y estable en los modelos de machine learning empleados.

F. Entrenamiento y selección de modelos

Para la tarea de clasificación de movimientos, se evaluaron varios algoritmos supervisados: Random Forest, SVM (con kernel RBF), k-Nearest Neighbors (k-NN) y XGBoost. Cada modelo fue entrenado utilizando el conjunto de entrenamiento previamente normalizado y su desempeño se evaluó mediante validación cruzada estratificada, así como en el conjunto de prueba independiente. El proceso de evaluación incluyó el cálculo de métricas como accuracy y F1-score, tanto en validación cruzada como en el conjunto de prueba, permitiendo comparar objetivamente la capacidad de generalización de cada modelo. Los resultados obtenidos para cada algoritmo se resumen en la siguiente figura:

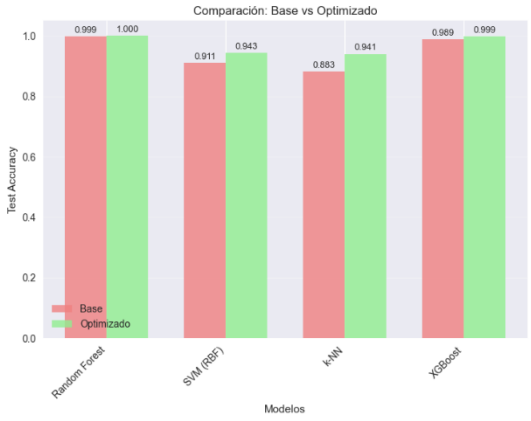


A partir de este análisis comparativo, se seleccionó Random Forest como modelo principal para el sistema, debido a su excelente equilibrio entre precisión, robustez y capacidad de generalización frente a datos ruidosos y complejos.

G. Optimización de hiperparametros

Para maximizar el rendimiento de los modelos, se realizó una optimización exhaustiva de los hiperparámetros clave de cada algoritmo mediante búsqueda en malla (GridSearchCV) y validación cruzada estratificada. Este proceso permitió explorar diferentes combinaciones de parámetros, como la profundidad máxima, el número de árboles y los criterios de división en Random Forest y XGBoost, así como los parámetros principales de SVM y k-NN.

La selección de hiperparámetros se basó exclusivamente en el desempeño sobre el conjunto de entrenamiento, garantizando así la validez de la evaluación en el conjunto de prueba.



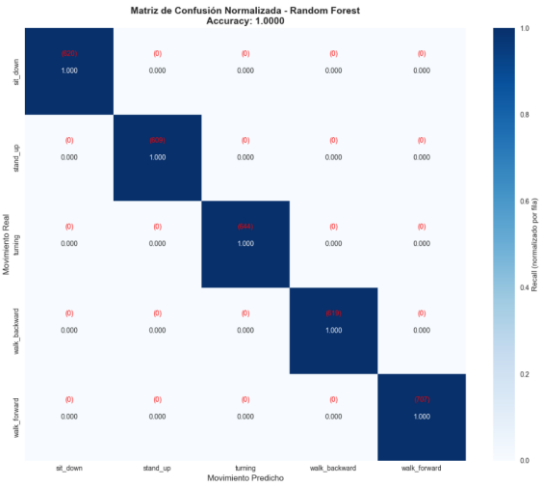
IV. RESULTADOS Y SU ANALISIS

A. Desempeño global del modelo final

El modelo Random Forest optimizado demostró un desempeño excepcionalmente alto en la tarea de clasificación de movimientos, superando a otros algoritmos evaluados como XGBoost, SVM y k-NN. Después de la optimización de hiperparámetros, Random Forest alcanzó una precisión perfecta (1.0000) en el conjunto de prueba, con una mejora mínima de +0.0006 respecto a su rendimiento en validación cruzada. XGBoost se posicionó como el segundo mejor modelo, seguido por SVM (RBF) y k-NN.

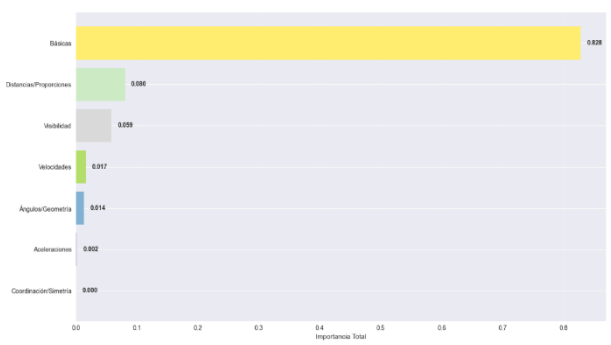
B. Matriz de confusión

La matriz de confusión del modelo muestra valores perfectos (1.0) en su diagonal principal, indicando que el modelo clasifica correctamente el 100% de los casos para cada tipo de movimiento, sin ningún error de clasificación entre las diferentes clases. Estos resultados excepcionales demuestran que el modelo ha logrado capturar completamente los patrones distintivos de cada tipo de movimiento. La ausencia total de confusión entre clases sugiere que las características seleccionadas y el modelo optimizado son capaces de distinguir claramente entre los diferentes tipos de movimientos, incluso aquellos que podrían parecer similares a primera vista.



C. Importancia de características

El análisis de importancia de características del modelo final permite identificar cuáles variables aportan mayor valor a la tarea de clasificación. En la figura siguiente se muestra la importancia total agrupada por categoría de características: coordenadas originales, ángulos biomecánicos, distancias, dinámicas (velocidades y aceleraciones) y medidas de simetría o coordinación.

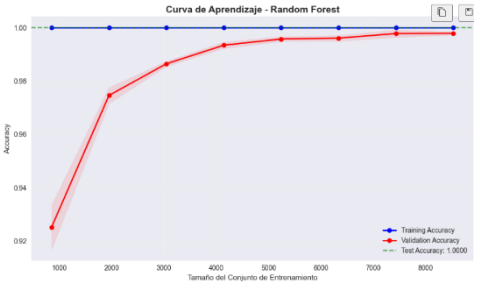


Se observa que las características básicas, es decir, las coordenadas originales de los landmarks, representan la mayor parte de la importancia total del modelo. Esto indica que la posición espacial de

los puntos clave del cuerpo es fundamental para distinguir entre los diferentes movimientos. Sin embargo, las características derivadas, como las distancias entre las articulaciones o los ángulos articulares, también contribuyen de manera significativa, especialmente para diferenciar movimientos con patrones similares, pero trayectorias distintas, como caminar y girar.

D. Curva de aprendizaje

La curva de aprendizaje del modelo final muestra tres métricas clave: accuracy de entrenamiento, validación y prueba.



El comportamiento de estas curvas revela varios aspectos importantes:

1. La accuracy de entrenamiento se mantiene constante en 1.0, indicando un aprendizaje muy efectivo desde el inicio.
2. La accuracy de prueba alcanza el valor perfecto de 1.0, con un gap mínimo de 0.0021 respecto a la validación.
3. La accuracy de validación muestra una curva ascendente que converge hacia un valor muy alto.

Este patrón, junto con las bandas de error pequeñas en la validación, sugiere un modelo robusto y estable. La consistencia entre las métricas de validación y prueba, y el gap mínimo entre ellas, indica que el modelo ha encontrado patrones consistentes y generalizables en los datos.

E. Visualización de resultados en tiempo real

Los niveles de confianza más moderados en la aplicación real proporcionan una visión realista del desempeño del modelo en condiciones del mundo

real, donde factores como la iluminación, el ángulo de la cámara y la variabilidad natural del movimiento humano influyen en la clasificación. Estos resultados demuestran la robustez del modelo y su capacidad para generalizar efectivamente a situaciones prácticas.

V. CONCLUSIONES Y TRABAJO FUTURO

El sistema desarrollado demuestra que es posible construir un clasificador de movimientos humanos en tiempo real utilizando tecnologías accesibles como una cámara convencional, MediaPipe y modelos de aprendizaje automático como Random Forest. Los resultados obtenidos, con una precisión perfecta en el conjunto de prueba y una robusta capacidad de generalización, validan la efectividad de la metodología seguida, desde la ingeniería de características hasta la optimización de hiperparámetros. Además, la integración en una aplicación web funcional resalta el potencial de la solución para ser utilizada en aplicaciones reales en campos como la rehabilitación física, el entrenamiento deportivo, la asistencia a personas mayores y la interacción hombre-máquina.

Sin embargo, a pesar de los resultados positivos, existen áreas claras para mejorar y expandir el trabajo:

- Ampliación del dataset: Aunque el modelo mostró un excelente desempeño, este se obtuvo con un conjunto limitado de movimientos y de sujetos. Incorporar mayor diversidad demográfica (edades, complexiones, estilos de movimiento) y nuevos tipos de movimiento permitiría construir un modelo aún más generalizable.

- Incremento en la complejidad de los escenarios: Las pruebas se realizaron en condiciones controladas. Futuras iteraciones deben considerar entornos no estructurados con distintos fondos, niveles de iluminación y múltiples personas en escena.

- Evaluación temporal: Incluir modelos que aprovechen la secuencia temporal de los frames, como redes neuronales recurrentes (LSTM) o modelos basados en Transformers, podría capturar mejor la dinámica del movimiento y mejorar la robustez ante ruido en frames individuales.

- Interacción multimodal: Combinar los datos visuales con otras fuentes como audio o sensores inerciales (en smartphones o wearables) podría enriquecer la clasificación y abrir nuevas líneas de aplicación.

Este proyecto constituye una base sólida sobre la cual se pueden construir soluciones más avanzadas y adaptativas, con gran potencial en áreas interdisciplinarias como la salud, la educación física y la inteligencia artificial aplicada al comportamiento humano.

VI. REFERENCIAS BIBLIOGRAFICAS

- Elakiya Sekar, "*Feature Engineering for Machine Learning: A step by step Guide*", Jul. 25, 2023. https://medium.com/kgxperience/feature-engineering-for-machine-learning-a-step-by-step-guide-part-1-33b520c67137
- Google Inc, "*Guía de detección de puntos de referencia de posiciones*", Ene. 13, 2025. https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=es-419
- Rushil Thareja, "*Video Analysis Engine for Predicting Effectiveness*", Dec. 5, 2024. https://dl.acm.org/doi/10.1007/978-3-031-78312-8_7