An AI Approach for Analyzing Driving Behaviour in Simulated Racing Using Telemetry Data

Chapter · November 2023

DOI: 10.1007/978-3-031-49065-1_19

4 authors:

Fazilat Hojaji
University of Limerick

Adam Joseph Toth
University of Limerick

John Joyce
University of Limerick

Mark J Campbell
University of Limerick

An AI Approach for Analyzing Driving Behaviour in Simulated Racing Using Telemetry Data

Fazilat Hojaji(✉), Adam J. Toth, John M. Joyce, and Mark J. Campbell

Esports Science Research Laboratory, Lero Irish Software Research Centre, University of Limerick, Limerick, Ireland
{Fazilat.Hojaji,Adam.Toth,John.M.Joyce,Mark.Campbell}@ul.ie

Abstract. The emerging and rapid progress of esports (competitive computer gaming) currently lacks approaches for ensuring high-quality analytics to augment performance in professional and amateur esports teams. In this paper, we demonstrate the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques in the esports domain, particularly in simulation (sim) racing, for analyzing drivers' behaviour based on telemetry data from race drivers. To achieve this, we used a professional racing simulator to collect a wide range of feature-rich telemetry data from 93 participants through MoTec telemetry software and the ACC sim racing gaming platform. An objective assessment of the characteristics of the driver's behaviour was then obtained through a set of predefined lap-based metrics derived from telemetry data. Additionally, a comparison of driving styles was carried out using machine learning approaches for grouping the acquired laps based on performance (lap time). The findings from our analysis contribute to a better understanding of how elite drivers differ from low skilled drivers based on their telemetry. Furthermore, our findings provide researchers with key metrics to develop more efficient training tools and techniques to improve sim racing performance.

Keywords: MoTec · Sim racing · Machine Learning · Artificial Intelligence

1 Introduction

Esports analytics, which may be considered a part of game analytics or sports analytics [1, 2], refers to the method of analyzing esports-related data to uncover significant patterns and trends in the data and then transmitting these patterns using visualization approaches to support decision-making processes [3]. Esports analytics has benefited from a variety of community-based projects aimed at helping players access, analyze, and derive meaning from data. In recent years, many new data analysis techniques have been used for processing and analyzing data to extract new knowledge, and these techniques have been well leveraged for improving players' performance [4]. However, there is a lack of tools that provide player performance feedback and suggestions for how to improve [5]. This opens plenty of new opportunities for esports research to determine what makes a gamer successful.

P. Dondio et al. (Eds.): GALA 2023, LNCS 14475, pp. 194–203, 2024.
https://doi.org/10.1007/978-3-031-49065-1_19

In this paper, we focus on simulated motorsport, which is one of the oldest genres in the world of esports. The phrase "sim racing" refers to any computer programs that attempt to properly simulate car racing, along with real-world elements like gasoline consumption, damage, tyre wear and grip, and suspension settings [6]. To be successful in sim racing, a driver must be knowledgeable about all the aspects of vehicle handling that make real-world racing so challenging [7], including threshold braking, maintaining control of a vehicle as the tires lose traction, and correctly entering and exiting turns without losing speed. Technological enhancement in the computer-based simulator domain of simulated driving contributes to the direct improvement of the team and sim racer performance [8]. In this case, solutions, and strategies for becoming the fastest driver are of utmost importance, with various methods of data analysis and data collection tools being used for sim racers. (See [9–12] as examples).

One challenge that researchers face in understanding how elite racers control vehicles at the limits of handling is the lack of ability to obtain data from highly dynamic driving. Racing organisations may have access to this information for internal analysis, but it is not publicly available. Until relatively recently, researchers only used in-game data to analyze driver behaviour and estimate a driver's performance. In this case, the ability to compare experimental and simulated driving behaviour is limited by available sensor information. The public may now obtain telemetry data from sim racing video games thanks to APIs and telemetry tools (e.g., vTelemetry PRO [13] and Motec [14]). Using such tools, the physical and control parameters of the simulation may be tracked and saved as telemetry files [8]. It allows sim racers to gather all the information provided by the vehicle and analyze the data captured during a race or session [15]. Insights from telemetry data lead to a better understanding of the corresponding strengths and weaknesses of the drivers' behaviour and can foster performance improvements through more accurate tuning of their car setup as well as informing them on driving strategies and techniques [16]. By leveraging sim telemetry data collected from sim racers and utilising the state of-the-art AI approach, the purpose of this study is to model and analyze race driver behaviour that may reflect a wide range of driver control features and, as a result, provide adaptations for many of the characteristics of human drivers.

While there are number of studies that investigate simulated driving style for general road cars [17, 18], few efforts have been made to analyze driving behaviour in a simulated racing environment. Some models, such as those employed in Formula One driving [19, 20] or related types of automated or more idealised driver performance, are primarily concerned with achieving perfect performance [15, 21]. These studies rarely include sim telemetry data from entire racing circuits. Even if they rely on telemetry data in their analysis [15, 22], they utilize only a limited number of parameters (i.e., steering wheel, throttle, pedal, and brake pedal) in their analysis and most of the telemetry in-game data is omitted.

This study combines sim telemetry data for entire laps from 93 sim racers through MoTec telemetry software and the ACC sim racing game and presents a structural method to define a set of lap-based metrics that objectively characterise various driving styles. Using ML techniques, these metric sets were then used to cluster elite and low skilled driver behaviour. Insights from this study will lead to a better understanding of the driver's behaviour and the key components that augment learning and performance.

2 Data and Methods

2.1 Apparatus

To address the research goal, we applied the following approach to data collection in the experimental design. Through the ACC simulated racing game, we employed a professional racing simulator to gather a variety of feature-rich telemetry data from 93 participants. The steering wheel, manual gearbox, and control pedals (brake, clutch, and accelerator; Logitech the Logitech Pro wheel and pedals) formed up the simulator's vehicle control interface. The participants were instructed to drive as quickly as they could while keeping the car on the track during the driving task. Before the driving session, general questions were posed about driving passenger cars. 85% had a driving licence and drove on average 10.14 (SD = 10.04) hours a week. In addition, the participants provided a self-assessment of their driving skills in video racing games, which resulted in an average of 12.32 (SD = 9.29) hours a week.

We augmented ACC with MoTec i2 Pro (v1.1.5, Melbourne, Victoria, Australia), a professional telemetric data analysis application, and routed all telemetric data to the data analysis package. The vehicle telemetry was recorded for all participants. Vehicle telemetry includes numerous channels, from which we chose those related to vehicle control. To extract telemetry data, we used Motec i2 Pro available from the MoTec website. We used Python (3.9) as our programming language on the Anaconda3 (Spyder 5.2) platform for the development of the pre-processing and analysis processes.

2.2 Data Analysis

Data preprocessing, we extracted all data channel statistics corresponding to in-game driver and vehicle features. Note that Motec records data for up to 84 metrics, including the driver's input data, chassis and suspension data, brake and tyre data, engine data, wheel data, balancing data, g-force, and car position data. To capture the driver's behaviour, time series data with a sample rate of 50 Hz were extracted. As a result, three types of data files, including a summary file of lap times, channel statistic data, and time series driving data, were obtained for each participant. Following a preliminary screening and cleaning of the data in accordance with Tabachnick and Fidell's [23] recommendations, 12 invalid laps, including laps with zero lap time caused by MoTec disconnection, were omitted from the analysis. We also excluded out laps—the first lap taken after the driver exits the pit lane—and in laps—the laps taken just before the driver returns to the pits, as both are slow laps not representative of the instructed task. A total of 571 laps remained that were subjected to additional criteria for outlier removal using the z-score normalization method [24]. After removing the outliers, 557 laps remained for further analysis. In addition, we did some general descriptive analysis to determine the distribution of data and identify any trends.

Vehicle Control Features, being that all participants used the same car and track, a driver's lap time is a representation of the activities they performed during that lap as well as their driving patterns (braking, throttling, steering, and car speed). Thus, the performance during a lap may be anticipated by evaluating the driving patterns and the driver's behaviour during the lap. Thus, for assessing the driving patterns, which was the main objective of this study, we relied on telemetry data retrieved from MoTec that contained data channels such as speed, RPMS, the steering angle, acceleration, the brake pedal position, and the throttle as a function of the travelled distance in the lap. RPMS describes the engine's rotations per minute and is a function of the gear used by the driver as an indication of when to change gear. Steering angle displays the angle of the steering wheel that is being input into the car at any given time. Lateral acceleration and longitudinal acceleration indicate the level of acceleration of the car in a specific direction, whether longitudinal (forward and backward) or lateral (side to side). More accurately, the higher the longitudinal acceleration, the more extreme acceleration the car has undergone, which means the car has more grip when accelerating [25]. The same can be applied through a corner using the lateral acceleration; the acceleration rate is higher for a car under the absolute limit of cornering grip.

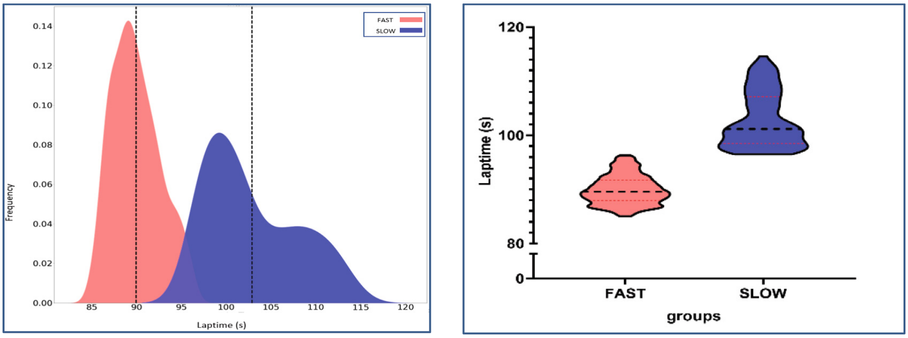
Furthermore, to assess a participant's performance during the session, we obtained new behavioural and car control measures from telemetry data. Based on the steering angle data channel, we defined lane deviation, oversteer, and understeer. Lane deviation is the value of the difference between the car's lateral position and the centre of the lane. Understeer and oversteer are two car control factors, especially in corners. The most common sign of understeer is when the front wheels lose grip due to excessive speed or hard braking through a corner, which can lock the front wheels. The same active variables that contribute to understeer also contribute to oversteer. Oversteer is typically brought on by strong acceleration in a rear-wheel-drive car or braking through a corner [22]. Steering reversal rate is the number of times the driver crossed the centred position of the wheel. Based on brake pedal position, we defined Trail braking application as the percentage of the distance between maximum brake and brake release divided by the length of braking zone. Similarly, we defined throttle release application as the distance covered from throttle application to throttle release.

Time-series telemetry data for each participant were used to produce the measurements for the new metrics and the data channels as well. We employed a sliding window of 200 frames (at a frequency of 50 Hz) for each of the data channels in order to record patterns of such variations in controls, and we then calculated the mean and the standard deviation for those windows. It is worth nothing that we had to apply a custom normalization method to all driver time-series data prior to taking these measurements. Because the number of rows for each lap varies depending on the length of the laps (i.e., a shorter lap has fewer rows in telemetry data), we constructed one row per lap distance indicator by averaging all the data related to that lap distance. Finally, a dataset of 93 time series data files was produced and exported as csv files, each referring to a participant data file that contains all of the participant's valid laps.

3 Results

3.1 Performance Level Analysis

In line with the aim of this study to understand key differences between elite and low skilled racers, we first attempted to categorize the laps into performance levels. To do this, we used a dataset containing 557 laps resulting from the pre-processing step for analyzing performance throughout the course of a lap. We analyzed two K-value selection algorithms, namely the elbow method [26] and Silhouette Coefficient [27], to determine the optimal number of clusters for the data set, then used the k-means method, the most commonly used clustering algorithm in both sport science (e.g., [15, 28]) and research conducted outside of the context of gaming literature [29, 30]. Table 1 presents the results of clustering as well as the statistics of the corresponding groups. The cluster names refer to the lap time, i.e., SLOW points to a low skilled's laptime and FAST points an elite racer's lap time. These clusters give an acceptable model with an accuracy of **81.42%**. Figure 1 shows the distribution of groups using a density plot as well as violin plots displaying the means and distributions within each group. Violin plots display the means and distributions of the two clusters. Wider areas of the violin plot represent a higher density of laps in the cluster for the given value, while a smaller population is represented by smaller sections. Both groups have a normal distribution, as we observe similar mean and median values.



a b

Fig. 1. Clustering visualization showcasing (a) density plot of lap-time across clusters, each density plot shows the data point's cluster distribution (filled area) and the mean of each cluster data point (vertical dash line). (b) violin plots displaying the means and distributions, within each cluster. The thick line in the centre of each plot represents the median, while the two blue lines represent the interquartile range.

Table 1. Lap time statistics for performance levels for the Brands Hatch track.

Group	Number of laps	mean	std	min	max	median
SLOW	115	102.86	5.09	96.47	114.61	101.16
FAST	441	89.93	2.65	85.04	96.32	89.58

3.2 Driving Patterns Analysis

Using the clustering results, the average of the time series lap data, explained in Sect. 2.2.2, for each metric was calculated for both the slow and fast groups. Figures 2–3 show driving patterns for different metrics for the Brands Haches circuit. They provide a comparison between human driver laps from groups FAST (red) and SLOW (blue). In order to make a better sense of the track, we incorporated section lengths of the track that we obtained from professional drivers. The vertical grey lines in Figs. 2–3 depict the track sections' boundaries. (e.g., S1 and T1 refer to the first straight and turn, respectively). It is clear that all groups of drivers race in the same manner on straight sectors while performing differently in corners. We can identify the distinction between various groups of drivers in terms of vehicle control features. To be more precise, the fast drivers accelerate earlier and more quickly after each corner, with a sharp throttle, higher brakes, and stable steering control. We can also observe how quickly the steering decreases when the throttle is increased in the fast laps. Additionally, how much turning is done while the brakes are fully applied? Moreover, it is obvious that fast drivers have more *throttle release application*, which is the act of applying the throttle more strongly and earlier while releasing the brake later. Similarly, when looking at the brake trace, there is greater *trail braking application* for fast drivers. They reach the peak very quickly, then modulate the brake and release the brake slowly as they get closer to the corner. The brake maximum and median are greater for laps with shorter times. It is interesting to see that the braking and longitudinal acceleration change proportionally for each driver group. However, there is no discernible trend in the acceleration feature (i.e., lateral and longitudinal acceleration).

Looking at Fig. 3, it appears that fast drivers apply a higher amount of steering angle when approaching corners, resulting in more amplitude in lane deviation. Regarding understeering and oversteering, as we observe, fast laps experience *understeer* situations almost twice as often as slow laps, while fast laps have far less *oversteer* than slow laps. While the results indicate greater turning in fast laps, they also imply less normal variation in steering angle.

As we examine the traces in greater detail, we can observe that the beginning part of the track has the most differences between the groups. The difference is more obvious in the first three corners, especially T2, which is kind of a hairpin corner that requires threshold braking in a straight line. Looking at the trace line for both groups, we see that when Fast drivers apply the brakes earlier before the entry of the corner, they trail off just as they begin to turn. They are doing this to aid in the initial rotation of the car and to help them reach the first apex of the corner so they can turn. Thus, the Fast driver keeps a tidier, tighter line through this section and then manages to roll on the throttle a little bit smoother and more consistently than the other one. Looking at T3, we see the Slow drivers get on the brakes earlier and off the throttle earlier than the Fast drivers, ending up with less speed and slowing the corner. Regardless of the big difference in the first section of the track, there is a little variation in the last section of the track (i.e., the last three corners), showing that both groups of drivers got off the throttle at about the same time and used roughly the same amount of brake pressure entering.

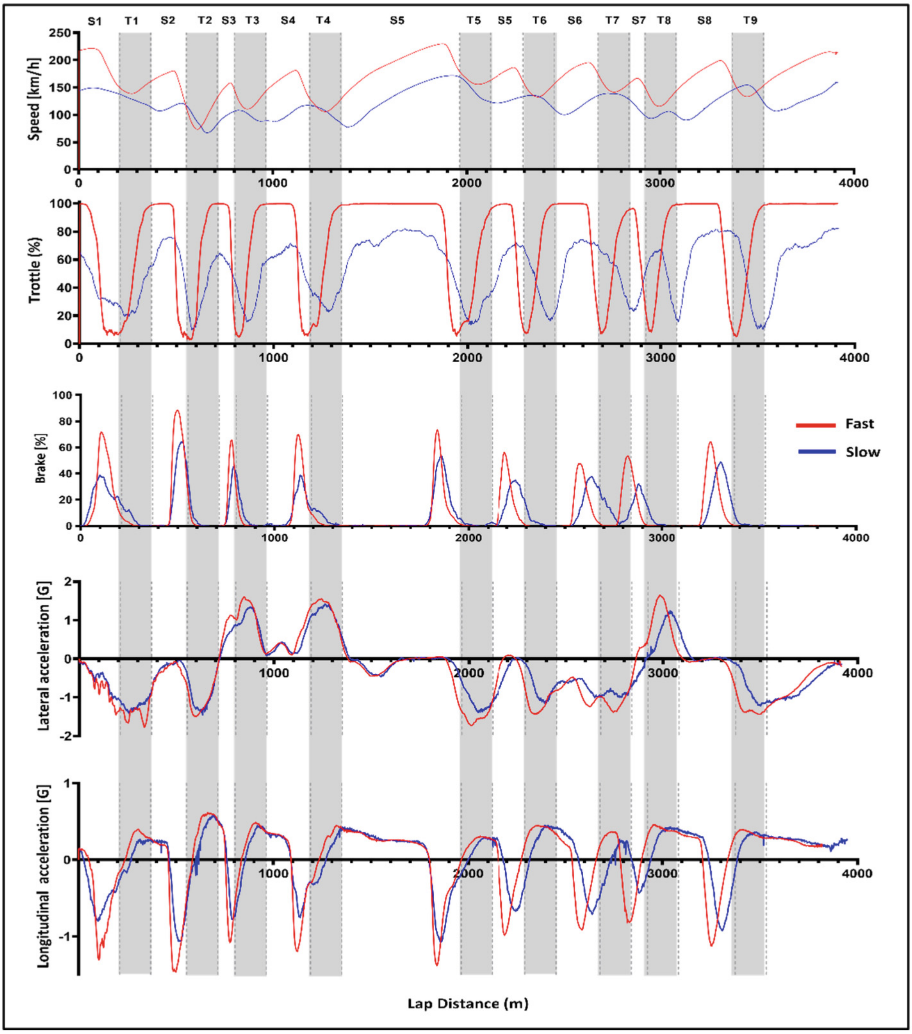


Fig. 2. Different features of driving behaviour for Fast and Slow performance group. The Y-axis (from top to bottom): speed, throttle position, brake position, lateral acceleration, longitudinal acceleration (Color figure online).

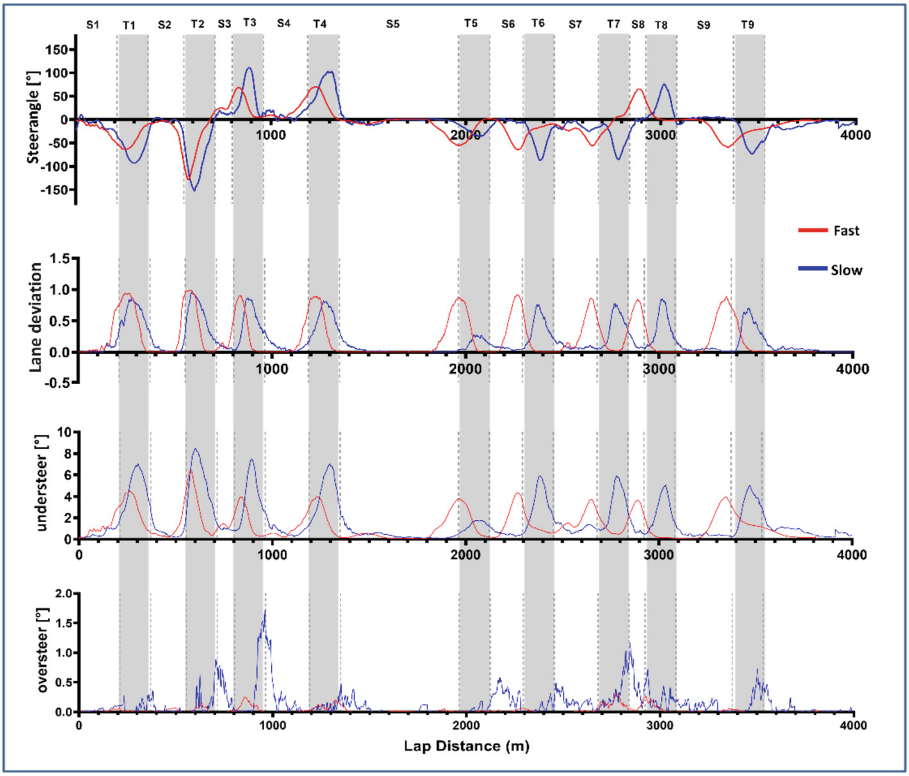


Fig. 3. Different features of driving behaviour for Fast and Slow performance group. The Y-axis (from top to bottom): steering angle, lane deviation, understeer, oversteer.

4 Conclusion

In this work, we provided an AI-enabled solution for analysing driver behaviour in sim racing. Given telemetry data from 93 race drivers, a cluster analysis was used to divide the resulting laps into two groups based on performance (lap-time), and then an in-depth analysis was conducted by defining a set of lap-based metrics derived from telemetry data and examining the car trajectory, steering behaviour, and car control features of different racer groups. From the results, we can conclude that over the course of the whole lap, a higher value of the speed mean, a longer throttle release application, and a higher lane deviation led to a shorter laptime. By deeply analyzing telemetry data for different sections of the track, we were able to determine indicators on how much turning is increased during cornering, how quickly the steering decreases, and how strong acceleration or braking is through a corner. The findings of this research might be used to improve the effectiveness and efficiency of sim racing performance, including software tools to train the drivers and manufacturers to evaluate the configurations and various parameters of the car. Furthermore, the findings from our analysis contribute to a better understanding of the human driver and pave the way to enhance sim racing performance with consideration of elite-racer driving styles.

Our work opens future research directions, including analyzing driving behaviours on various types of tracks and utilising the model to identify the channels that are more challenging to professional racers. Furthermore, time-series telemetry data can be used to forecast several features such as the lap time. The knowledge from this analysis can be transferred to self-driving vehicles to apply reinforcement learning or to offer a technique to enhance their driving style can be applied for self-driving vehicles.

References

1. Lindsey, G.R.: Tatistical data useful for the operation of a baseball team. Oper. Res. **7**(2), 197–207 (1959)
2. Drachen, A., El-Nasr, M.S., Canossa, A.: Game Analytics: Maximizing the Value of Player Data. Springer, London (2013)
3. Schubert, M., Drachen, A., Mahlmann, T.: Esports analytics through encounter detection. In: MIT Sloan Sports Analytics Conference. MIT Sloan (2016)
4. Chu, W.C.-C., et al.: Artificial intelligence of things in sports science: weight training as an example. Computer **52**(11), 52–61 (2019)
5. Roose, K.M., Veinott, E.S.: Leveling up: using the tracer method to address training needs for esports players. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications Sage CA: Los Angeles, CA (2020)
6. Bob, B.: Game Design (2004)
7. Ricmotech. *What is Sim Racing?* (2016). https://www.ricmotech.com/What is Sim Racing s/1553.htm.
8. de Frutos, S.H., Castro, M.: Assessing sim racing software for low-cost driving simulator to road geometric research. Transp. Res. Procedia **58**, 575–582 (2021)
9. Tango, F., et al.: Field tests and machine learning approaches for refining algorithms and correlations of driver's model parameters. Appl. Ergon. **41**(2), 211–224 (2010)
10. Baldisserri, L., et al.: Motorsport driver workload estimation in dual task scenario. In: Sixth International Conference on Advanced Cognitive Technologies and Applications. Citeseer (2014)
11. Tjolleng, A., et al.: Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. Appl. Ergon. **59**(Pt A), 326–332 (2017)
12. Chihara, T., Kobayashi, F., Sakamoto, J.: Evaluation of mental workload during automobile driving using one-class support vector machine with eye movement data. Appl. Ergon. **89**, 103201 (2020)
13. *vTelemetry PRO*. https://www.renovatio-dev.com/vtelemetry-pro
14. Forum, K.: MoTeC telemetry and dedicated ACC workspace. https://www.assettocorsa.net/forum/index.php?threads/motec-telemetry-and-dedicated-acc-workspace.55103/
15. Remonda, A., Veas, E., Luzhnica, G.: Comparing driving behaviour of humans and autonomous driving in a professional racing simulator. PLoS ONE **16**(2), e0245320 (2021)
16. Sim Racing Telemetry. https://www.simracingtelemetry.com/
17. Constantinescu, Z., Marinoiu, C., Vladoiu, M.: Driving style analysis using data mining techniques. Int. J. Comput. Commun. Control **5**(5), 654–663 (2010)
18. Erséus, A., Driver-vehicle interaction: identification, characterization and modelling of path tracking skill. KTH (2010)
19. Casanova, D., Sharp, R.S., Symonds, P.: On minimum time optimisation of formula one cars: the influence of vehicle mass. Proc. AVEC **2000**, 22–24 (2000)
20. Metz, D., Williams, D.: Near time-optimal control of racing vehicles. Automatica **25**(6), 841–857 (1989)

21. Kegelman, J.C.: Learning from Professional Race Car Drivers to Make Automated Vehicles Safer. Stanford University, Stanford (2018)
22. Bugeja, K., Spina, S., Buhagiar, F.O.: Telemetry-based Optimisation for User Training in Racing Simulators. In: 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). IEEE (2017)
23. Tabachnick, B.G., Fidell, L.S., Ullman, J.B.: Using Multivariate Statistics, vol. 5. Pearson, Boston (2007)
24. Smiti, A.: A critical overview of outlier detection methods. Comput. Sci. Rev. **38**, 100306 (2020)
25. Lopez, C.: Going Faster! Bentley Publishers, Cambridge (1997)
26. Géron, A.: Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Unsupervised Learning Techniques. O'Reilly Media, Sebastopol (2019)
27. McCallum, A., Nigam, K., Ungar, L.H.: Efficient clustering of high-dimensional data sets with application to reference matching. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2000)
28. Odierna, B.A., Silveira, I.F.: MMORPG player classification using game data mining and k-means. In: Arai, K., Bhatia, R. (eds.) FICC 2019. LNNS, vol. 69, pp. 560–579. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-12388-8_40
29. Abdullah, D., et al.: The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Qual. Quant. **56**(3), 1283–1291 (2022)
30. Ashari, I.F., et al.: Application of data mining with the k-means clustering method and davies bouldin index for grouping IMDB movies. J. Appl. Inform. Comput. **6**(1), 07–15 (2022)