The Role of Big Data in the Development of Safety in Fully Automated Vehicles: A Survey

Ashkan Yousefi Zadeh, Alireza Davoud Pasha

1st Automotive Industries Recent Advances and Future Trends (AIRAFT2020)
October 2020
Iran University of Science and Technology, Tehran, Iran

The Role of Big Data in the Development of Safety in Fully Automated Vehicles: A Survey

Ashkan Yousefi Zadeh[1], Alireza Davoud Pasha[2]

[1]MSc Student, Department of Mechanical Engineering, University of Guilan, Iran; a_yousefizadeh@msc.guilan.ac.ir
[2]MASc Student, Department of Mechanical Engineering, University of Windsor, Canada; Davoudp@uwindsor.ca

Abstract
Cars are the most used vehicles by people for transportation in the world that their use endangers the lives and safety of millions of people every year. It seems that the solution presented in recent years, given the advances in computer science and artificial intelligence to overcome these dangers and have a safer driving is the use of fully automated vehicles. In this paper, first, autonomous vehicles and types of them are introduced and the safety challenges of vehicles driven by humans and factors of insecurity in these vehicles have been studied. In the following, the current trends and future challenges of autonomous vehicles have been reviewed, as well as the importance and safety position of these vehicles. Eventually, the role of big data to promote safety in fully automated vehicles and its importance in developing safety and maintaining public health in this industry for the following years has been studied.

Keywords: Autonomous Vehicles, Big data, Driving Safety, Artificial Intelligence, Deep Learning

1.0 Introduction
Nowadays, one of the most notable trends in the application of artificial intelligence in mechanical engineering is autonomous vehicles (AVs), which can be one of the most challenging of these applications due to the straight and close relation to public health. Based on the J3016 standard issued by the Society of Automotive Engineers or “SAE”, levels of driving automation is defined in six levels, from SAE Level Zero (no automation) to SAE Level 5 (full vehicle autonomy) [1]. “Level 0” or “No automation level”, indicates that the driver performs all the driving tasks. “Level 5” or “Full automation level”, indicates the vehicle carries out all of the driving duties under all conditions, and the driver may have the options to control the vehicle; this level of automation is the case study of this article [2]. When the vehicle is fully autonomous, it is capable of longitudinal and lateral control; also, it can perform “Object and Event Detection and Response” or “OEDR” and can handle emergencies autonomously. The final property which distinguished fully automated vehicles versus high automated vehicles is unlimited “Operational Design Domain” or “ODD” which means the system can operate under various conditions [3]. Until 2018, only Waymo had deployed vehicles for public transportation with the highest autonomy among AVs, which was "level 4" or highly automated [4].

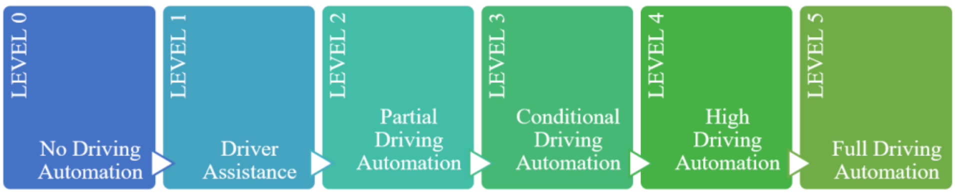

Figure 1: Standard levels of driving automation by SAE [1].

Based on the Waymo Safety Report in 2018, every year, 1.2 million people die in traffic crashes in the world, and these numbers are annually rising around the world [5]. Also, The World Health Organization (WHO) estimates that around the world, the yearly road deaths are 1.35 million, with about 50 million injuries [6]. These reports determine that about 95% of road accidents are because of human errors [5]. These statistics determine that. However, companies are improving the quality of the safety features on their vehicles, such as seat belts, airbags, impact absorption etc. However human errors still exist, and they are inevitable. Thus, it is obvious that using an autonomous system which can provide the least number of errors in driving with intelligent computations, can help people to drive safely. Recently, with the development of self-driving cars, safety measures take important steps, preventing collisions by using report information received from the environment outside the vehicle [7]. Since fully automated vehicles ultimately use data to control the vehicle, accuracy, and validation of these data are crucial for safety. Big data is principally applied to represent enormous datasets, and they include masses of unstructured data that require real-time analysis. Big data makes new chances for exploring new values, helps obtain an in-depth perception of the disappeared and hidden values, and brings new difficulties and challenges. Such as how to arrange and handle such datasets efficiently [8]. To have more accurate and useful data to use for controlling a fully automated vehicle, using big data is essential. In the next sections, details of this topic have been discussed.

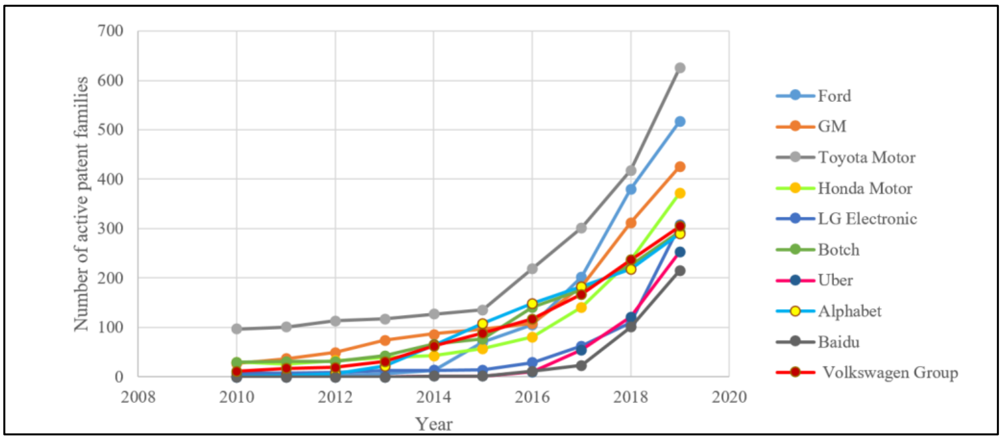

Figure 2: Number of active patent issues from 2010 to 2019 for the biggest autonomous driving patent owners worldwide [9]

The above figure shows that manufacturers gradually increased their focus on AVs, especially since 2015. Figure 3 displays that autonomous vehicles level 4, will allocate a huge volume of ground transportation investments until 2030. This figure accurately shows that a possible reason for lack of AVs level 5 before the year 2030 is the importance of developing safety factors for fully automated vehicles, which is a time-consuming and delicate process. As a result, research and development of these AVs must proceed in a precise form.

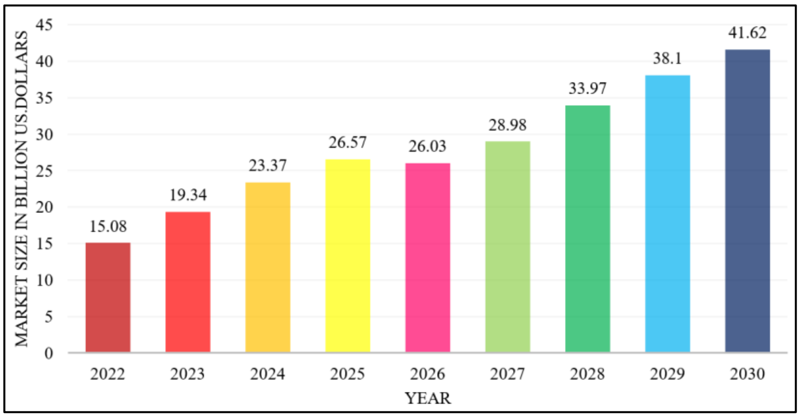

Figure 3: The predicted market size of Level 4 autonomous driving worldwide between 2022 and 2030 [10]

2.0 Importance of Safety in Fully Automated Vehicles

It can be said with certainty that the most important reason for using AVs is to drive safer. As mentioned, the leading cause of road accidents is human errors. Table 1 shows the detailed analysis of accidents in Iran from 2006 to 2015. As it is obvious, the numbers of deaths and injuries are remarkable, and yearly, many people are killed while they are using cars for transportation.

Table 1: Analysis of road accidents in Iran [11]

Year	*Total accidents*	*Number of fatal accidents*	*Number of people killed*	*Number of accidents resulting in injuries*	*Number of people injured*	*Number of accidents resulting in damage*
2006-2007	795205	5841	7590	123923	160152	665441
2007-2008	758801	6460	8416	125762	158573	626579
2008-2009	780352	6554	8645	139370	174284	634428
2009-2010	702512	11854	18159	166456	231972	524202
2010-2011	776517	26239	37893	388866	472132	361412
2011-2012	396893	13665	16884	200223	254699	183005
2012-2013	429544	15123	18569	226565	293665	187856
2013-2014	305787	7038	8940	135008	174064	163741
2014-2015	264557	6609	8248	126876	163450	131072

Based on the US National Highway Traffic Safety Administration (NHTSA) report published in 2015, the critical driver-related causes in an accident are as follows:

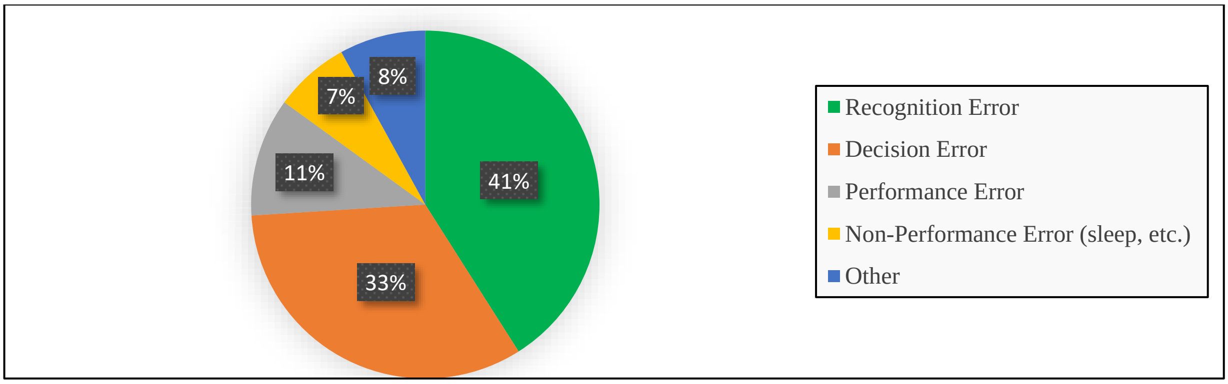

Figure 4: Critical driver-related causes [12]

As shown in this figure, the two main reasons for human errors in driving are, recognition error and decision error. To reduce these numbers and have a safer driving by AVs instead of humans, safety parameters are required to be examined and implemented properly. Based on NHTSA autonomous driving systems (ADS) safety elements, there are 12 elements that each company must consider to incorporate into their vehicles [13]: System Safety, Operational Design Domain (ODD), Object and Event Detection and Response (OEDR), Fallback (Minimal Risk Condition), Validation Methods, Human Machine Interface (HMI), Vehicle Cybersecurity, Crashworthiness, Post-Crash ADS Behavior, Data Recording, Consumer Education and Training, Federal, State, and Local Laws.

2.1 Some Examples of Safety Considerations of Autonomous Vehicles by Manufacturers

2.1.1 Ford Motor Company

Argo AI is a collaboration company which works with Ford in the development of AVs; this company uses a three-phase training plan for safety operators [14]. Ford's development team record and extract massive daily data from simulations and test vehicles from onboard data systems like braking and steering. in addition, they use Virtual Driver System as well as camera and sensor data. Using an autonomous vehicle data recording device and its Event Data Recorders (EDR), data from the crash will be logged into the vehicle system automatically. Ford's validation and testing methods apply data from testing to develop and enhance the system, sub-system, and component performance. In real-world testing, they produce the distances logged and near misses faced to either improve requirements or validate system safety with new data for simulations. Finally, they will test the vehicle's capacity to independently make safety fallback maneuvers to remove safety operators when the Virtual Driver System is available [14].

Figure 5: Argo AI training program for safety operators [14]

2.1.2 General Motors Company

General Motors (GM) safety is through an iterative design. They design and build technologies and systems, test them in simulators as well as in the field, and then check the results and feedback to the design the process. they include learnings, mainly safety data, into future productions so they will be even safer. GM does this with many iterations, which leads to new technologies and systems for autonomous vehicles. In GM's controlled deployment, autonomous vehicles drive only within defined boundaries, and only on roads for which they have acquired high-definition map data, also, they operate just under known operational situations and constraints that apply to the entire fleet. GM's autonomous vehicle has two data recording features a second robust data logging system and a conventional Event Data Recorder (EDR) as well as their vehicle's robust data recording ability, gives information on vehicle performance while normal driving and collision avoidance situations [15].

3.0 Role of big data in the safety of fully autonomous driving

AVs entirely work by the data that they obtain by sensors such as IR (Infra-Red) Sensors, radar, Lidar, Global Positioning System (GPS), and the information they concoct through 3D cameras, etc. the data that cars collect within these references lead them to make safe driving decisions and control them [16]. Table 2 indicates that each AV collects approximately 120,000 GB data per month and 1,440,000 GB data per year. It is obvious that we are left with a massive dataset and extracting useful information from this data set is very challenging.

Table 2: An approximate volume of data into Autonomous vehicles [17]

Type of data resource	Volume of data
Cameras	~20-40 MB (Per Second)
Radar	~10-100 KB (Per Second)
Sonar	~10-100 KB (Per Second)
GPS	~50 KB (Per Second)
LIDAR	~20-70 MB (Per Second)
Total	4,000 GB (Per Day)

The above table indicates that a fully automated vehicle requires many infrastructures such as artificial software and a cloud server to store massive datasets from sensors. Therefore, a driverless vehicle can build strategies for many possible situations such as different weather, lighting levels on the road, driver behavior, etc. by using collected data [18]. As a result, data volumes and storing, real-time and fast analyzing, and cybersecurity are the most significant challenges that big data make.

Currently, many famous companies and universities worldwide are working to address these issues. For instance, Amazon Web Services S3, Microsoft Azure Data Lake, Google Cloud Storage, and IBM Cloud are the most popular companies providing cloud storage for big data [19]. So scientists and engineers can use datasets faster and enhance the performance of their designs for AVs. Moreover, in the Cybersecurity section, companies like APTIV and NVIDIA are developing software and computing platforms for cyber safety in AVs. NVIDIA utilizes AI-powered data processors and chips to produce and protect AVs. The software and cloud-based technologies assist AVs relay driving data and in securely learn [20]. Tackling and addressing the entire big data challenges in fully autonomous vehicles is a time-consuming task that will gradually be developed and modified.

Hence, to acquire and develop AVs functions, manufactures must stock, manage, and process vast and massive volumes of data. The primary research and development challenges in this field is an investigation on how AVs should use the advantages of big data to behave in possible crash conditions to protect human and public health. On the other hand, for an appropriate decision making how broad the scope of data should be. Accessing open data is essential to the use of big data in fully automated vehicles. Because AVs, in addition to learning from neural networks of other vehicles manufactured by the same corporation, must proceed learning process from all other AVs as well. By using big data, AVs allow the implementation of safety and public health policies on prioritization of protection in accident situations, and these policies have to be considered and studied thoroughly [21]. The synchronization and coordination of AVs in the networks of vehicles are one of the vital usages of big data in autonomous vehicles. Also, one of the most significant importance of big data refers to controller design engineer's task part which is related to using big data for autonomous vehicle control systems [22].

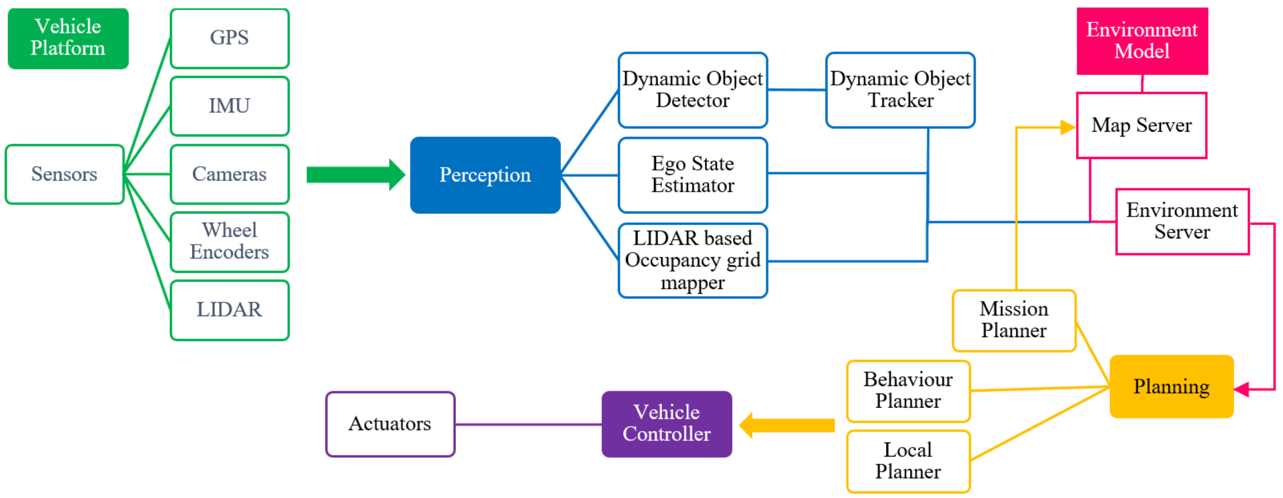

Figure 6: Baseline architecture of autonomous driving [23]

Big data in control systems can be applied for prediction and estimation goals. Also, it can be a section of the control strategy, such as an optimal trajectory adaptation approach concentrating on the safety factors of AVs using a cloud database [22]. Model predictive control (MPC) is one of the famous controlling methods in this area. Since the feedforward controllers implement a predictive response as they provide a reference output to obtain an appropriate tracking response, especially when the expected inputs are non-zero are the main reason that feedback and feedforward controllers are used to control a system. To keep a constant-radius turn or brake command and a constant throttle to maintain a constant velocity or deceleration measures, AVs need non-zero steering commands. Therefore, feedforward control is very advantageous in developing tracking performance in fully automated driving [24].

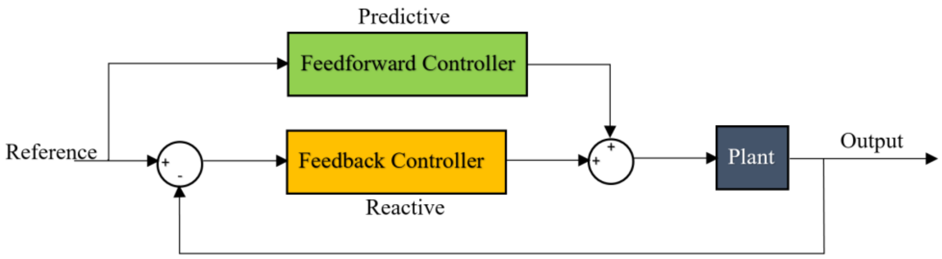

Figure 7: Combined feedforward and feedback control for MPC, general Scheme [25]

As mentioned and is evident from figure 6, data-driven control for a fully automated vehicle relies on data that various sensors of the AV system obtain from the environment. To help AVs to correctly understand the environment and make better decisions compared to humans in situations such as accelerating, decelerating, steering, and path planning, many artificial intelligence strategies like machine learning (ML) and Reinforcement Learning (RL) solutions could be applied [26]. On the other hand, this volume of data is massive and unstructured. As a result, using traditional database models and tools are useless and cannot be managed, processed and analyzed properly. Therefore to explore, control, and handle this enormous volume of data, data-driven models are required [27]. A machine learning method that uses automated feature engineering for large amounts of data to train Deep Neural Networks (DNNs) is Deep learning (DL). DNNs could attain human-level precision by real-time decisions and outcomes from real-time data input. Until today, over 80 % of the whole analysis process time is approximately allocating to preparing the data, which involves labeling, feature engineering, etc. [28]. Regardless of which data-driven controller approach is used, It would be essential to work with proper tools and modern techniques to manage and analyze big data in order to have a more robust and faster learning process for the system and eventually reach a better and high-performance controller that lead the vehicle to a proper decision making to have a safer driving.

4.0 Conclusion

As discussed, almost 95% of road accidents are happening due to human errors, and everyday thousands of people around the world lose their lives or get physically or mentally injured because of these accidents. by the increasing development and promotion of artificial intelligence and their applications in various technologies such as autonomous vehicles, the number of these driver related accidents can be reduced remarkably, and the number would be down to acceptable figures. These goals can be achieved when multiple parameters involved in the control of fully autonomous vehicles, both internal and external, are identified adequately by the system. The system must also be well-trained and able to know how to deal with the various scenarios that may occur while driving or be able to correctly predict the outcome for a safer decision making. Since the basis of controlling fully autonomous vehicles is dependent on the training and testing of the controller using the data received from different sensors installed in the car using different scenarios and situations, it is discussed that these vehicles must undergo a huge and massive data which are unstructured. Managing and analyzing the data to obtain an optimum and proper controller is a challenging and complicated task which must be done by modern tools and updated data science methods. As a result, big data has a significant and crucial role in having a safer driving with fully autonomous vehicles and it is essential that control developer scientists and engineers of these vehicles suitably apply these big data into designing and developing AVs to ensure the safety of the occupants of these vehicles and the pedestrians who encounter them.

References

[1] International, S. *SAE Standards News: J3016 automated-driving graphic update*. 2019; Available from: https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic.

[2] NHTSA. *Automated Vehicles for Safety*. 2020; Available from: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

[3] Calvert, S.C., et al., *Gaps in the control of automated vehicles on roads.* IEEE intelligent transportation systems magazine, 2019. **1,(2020)**: p. 147-158.

[4] Waslander, S. *Introduction to Self-Driving Cars*. 2020; Available from: https://www.coursera.org/learn/intro-self-driving-cars.

[5] Waymo. *Waymo Safety Report: On the Road to Fully Self-Driving*. 2018; Available from: https://waymo.com/safety/.

[6] Threlfall, R. *Autonomous Vehicles Readiness Index (AVRI)*. 2020; Available from: https://home.kpmg/xx/en/home/insights/2020/06/autonomous-vehicles-readiness-index.html.

[7] Mallozzi, P., et al., *Autonomous Vehicles: State of the Art, Future Trends, and Challenges*, in *Automotive Systems and Software Engineering*. 2019, Springer. p. 347-367.

[8] Chen, M., S. Mao, and Y. Liu, *Big data: A survey.* Mobile networks and applications, 2014. **19**(2): p. 171-209.

[9] Wunsch, N.-G. *Largest autonomous driving patent owners worldwide from 2010 to 2019, by number of active patent families*. 2020; Available from: https://www.statista.com/statistics/1016110/worldwide-autonomous-driving-patent-owners-trend/.

[10] Wagner, I. *Projected L4 autonomous driving market size worldwide between 2022 and 2030*. 2018; Available from: www.statista.com.

[11] Iran, S.C.o. *Number of accidents and human casualties of vehicles leading to death, injury and damage in road transport*. 2017; Available from: https://www.amar.org.ir.

[12] NHTSA, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. 2015, National Highway Traffic Safety Administration.

[13] NHTSA, *AUTOMATED DRIVING SYSTEMS: A Vision for Safety*. 2017.

[14] Company, F.M. *A MATTER OF TRUST: FORD'S APPROACH TO DEVELOPING SELF-DRIVING VEHICLES*. 2020; Available from: https://media.ford.com/content/dam/fordmedia/pdf/Ford_AV_LLC_FINAL_HR_2.pdf.

[15] Motors, G. *SELF-DRIVING SAFETY REPORT*. 2018; Available from: https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf.

[16] Potter, M. *Big Data's Role in Self-Driving Car Development*. 2019; Available from: www.insidebigdata.com.

[17] Krzanich, B. *Data is the New Oil in the Future of Automated Driving*. 2016; Available from: https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/#gs.d9hj6d.

[18] Haydin, V. *How Big Data in Autonomous Vehicles Defines the Future*. 2019; Available from: https://www.intellias.com/how-big-data-in-autonomous-vehicles-defines-the-future/.

[19] Marr, B. *What Are The Best Big Data Cloud Storage Providers? Here Are The Top 6*. 2020; Available from: https://www.forbes.com/sites/bernardmarr/2020/05/11/what-are-the-best-big-data-cloud-storage-providers-here-are-the-top-6/#7aa016ff5cd2.

[20] Daley, S. *NICE RIDE: 10 AUTOMOTIVE CYBERSECURITY COMPANIES MAKING VEHICLES SAFER AND MORE SECURE*. 2019; Available from: https://builtin.com/cybersecurity/automotive-cyber-security.

[21] Shaw, D., B. Favrat, and B. Elger, *Automated vehicles, big data and public health.* Medicine, Health Care and Philosophy, 2020. **23**(1): p. 35-42.

[22] Fényes, D., B. Németh, and P. Gáspár, *A predictive control for autonomous vehicles using big data analysis.* IFAC-PapersOnLine, 2019. **52**(5): p. 191-196.

[23] Waslander, S., *autonomoose: Toward All-Weather Autonomous Driving*. 2018: University of Toronto, University of Waterloo. Available from: https://www.4shared.com/s/fU59YMdZAea.

[24] Sailan, K. and K.D. Kuhnert, *Modeling and design of cruise control system with feedforward for all terrian vehicles*, in *Computer Science & Information Technology (CS & IT)*. 2013. p. 339–349.

[25] Waslander, S. *Feedforward Speed Control*. 2018; Available from: https://www.coursera.org/learn/intro-self-driving-cars.

[26] Ma, Y., et al., *Artificial intelligence applications in the development of autonomous vehicles: a survey.* IEEE/CAA Journal of Automatica Sinica, 2020. **7**(2): p. 315-329.

[27] Kaffash, S., A.T. Nguyen, and J. Zhu, *Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis.* International Journal of Production Economics, 2020. **231**: p. 107868.

[28] Ziegler, L., F. Krämer, and N. Haustein, *AI and Big Data Management for Autonomous Vehicles.* ATZ Electron Worldw, 2020. **15, 40–45 (2020)**.