

## Control longitudinal de un vehículo mediante aprendizaje por refuerzo profundo

Felipe Barreno<sup>a,\*</sup>, Matilde Santos<sup>b</sup>, Manuel G. Romana<sup>c</sup>

<sup>a</sup> Facultad de Ingeniería Informática, Universidad de Complutense, Madrid, España

<sup>b</sup> Instituto de Tecnología del Conocimiento, Universidad de Complutense, Madrid, España

<sup>c</sup> Escuela de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Madrid, España.

**To cite this article:** Barreno F., Santos, M., Romana, M.G.. 2023. Deep reinforcement learning vehicle longitudinal control. XLIV Jornadas de Automática, 127-131.  
<https://doi.org/10.17979/spudc.9788497498609.127>

### Resumen

Este artículo presenta un sistema inteligente para el control longitudinal de un vehículo mediante aprendizaje por refuerzo profundo basado en la influencia de la curvatura de la carretera. El sistema inteligente consiste en un agente que usa el algoritmo de gradiente de política determinista profundo (DDPG) para controlar la velocidad. Para entrenar al agente del modelo, el efecto de la curvatura de la carretera se considera a través de la aceleración percibida obtenida a partir de la aceleración lateral y velocidad angular debida a la propia carretera. Los resultados del sistema inteligente son valores continuos de aceleración. El modelo propuesto ofrece resultados prometedores, lo que sugiere que este sistema inteligente puede ayudar al conductor y que el sistema de control del vehículo puede aplicarse a la conducción semiautónoma o autónoma haciendo que la conducción sea más segura y cómoda.

*Palabras clave:* Control, aprendizaje por refuerzo, aprendizaje profundo, vehículos autónomos, control adaptativo.

### Deep reinforcement learning vehicle longitudinal control

#### Abstract

This paper presents an intelligent system for longitudinal control of a vehicle using deep reinforcement learning based on the influence of road curvature. The intelligent system consists of an agent based on the deep deterministic policy gradient (DDPG) algorithm for speed control. To train the model agent, the road curvature effect is considered through the perceived acceleration obtained from the lateral acceleration and angular velocity due to the road itself. The results of the intelligent system are continuous acceleration values. The proposed model offers promising results, suggesting that this intelligent system can assist the driver and that the vehicle control system can be applied to semi-autonomous or autonomous driving making driving safer and more comfortable.

*Keywords:* Control, reinforcement learning, deep learning, autonomous vehicles, adaptive control.

## 1. Introducción

Los sistemas avanzados de asistencia al conductor (ADAS, por las siglas en inglés de *Advanced Driver Assistance Systems*) ayudan al conductor a controlar el vehículo, y pueden prestar asistencia en maniobras peligrosas, distracciones del conductor o situaciones de alto riesgo. La homologación de los sistemas ADAS y la introducción gradual de coches autónomos en la carretera cambiarán la experiencia de conducción en muchos y variados aspectos, como el control de

la dirección, la aceleración y el frenado, y por lo tanto en la seguridad en general (Mekala, et al, 2021).

La forma en que el conductor percibe las distintas propiedades geométricas de la carretera puede dar lugar a comportamientos de conducción diferentes en función del tipo y las características de la carretera (Martín et al., 2016), Barreno et al., 2022a). La dinámica del vehículo permite reconocer determinadas características de la geometría de la carretera. Estas características pueden derivarse de la información inercial, como las lecturas del acelerómetro y el giroscopio, los datos del GPS y los datos del bus CAN, que han

\*Autor para correspondencia: febarren@ucm.es

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

demostrado su utilidad en este campo (Van Ly, et al., 2013). Estos datos se pueden utilizar para obtener ciertas características de la conducción y reconocer implícitamente cómo es la geometría de la carretera.

El objetivo de este trabajo es diseñar un sistema de control inteligente de un vehículo utilizando las características extraídas de la geometría de la carretera durante la conducción, y obtener los valores adecuados de aceleración longitudinal y ángulo de giro. Se ha utilizado un modelo experimental de vehículo de 3 grados de libertad. El sistema de control se ha diseñado con aprendizaje profundo por refuerzo. El sistema inteligente aquí propuesto para la asistencia a la conducción se centra en caracterizar la conducción del vehículo en función del efecto de la curvatura de la carretera, de forma que incorpora la geometría de la misma a través de ciertas características relacionadas con ella, como la aceleración longitudinal y lateral del vehículo, percibidas por el conductor al tomar las curvas (Barreno et al., 2022b). Esto contribuye a una conducción más segura, ya que una velocidad inadecuada en una curva podría ser insegura para un determinado tramo de carretera.

Los métodos de aprendizaje profundo se han mostrado muy prometedores no solo al proporcionar un rendimiento excelente para problemas de control complejos y no lineales, sino también al generalizar reglas previamente aprendidas a nuevos escenarios (Kuutti, et al, 2020). En (Zhang, et al., 2022) se propone un modelo de seguimiento construido con un proceso de decisión de Markov, que incorpora algoritmos de aprendizaje por refuerzo profundo y utiliza una red Double Deep Q para establecer una estrategia de seguimiento de automóviles. En (Qin et al., 2022) se propone un modelo multiagente de control longitudinal y lateral de un vehículo para proporcionar una referencia a un sistema de control de cruce adaptativo que tenga en cuenta la estabilidad lateral del vehículo. (Pérez et al., 2010) propone un sistema de control longitudinal de un vehículo ajustado mediante técnicas neuro-difusas. Los autores en (Gongsheng, et al., 2020) se basan en escenas de carretera reales haciendo uso de un simulador de conducción autónoma (AirSim), utilizando técnicas de aprendizaje profundo para realizar una conducción autónoma. En (Berahman et al., 20) se utiliza una combinación de fuerzas artificiales (fuerzas laterales y longitudinales del vehículo) y un enfoque de aprendizaje por refuerzo utilizando el simulador SUMO con el objetivo de generar una estrategia de conducción para vehículos sin carriles.

## 2. Identificación de la dinámica de un vehículo

Durante la conducción, un vehículo está sometido a fuerzas que afectan a su dinámica (ASSHTO, 2011). Si se accionan los frenos, el vehículo decelera; si se pisa el acelerador, el vehículo aumenta de velocidad. Si el vehículo gira mediante una maniobra de la dirección, cuanto mayor sea la velocidad del vehículo, mayor será la fuerza centrífuga sobre él. Esto configura su comportamiento dinámico.

Por otro lado, las carreteras se definen por sus características geométricas, que determinan la velocidad a la que un vehículo puede circular con un grado adecuado de comodidad y seguridad. La velocidad de diseño de un tramo de carretera ( $V$ ), independientemente de la velocidad máxima

permitida por la administración, está relacionada con sus características geométricas, principalmente con el radio de curvatura ( $R$ ), según:

$$V^2 = 127R(f_t + \frac{\rho}{100}) \quad (1)$$

La aceleración medida en la dirección de avance del vehículo es la aceleración longitudinal, paralela a la velocidad lineal. La aceleración medida en la dirección ortogonal del vehículo es la aceleración lateral. La aceleración debida a la carretera se calcula (Rajamani, 2011):

$$a_{road} = g(\rho + f_t) \quad (2)$$

donde  $a_{road}$  ( $m/s^2$ ), es la aceleración crítica debida a los efectos de la geometría de la carretera,  $g$  ( $m/s^2$ ) es la aceleración de la gravedad,  $\rho$  ( $m$ ) es la pendiente transversal de la carretera y  $f_t$  es el coeficiente máximo de fricción transversal movilizada. Esta aceleración de la carretera es la aceleración límite de una curva horizontal para mantener la seguridad y el confort. Independientemente de la pendiente transversal y del coeficiente de fricción movilizado máximo, que son difíciles de obtener,  $a_{road}$  puede determinarse a partir de sensores inerciales. La aceleración debida al efecto de la carretera puede definirse como:

$$a_{road} = |\omega| \cdot v_l \quad (3)$$

donde  $\omega$  ( $rad/s$ ) es la velocidad angular y  $v_l$  ( $m/s$ ) es la velocidad lineal del vehículo. La aceleración lateral percibida por el conductor es (Barreno, F. et al, 2022c):

$$a_p = |a_m| - a_{road} \quad (4)$$

Donde  $a_p$  ( $m/s^2$ ) es la aceleración percibida por el conductor y  $a_m$  ( $m/s^2$ ) es la aceleración lateral medida. Desde el punto de vista del conductor, esta aceleración refleja la "sensación" de aceleración debida al efecto de la geometría de la carretera.

## 3. Control longitudinal del vehículo considerando la influencia de la geometría de la carretera

El sistema de control longitudinal de vehículo está basado en modelo de aprendizaje por refuerzo profundo que se implementa teniendo en cuenta el efecto de la curvatura de la carretera. El efecto de la geometría de la carretera se obtiene a partir de las fuerzas longitudinales y laterales obtenidas a través de medidas de aceleración longitudinal y lateral, lecturas de guiñada y ángulo de dirección.

La caracterización del estilo de conducción al aproximarse a una curva se basa en la siguiente consideración. Si la velocidad de giro actual es demasiado alta, la aceleración lateral será elevada, lo que provocará una mala maniobrabilidad del vehículo. De hecho, la aceleración lateral percibida, que evalúa la sensación del conductor durante la conducción, será perceptible, por lo que el efecto de la curvatura de la carretera se extrae indirectamente a través de la dinámica del vehículo mediante la aceleración longitudinal y lateral, y la velocidad de guiñada.

El aprendizaje por refuerzo profundo con el vehículo autónomo se aplica utilizando el software Matlab/Simulink. El entorno consiste en un modelo 3-DOF para el vehículo ego y un modelo longitudinal simple para el vehículo líder. El objetivo del entrenamiento es conseguir que el ego-car circule a una velocidad de referencia manteniendo una distancia de seguridad con el vehículo líder mediante el control de la aceleración longitudinal, sin sobrepasar en media de lo posible la velocidad del vehículo líder. El controlador longitudinal tiene en cuenta la interacción del ego-vehículo con los vehículos que circulan por delante, mientras que el controlador lateral sólo tiene en cuenta los elementos que afectan al movimiento lateral del ego-vehículo.

La interacción entre el ego-vehículo y el vehículo que circula delante se define del siguiente modo. Si la distancia relativa es inferior a la distancia de seguridad, el vehículo sigue el mínimo de la velocidad del vehículo precedente y la  $v_{ref}$  establecida. De este modo, el vehículo mantiene una cierta distancia con el vehículo que le precede. Si la distancia relativa es superior a la distancia de seguridad, el vehículo sigue la  $v_{ref}$  establecida. Se implementa la curvatura variable de la carretera para un enfoque más realista.

El ciclo UDC (por las siglas en inglés *Urban Driving Cycle*) es un ciclo urbano o interurbano, diseñado para representar las condiciones de conducción. Se elige como ciclo de referencia para simular la prueba en carretera de los vehículos comerciales ligeros. El ciclo de conducción UDC estándar consta de varias fases caracterizadas por aceleración, velocidad constante y deceleración (Xie, C., et al, 2013). Para este trabajo, se han implementado dos perfiles UDC para evaluar el rendimiento del modelo de seguimiento de trayectoria propuesto (Figura 1).

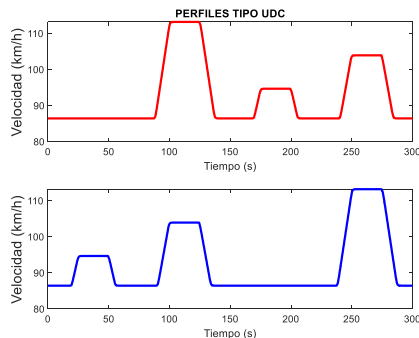


Figura 1: Perfiles UDC de conducción usados.

Así, a partir de las características medidas de la dinámica del vehículo (Tabla 1) es posible obtener la aceleración percibida por el usuario y, de esta forma, el efecto de curvatura de la carretera en cada transición de curva.

Tabla 1: Características de la dinámica del vehículo.

| Variables  | Descripción              | Unidades |
|------------|--------------------------|----------|
| $a_{long}$ | Aceleración longitudinal | $m/s^2$  |
| $a_{lat}$  | Aceleración lateral      | $m/s^2$  |
| $a_p$      | Aceleración percibida    | $m/s^2$  |
| $\omega$   | Velocidad angular        | rad/s    |
| $J_k$      | Sobre-aceleración (Jerk) | $m/s^3$  |
| $v_l$      | Velocidad del vehículo   | km/h     |

En este trabajo, se utiliza el aprendizaje profundo por refuerzo para entrenar políticas utilizando el agente DDPG (por las siglas en inglés de *Deep Deterministic Policy Gradient*), permitiéndole interactuar con un entorno en el que el vehículo está en movimiento. Estas políticas se utilizan para implementar los sistemas de control de aceleración longitudinal. La plataforma Simulink se utiliza para evaluar el algoritmo DDPG, así como, probarlo con ajustes de parámetros y supervisar el progreso del entrenamiento.

### 3.1. Agente DDPG

El control longitudinal es un agente DDPG. El DDPG estima la recompensa de las observaciones y las acciones utilizando una representación de la función de valor crítico, y selecciona las acciones aplicando una representación de la política del agente. Las observaciones del entorno son las medidas longitudinales: la velocidad longitudinal del ego-car, el error de velocidad  $e_v$ , es decir, la diferencia entre la velocidad de referencia del ego-car y la velocidad lineal del ego-car, su integral, la aceleración lateral percibida y el jerk longitudinal. La señal de acción consiste en valores continuos de aceleración entre  $-3$  y  $2$   $m/s^2$ . La recompensa  $r_t$  se obtiene en cada tiempo  $t$  de la siguiente manera (5):

$$r(t) = - \left( \frac{e_v^2}{100} + \frac{a_{(t-1)}^2}{10} + \frac{a_{p(t-1)}^2}{10} + \frac{j_k^2(t-1)}{10} \right) + C_t + D_t \quad (5)$$

donde,  $e_v$  es el error de velocidad (entre el coche líder y el ego),  $a_{(t-1)}$  es la aceleración longitudinal del paso de tiempo anterior,  $a_{p(t-1)}$  es la aceleración lateral percibida,  $j_k(t-1)$  es el tirón longitudinal (sobre-aceleración) para medir cambios repentinos en la aceleración longitudinal, y C y D son constantes que dependen de:

$$\begin{aligned} C_t &= 1 \text{ si } e_v < 0.1, \text{ en otro caso } C_t = 0 \\ D_t &= 1 \text{ si } a_p < 0.25, \text{ en otro caso } D_t = 0 \end{aligned} \quad (6)$$

Los parámetros configurados para el entrenamiento del agente DDPG para el control longitudinal se muestran en la Tabla 2.

Tabla 2: Parámetros para entrenamiento DDPG.

| Descripción                    | Valor   |
|--------------------------------|---------|
| Ratio de aprendizaje           | 0.001   |
| Método del umbral de gradiente | L2 norm |
| Factor de descuento            | 0.99    |
| Tamaño del mini-lote           | 64      |
| Tiempo simulación              | 60s     |
| Tiempo de muestreo             | 0.1s    |

## 3. Resultados

La experiencia del usuario de la carretera, tanto en términos de comodidad como de seguridad, depende en gran medida de las condiciones de la vía. Así pues, la geometría de la carretera influye considerablemente en la seguridad o la percepción de la conducción.

En los experimentos de simulación se muestra la velocidad obtenida por el sistema de control longitudinal junto con otras características de la carretera. La velocidad obtenida tiene en cuenta las condiciones reales de la geometría de la carretera mediante la curvatura variable de la misma.

El tiempo de entrenamiento del agente DDPG, para cada experimento las simulaciones duraron más de una hora cada una. Las características del ordenador utilizado son procesador IntelCore i7 1,2 GHz con 8 Gb RAM.

Las siguientes figuras muestran los resultados obtenidos durante el entrenamiento del agente DDPG (Figura 2) y los resultados obtenidos para el control longitudinal del vehículo implementado con el agente entrenado con un perfil de velocidad tipo UDC para el vehículo (Figura 4).

En cada figura, el gráfico superior muestra la velocidad del vehículo ego (rojo), la del vehículo líder (naranja) y la  $v_{ref}$  del vehículo ego (azul) en km/h. El segundo gráfico representa la aceleración longitudinal instantánea ( $m/s^2$ ); el tercer gráfico muestra la aceleración lateral percibida ( $m/s^2$ ); el cuarto gráfico inferior muestra la sobre-aceleración o jerk del vehículo ( $m/s^3$ ). Este jerk es la tasa de cambio de la aceleración de un objeto con respecto al tiempo. Por lo tanto, es importante evitar los tirones bruscos del vehículo para mejorar el confort. El gráfico inferior muestra la curvatura de carretera en  $m^{-1}$ .

La Figura 2 representa los resultados obtenidos por el sistema de control longitudinal del vehículo durante el entrenamiento. El primer gráfico muestra la velocidad del vehículo ego (m/s), la velocidad de referencia  $v_{ref}$  para el vehículo ego y la velocidad del vehículo líder. En este caso para el entrenamiento ha sido generada una señal senoidal para el vehículo líder. Los valores de aceleración longitudinal ( $m/s^2$ ) se representan en el 2º gráfico y aceleración lateral percibida ( $m/s^2$ ) en el 3º. Se puede observar cómo se van atenuando a lo largo del tiempo de entrenamiento. Por último, en el gráfico inferior se muestra el jerk ( $m/s^3$ ), que también se va atenuando a lo largo del tiempo de entrenamiento.

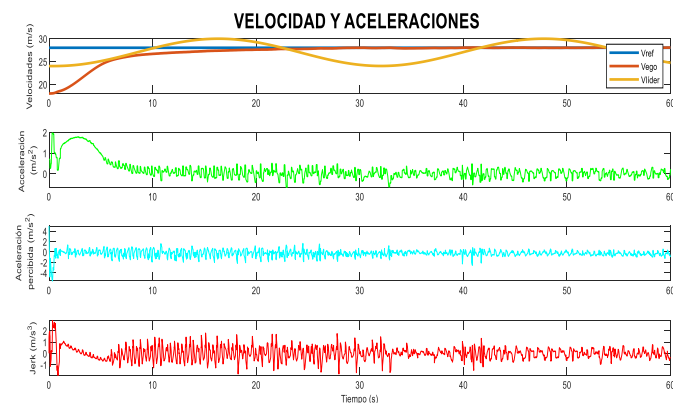


Figura 2: Velocidades y aceleraciones obtenidas durante el entrenamiento del agente DDPG.

La Figura 3 muestra las métricas de recompensa por cada episodio de simulación. La figura muestra la recompensa en cada episodio de simulación y la recompensa media sobre los episodios acumulados. El final de entrenamiento llega cuando la recompensa alcanza a un valor de 600 unidades.

La Figura 4 representa los resultados obtenidos por el sistema de control longitudinal del vehículo. El primer gráfico

muestra la velocidad del vehículo ego (km/h), la velocidad de referencia  $v_{ref}$  para el vehículo ego y la velocidad del vehículo líder. El vehículo ego aumenta su velocidad para alcanzar la velocidad de referencia o hasta igualar la velocidad del vehículo líder si ésta es inferior a  $v_{ref}$ . Los valores de aceleración longitudinal ( $m/s^2$ ) y aceleración lateral percibida ( $m/s^2$ ) se muestran en los gráficos 2º y 3º, respectivamente. Si la curvatura de la carretera aumenta, el radio de curvatura es menor, por lo que se producen cambios en la aceleración longitudinal y la sacudida para reducir o aumentar la velocidad del vehículo para adaptar su velocidad a la curvatura de la carretera. Los cambios en la aceleración longitudinal no suelen ser bruscos y, si hay cambios repentinos, se producen en la sección de la carretera en la que aumenta la curvatura. Además, la aceleración lateral percibida se mantiene dentro del intervalo de  $-0,5$  a  $0,5 m/s^2$ , lo que hace que la conducción sea cómoda y agradable (De Winkel, K. N., et al, 2023). Como se muestra en la cuarta gráfica, el jerk ( $m/s^3$ ) es cercano a cero excepto en la zona donde cambia la curvatura. El radio de curvatura en esta zona es de 100 m entre los 150-180 segundos aproximadamente, el resto del trazado tiene un radio de curvatura grande, lo que aproximaría a una carretera con un trazado recto.

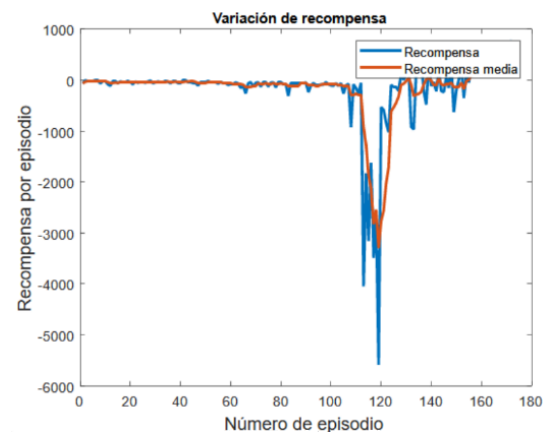


Figura 3: Variación de recompensa durante el entrenamiento.

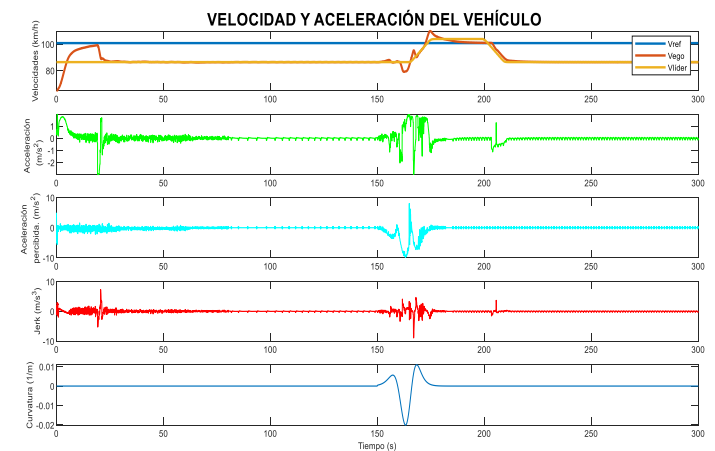


Figura 4: Velocidades y aceleraciones obtenidas del ego vehículo durante el seguimiento a otro vehículo líder.

#### 4. Conclusiones y trabajos futuros

En este trabajo se ha diseñado un sistema inteligente de seguimiento de trayectorias que tiene en cuenta el efecto de la geometría de la carretera mediante el refuerzo profundo. Utiliza como entradas las aceleraciones obtenidas de los sensores del coche. A partir de estas medidas, es posible obtener la aceleración debida a la carretera y la aceleración percibida por el conductor. Esta información permite entrenar al agente DDPG para que proporcione valores de aceleración al sistema de control longitudinal del vehículo.

El resultado obtenido con este sistema de control longitudinal mediante aprendizaje por refuerzo es alentador y útil para complementar otros sistemas de asistencia al conductor instalados en el vehículo.

Se pueden extraer varias conclusiones. El sistema utiliza como entradas las lecturas de aceleración y giroscopio, medidas que pueden tomarse de los sensores de medición inercial que incorporan todos los vehículos actuales. Por otro lado, el efecto de la curva de la carretera durante la conducción se tiene en cuenta mediante las percepciones subjetivas del conductor para adoptar un comportamiento de conducción seguro y cómodo.

Los trabajos futuros podrían tener en cuenta otros factores que empiezan a ser medidos por sensores inteligentes, como la somnolencia, la distracción, la fatiga, la interacción rueda/carretera, las condiciones meteorológicas, etc.

#### Referencias

- Barreno, F., Romana, M. G., & Santos, M. (2022a). Fuzzy expert system for road type identification and risk assessment of conventional two-lane roads. *Expert Systems*, 39(9), e12837. DOI:10.1111/exsy.12837.
- Barreno, F., Santos, M., & Romana, M. G. (2022b). A novel adaptive vehicle speed recommender fuzzy system for autonomous vehicles on conventional two-lane roads. *Expert Systems*, e13046. DOI:10.1111/exsy.13046.
- Barreno, F., Santos, M., & Romana, M. (2022c). Fuzzy Logic System for Risk and Energy Efficiency Estimation of Driving Maneuvers. In 14th International Conference on Computational Intelligence in Security for Information Systems and 12th International Conference on European Transnational Educational (CISIS 2021 and ICEUTE 2021) 14 (pp. 94-104). Springer International Publishing.
- Berahman, M., Rostmai-Shahrbabaki, M., & Bogenberger, K. (2022). Driving Strategy for Vehicles in Lane-Free Traffic Environment Based on Deep Deterministic Policy Gradient and Artificial Forces. *IFAC-PapersOnLine*, 55(14), 14-21.
- De Winkel, K. N., Irmak, T., Happee, R., & Shyrokau, B. (2023). Standards for passenger comfort in automated vehicles: Acceleration and jerk. *Applied Ergonomics*, 106, 103881.
- Gongsheng, Z., Chunmei, P., Jiang, D., & Junfeng, S. (2020, December). Deep Deterministic Policy Gradient Algorithm based Lateral and Longitudinal Control for Autonomous Driving. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) (pp. 740-745). IEEE.
- Kuutti, S., Bowden, R., Jin, Y., Barber, P., & Fallah, S. (2020). A survey of deep learning applications to autonomous vehicle control. *IEEE Transactions on Intelligent Transportation Systems*, 22(2), 712-733.
- Martin, S., Romana, M. G., & Santos, M. (2016). Fuzzy model of vehicle delay to determine the level of service of two-lane roads. *Expert Systems with Applications*, 54, 48-60.
- Mekala, M. S., Dhiman, G., Patan, R., Kallam, S., Ramana, K., Yadav, K., & Alharbi, A. O. (2021). Deep learning-influenced joint vehicle-to-infrastructure and vehicle-to-vehicle communication approach for internet of vehicles. *Expert Systems*, e12815, pp. 15. DOI:10.1111/exsy.12815
- Pérez, J., Gajate, A., Milanés, V., Onieva, E., & Santos, M. (2010, July). Design and implementation of a neuro-fuzzy system for longitudinal control of autonomous vehicles. In *International Conference on Fuzzy Systems* (pp. 1-6). IEEE.
- Qin, P., Tan, H., Li, H., & Wen, X. (2022). Deep Reinforcement Learning Car-Following Model Considering Longitudinal and Lateral Control. *Sustainability*, 14(24), 16705.
- Rajamani, R. (2011). *Vehicle dynamics and control*. Springer Science & Business Media.
- Transportation Officials. (2011). *A Policy on Geometric Design of Highways and Streets*, 2011. AASHTO.
- Van Ly, M., Martin, S., & Trivedi, M. M. (2013, June). Driver classification and driving style recognition using inertial sensors. In 2013 IEEE Intelligent Vehicles Symposium (IV) (pp. 1040-1045). IEEE.
- Xie, C., Ogden, J. M., Quan, S., & Chen, Q. (2013). Optimal power management for fuel cell-battery full hybrid powertrain on a test station. *International Journal of Electrical Power & Energy Systems*, 53, 307-320.
- Zhang, C., Zhang, X., Ma, P., Dai, S., Lu, Y., & Jiang, L. (2022, January). Vehicle Driving Longitudinal Control Based on Double Deep Q Network. In 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 273-275). IEEE.