

# **Sistema de reconocimiento de movimientos**

## **Abstract**

Este proyecto presenta un sistema para la clasificación de actividades humanas (caminar hacia la cámara, caminar de regreso, girar, sentarse y levantarse) en tiempo real, utilizando tecnologías accesibles como MediaPipe y un modelo SVM. El sistema integra un seguimiento de puntos articulares clave, con técnicas avanzadas de preprocesamiento y reducción de dimensionalidad mediante PCA. Al implementar modelos supervisados, el sistema alcanzó una precisión global del 93% y un F1-score ponderado de 0.93, destacándose en actividades comunes como caminar y sentarse. Sin embargo, evidenció limitaciones significativas al clasificar "girar", lo cual se atribuye a la variabilidad de este movimiento. Este trabajo demuestra la viabilidad de sistemas escalables y eficientes para el análisis biomecánico, proponiendo como mejoras futuras la expansión del dataset y la integración de modelos más avanzados como redes neuronales ligeras.

## **Introducción**

El análisis automático de actividades humanas tiene aplicaciones importantes en rehabilitación médica, análisis deportivo y vigilancia, donde la precisión y la rapidez son esenciales. Este proyecto aborda el desafío de clasificar actividades humanas en tiempo real mediante un sistema que combina MediaPipe para el seguimiento articular y modelos supervisados como SVM, Random Forest y XGBoost.

A diferencia de soluciones que requieren hardware especializado, este sistema utiliza herramientas accesibles y computacionalmente eficientes, permitiendo su implementación en dispositivos de bajo costo. El proyecto se centra en actividades como caminar, girar, sentarse y levantarse, con un enfoque en analizar movimientos articulares clave y patrones posturales. Además, plantea el desafío de operar en condiciones variadas, incluyendo diferencias en iluminación, perspectivas y velocidad de movimiento, lo que lo convierte en una solución interesante y versátil.

## **Teoría**

El diseño del sistema combina elementos clave de dos herramientas: visión por computadora y aprendizaje automático. MediaPipe es una herramienta avanzada que permite el seguimiento en tiempo real de puntos articulares clave, como caderas, rodillas, tobillos, hombros y muñecas. Utilizando modelos de aprendizaje profundo preentrenados, MediaPipe analiza flujos de video para identificar y rastrear estos puntos, generando coordenadas tridimensionales (x, y, z) para cada articulación. Este enfoque proporciona información detallada sobre las posiciones

relativas del cuerpo, convirtiéndose en la base del sistema. Entre las ventajas de MediaPipe se destacan su eficiencia computacional, ya que puede ejecutarse en dispositivos de bajo rendimiento como cámaras web o teléfonos inteligentes..

Para clasificar las actividades, se seleccionaron tres modelos de aprendizaje supervisado: SVM, Random Forest y XGBoost. Cada modelo aporta características específicas al sistema. SVM es ideal para manejar datos de alta dimensionalidad y problemas con límites de decisión complejos, permitiendo clasificaciones precisas mediante el uso de funciones kernel. Por otro lado, Random Forest combina múltiples árboles de decisión entrenados con subconjuntos aleatorios de datos y características, mejorando la generalización y reduciendo el sobreajuste. Finalmente, XGBoost se basa en boosting, donde múltiples árboles de decisión se entrenan iterativamente para corregir errores de iteraciones previas. Este modelo destaca por su regularización integrada, que evita el sobreajuste, su eficiencia computacional y su capacidad para manejar datos desbalanceados.

La evaluación de los modelos se realizó utilizando métricas estándar como precisión, recall y F1-Score. Estas métricas garantizan que el sistema pueda clasificar actividades humanas con alta precisión y generalización. Además, se implementó validación cruzada para evitar el sobreajuste y asegurar la robustez del sistema en condiciones reales. Este enfoque teórico asegura que el sistema pueda operar de manera eficiente y confiable en una variedad de escenarios operativos.

## Metodología

El desarrollo del sistema siguió un enfoque estructurado para abordar cada etapa del proyecto de manera eficiente. Inicialmente, se definió claramente el problema, estableciendo como objetivo principal desarrollar un sistema capaz de detectar y clasificar actividades humanas en tiempo real. Las actividades objetivo incluyeron caminar hacia la cámara, caminar de regreso, girar, sentarse y levantarse. Esto guió el diseño del sistema, la selección de herramientas y las métricas de evaluación. Se realizó una revisión bibliográfica para fundamentar el marco teórico, destacando la relevancia de herramientas como **MediaPipe** y modelos de aprendizaje supervisado.

Para construir el sistema, se creó un conjunto de datos personalizado a partir de videos que capturaron las actividades objetivo realizadas por participantes con diversas características físicas. Estos videos incluyeron variaciones en velocidades de movimiento (rápido y lento) y diferentes condiciones de iluminación, desde entornos bien iluminados hasta escenarios con poca luz, asegurando que los datos fueran representativos de situaciones reales.

La anotación de datos se realizó utilizando un enfoque combinado. Se emplearon herramientas de serialización como **joblib** y **pickle** para asignar etiquetas específicas a las actividades de forma eficiente y reproducible. Estas etiquetas se

generaron basándose en patrones detectados por **MediaPipe**, verificándose manualmente para garantizar su calidad y consistencia. Este conjunto de datos anotados sirvió como base para entrenar y evaluar los modelos de aprendizaje automático.

El preprocesamiento de los datos incluyó la normalización de las coordenadas articulares (x, y, z) para evitar dependencias de la altura del sujeto o la posición de la cámara. Además, se aplicaron técnicas de filtrado para minimizar el ruido en el seguimiento articular y mejorar la calidad de las características extraídas. Estas características procesadas constituyeron la entrada para el entrenamiento de los modelos.

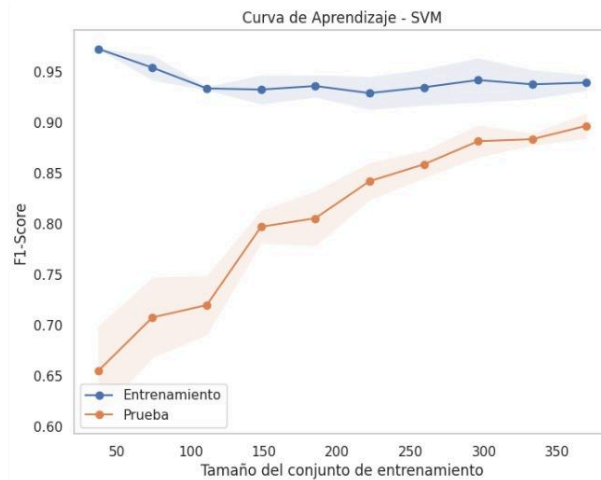
Para el entrenamiento de los modelos el dataset fue dividido en un 80% para entrenamiento y 20% para prueba. Los hiperparámetros se ajustaron utilizando técnicas automatizadas para maximizar la precisión y la generalización del modelo.

Finalmente, el modelo SVM entrenado fue integrado con MediaPipe y OpenCV en un sistema interactivo de análisis en tiempo real. Se diseñó una interfaz gráfica que permite al usuario visualizar la actividad detectada. El sistema fue probado en condiciones reales y adversas, como iluminación baja y movimientos rápidos. Aunque el modelo mostró un desempeño robusto en la mayoría de las actividades, presentó dificultades al clasificar correctamente la acción de "girar", evidenciando la necesidad de mejorar la representación de esta actividad en los datos de entrenamiento.

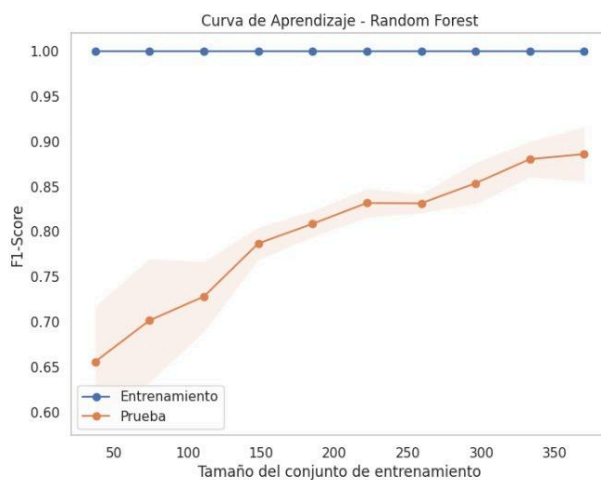
## Resultados

Después de haber entrenado y evaluado los modelos obtuvimos los siguientes resultados:

El modelo **SVM** demostró ser la opción más robusta y efectiva, alcanzando una precisión global del **93%** y un F1-score ponderado de **0.93**. Las curvas de aprendizaje muestran un excelente balance entre el F1-score de entrenamiento y prueba, sin diferencias significativas entre ambas, lo que indica que el modelo **no presenta sobreajuste**. Además, a medida que se incrementa el tamaño del conjunto de entrenamiento, el F1-score de prueba mejora consistentemente, reforzando su capacidad de generalización. SVM sobresalió en actividades comunes como "caminar" y "sentarse", manteniendo un desempeño aceptable incluso en actividades más desafiantes como "girar".

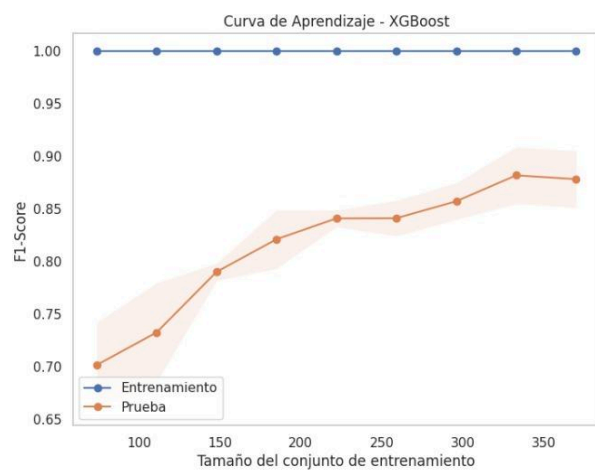


El modelo **Random Forest**, por otro lado, alcanzó una precisión global de **89%** y un F1-score ponderado de **0.88**. Sin embargo, su curva de aprendizaje refleja un **claro sobreajuste**, ya que la curva de entrenamiento permanece plana y cercana a 1, mientras que la curva de prueba es significativamente más baja. Esto indica que el modelo se adapta excesivamente a los datos de entrenamiento, limitando su capacidad de generalización. Aunque tuvo un buen desempeño en actividades simples como "caminar", mostró dificultades significativas en actividades más complejas como "girar" y "levantarse", con precisiones inferiores al 80%. Para mitigar el sobreajuste, sería necesario ajustar los hiperparámetros, como reducir la profundidad de los árboles o aumentar la aleatoriedad en la selección de características.



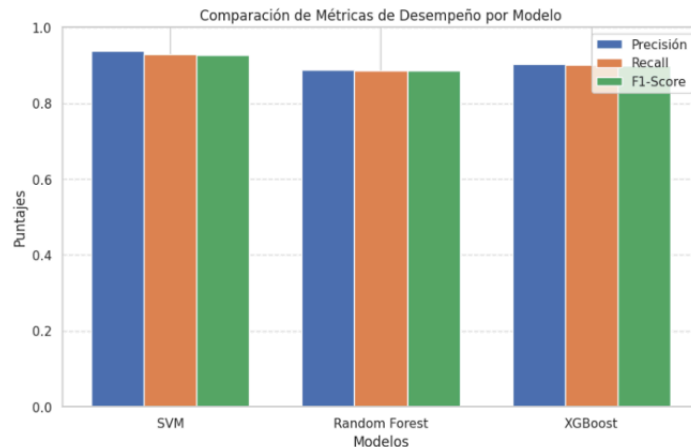
En cuanto al modelo **XGBoost**, logró una precisión global de **90%** y un F1-score ponderado de **0.89**. Aunque este modelo cuenta con regularización integrada, su curva de aprendizaje también evidencia **signos de sobreajuste**, aunque menos pronunciados que en Random Forest. La curva de entrenamiento se mantiene alta, mientras que la curva de prueba mejora gradualmente pero con una brecha notable. XGBoost mostró un desempeño sólido en actividades comunes, pero

presentó limitaciones en actividades más desafiantes como "girar", con precisiones en el rango del 75-80%. Ajustes en los parámetros de regularización, como lambda y alpha, podrían mejorar su capacidad de generalización.



El rendimiento superior del SVM puede atribuirse a su capacidad de encontrar un margen óptimo de decisión y manejar datos multidimensionales con mayor eficacia. En contraste, los modelos basados en árboles, aunque efectivos para muchas tareas de clasificación, parecen ser menos sensibles a las características específicas de las clases menos representadas o más complejas, lo que resulta en una disminución de su precisión.

La acción de "girar" presentó dificultades significativas en su clasificación, con un desempeño inferior al de otras actividades en términos de precisión. Esto puede deberse a varios factores. En primer lugar, la similitud en patrones articulares con otras actividades, como "caminar de regreso", puede confundir al modelo, ya que ambas comparten movimientos transitorios y cambios de posición que no son fácilmente distinguibles mediante coordenadas x, y, z. Además, la forma en la que se estructuró el dataset, basada únicamente en coordenadas articulares sin considerar la temporalidad o patrones más complejos entre frames consecutivos, podría limitar la capacidad del modelo para identificar características únicas del giro.



## Conclusiones

En este proyecto, desarrollamos un sistema de clasificación de actividades humanas en tiempo real que combina la tecnología de seguimiento articular de MediaPipe con modelos de aprendizaje supervisado, destacando al SVM como la mejor solución para el problema planteado. El sistema logró una precisión global del 93%, con un rendimiento notable en actividades comunes como caminar y sentarse. A través de un flujo metódico, se demostró la importancia del preprocesamiento (normalización, reducción de ruido y PCA) para mejorar la eficiencia computacional y la capacidad de generalización del modelo.

### Lecciones aprendidas:

**Eficiencia del SVM:** Este modelo sobresalió por su capacidad para manejar datos multidimensionales y límites de decisión complejos, especialmente en clases con patrones consistentes.

**Limitaciones de los datos:** El desempeño en la actividad de "girar" evidenció cómo la falta de representatividad en el dataset afecta la precisión. Esto subraya la necesidad de datasets equilibrados y diversos.

**Flexibilidad del sistema:** La integración de MediaPipe demostró ser una solución práctica y escalable, adecuada para dispositivos de bajo rendimiento y condiciones de uso reales.

### Áreas de mejora:

**Recolección de datos:** Incluir ejemplos más diversos y representativos de actividades como "girar" para mejorar el desempeño en estas clases menos comunes.

**Aumento de datos:** Implementar técnicas avanzadas como redes generativas adversarias (GANs) para generar ejemplos sintéticos que enriquezcan el conjunto de datos.

**Modelos más avanzados:** Explorar arquitecturas ligeras de redes neuronales (por ejemplo, MobileNet o redes basadas en transformadores) para mejorar la precisión en escenarios más complejos y dinámicos.

**Interfaz y visualización:** Optimizar la presentación de resultados al usuario final, incorporando métricas adicionales y personalización de análisis biomecánicos.

## Future Work

Este proyecto sienta las bases para una amplia gama de aplicaciones prácticas, pero también abre la puerta a futuras mejoras en las siguientes áreas:

**Optimización del modelo:** Ajustar hiperparámetros del SVM y evaluar nuevas estrategias para balancear clases y aumentar la robustez del modelo.

**Ampliación del sistema:** Implementar funcionalidades adicionales, como la detección de anomalías en movimientos, que podrían ser útiles en rehabilitación médica y prevención de lesiones.

**Desarrollo de hardware portátil:** Integrar el sistema en dispositivos como teléfonos inteligentes o wearables para análisis en tiempo real directamente en el punto de uso.

**Aplicaciones específicas:** Crear módulos personalizados para entrenamiento deportivo, rehabilitación y vigilancia, basados en las métricas biomecánicas generadas.

Este trabajo no solo demuestra la viabilidad de sistemas accesibles y eficientes, sino que también ofrece un marco sólido para futuras innovaciones en el análisis automático de actividades humanas.

## Referencias bibliográficas

1. MediaPipe. "MediaPipe Solutions Guide." [En línea]. Disponible: <https://ai.google.dev/edge/mediapipe/solutions/guide?hl=es-419>
2. LabelStudio. "Annotation Tool." [En línea]. Disponible: <https://labelstud.io/>
3. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.
4. Breiman, L. "Random Forests." *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
5. Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, 1995.
6. OpenPose vs. MediaPipe: A Comparative Review. Medium, 2023. Disponible: <https://quickpose.ai/faqs/mediapipe-vs-openpose>

