**Reconocimiento automático de actividades humanas mediante extracción de puntos clave corporales y modelos predictivos avanzados**

Daniel Escobar

Juan Calderón

Alejando Castro

**Resumen**

Este informe presenta el desarrollo de un sistema automático para el reconocimiento de actividades humanas utilizando técnicas avanzadas de visión por computadora y aprendizaje automático. Los datos se obtuvieron mediante la grabación de videos de participantes realizando actividades cotidianas específicas y fueron ampliados mediante técnicas de generación sintética controlada. Posteriormente, los datos se procesaron mediante MediaPipe para la extracción de puntos clave corporales (landmarks), normalizados respecto a la posición de la cadera. Se implementó un modelo de Random Forest optimizado con búsqueda sistemática de hiperparámetros utilizando validación cruzada. Los resultados mostraron una precisión global del 92.15%, revelando ligeras confusiones en actividades con movimientos similares. Este enfoque representa un significativo avance en la detección precisa y robusta de actividades físicas en entornos reales.

**Abstract**

This report presents the development of an automatic system for recognizing human activities using advanced computer vision and machine learning techniques. The data was obtained by recording videos of participants performing specific daily activities and was augmented through controlled synthetic generation techniques. The data was then processed using MediaPipe for the extraction of body keypoints (landmarks), normalized with respect to the hip position. An optimized Random Forest model was implemented using systematic hyperparameter search with cross-validation. The results showed an overall accuracy of 92.15%, revealing slight confusions in activities with similar movements. This approach represents a significant advancement in the accurate and robust detection of physical activities in real-world environments.

**Introducción**

La capacidad para reconocer automáticamente actividades humanas a partir de datos visuales es crítica para múltiples aplicaciones en salud, deporte y ergonomía. No obstante, este reconocimiento plantea desafíos considerables debido a la variabilidad inherente de los movimientos humanos. Este proyecto aborda dicho desafío mediante la implementación de técnicas robustas de extracción de puntos clave y algoritmos predictivos optimizados, justificándose por la necesidad de sistemas capaces de operar con fiabilidad en condiciones reales y diversas.

**Marco teórico**

El reconocimiento de actividades humanas mediante visión computacional se fundamenta en la extracción y análisis detallado de puntos clave corporales, conocidos como landmarks. Estos puntos clave, obtenidos mediante técnicas como MediaPipe Pose, proporcionan coordenadas espaciales precisas de articulaciones humanas, permitiendo la representación cuantitativa y cualitativa de movimientos complejos. La utilización de landmarks facilita una descripción robusta e invariante de movimientos frente a variaciones espaciales y lumínicas.

El algoritmo MediaPipe Pose se basa en redes neuronales convolucionales optimizadas para inferir de forma eficiente las posiciones de múltiples puntos clave corporales a partir de imágenes bidimensionales, lo que lo hace ideal para aplicaciones en tiempo real con recursos limitados. Por otra parte, los métodos de ampliación de datos mediante ruido gaussiano permiten mejorar la generalización y prevenir el sobreajuste, generando datos sintéticos que simulan variaciones naturales observables en situaciones reales.

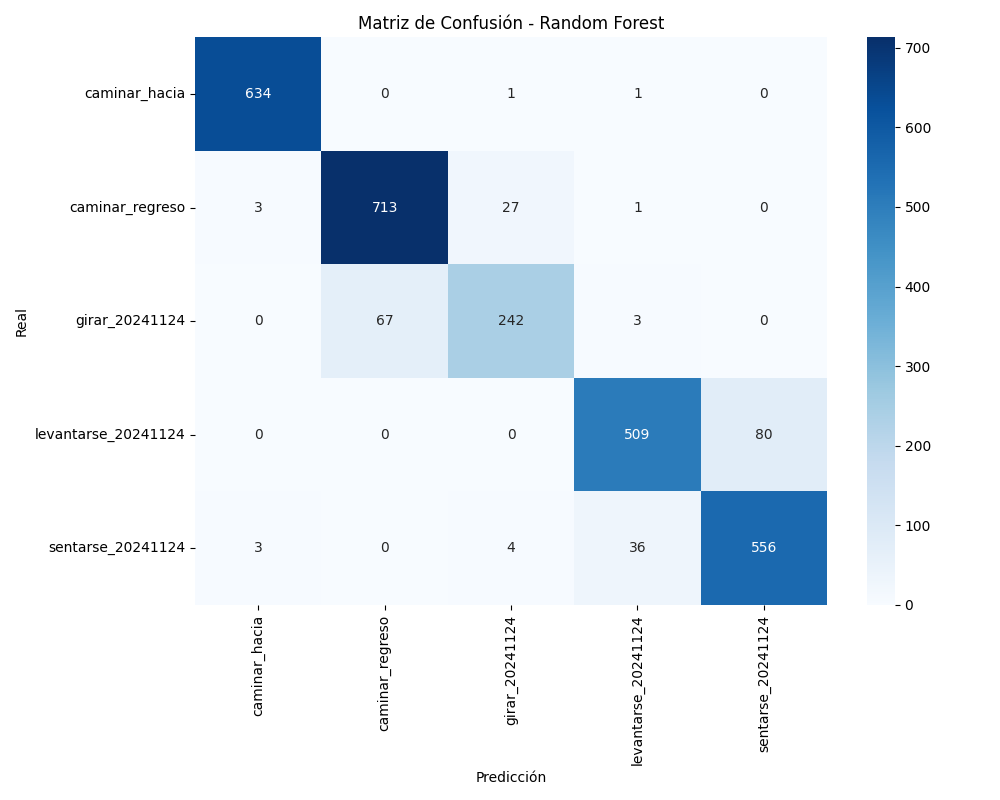
El algoritmo de clasificación Random Forest, propuesto por Breiman, constituye una técnica de aprendizaje supervisado basado en ensamble que aprovecha múltiples árboles de decisión independientes para mejorar la precisión y estabilidad del modelo predictivo. La selección óptima de hiperparámetros, como profundidad del árbol, número de estimadores y mínimos por división, es crucial para equilibrar la complejidad del modelo y su capacidad predictiva, especialmente en tareas complejas y multiclase como la clasificación de actividades humanas.

**Metodología**

Inicialmente, se capturaron videos en múltiples condiciones ambientales con personas realizando acciones definidas (caminar hacia la cámara, regresar, girar, sentarse y levantarse). Posteriormente, se generaron muestras adicionales mediante variaciones gaussianas sutiles (nivel de ruido: 0.005), duplicando así el conjunto inicial. Los puntos clave fueron extraídos usando MediaPipe Pose, generando conjuntos organizados y normalizados respecto al landmark de la cadera para asegurar invariancia espacial. Se realizó una selección rigurosa de hiperparámetros mediante GridSearchCV, evaluando combinaciones de profundidad máxima, número de estimadores y parámetros mínimos por hoja y división con validación cruzada de 5 pliegues.

**Resultados**

El modelo final obtuvo una precisión global de 92.15%, precisión específica de 92.17%, recall de 92.15%, y F1-score ponderado de 92.07%. La matriz de confusión mostró alta efectividad en la clasificación de actividades como "caminar\_hacia" y "caminar\_regreso". Sin embargo, presentó ligeras confusiones entre acciones similares como "girar" con "caminar\_regreso" (67 casos) y entre "sentarse" y "levantarse" (80 y 36 casos respectivamente).



**Análisis de resultados**

El desempeño general del modelo es robusto, alcanzando métricas alineadas o superiores a las reportadas en literatura reciente sobre reconocimiento visual de actividades. La confusión observada es consistente con estudios previos, sugiriendo la necesidad de incorporar características temporales más sofisticadas para diferenciar movimientos que comparten patrones comunes. La generalización del modelo fue reforzada eficazmente mediante las técnicas de normalización y ampliación, aunque persiste margen de mejora en la distinción precisa de actividades muy similares.

**Conclusiones y trabajo futuro**

Este estudio demuestra la viabilidad técnica del reconocimiento automático de actividades humanas mediante extracción y análisis de landmarks corporales, con alta precisión y robustez operacional. Las limitaciones observadas, como la confusión entre actividades con patrones motores similares, subrayan la importancia de profundizar en métodos complementarios que aprovechen secuencias temporales o técnicas avanzadas de aprendizaje profundo.

Como trabajo futuro, se propone la integración de modelos secuenciales como las redes neuronales recurrentes (LSTM) o arquitecturas tipo Transformer, capaces de capturar patrones dinámicos y contextuales más complejos. Además, sería relevante evaluar el rendimiento del sistema en grupos poblacionales específicos con variabilidades motoras propias, como adultos mayores o personas con movilidad reducida, para expandir el alcance práctico y mejorar la adaptabilidad del sistema.

Asimismo, podrían explorarse técnicas adicionales de aumento de datos más sofisticadas, incluyendo métodos generativos adversariales (GANs), para una representación más realista y diversa de variaciones naturales del movimiento humano.

**Referencias bibliográficas**

[1] Lugaresi, C., Tang, J., Nash, H., et al., "MediaPipe: A Framework for Building Perception Pipelines," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1-8.

[2] Breiman, L., "Random forests," Machine Learning, vol. 45, no. 1, 2001, pp. 5-32.