

## CLASIFICACIÓN DE ESTUDIOS DE MARCHA DE PACIENTES CON LESIÓN MEDULAR USANDO K-VECINOS MÁS CERCANOS

Diana Herrera-Valenzuela, dherrerav@externas.sescam.jccm.es, Isabel Sinovas-Alonso, msinovas@sescam.jccm.es, Ana de los Reyes, adlos@sescam.jccm.es, Ángel Gil Agudo, amgila@sescam.jccm.es, Hospital Nacional de Paraplégicos, España; Juan Camilo Moreno, jc.moreno@csic.es, Instituto Cajal-Consejo Superior de Investigaciones Científicas, España; Antonio J. del-Ama, antonio.delama@urjc.es, Universidad Rey Juan Carlos. España

### Resumen

Los análisis tridimensionales de la marcha permiten analizar objetivamente la marcha de cada individuo para cuantificar los trastornos que presenten. La gran cantidad de información cinemática, cinética y espacio-temporal que brindan estos estudios puede ser tratada con herramientas de aprendizaje de máquina para identificar las variables más importantes que representan a determinada población y que permiten diferenciarla de otras. En este trabajo se implementa un modelo de k-vecinos más cercanos para clasificar la marcha de pacientes con lesión medular y la marcha de sujetos sanos, usando variables cinemáticas y espacio-temporales de 776 pasos de pacientes y 496 pasos de sujetos sin alteraciones de la marcha, obtenidos en estudios de marcha. Se obtuvo un 97.9% de exactitud y un 98.9% de precisión en el algoritmo. Resultados similares, por encima del 92% para ambas métricas, fueron alcanzados al entrenar el modelo únicamente con las variables cinemáticas o con las espaciotemporales. La información cinemática y espacio-temporal recopilada en estudios de marcha permite discriminar la marcha de pacientes con lesión medular de la de sujetos sanos con exactitud y precisión al ser usada para entrenar un modelo de k-vecinos más cercanos.

### Abstract

Biomechanical gait analysis studies allow to objectively study the gait of patients to quantify the alterations they present. The large amount of kinematic, kinetic and spatiotemporal information provided by these studies is useful to train machine learning algorithms to identify the most important variables that represent a certain population and that differentiate it from others. In this paper, a model of K-Nearest Neighbors is implemented with data obtained from gait analyses, to classify the gait of patients with spinal cord injury and the gait of healthy subjects, using kinematic and spatiotemporal variables from 776 steps belonging to patients and 496 steps of subjects without gait alterations. The model reached an accuracy of 97.9% and precision of 98.9%. Similar results, above 92% for both metrics, were achieved by training the model only with the set of either kinematic or spatiotemporal variables. The kinematic and spatiotemporal information collected in gait studies allows to discriminate the gait of patients with spinal cord injury from that of healthy subjects with accuracy and precision when used to train a model of K-Nearest Neighbors.

**Palabras clave:** análisis biomecánico de la marcha, k-vecinos más cercanos, lesión medular.

**Keywords:** biomechanical gait analysis, K-Nearest Neighbors, spinal cord injury.

## **1. Introducción**

La pérdida de la función motora es una consecuencia común en la lesión medular (LM), por lo cual uno de los principales objetivos de rehabilitación para los pacientes que la padecen, es la marcha (Simpson et al., 2013). Para lograr una rehabilitación eficiente y eficaz, ésta debe ser adaptada según el estado funcional y las características de la lesión de cada sujeto (Stampacchia et al., 2020), pues la variabilidad en estos aspectos y en las estrategias compensadoras de los patrones de marcha entre pacientes es muy elevada. Una forma de identificar dichos requerimientos es a través de los estudios de marcha, ya que estos permiten analizar objetivamente la marcha de cada individuo para de esta manera detectar y cuantificar los trastornos que presenten (Krawetz & Nance, 1996).

Dada la gran cantidad de información cinemática, cinética y espacio-temporal que brinda un estudio de marcha (Krawetz & Nance, 1996), las herramientas de aprendizaje de máquina resultan de utilidad para identificar las variables más importantes que representan a determinada población y que permiten diferenciarla de otras (Prakash et al., 2016; Xu et al., 2006). En este sentido, los algoritmos de clasificación pueden ser usados para diferenciar la marcha patológica de la no patológica, e incluso para identificar las principales variables que representan la variabilidad de la marcha en una población. Estas se podrían traducir en necesidades funcionales específicas para la marcha de cada paciente, que permitan desarrollar estrategias y herramientas personalizadas para la rehabilitación de la marcha.

## **2. Desarrollo**

### **2.1. Marco teórico**

Distintos algoritmos de aprendizaje de máquina supervisado y no supervisado se han usado con éxito para agrupar estudios de marcha de pacientes con parálisis cerebral con fines diagnósticos (Prakash et al., 2018; O'Malley et al., 1997; Wong et al., 2007; Xu et al., 2006). En lesión medular, se han desarrollado trabajos para identificar los parámetros espaciotemporales de la marcha que permiten discriminar mejor entre un grupo de ratas con lesión incompleta y un grupo control (Timotius et al., 2021), y se han implementado algoritmos de aprendizaje no supervisado para el pronóstico de la rehabilitación de la marcha en pacientes con lesión medular (DeVries et al., 2019).

### **2.2. Planteamiento del problema**

A conocimiento de los autores, no se han realizado estudios que exploren el uso de algoritmos de clasificación basados en aprendizaje de máquina para discriminar la marcha no patológica de la de pacientes con LM. Es por ello que el objetivo de este trabajo es implementar un modelo clásico de clasificación, conocido como k-vecinos más cercanos, para clasificar la marcha de pacientes con LM y la marcha de sujetos sanos, usando variables cinemáticas y espacio-temporales obtenidas de estudios de marcha. Adicionalmente, se identifican los componentes que contribuyen en mayor medida al desempeño favorable del clasificador.

### **2.3. Método**

Se recopiló de forma retrospectiva 18 variables espacio-temporales y cinemáticas de los pacientes con diagnóstico de síndrome de LM que realizaron un estudio de marcha en el Hospital Nacional de Paraplégicos de Toledo, España, como parte de su seguimiento clínico entre agosto de 2019 y junio de 2021. Los estudios de marcha se realizaron con el sistema de análisis de movimiento CODAmotion y

fueron procesados con el software ODIN v. 2.02 (Codamotion Ltd., Inglaterra, Reino Unido). Todos los sujetos o sus acudientes firmaron el consentimiento informado previo a la realización del estudio. En total se recopilaron registros de 776 pasos, correspondientes a 96 estudios de marcha de 81 pacientes, con edades entre los 5 y 70 años (mediana: 15 años, promedio: 20.84 años  $\pm$ 16.51). Los registros del grupo control se recopilaron de la misma manera y están conformados por 496 pasos de 50 sujetos, con edades entre los 18 y 63 años (mediana: 28 años, promedio: 34.54 años  $\pm$ 15.02).

Las 18 variables utilizadas para entrenar el clasificador fueron: nueve cinemáticas, correspondientes a los valores máximo, mínimo y el rango articular en el plano sagital de la cadera, la rodilla y el tobillo, y nueve parámetros espacio-temporales que son la velocidad, la cadencia, las zancadas por minuto, el porcentaje de la fase de apoyo, el tiempo de paso, tiempo de zancada y los valores de ancho de paso, longitud de paso y longitud de zancada normalizados por la estatura de cada paciente.

En total se realizaron cuatro experimentos en los que se modificaron los datos utilizados para entrenar y evaluar el modelo de k-vecinos más cercanos. En el primero de ellos se utilizaron la totalidad de las 18 variables disponibles. En el segundo se aplicó el análisis de componentes principales (ACP) antes de entrenar el modelo, de modo que el ACP indicara el número de componentes principales mínimo requerido para describir las diferencias entre los datos. Posteriormente se entrenó el clasificador únicamente con las nueve variables cinemáticas y finalmente, se entrenó únicamente con las nueve variables espacio-temporales. Estas últimas pruebas tenían como finalidad identificar si uno de los dos tipos de variables resultaba más importantes para discriminar los grupos, reflejado en el desempeño del clasificador.

Para todos los experimentos cada variable fue normalizada con el método de puntaje z y los datos se dividieron aleatoriamente en dos grupos para el entrenamiento y evaluación del clasificador, en un 70 y 30 por ciento, respectivamente (899 pasos de entrenamiento, 383 pasos de evaluación). Se entrenó un clasificador con base en el algoritmo de k-vecinos más cercanos con cinco iteraciones de validación cruzada. El desempeño del clasificador se evaluó con una matriz de confusión y tomando como métricas principales de evaluación la exactitud y la precisión, calculadas de la siguiente manera:

$$\text{Exactitud} = (\text{Verdaderos positivos} + \text{Verdaderos negativos}) / \text{Total datos}$$

$$\text{Precisión} = \text{Verdaderos positivos} / (\text{Verdaderos positivos} + \text{Falsos positivos})$$

Todas las pruebas se realizaron con el software MATLAB R2019a (The MathWorks, Inc., Natick, Massachusetts, Estados Unidos).

## 2.4. Resultados

En la tabla 1 se muestran los resultados obtenidos en las bases de datos de entrenamiento y evaluación para las cuatro condiciones experimentales evaluadas. Al incluir el ACP, se encontró que seis componentes son suficientes para representar el 95% de la variabilidad de la base de datos.

**Tabla 1.** Resultados de la clasificación usando k-vecinos más cercanos en cuatro experimentos distintos

Base de datos	Entrenamiento (899 pasos)		Evaluación (383 pasos)	
	Exactitud	Precisión	Exactitud	Precisión
Datos Usados				
18 variables	97.9%	98.9%	100%	100%
ACP (6 componentes)	94.2%	95.4%	96.6%	98.7%
9 variables cinemáticas	93.7%	95.4%	92.1%	94.7%
9 variables espacio-temporales	94.2%	96.1%	96.8%	98.2%

## 2.5. Discusión

La exactitud y precisión obtenidas al usar un algoritmo básico de clasificación, como es k-vecinos más cercanos, son elevadas. Esto indica que las 18 variables cinemáticas y espacio-temporales usadas son indicadores suficientes para discriminar la marcha de sujetos con LM de la de sujetos sanos. De hecho, al entrenar el clasificador con los primeros 6 componentes principales de la base de datos, las métricas obtenidas son casi equivalentes, demostrando una alta correlación entre más de la mitad de las variables originales.

Al entrenar el mismo algoritmo únicamente con la información espacio-temporal o la cinemática, los resultados que se obtienen son ligeramente inferiores a los de los experimentos anteriores pero por encima de 92%, que puede considerarse como un desempeño satisfactorio. Esto indica que, de forma independiente, cualquiera de los dos tipos de información permite discriminar la marcha patológica de la no patológica, si bien las variables restantes aportan alguna información adicional que no refleja uno solo de los conjuntos de forma individual. Sin embargo, dado que el resultado obtenido al usar únicamente la información espacio-temporal es ligeramente superior al obtenido al usar solo la información cinemática, se puede sugerir que las primeras son más efectivas para discriminar la marcha de las poblaciones. Esto puede atribuirse a que es a nivel funcional donde se observa mayor diferencia entre ambos grupos (Thibaudier et al., 2020; Nair et al. 2012). La mejora de indicadores espaciotemporales como la cadencia, la longitud de paso y la velocidad se asocian con objetivos de rehabilitación comunes directamente relacionados con pruebas para la evaluación de la marcha usados en la práctica clínica como el Time Up and Go, el test de los diez metros o el test de los seis minutos. Éstos reflejan cambios a nivel fisiológico como mejoras en el equilibrio y en el uso eficiente de la energía invertida para caminar o reducciones en el riesgo de caída (Rini et al., 2018; Pérez-Sanpablo et al., 2017; Amatachaya et al., 2014). En la vida cotidiana, estos indicadores se relacionan con factores que afectan la calidad de vida de los pacientes (Nilsagard et al., 2007).

## 3. Conclusiones

La información cinemática y espacio-temporal recopilada en análisis tridimensionales de la marcha permite discriminar la marcha de pacientes con LM de la de sujetos sanos con exactitud y precisión al ser usada para entrenar un modelo de k-vecinos más cercanos. Un desempeño similar se consigue usando únicamente las variables espacio-temporales o cinemáticas para entrenar el mismo algoritmo. Los avances en el uso de técnicas de aprendizaje de máquina en el análisis de la marcha podrían permitir que estas brinden un apoyo al proceso de diagnóstico de patologías de la marcha. Adicionalmente, estas herramientas permiten identificar de forma precisa los aspectos en los que hay más diferencia entre la marcha patológica y la no patológica, lo que podría ser de utilidad en el desarrollo de herramientas y estrategias de rehabilitación, así como en métricas orientadas a la evaluación integral de la marcha.

## Agradecimientos

Las instituciones pertenecientes al consorcio TAILOR agradecen la financiación concedida por la Agencia Estatal de Investigación en la convocatoria 2018, con referencia RTI2018-097290-B-C31.

## Referencias

1. Amatachaya, S., Naewla, S., Srisim, K., Arrayawichanon, P., & Siritaratiwat, W. (2014). Concurrent validity of the 10-meter walk test as compared with the 6-minute walk test in patients with spinal cord injury at various levels of ability. *Spinal cord*, 52(4), 333–336. <https://doi.org/10.1038/sc.2013.171>
2. DeVries, Z., Hoda, M., Rivers, C., Maher, A., Wai, E., Moravek, D., Stratton, A., Kingwell, S., Fallah, N., Parquet, J., Phan, P. (2019). Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *The Spine Journal*. doi:10.1016/j.spinee.2019.09.007
3. Krawetz, P., & Nance, P. (1996). Gait analysis of spinal cord injured subjects: effects of injury level and spasticity. *Archives of physical medicine and rehabilitation*, 77(7), 635–638. [https://doi.org/10.1016/s0003-9993\(96\)90000-3](https://doi.org/10.1016/s0003-9993(96)90000-3)
4. Nair, P. M., Hornby T, G., & Behrman, A. L. (2012). Minimal detectable change for spatial and temporal measurements of gait after incomplete spinal cord injury. *Topics in spinal cord injury rehabilitation*, 18(3), 273–281. <https://doi.org/10.1310/sci1803-273>
5. Nilsagard, Y., Lundholm, C., Gunnarsson, L. G., & Dcnison, E. (2007). Clinical relevance using timed walk tests and 'timed up and go' testing in persons with multiple sclerosis. *Physiotherapy research international: the journal for researchers and clinicians in physical therapy*, 12(2), 105–114. <https://doi.org/10.1002/pri.358>
6. O'Malley, M. J., Abel, M. F., Damiano, D. L., & Vaughan, C. L. (1997). Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE Transactions on Rehabilitation Engineering*, 5(4), 300–309. doi:10.1109/86.650282
7. Pérez-Sanpablo, A. I., Quinzaños-Fresnedo, J., Loera-Cruz, R., Quiñones-Uriostegui, I., Rodríguez-Reyes, G., & Pérez-Zavala, R. (2017). Validation of the instrumented evaluation of spatio-temporal gait parameters in patients with motor incomplete spinal cord injury. *Spinal cord*, 55(7), 712. <https://doi.org/10.1038/sc.2017.51>
8. Prakash, C., Kumar, R., & Mittal, N. (2016). Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review*, 49(1), 1–40. doi:10.1007/s10462-016-9514-6
9. Prakash, C., Kumar, R., & Mittal, N. (2018). Optimized Clustering Techniques for Gait Profiling in Children with Cerebral Palsy for Rehabilitation. *The Computer Journal*, 61(11), 1683–1694, <https://doi.org/10.1093/comjnl/bxy022>
10. Rini, D., Senthilvelkumar, T., Noble, K., & Magimairaj, H. (2018). Test–retest reliability of the 10-meter walk test in ambulatory adults with motor-complete spinal cord injury. *International Journal of Therapy and Rehabilitation*, 25(7), 335–339. doi:10.12968/ijtr.2018.25.7.335
11. Simpson, L., Eng, J., Hsieh, J., and Wolfe, D., & the Spinal Cord Injury Rehabilitation Evidence (SCIRE). (2013). The health and life priorities of individuals with spinal cord injury: A systematic review. *J. Neurotrauma*, 29(8), 1548–1555. <http://doi.org/10.1089/neu.2011.2226>

12. Stampacchia, G., Olivieri, M., Rustici, A., D'Avino, C., Gerini, A., & Mazzoleni, S. (2020). Gait rehabilitation in persons with spinal cord injury using innovative technologies: an observational study. *Spinal cord*, 58(9), 988–997. <https://doi.org/10.1038/s41393-020-0454-2>
13. Thibaudier, Y., Tan, A. Q., Peters, D. M., & Trumbower, R. D. (2020). Differential deficits in spatial and temporal interlimb coordination during walking in persons with incomplete spinal cord injury. *Gait & posture*, 75, 121–128. <https://doi.org/10.1016/j.gaitpost.2019.10.023>
14. Timotius, I. K., Bieler, L., Couillard-Despres, S., Sandner, B., Garcia-Ovejero, D., Labombarda, F., Estrada, V., Müller, H. W., Winkler, J., Klucken, J., Eskofier, B., Weidner, N., & Puttagunta, R. (2021). Combination of Defined CatWalk Gait Parameters for Predictive Locomotion Recovery in Experimental Spinal Cord Injury Rat Models. *eNeuro*, 8(2), ENEURO.0497-20.2021. <https://doi.org/10.1523/ENEURO.0497-20.2021>
15. Wong, M. A., Simon, S., & Olshen, R. A. (2007). Statistical analysis of gait patterns of persons with cerebral palsy. *Statistics in Medicine*, 2(3), 345–354. doi:10.1002/sim.4780020306
16. Xu, G., Zhang, Y., & Begg, R. (2006). Mining Gait Pattern for Clinical Locomotion Diagnosis Based on Clustering Techniques. *Lecture Notes in Computer Science*, 296–307. doi:10.1007/11811305\_33