IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 19, 2020, accepted October 23, 2020, date of publication October 29, 2020, date of current version November 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034588

Data Analysis Techniques in Vehicle Communication Networks: Systematic Mapping of Literature

LUCAS GOMES DE ALMEIDA[1], ADLER DINIZ DE SOUZA[2], BRUNO TARDIOLE KUEHNE[1], AND OTAVIO S. M. GOMES[1], (Member, IEEE)

[1]Instituto de Engenharia de Sistemas e Tecnologia da Informação (IESTI), Universidade Federal de Itajubá, Itajubá 37500-903, Brazil
[2]Instituto de Matemática e Computação (IMC), Universidade Federal de Itajubá, Itajubá 37500-903, Brazil

Corresponding author: Lucas Gomes de Almeida (lucas.ga@unifei.edu.br)

This work was supported in part by CAPES and in part by UNIFEI.

ABSTRACT Vehicles are becoming more intelligent and connected due to the demand for faster, efficient, and safer transportation. For this transformation, it was necessary to increase the amount of data transferred between electronic modules in the vehicular network since it is vital for an intelligent system's decision-making process. Hundreds of messages travel all the time in a vehicle, creating opportunities for analysis and development of new functions to assist the driver's decision. Given this scenario, this article presents the results of research to found out which data analysis techniques in vehicular communication networks and for which purposes they are designed. The research method adopted was the systematic mapping of literature, where 196 articles were found using a search protocol. All papers were classified according to the established inclusion and exclusion criteria, and the main results contained were discussed. To obtain a clear view of the generated information and support the identification of possible gaps in this field, correlation graphs, and a systematic map was developed. It was possible to verify that the identification of the driver's profile was the most studied application, with the use of neural network techniques to correlate the gathered data.

INDEX TERMS Artificial intelligence techniques, In-vehicle network, systematic mapping, vehicle data analysis.

I. INTRODUCTION

Automotive electronics are continuously evolving. In the last few decades, it was possible to observe an increase in the use of electronic modules inside the vehicles to support a connected world's new technologies. It is estimated that there are about 70 electronic control units (ECU) in a modern vehicle [1], communicating through different vehicle networks, and the Controlled Area Network - CAN is the most used one.

The CAN protocol was first introduced by Robert Bosch company GmbH in 1983. It has been widely used in several applications in the automotive industry and even in other application fields, such as domestic and medical [2]. Its great use could assign to high scalability, robustness, and ease configuration added to the system. This protocol travels relevant vehicle data that can be applied to support the driver's navigation system and perform critical driving decisions. Therefore, data availability is no longer a restrictive aspect, as data could be simply and quickly collected.

In the last few years, researchers have been developing several methods to analyze the transmitted data to add value to the product and increase vehicle safety. This article proposes a systematic mapping of the best techniques and applications for data that travels in vehicular networks. The main purpose is to find future opportunities and study gaps that could be better solved shortly. The paper was subdivided into five main sections. Section II describes the research methodology used to collect and compare the different found articles and research, as well as the criteria used to select them. In section III, the main results are presented and discussed, with the introduction of a conceptual mapping concentrating the selected results. In section IV, are answered the research questions. Finally, in section V a conclusion with future opportunities will be presented.

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

TABLE 1. Set of inclusion criteria.

Criteria	Description
CI-01	It could be selected articles that are related to the automotive world.
CI-02	It could be selected papers that introduce any technique for data analysis.
CI-03	It could be selected only articles that have an application for the data analysis.

II. RESEARCH METHODOLOGY

This systematic mapping of literature recognizes and categorizes the best existing techniques and applications of data analysis in vehicular communication networks. Systematic mapping supports the identification, recognition, and classification of the existing state-of-the-art. All data collected from found primary studies will be categorized. The methodology used in this article was based on Petersen [3] and will address the research questions, search protocol, selection of relevant studies, quality assessment, and data extraction.

A. DEFINITION OF RESEARCH QUESTIONS

The first step of a systematic mapping is the definition of the research questions. It is extremely important to define these questions to lead to a good selection of articles. This article 's main objective is to understand the techniques and applications of data analysis in vehicular communications networks. Based on this objective, the following questions were elaborated:

1. QP1: Which are the main concepts and techniques presented in research involving data analysis and artificial intelligence in vehicular communication networks?
2. QP2: The techniques found are used for which types of tasks within the context of data analysis?
3. QP3: Which are the existing gaps and opportunities in data analysis applications in vehicular communication networks?
4. QP4: Is there a most efficient technique in the found applications?

The main purpose sought by the authors with these research questions is to understand which key concepts and techniques are used for data analysis in a vehicular communication network. The first research question is the starting point for gathering information from the articles. After finding the techniques used, it will be searched for which tasks these techniques refer to (QP2). Inside this topic, it will also be checked if there is any relationship between the techniques and the applications, to determine the most used technique in vehicular network analysis. The third research question will investigate if there is any gap or future research opportunity in an area that is expanding and increasing the presence in the driver's daily lives. Finally, a comparison between the found techniques will be performed to find the most efficient one for each application.

TABLE 2. Set of exclusion criteria.

Criteria	Description
CE-01	It shall not be selected articles in which Keywords are not included in the title, abstract and / or publication text.
CE-02	It shall not be selected publications in which the authors do not use data analysis of an automotive network.
CE-03	It shall not be considered the same article in different repositories.
CE-04	It shall not be selected publications in which the data analysis is being carried out only with information provided outside the vehicle.
CE-05	It shall not be considered articles in which data analysis is performed for an application outside the automotive world.
CE-06	It shall not be selected articles that are not published in a relevant journal or conference.

B. SEARCH PROTOCOL AND SELECTION

The authors conducted a systematic and elaborated approach to extract meaningful and relevant information of the analyzed data. The first step to find articles related to the studied topic was the definition of a suitable search protocol. The right choice of keywords had significant importance in the results. As the techniques and applications are not yet known, the protocol considers the use of any data analysis technique in a vehicular network. Although the authors are familiar that the CAN network is the most used in the automotive world, other types of vehicle networks and their variants were also considered to ensure that new technologies are also investigated in this article.

- (''Automotive'' OR ''Automobile'' OR ''Vehicle'') AND
- (''Vehicle Network'' OR ''CAN Bus'' OR ''Controller Area Network'' OR ''LIN Bus'' OR ''Ethernet'') AND
- (''Data Analysis'')

The protocol was then adapted to perform the research in different article repositories. For instance, in Scopus, the words ''TITLE-ABS-KEY'' were added at the beginning of the protocol to return results based on the searched keywords in the titles, abstracts, and keywords of the articles in the repository. In IEEE, it was necessary to add the words (''All Metadata'') for the same research type. The research was done in the IEEE Explore, Scopus, and ACM repositories. The search returned 206 papers (IEEE: 29 articles, Scopus: 40 articles and ACM: 137 articles). Then, the inclusion and exclusion criteria described in Tables 1 and 2 were applied to these 206 articles, which selected 38 articles for the complete reading and application of the quality assessment prepared by the authors.

During the search analysis, the quality and reliability of the found papers could be questioned. To avoid this, it was

then necessary to apply a quality assessment, considering the following questions:

- Is this work relevant to the research field studied?
- Was the article published in a relevant journal or conference?

A score was assigned to each answer, where we gave "0" if the answer is negative and "1" if it is affirmative. Only papers with a score equal to "2" were considered. For an article to be considered relevant, the percentile criteria and conference level were adopted. The paper shall be included in the list of best conferences, or it shall have a percentile equal or greater than the fiftieth. The percentile was checked on the Scopus repository, whereas the conference level was checked on Google's H5 index. Considering this approach, six articles were removed from the selected ones. After a detailed reading of the remaining articles, seven false positives were found, thus leaving twenty-five papers to be studied and discussed throughout this article. An example of a false positive found was a paper where a vehicle data analysis was performed. However, the used data were not provided within the vehicle network, yet with the support of an external sensor coupled to the vehicle. Fig. 1 summarizes the authors' approach and considered methodology.

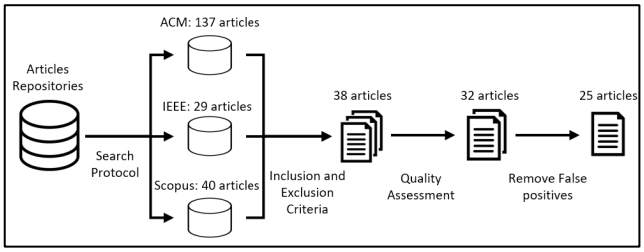


FIGURE 1. Search methodology used for systematic mapping. After the inclusion and exclusion criteria applied, along with the quality assessment, 25 works remained for analysis by the authors.

C. DATA EXTRACTION

After selecting the twenty-five articles that will compose this article, a data extraction was carried out to answer the research questions previously elaborated. The main extracted information was:

- Title
- Authors
- Publication Year
- Published in
- Relevant Category and Percentile
- Applied data analysis technique
- Applied application
- Keywords

Selected articles were arranged in Table 3, where it is possible to check the publication year, the category, and percentile relevance for the studied field of each paper. The tabulation was performed using a spreadsheet, to be available in a folder into Google Drive [38]. Answers to research questions were generated, as well as a discussion and conclusion of the found results.

III. RESULTS

This section will discuss the main results found after the analysis of the selected and presented articles in section II. A first analysis identified which were the main applications obtained from the applied techniques. Fig. 2 illustrates the graph of the comparison of all observed articles, whereas Fig. 3 illustrates the conceptual map developed by the authors. We observed that the determination of a profile for the driver represents 48% of the analyzed studies, followed by security against cyberattacks, which represents 20%. Other applications, such as collision alert, traffic estimation, road conditions, and others, complete the remained searched articles. It was also analyzed the main keywords used in each article and researchers focus throughout the years. In order to have a clear picture of data analysis evolution during the last years, Fig. 4 was generated. Driver behavior was the focus of all researchers until 2020, when autonomous vehicle became the most used keyword in the selected articles. One possibility for this behavior is the recent trending of the autonomous industry with Tesla and Google and the fact that in a future where all vehicles are autonomous, driver behavior will no longer be needed since there will be no drivers in charge of the vehicle. Regarding the applied techniques for data analysis, neural network and machine learning have become the most used ones since 2018. Big Data also appear as a trending keyword in 2019, which may indicate that it could be used as a solution for data analysis in vehicle communication networks in the next few years.

Since different techniques have different application scope, this chapter will be subdivided according to the found applications of vehicle data analysis. The efficiency of each technique will also be analyzed, according to the evaluation of the author's article.

A. DETERMINATION OF DRIVER'S PROFILE

It is public knowledge that each driver has a unique way of driving. Every day it is possible to watch people in traffic who use their vehicle differently. Some people use the accelerator pedal more abruptly, whereas others use the brake more smoothly, which makes it possible to determine the driver's fingerprint. According to Ezzini *et al.* [11], the idea of a driver's fingerprint obtained from the analysis of existing data on the CAN network is not new, since many articles using machine learning techniques have shown that this is feasible. Therefore, identifying the driver's profile and behavior during traffic is possible and valuable information to be achieved. According to the driver behind the wheel, its applications could be numerous, such as security authentication or vehicle customization.

However, which are the most used techniques for determining this behavior? According to the related studies, there is no common understanding on a single technique that is highly efficient. In the papers of Zhang *et al.* [15] and

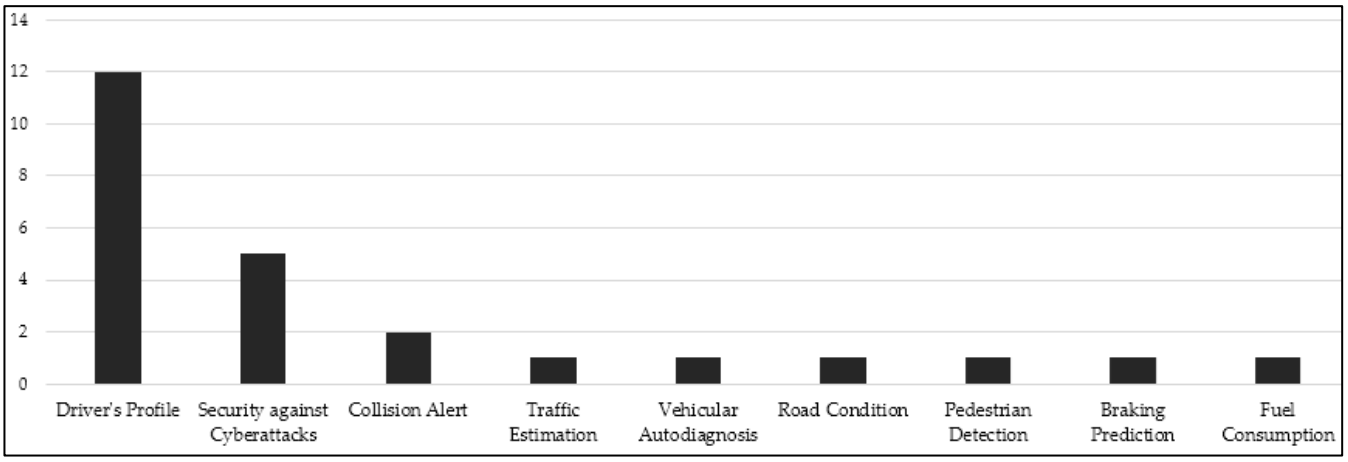


FIGURE 2. Comparison of the most found applications for data analysis in a vehicular network. Graph was obtained with the extracted information of the articles on Section II.

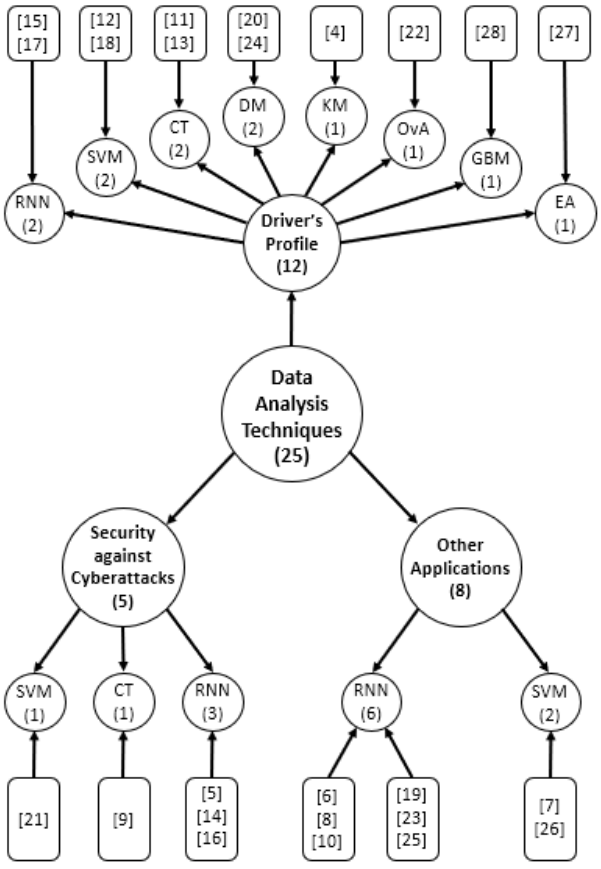


FIGURE 3. Conceptual map of data analysis techniques based on the articles selected according to section II and references on Table 3. The acronym used for data analysis techniques were: RNN (Recursive Neural Network), SVM (Vector Support Machine), CT (Comparative of Techniques), DM (Data Mining), KM (K-means), OvA (One versus All), GBM (LightGBM) and EA (Statistical Analysis).

Narayanan *et al.* [17], a neural network was the chosen technique. Zhang developed a recurrent neural network and analyzed 51 different functions existing in the vehicle's CAN network to determine the driver's profile. Although Narayanan *et al.* [17] also developed a neural network to determine the driver's behavior and direction pattern. A sensor fusion was also implemented with highway lanes images and steering wheel data regression, thereby achieving a 10% increase precision in the results comparing to the state-of-the-art.

Another technique used for this application was the support vector machine, proposed in the papers of Fuwu *et al.* [12] and Burton *et al.* [18]. Fuwu developed an algorithm and compared the obtained CAN network data with the drivers' brain waves, aggrouping it into three main profiles: aggressive, moderate, and conservative. The main analyzed data in the network were vehicle speed, total driving time, number of lane changes, steering rotation, angular acceleration, and collisions. The results were satisfactory, however, some limitations were pointed in this research, such as a simple route to collect data and use of young drivers, with an average age of 23.6 $\pm$ 1.6 years. In Burton *et al.* [18] article, the support vector machine had one of the best results found in the studied literature: 95% confidence to obtain the driver profile under 2.5 minutes. To achieve this, five parameters on the network were considered: distance traveled, average vehicle speed, the standard deviation of the steering wheel position, the average chance of brake pedal position, and the average change of accelerator pedal position.

Ezzini *et al.* [11] and Lin *et al.* [13] not only applied one technique, but they also compared several techniques in order to determine which one would be the most efficient. Ezzini concluded that Random Forest and Extra Trees had the most relevant results. An interesting aspect of this study was the analyzed data classification according to its importance in the vehicle, where weight was assigned to each considered parameter. The most critical data in the algorithm was the engine's condition and fuel consumption during the driving time. In Lin *et al.* [13], four different methods of data analysis were compared: neural networks, hidden Markov chain, fuzzy logic, and Gaussian mixture model. The pros and cons were illustrated in a table, which was adapted in Table 5.

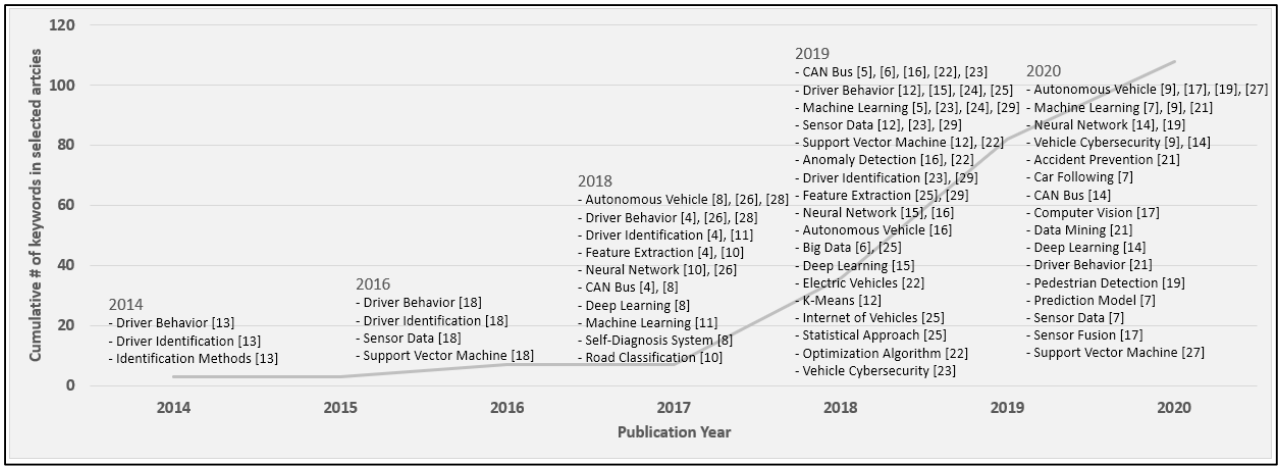


FIGURE 4. Timeline of the selected articles on section II. The main keywords with the respective articles reference are projected in each year.

Lin *et al.* [13] also point out that the driver is a complex and uncertain individual who introduces different driving patterns according to the situation that he faces, adding enormous complexity to the definition of the most effective technique. Several tests are required in real environments and with different types of samples to achieve a reliable result.

For Sun *et al.* [20] and Chen *et al.* [24], the technique known as data mining was the most promising for identifying the driver's profile. Sun performed the analysis using four parameters in the vehicles (distance traveled, average speed, fast acceleration, and deceleration). With an in-depth analysis of these data, he assigned a driver's score to classify them according to the probability of causing an accident on a highway. Chen *et al.* [24] introduce the Internet of Vehicles concept, where vehicles would be connected and continuously interacting with each other. Eight vehicle functions were analyzed, with the respective degree of importance to calculate the driver behavior. For instance, engine speed was the most important, followed by vehicle speed. Data from approximately 150 drivers on 20,000 trips on pre-defined routes were collected.

Finally, other techniques were also used in other papers. In Fugiglando *et al.* [4], unsupervised analysis of the signals obtained from the CAN network was proposed. Eight signals that are directly or indirectly related to the driver's relationship with the vehicle are analyzed. The K-means technique was used to determine the minimum set of data to achieve significant results. In Zardosht *et al.* [27], a different approach was taken. The profile is obtained according to maneuvers in curves, where the signals were collected moments before the driver made a conversion. Data were gathered and analyzed using a statistical approach. Drivers were classified into two groups: moderate and aggressive. Yan *et al.* [28] decided to use an improved decision tree and gradient boosting technique, called LightGBM. It was analyzed data network combined with lane images of the road. The average accuracy of the model was 83%. In Lestyán *et al.* [22], it was discussed about the fact that each automaker uses its protocol for communication between electronic modules in the CAN network, which is not in a public domain. The non-standardization of network messages makes it difficult to extract data since when analyzing these messages, it is not possible to determine the content of each one. The "one versus all" technique was then proposed to re-identify the driver's profile without knowing the vehicle protocol in question. Although the algorithm has been analyzed in only thirty-three drivers and requires further validation with larger samples, it has proved to be a useful technique for performing data extraction. It was also verified that using a simple technique, the identification of each message as possible, thus questioning the efficiency of adopting proprietary protocols that could be easily decoded.

Finally, the efficiencies found in each of the articles were summarized in Table 4. Note that each of the authors conducted a research using different data and environments.

B. SECURITY AGAINST CYBERATTACKS

It was observed in the last paper the fragility of using CAN Bus to send important vehicle data and the importance of developing alternatives to ensure that data could be safely sent on the network. With the increase of vehicle connectivity through technologies such as Bluetooth, Wi-Fi, and smartphones, the system vulnerability becomes clear. If an electronic module is hacked, communication messages on the network could be simulated, allowing undesired braking or acceleration in the vehicle, causing then an accident.

According to Hanselmann *et al.* [14], with the development of deep learning techniques in recent years, new tools could be introduced to detect potential attacks into the vehicle. In their paper, Hanselmann *et al.* [14] proposed the use of neural networks to create a system called CANet, capable of detecting cyberattacks and even predicting technical network failures. The main idea behind this system is to manage the

TABLE 3. Articles selected through systematic mapping of literature.

N°	Ref.	Title	Published Year	Category	Percentile
1	[4]	Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment	2018	Computer Science	96th
2	[5]	Internet of Things Meets Vehicles Sheltering in Vehicle Network through Lightweight Machine Learning	2019	Computer Science	75th
3	[6]	Multi-Dimensional and Multi Scale Modeling of Traffic State in Jiangxi Expressway based on Vehicle Network	2019	Engineering	50th
4	[7]	A Drivers Car Following Behavior Prediction Model Based on Multi Sensors Data	2020	Computer Science	63rd
5	[8]	An Integrated Self-Diagnosis System for an Autonomous Vehicle Based on an IoT Gateway and Deep Learning	2018	Computer Science	70th
6	[9]	Malware Detection in Self Driving Vehicles Using Machine Learning Algorithms	2020	Computer Science	60th
7	[10]	Road Surface Classification Using a Deep Ensemble Network with Sensor Feature Selection	2018	Engineering	84th
8	[11]	Who is Behind the Wheel Driver Identification and Fingerprinting	2018	Computer Science	97th
9	[12]	Driving Style Recognition Based on Electroencephalography Data from a Simulated Driving Experiment	2019	Psychology	76th
10	[13]	An Overview on Study of Identification of Driver Behavior Characteristics for Automotive Control	2014	Engineering	62nd
11	[14]	CANet: An Unsupervised Intrusion Detection System for High Dimensional CAN Bus Data	2020	Computer Science	95th
12	[15]	A Deep Learning Framework for Driving Behavior Identification on In Vehicle CAN Bus Sensor Data	2019	Engineering	84th
13	[16]	Anomaly Detection of CAN Bus Messages Using a Deep Neural Network for Autonomous Vehicles	2019	Computer Science	70th
14	[17]	Gated Recurrent Fusion to Learn Driving Behavior from Temporal Multimodal Data	2020	Computer Science	88th
15	[18]	Driver Identification and Authentication with Active Behavior Modelling	2016	Conference	22 (H5 Index)
16	[19]	Pedestrian Detection in Severe Weather Conditions		Computer Science	95th
17	[20]	Research on Safe Driving Behavior of Transportation Vehicles Based on Vehicle Network Data Mining	2020	Engineering	57th
18	[21]	An Intelligent Secured Framework for Cyberattack Detection in Electric Vehicles CAN Bus Using Machine Learning	2020	Computer Science	95th
19	[22]	Extracting Vehicle Sensor Signals from CAN Logs for Driver Re-Identification	2019	Conference	29 (H5 Index)
20	[23]	On Combining Big Data and Machine Learning to Support Eco-Driving Behaviors	2019	Computer Science	97th
21	[24]	Driving Behaviors Analysis Based on Feature Selection and Statistical Approach: A Preliminary Study	2019	Computer Science	73rd
22	[25]	Development of a Driving Behavior-Based Collision Warning System Using a Neural Network	2018	Engineering	76th
23	[26]	Fuel Consumption Using OBD-II and Support Vector Machine Model	2020	Computer Science	55th
24	[27]	Identifying Driver Behavior in Pre-turning Maneuvers Using In-Vehicle CANbus Signals	2018	Computer Science	60th
25	[28]	Research on Classification Model of Natural Driving Scenario Based on LightGBM	2019	Conference	61 (H5 Index)

network data structure by introducing several recurrent neural networks at input signals. The system achieved an efficiency of 99% in some conditions, higher than classic methods of detection, and tested in a simulated data environment. In Zhou *et al.* [16], neural networks were also applied to solve anomalies in the CAN Bus. In their research, data were gathered in a set of three and compared to an anchor set through three deep neural networks. After the algorithm training step, the detection accuracy, along with other techniques, was evaluated, in which their algorithm presented results with accuracy above 90%. Also, using this artificial intelligence technique, Xiao *et al.* [5], developed a recurrent neural network called SECCU. The SIMATIC simplified care model was introduced in their paper, to be incorporated with SECCU. This model requires a low computational cost and significantly improves the accuracy in detecting anomalies in the network.

In Avatefipour *et al.* [21], the support vector machine was the chosen technique to detect cyberattacks in the network. The technique was improved with the modified bat algorithm metaheuristic in order to find the optimal model parameters created by the support vector machine, avoiding then a possible optimal location and premature convergence. The applied technique introduced a detection efficiency of 97%, higher than the existing analyzed techniques (support vector machine and isolated forest) with an average of 85% detection. Furthermore, Park *et al.* [9] article, introduced a machine learning technique and performed a comparison with other existing techniques to prove the effectiveness of the proposed algorithm. According to the simulations, it was possible to

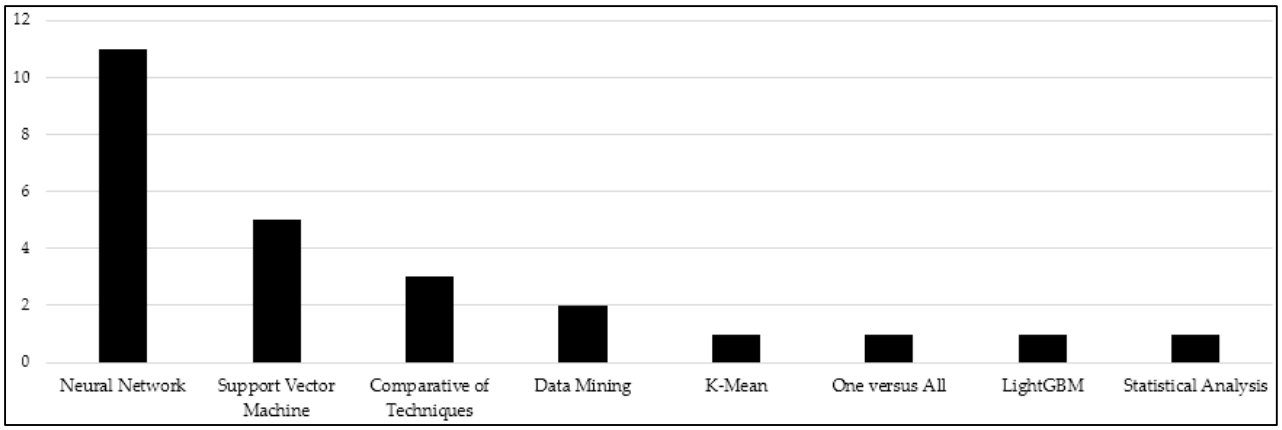


FIGURE 5. Comparison of the most used techniques for data analysis in a vehicular network. Graph was obtained with the information extracted in section II of this article.

TABLE 4. Comparison of the efficiency of driver profile techniques.

Ref.	Technique Applied	Efficiency
[4]	K-means	99%
[11]	Comparative of Techniques	Not provided
[12]	Support vector machine	80%
[13]	Comparative of Techniques	Not provided
[15]	Recursive neural network	98,36%
[17]	Recursive neural network	Not provided
[18]	Support vector machine	95%
[20]	Data mining	Not provided
[22]	One versus All	77%
[24]	Data mining	Not provided
[27]	Statistical analysis	Not provided
[28]	LightGBM	83%

achieve an accuracy of 92,90% in 0.049 seconds, making it suitable for real-time vehicular applications. The comparison of found techniques was clustered in Table 6.

C. OTHER APPLICATIONS

Besides driver's identification and security against cyberattacks, it was also identified on the found papers other applications for data analysis in the CAN network:

- Collision alert
- Estimated traffic on a highway
- Vehicle self-diagnosis
- Road condition identification
- Pedestrian detection
- Prediction of regenerative braking in electric vehicles
- Fuel consumption prediction

The collision alert system was proposed in Wang *et al.* [7] and Lee *et al.* [25]. According to Wang, hitting a vehicle rear is the most common traffic accident. One of the reasons that accidents could occur is due to the distance between vehicles. They could be close to each other and drivers do not have enough reaction time, or the rear vehicle driver shows lack of attention and does not notice the front vehicle braking. The work proposed by Wang *et al.* [7] discussed an improvement in the collision alert system based on vehicle's network existing data, such as vehicle speed, use of brake pedal position, use of acceleration pedal position and steering angle. Their improved proposal consists of anticipating possible braking moments and increase alert sound signal in order to prevent driver inattention moments. In Lee *et al.*[25], an artificial neural network learning algorithm was used to determine the collision risk according to driver's behavior. In this way, collision alerts would not always be given in the same condition, but only when the network detects a dangerous direction for that driver, avoiding then unnecessary alerts that may decrease driver confidence with the system.

In the research carried out by Chen *et al.* [6], neural networks were also used, however, with the aim of predicting traffic estimation on a highway in China. The vehicle data is collected and sent to the cloud, to be fused with data gathered from the highway and applied to a neural network to better manage traffic. In Park *et al.* [10], the road was also the focus of data use application, however, in order to determine its condition. A deep neural network was used on selected vehicle data in order to classify the road in four conditions: flat, sinusoidal, manhole, and pothole and bumps on a flat road. The algorithm showed 94.6% accuracy in the proposed classification.

Another possible found application is Delnevo *et al.* [23] paper. In their published article, an electric vehicle CAN network data are used to identify the right moment for the driver to regenerate the vehicle, in order to achieve higher autonomy. Different machine learning techniques are analyzed, and neural networks had the best result for the desired application. Also, considering an improvement in vehicle autonomy, Abukhalil *et al.* [26] proposed the support vector machine use to predict combustion engine consumption through engine speed and injection sensor data analysis.

TABLE 5. Comparative of Techniques to Determine Driver's Profile, adapted from Lin *et. al* [13].

Identification Methods	Neural Network	Hidden Markov Model	Fuzzy Control theory	Gaussian Mixture Model
Model accuracy	Very high	Very high	High	High
Real-time performance	Fair	Very good	Fair	The traditional GMM is poor, and the advanced GMM is good.
Disadvantages	There is not a unified feasible method to adjust parameters (e.g., the number of network layers) but generally subjective adjustments based on the simulation results of the models. Training time is long.	It is not suitable for long-term forecasting system and requires artificial hypothesis for the sequence distribution of the current states.	Since its fuzzy rules are formulated based on a *priori* knowledge, the simulation results may deviate from the actual values.	It cannot obtain more efficient modeling of the time series of feature vectors than other methods do.
Applications	It is suitable for pattern recognition that is easy to access to acquire the feature parameters, such as music recognition and speech recognition.	It is suitable for pattern recognition with strong time series data, such as driver's intention recognition and speech recognition.	Fuzzy control theory is suitable for pattern recognition whose parameter range is difficult to determine.	It is expert in identifying short-term driving behaviors but not in long-term driving behaviors. If combined with PWARX, the model can have a good performance both in the short- and long-term driving behaviors.

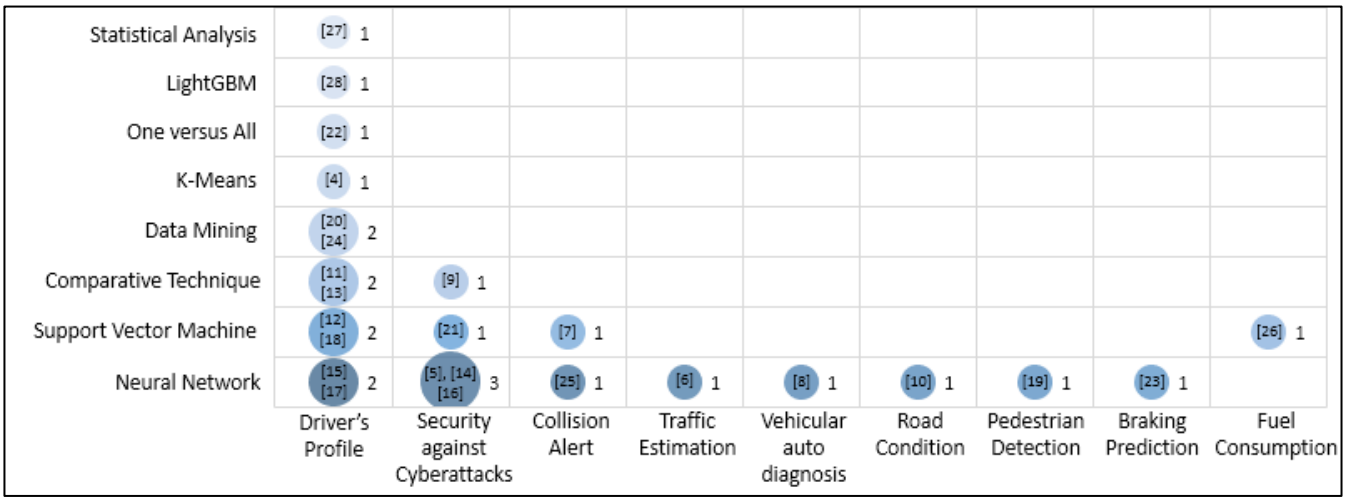


FIGURE 6. Frequency relation between the applications and found techniques for data analysis in a vehicular network. Graph was obtained using the extracted information in section II.

Furthermore, Tumas *et al.* [19] and Jeong *et al.* [8] also proposed neural networks use for their applications. Tumas *et al.* [19] developed a pedestrian detection system using vehicle cameras fused with CAN Bus data, whereas Jeong *et al.* [8] proposed a system that performs autonomous vehicle self-diagnosis to inform the driver. Both solutions presented acceptable results and possible applications to improve vehicle safety.

IV. DISCUSSION

The results and answers to the research questions presented in section II will be discussed below.

QP1: Which are the main concepts and techniques presented in research involving data analysis and artificial intelligence in vehicular communication networks?

The main techniques found for data analysis in a vehicular communication network are condensed in Fig. 5. Neural networks were responsible for 44% of the selected articles, followed by a support vector machine, with 20% of the published papers. Although neural networks have been used more in a scenario considering all applications, when the intention is to identify the driver's profile or behavior, this statement is no longer true, as other techniques have also been used showing promising results. This demonstrates that there is still no consensus among researchers to use an optimal technique for the driver's identification.

QP2: The techniques found are used for which types of tasks within the context of data?

The techniques found for data analysis were used mainly to identify the driver's profile and security against cyberattacks,

TABLE 6. Comparison of cyberattacks security techniques efficiency.

Ref.	Technique Applied	Efficiency
[5]	Recursive neural network	94,40%
[9]	Comparative of Techniques	92,90%
[14]	Recursive neural network	99,60%
[16]	Recursive neural network	95,00%
[21]	Support vector machine	97,01%

as discussed during section III and summarized in Fig. 2 in graphic format. Fig. 6 displays the frequency relation of the applications and analysis techniques. Other applications were also found, as explained in detail in subsection III.C.

QP3: Which are the existing gaps and opportunities in data analysis applications in vehicular communication networks?

According to this article, there is an opportunity to identify which are the best parameters to be considered in the network for the driver's identification. We observed that there is no consensus among the analyzed studies of which parameters have better efficiency, nor whether it would be better to use a combination of techniques. Although the searched articles results were above 80% accuracy, the techniques were not applied in real driving scenarios, with different types of drivers for greater data validation.

QP4: Is there a most efficient technique in the found applications?

The comparison of the techniques' accuracy was condensed in Tables 4 and 6. Table 4 contains efficiencies for determining the driver's profile, whereas Table 6 for security against cyberattacks. It is observed that Fugiglando *et al.* [4] paper using K-means provided 99% efficiency when the algorithm is used to determine the driver's profile. Table 6 shows that the recursive neural network had higher efficiency in Hanselmann *et al.* [14] article for security against cyberattacks.

V. CONCLUSION

The main idea of this systematic mapping of literature was to find out the state-of-the-art data analysis techniques in vehicular communication networks. It was verified that the CAN network is the most used network for this kind of analysis since it is present in practically all automotive vehicles. We observed through this article that the driver's profile identification was the most used application for data analysis, followed by security against cyberattacks. As for the techniques, it was found that neural networks are the ones that have the greatest focus of recent studies, although some other techniques mentioned also present consistent and high-performance results, such as the support vector machine and K-means.

This research was carried out following the systematic mapping methodology, where 196 articles were refined with a search protocol, which after the inclusion and exclusion criteria and a paper's quality assessment, twenty-five remained. These papers were the analyzed content by the authors to prepare this review. This article contributed to the state-of-the-art analysis techniques in automotive networks by providing an overview of the main techniques used by researchers around the world. As future opportunities, the identification of the best parameters within the network to compose the driver's profile is a niche to be further studied since there is no consensus on this topic in the analyzed papers.

REFERENCES

[1] O. Avatefipour and H. Malik, *State-of-the-Art Survey on In-Vehicle Network Communication CAN-Bus Security and Vulnerabilities*. Ithaca, NY, USA: Cornell University, 2018.

[2] *Postfach 50, D-7000. Stuttgart 1Print*, CAN-Bus Specifications Rep., Robert Bosch GmbH, Stuttgart, Germany, 2010.

[3] K. Petersen, "Systematic mapping studies in software engineering," *EASE. v.*, vol. 8, pp. 68–77, Dec. 2008.

[4] U. Fugiglando, E. Massaro, P. Santi, S. Milardo, K. Abida, R. Stahlmann, F. Netter, and C. Ratti, "Driving behavior analysis through CAN bus data in an uncontrolled environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 737–748, Feb. 2019.

[5] J. Xiao, H. Wu, and X. Li, "Internet of Things meets vehicles: Sheltering in-vehicle network through lightweight machine learning," *Symmetry*, vol. 11, no. 11, p. 1388, Nov. 2019.

[6] C. Zhaozheng, W. Yuanyuan, T. Zhengyu, and Z. Yuejin, "Multi-dimensional and multi-scale modeling of traffic state in jiangxi expressway based on vehicle network," *Int. J. Performability Eng.*, vol. 15, no. 12, p. 3287, 2019.

[7] H. Wang, M. Gu, S. Wu, and C. Wang, "A driver's car-following behavior prediction model based on multi-sensors data," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–2, Jan. 2020.

[8] Y. Jeong, S. Son, E. Jeong, and B. Lee, "An integrated self-diagnosis system for an autonomous vehicle based on an IoT gateway and deep learning," *Appl. Sci.*, vol. 8, no. 7, p. 1164, Jul. 2018.

[9] S. Park and J.-Y. Choi, "Malware detection in self-driving vehicles using machine learning algorithms," *J. Adv. Transp.*, vol. 2020, pp. 1–9, Jan. 2020.

[10] J. Park, K. Min, H. Kim, W. Lee, G. Cho, and K. Huh, "Road surface classification using a deep ensemble network with sensor feature selection," *Sensors*, vol. 18, no. 12, p. 4342, Dec. 2018.

[11] S. Ezzini, I. Berrada, and M. Ghogho, "Who is behind the wheel? Driver identification and fingerprinting," *J. Big Data*, vol. 5, no. 1, p. 9, Dec. 2018.

[12] F. Yan, M. Liu, C. Ding, Y. Wang, and L. Yan, "Driving style recognition based on electroencephalography data from a simulated driving experimentv frontiers in psychology," *Frontiers Psychol.*, vol. 10, p. 29, Mar. 2019.

[13] N. Lin, C. Zong, M. Tomizuka, P. Song, Z. Zhang, and G. Li, "An overview on study of identification of driver behavior characteristics for automotive control," *Math. Problems Eng.*, vol. 2014, pp. 1–15, Dec. 2014.

[14] M. Hanselmann, T. Strauss, K. Dormann, and H. Ulmer, "CANet: An unsupervised intrusion detection system for high dimensional CAN bus data," *IEEE Access*, vol. 8, pp. 58194–58205, 2020.

[15] J. Zhang, Z. Wu, F. Li, C. Xie, T. Ren, J. Chen, and L. Liu, "A deep learning framework for driving behavior identification on in-vehicle CAN-BUS sensor data," *Sensors*, vol. 19, no. 6, p. 1356, Mar. 2019.

[16] Zhou, Li, and Shen, "Anomaly detection of CAN bus messages using a deep neural network for autonomous vehicles," *Appl. Sci.*, vol. 9, no. 15, p. 3174, Aug. 2019.

[17] A. Narayanan, A. Siravuru, and B. Dariush, "Gated recurrent fusion to learn driving behavior from temporal multimodal data," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1287–1294, Apr. 2020.

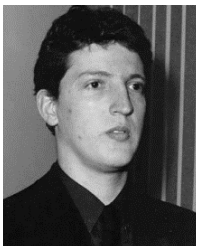
[18] A. Burton, T. Parikh, S. Mascarenhas, J. Zhang, J. Voris, N. Artan, L. Sertac, "Driver identification and authentication with active behavior modeling," in *Proc. 12th Int. Conf. Netw. Service Manage. (CNSM)*, Out. 2016, pp. 1–8.

[19] P. Tumas, A. Nowosielski, and A. Serackis, "Pedestrian detection in severe weather conditions," *IEEE Access*, vol. 8, pp. 62775–62784, 2020.

[20] Y. Sun, Y. Bi, Y. Han, D. Xie, and R. Li, "Research on safe driving behavior of transportation vehicles based on vehicle network data mining," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 5, p. e3772, May 2020.
[21] O. Avatefipour, A. S. Al-Sumaiti, A. M. El-Sherbeeny, E. M. Awwad, M. A. Elmeligy, M. A. Mohamed, and H. Malik, "An intelligent secured framework for cyberattack detection in electric Vehicles' CAN bus using machine learning," *IEEE Access*, vol. 7, pp. 127580–127592, 2019.
[22] S. Lestyán, G. Acs, G. Biczók, and Z. Szalay, "Extracting vehicle sensor signals from CAN logs for driver re-identification," in *Proc. 5th Int. Conf. Inf. Syst. Secur. Privacy*, 2019, pp. 1–15.
[23] G. Delnevo, P. Di Lena, S. Mirri, C. Prandi, and P. Salomoni, "On combining big data and machine learning to support eco-driving behaviours," *J. Big Data*, vol. 6, no. 1, p. 64, Jul. 2019.
[24] M.-S. Chen, C.-P. Hwang, T.-Y. Ho, H.-F. Wang, C.-M. Shih, H.-Y. Chen, and W. K. Liu, "Driving behaviors analysis based on feature selection and statistical approach: A preliminary study," *J. Supercomput.*, vol. 75, no. 4, pp. 2007–2026, Apr. 2019.
[25] S. H. Lee, S. Lee, and M. H. Kim, "Development of a driving behavior-based collision warning system using a neural network," *Int. J. Automot. Technol.*, vol. 19, no. 5, pp. 837–844, Oct. 2018.
[26] T. Abukhalil, H. AlMahafzah, M. Alksasbeh, and B. A. Y. Alqaralleh, "Fuel consumption using OBD-II and support vector machine model," *J. Robot.*, vol. 2020, pp. 1–9, Jan. 2020.
[27] M. Zardosht, S. S. Beauchemin, and M. A. Bauer, "Identifying driver behavior in preturning maneuvers using in-vehicle CANbus signals," *J. Adv. Transp.*, vol. 2018, pp. 1–10, Nov. 2018.
[28] B. Yan, Y. Pei, Z. Shuai, Z. Yang, and C. Jianhua, "Research on classification model of natural driving scenario based on LightGBM," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2019, pp. 1688–1693.
[29] K. Park, J. Kwahk, S. H. Han, M. Song, D. G. Choi, H. Jang, D. Kim, Y. D. Won, and I. S. Jeong, "Modelling the intrusive feelings of advanced driver assistance systems based on vehicle activity log data: Case study for the lane keeping assistance system," *Int. J. Automot. Technol.*, vol. 20, no. 3, pp. 455–463, Jun. 2019.
[30] L. Nkenyereye, Y. Park, and K.-H. Rhee, "Secure vehicle traffic data dissemination and analysis protocol in vehicular cloud computing," *J. Supercomput.*, vol. 74, no. 3, pp. 1024–1044, Jun. 2016.
[31] M. Remeli, S. Lestyan, G. Acs, and G. Biczok, "Automatic driver identification from in-vehicle network logs," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1150–1157.
[32] L. Zhu, S. Zhao, L. Zhang, B. Zhou, Y. Li, J. Hao, and Y. Wu, "A study on driving behavior intelligence detection based on discrete wavelet transform and support vector machine algorithmv," in *Proc. IEEE 4th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Dec. 2018, pp. 1042–1047.
[33] U. Fugiglando, P. Santi, S. Milardo, K. Abida, and C. Ratti, "Characterizing the driver DNA through CAN bus data analysis," in *Proc. 2nd ACM Int. Workshop Smart, Auton., Connected Veh. Syst. Services*, 2017, pp. 1–5.
[34] N. Virojboonkiate, A. Chanakitkarnchok, P. Vateekul, and K. Rojviboonchai, "Public transport driver identification system using histogram of acceleration data," *J. Adv. Transp.*, vol. 2019, pp. 1–15, Feb. 2019.
[35] C. Schwarz, J. Gaspar, T. Miller, and R. Yousefian, "The detection of drowsiness using a driver monitoring system," *Traffic Injury Prevention*, vol. 20, no. 1, pp. 157–161, Jun. 2019.
[36] M. Wu, S. Zhang, and Y. Dong, "A novel model-based driving behavior recognition system using motion sensors," *Sensors*, vol. 16, no. 10, p. 1746, Oct. 2016.
[37] L. Huan and L. Chao, "FlexRay vehicle network predictive control based on neural network," *MATEC Web Conf.*, vol. 232, Dec. 2018, Art. no. 01042.
[38] *Tabulation of found results in Google Drive folder*. Accessed: Oct. 18, 2020. [Online]. Available: https://drive.google.com/drive/folders/1EmeHRAnOI56v9S9nl6nZeC-cXWOhj7hG?usp=sharing

LUCAS GOMES DE ALMEIDA received the B.S. degree in electrical engineering from the Federal University of Itajubá (UNIFEI), in 2016, where he is currently pursuing the master's degree in science and computer technology. Since 2018, he has been with Volkswagen Truck and Bus, Resende, Brazil, as a Product Engineer with the Electrical Engineering Department, where he is responsible for developing electronic modules and instrument cluster for trucks and buses. His research interests include artificial intelligence, automotive electronics, embedded hardware and software, and the Internet of Things.

ADLER DINIZ DE SOUZA received the degree in science computer, the Project Management Professional (PMP) and the M.B.A. degrees in software process improvement from the Federal University of Lavras (UFLA), in 2004, 2005, and 2005, respectively, the master's degree in system engineering and computer and the Ph.D. degree in system engineering and computer from PESC, COPPE, UFRJ, in 2008 and 2014, respectively. He is currently a Professor with the Federal University of Itajubá (UNIFEI)-Undergraduate and Postgraduate. His research interests include project management, software process improvement, portfolio management, and earned value management (EVM).



BRUNO TARDIOLE KUEHNE graduated in computer science from the Catholic University of Minas Gerais, in 2006, and the master's degree in computer science and computational mathematics and the Ph.D. degree in computer science from the University of São Paulo, in 2009 and 2015, respectively. Since 2014, he has been a Professor with the Federal University of Itajubá. He has experience in computer science, with emphasis on computer systems, working mainly on the following topics: QoS, web services, cloud computing, fog computing, the IoT, and performance evaluation.

OTAVIO S. M. GOMES (Member, IEEE) received the B.S. degree in computer engineering and the M.S. and Ph.D. degrees in electrical engineering from the Federal University of Itajubá, Brazil, in 2007. From 2012 to 2019, he was a Professor and a Researcher with the Federal Institute of Minas Gerais, Brazil. Since 2019, he has been an Assistant Professor with the Institute of Systems Engineering and Information Technology, Federal University of Itajubá. He has a solid knowledge in reconfigurable hardware, computer architecture, and cryptography. His research interests include embedded systems, reconfigurable computing, operating systems, and information security.