Conceptual Model for Connected Vehicles Safety and Security using Big Data Analytics

Noor Afiza Mat Razali[1]
Nuraini Shamsaimon[2], Muslihah Wook[3]
Faculty of Defence Science and Technology
National Defense University of Malaysia
Kuala Lumpur, Malaysia

Khairul Khalil Ishak[4]
Center of Cybersecurity and Big Data
Management and Science University
Shah Alam, Selangor

Abstract—The capability of Connected Vehicles (CVs) connecting to a nearby vehicle, surrounding infrastructure and cyberspace presents a high risk in the aspect of safety and security of the CV and others. Data volume generated from the sensors and infrastructure in CVs environment are enormous. Thus, CVs implementations require a real-time big data processing and analytics to detect any anomaly in the CVs's environment which are physical layer, network layer and application layer. CVs are exposed to various vulnerabilities associated with exploitations or malfunctions of the components in each layer that could result in various safety and security event such as congestion and collision. The safety and security risks added an extra layer of required protection for the CVs implementation that need to be studied and refined. To address this gap, this research aims to determine the basic components of safety and security for CVs implementation and propose a conceptual model for safety and security in CVs by applying the machine learning and deep learning techniques. The proposed model is highly correlated to safety and security and could be applied in congestion and collision prediction.

Keywords—Connected vehicles; safety and security monitoring; collision prediction; congestion prediction; machine learning; deep learning

I. INTRODUCTION

Connected Vehicles (CVs) is becoming more relevant in recent years after the realization of Industrial Revolution 4.0 (IR4.0), especially for the implementations in the smart cities and the Intelligent Transportation System (ITS) [1], [2]. CVs introduces a new concept of Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I) and Vehicle to Everything (V2X) concepts that has high potential to become a destructive technology that could change on how communities commute and will impact the economic landscape of logistic and transportation industries [3], [4]. Safety and security of CVs are one of the hot topics in research that had gained research attention. The basic characteristic for CVs is its ability to connect to other vehicles, to surrounding infrastructure and the internet through sensors such as Laser Detection and Ranging (LiDAR), Radio Detection and Ranging (Radar), Global Positioning System (GPS), Dedicated Short Range Communication (DSRC), Radio Frequency Identification (RFID), Advance Driver Assistance System (ADAS) and sensors that are embedded in the vehicles itself [5], [6]. Communications devices are either embedded in the vehicle or connected to the vehicle's power socket. Due to the capability of CVs connecting to a nearby vehicle, infrastructure and cyberspace, it presents a high risk in the aspect of safety and security of the CVs and others, particularly if any vulnerability is exploited or malfunction occurred in any of the sensors.

CVs implementations will also require real-time big data processing and analytics, for instance, to detect any anomaly for the CV's network communication or employing CV sensors for collision prediction. This will add an extra layer of required protection for the safety and security of CV implementation. Acknowledging the importance of CV's safety and security, countries, such as United States of America (USA) had introduced the framework for CVs deployment in the country. The framework focuses on the deployment of CV in regards to privacy, cybersecurity, safety regulation, ethical issues and more. Information-Technology Promotion Agency, Japan, has published a Vehicle Information Security Guide [7]. This guide presents potential threats faced by automotive systems and security measures against those threats, aiming at helping automotive system developers improve their security design.

Various studies [4], [6], [8] has discussed regarding the emergence and implementation of connected and autonomous vehicles, to improve the driving experience and reduce the risk of a crash, improve traffic control and provide real-time interactive communications between other vehicles as well as roadside infrastructure in a network.

This proves beneficial, as the future implementation of smart cities requires Intelligent Transportation Systems (ITS) to promote smart mobility in a city. But the implementation of a CV environment contracts several implications and challenges. Researchers also are discussing the improvement of traffic control to reduce traffic congestion.

Hence in this paper, the focus would be on discussing the safety and security of CVs based on physical, network and application layers. This paper proposes a conceptual model for safety and security in CVs by applying the machine learning and deep learning techniques. The proposed model is highly correlated to safety and security and could be applied in congestion and collision prediction.

II. BACKGROUND

Many studies are being done to investigates the implementations, implications and challenges of the CVs in the

Smart Cities or as the main component in the Intelligent Transportation System (ITS) [6], [9], [10]. However, many studies are focusing more on the utilization of individual or combinations of technologies in CVs applications. The ecosystem as a whole or the development of a model for the implementation of CV technologies in a particular targeted environment that take safety and security into deep consideration is still a new field that needed to be explored [4], [6].

A. Safety and Security for Connected Vehicle

The main challenges of CV implementation would be safety and security. In a CV environment, where communications occur between vehicles and infrastructure in a network could open up various possibilities and threats for attacks and misdemeanor. Moreover, there is a lack of clear guidelines and requirements for the usage of CV. Safety focuses on the physical aspects of an accident failure, while security focuses on the failures caused by malicious attackers [11].

1) Safety: Safety focuses on the physical aspect of the CV implementations, that includes several main aspects such as Driver Safety, Vehicle Safety, Road Safety, and Traffic Safety based on studies in various literatures [12]–[15].

Based on Fig. 1, we are focusing on these four aspects as the main pillars of safety in CV, which includes (i) Driver safety; where the condition of the driver maneuvering and handling the car contributes to the care of their safety. (ii) Road safety; the condition of the road infrastructure that can contribute to the occurrence of accidents. (iii) Vehicle safety; the condition of the vehicle, through its ability to operate in a well-mannered condition, and low risk of vehicle failure from occurring. (iv) Traffic safety; where the condition of traffic signs, stoplights can contribute to the lack of information for the driver, regarding the possible warnings and hazards on the road.

Studies by [16]–[22] discussed that safety components in CV environments include road friction, accident and collision prediction, road user detection, cluster identification, route planning and image signing.

Fig. 2 shows the components of safety in the CV. Road friction and conditions are one of the factors that can lead to accidents or collisions from happening. For example, when the condition of the road is slippery due to weather condition, the risk for collision is slightly increased due to the reduced amount of friction from vehicle tires and the road. Accident or collision prediction is a feature where if the proposed model can estimate that an accident would occur, an alert or notification can be sent to alert and notify the driver for action to be taken. Road User Detection is to provide a wider range of information regarding the surrounding of CV. Image Sign Board is a feature to detect road signs along the driveway. As an example, a speed bump is detected, the CV can be alerted to slow down their vehicle. Cluster identification is a feature to determine the cluster head or leader of a certain cluster of CVs in a driveway. Cluster heads are responsible to efficiently communicate road information for their clusters, alert for any situations and are chosen based on their CV capability. Route

Planning feature is for CVs to calculate all available routes which are safe and avoid any unwanted situation.

2) Security: As mentioned in [11], security is more leaning towards malicious attacks on the networks where it can be from external or internal, either with intentions or not. This can affect and interrupt the transmission of data and data analysis process. Fig. 3 shows the security principles in Information Security, where each of the principles are important aspects to ensure the security of operations in CV environment. (i) Availability; where only authorized user can reliably access to information they need. For example, a car owner is right to be assigned as an authorized user and respond to the CV information. (ii) Confidentiality; is to prevent unauthorized users from the disclosure of information or data for their benefit by limiting their access[23]. (iii) Integrity; means that data cannot be deleted, modified or added by an unauthorized user. (iv) Authentication; is about data or information that proves who you are. It is about username and password to identified user as a legal user. This authentication in CV is a key component to allow you to gain or responsible for all the action. (v) Authorization; when system approved a legal user that allows them to do the next action.

Studies by [23]–[25] discussed that security aspects in CV environment include cyberattack, intrusion detection and prevention and attack classification. Fig. 4 shows the aspects of safety in CV implementations. The list of security aspects where every of it has their task and operation to be executed in the CV environment to ensure that communication transmission in the network involving CV can be protected from any unwanted occurrences. Intrusion detection and prevention in a network are the possible types of intrusion that can be detected and prevented through the implementation of security measures. Meanwhile, attack classification is the method to classify types of attack which are possible to occur in a CV networking environment. Cyberattack is another unwanted type of attack that can breakdown networking infrastructures.
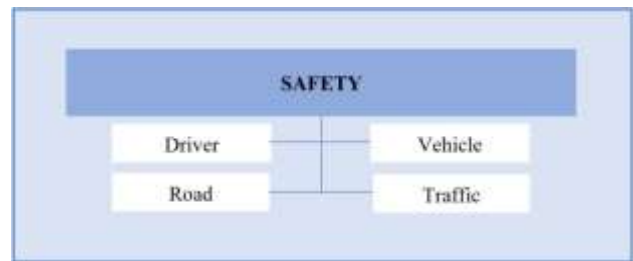

Fig. 1. Safety Aspects in CV.

Fig. 2. Safety in CV Implementations.
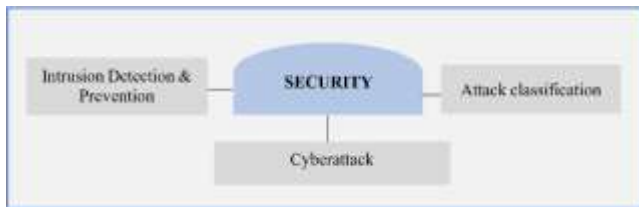
Fig. 3.   Security Aspects in CV.

Fig. 4.   Security in CV Implementations.

III.   Components of Safety and Security in Connected Vehicles

In the CV environment, safety and security need to be considered in three layers perspective which consists (i) Physical Layer; (ii) Network Layer; (iii) Application Layer.

A.   Physical Layer

The physical layer is including:

1) Road Side Infrastructure (RSI) and Road Side Units (RSU): an infrastructure-based communication meeting-point, where vehicles can communicate within the proximity of RSU installed in the nearby infrastructure. The higher number of RSU deployed in the infrastructure, the higher the capacity of vehicles able to communicate within the network. In a connected vehicle environment, RSU plays an important element in ensuring the reliability of both V2V and V2I communications [26], [27].

2) On-Board Unit (OBU): OBU and Onboard Sensor (OBS) is a fixed networking device in a vehicle, usually connected to the wireless network. The components of OBU typically consists of the GPS unit, human-machine interface unit, a wireless communication unit, and the central control unit. The central control unit is the main unit, consisting of data transceiver, serial port for information processing, memory, and decision and judgement making [28].

3) RSU and OBU Technologies: In recent years, there are rapid developments of the next-generation vehicle system that are based on connected vehicles (CVs) platform. CVs utilizes state of the art technologies such as Light Detection and Ranging (LiDAR), Global Positioning System (GPS), Dedicated Short Range Communication (DSRC), Advance Driver Assistance System (ADAS) and many more [5].

a) Dedicated Short-Range Communication (DSRC): The usage of DSRC protocol in the vehicular and automobiles industry has introduced innovations and technologies that utilize the DSRC communications service in vehicles to provide traffic safety and enhancement of mobility. DSRC provides dependable and quick information and data interchange for vehicle-based communications. Applications of DSRC help for the implementation of secure communication between RSU and OBU of a CV environment using V2I and V2V [29].

b) Advanced Driver Assistant (ADAS): The implementation of ADAS technologies is based on a vision/camera system, with several of researches have discussed regarding the implementation of camera-based ADAS [30], [31]. Combination of cameras with sensors for surround-view of safety vision detection applications such as to detect pedestrians and automatic braking functionalities have improved in recent years. The camera can be considered an essential part of the CV network to provide vision and imaging for the system.

c) Light Detection and Ranging (LiDAR): The usage of LiDAR is a remote sensing application able to create a three-dimensional (3D) representation of characteristics of vehicles in a CV environment. LiDAR technology uses light to estimate the parameters of a surface, in this case, vehicles. LiDAR is chosen compared to other mainstream sensor technologies due to its ability to acquire accurate calculations and measurements of the vehicle for its speed, type, and position. The data generated by LiDAR is highly accurate, due to its ability to cover an area with a view of 360 degrees, without depending on the light conditions [1], [32].

d) Global Positioning System (GPS): GPS is an essential unit in a CV network as it acts as one of the main units to accurately determine the position of a device/vehicle. GPS signals received from the satellite are interpreted and filtered by the GPS unit in the OBU. The recorded data are then taken to calculate the location, speed and the rate of change of speed of vehicles [28].

e) Traffic Light Controller (TLC): A TLC in a CV environment acts as a device that dictates instructions and provides a set of rules for drivers to rely on, for the main purpose of avoiding collisions, giving directions, and giving warnings. Several studies have proposed solutions on implementing fuzzy logic in TLCs, mainly to control traffic volume, hence reducing delays and increase data interchange. This proves to be beneficial in a CV environment, where maximum data throughput is recommended to maximize productivity [33]–[35].

f) Radio Frequency Identification (RFID): The implementation of RFID in a CV environment is mainly used for vehicle positioning purposes besides GPS. The OBU will be attached with an RFID tag, and the reader placed at the RSU. This creates an intra-vehicles sensor network. The usage of RFID in critical locations acts as a replacement solution, where GPS could not be utilized and used for positioning purposes [36], [37].

B.   Network Layer

Communication in CV is one of important element and component to be implemented. CV technology can also further

increase the efficiency and reliability of autonomous vehicles, though these vehicles could be operated solely with their on-board sensors, without communication [38].

There are several types of communications in the connected vehicle, that are mainly split into three parts; vehicle-to-vehicle communication (V2V) and vehicle-to-infrastructure communications (V2I) and vehicle-to-anything communication (V2X). The capability of a connected vehicle, to communicate with other vehicles, and the infrastructure (RSU) reliably open up various possibilities, including to enhance the safety and security of operating vehicles.

V2V communications in connected vehicle environment have contributed to providing important reassurance on the improvement of operational safety in vehicles such as collision warning, as vehicles can communicate actively within the area through message exchange, primarily for accident prevention and warnings [28], [39]. V2I is also an essential part of the CV communication for the road-side units (RSU) to provide continuous connection and communication in the network [40]. V2X is an entire network communication where all infrastructures and vehicles are internetworking amongst each other to communicate and transfer information and data. Vehicular cybersecurity attacks include the shutdown of engines, tampering and disabling of brakes is an example of how an attack on the security of CV implementations could prove dangerous. Various articles have discussed regarding this issue and proposed various methods to detect and prevent these attacks [23], [24], [41], [42].

Fig. 3 shows the components which are involved in the security aspect. Security focuses on network security connection, between road users and the RSU/OBU in the CV environment. A study [23] states that the cybersecurity has its concerns to protect CVs, especially in network communication, to avoid threats or attack that can compromise CV functions. The threats can be done remotely, where data communication can be stolen, altered, and destroyed. Studies have found out the attributes of CV's cybersecurity as follows; (i) it is difficult to estimate all potential attack before it occurs. Attackers would only need to determine a vulnerability gap for them to infiltrate, while defenders are required to ensure all vulnerability gap is secured from potential attacks. Network attack and prevention have its challenges to ensure full security. (ii) there are a variety of connection medium in the network, that includes DSRC such as Wi-Fi, Bluetooth and others. (iii) CVs have different sensors and technologies such as LiDAR, where each sensor has its capabilities, functionalities and types of data which is compatible with CV. (iv) CV environment consists of various components and functions, where there is a possibility if one component dysfunctions, would affect the performance of the whole system. If the system is being attacked it can give bad impact and consequences to CV road users.

C. Application Layer

CV is an application that lies within the concept of Smart City in IoT, that consists of technologies such as smart transportation, smart parking, smart building and others where everything can communicate amongst each other in a network [43]. The smart city is technically an urban high-tech city, that

enables people to improve the quality of life alongside technology. People would be able to utilize technology resources to further improve their daily life and expands the growth of an urban city in their country [44]. In a CV environment, varieties of transportation-related fields can be expanded, such as daily traffic monitoring, smart parking that enables users to conveniently locate the nearest parking spot available for their vehicle that will be using technologies and sensors such CCTV, LiDAR, mobile devices, GPS, accelerometers, gyroscope-based applications, weather sensor, ADAS, DSRC, TLC and RFID. All these technologies and devices for applications in CV contribute to the growth of IoT around the world.

D. Connected Road User

In a CV environment, the users of the network would be any pedestrians, cyclists, motorcyclists and other vehicle drivers that possess a personal mobile device such as tablets and smartphones with portable DSRC units. For a vehicle such as cars, buses, and lorries to be connected, they would require an On-board unit, such as a DSRC unit to receive real-time DSRC messages, as well as broadcasts their data and information to other vehicles and connected infrastructures in the CV environment. Through the connection of DSRC Unit to the personal device such as tablets is represented through an application, whereby the personal device would require connection to the DSRC Unit through Bluetooth. Once connected, the drivers would receive real-time traffic data, alert messages, collision warning and communication with other CVs in the network [4], [6], [39].

E. Big Data Analytics

Big data is a concept of data in which it is very massive, unorganized, unstructured data which could not be processed and analyzed by a traditional IT hardware and software in a considerable and tolerable amount of time [45], [46]. This is why techniques for data analytics that includes ML and DL have been introduced to provide a better solution for big data management. This opens up various opportunities in the advancement and development growth of technologies such as IoT in businesses and organizations. Big data analytics are also involved in the implementations of CV [47], [48], where a massive data collection and management occurs in a CV environment as a massive amount of vehicles are involved.

F. Data Analysis: Machine Learning and Deep Learning Techniques in Traffic Data Centre

The traffic data centre is a process of data analysis and determination of results for a certain condition based on collected data through the Machine Learning and Deep Learning techniques such as Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), Linear Regression (LR), Clustering, Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and Fuzzy Logic (FL).

1) Decision Tree (DT): DT is known as a prediction tree that uses a tree structure for sequences and consequences decision specification. Constructing the DT with test points and branches can achieve the prediction. At each test point, a decision is made to choose a specific branch and cross down

the tree. Test points require testing for particular input variables and each of the branches represents the decision being made. In [23], the research has implemented DT in CV, to classify and detect connected autonomous vehicles in cyber-attack. DT is one of the most-used classification models with good readability. In [24], DT was implemented as an intrusion detection for CV in smart cities. The usage of DT is a feature selection and attack classification purposes. This paper concludes that the accuracy for detection rate is effective after considering the false positive and false negative rate generated using this method. From these two authors, it can be seen that DT can make a good classification for prediction, especially on detecting which nodes are malicious and non-malicious, to prevent cyber-attack, mainly in a CV environment.

2) Random Forest (RF): RF is a machine learning method that operates by constructing a multitude of decision trees by using classification, regressing and other tasks. This shows that RF is simplicity and diversity which can be used for more than 1 task. RF operating at training time and outputting the classes into classification or mean prediction into the regression of individual trees. RF is categorized in supervised learning algorithms. Several studies [49], [50] discuss using RF implemented in CV environments to classify supervised time-series based on driver behaviour and to classify vehicle recognition to differentiate among road users such as pedestrians, bicycles, cars and others. As a result, these two studies show that RF is effective to prevent overfitting for driver behaviour and successfully integrate the data and processing to identify and differentiate road user.

3) Naïve Bayes (NB): Classification method based on Bayes theorem in which it has the capabilities to provide the relationship between two event probability and their conditional probabilities. NB makes strong assumptions for presence or absence features of a class that are not related to other features independently. The NB is easily implemented and can be executed efficiently without prior knowledge of the data. In CV environments, [17] used NB to predict accidents and congestion before it happens. The author experimented by completing 10-fold cross-validation on their dataset. All accident severity types including minor, intermediate, major and NULL has experimented. As a result, NB is not better compared to Distributed Random Forest (DRF) but NB has fewer features to collect and make the decision quickly. In [51], NB was used to identify a driver's identity. The classification was made through a voting mechanism (VM) to analyze data in details of automotive characteristics based on a large number of sample data. Henceforth, it has successfully recognized 10 driver identification. In comparison, we can see that NB provides high accuracy for making the right decision and can be calculated with fewer data or samples quickly.

4) Support Vector Machine (SVM): SVM is a classification or regression algorithm that is categorized as a supervised machine learning algorithm and it is mostly used in problem classification. In SVM, each data item has placed as in n-dimensional space (N is number features) with the value of each feature. Specifically, it coordinates to find a hyper-plane where it can be separated into 2 classes. SVM is an extended version of a linear regression which can give high accuracy and give a simple decision boundary. SVM has limitations where it can separate only 2 classes. Usually, implementation of SVM is suitable for text classification, spam detection and computer version identification. The author in [16] researched about the implementation of a CV environment that can predict the friction class for specific road segments by using SVM, logistic regression (LG) and artificial neural networks (ANN) method. Another research in [22], used SVM with the Distance to Border features of the segment blobs for classification to detect and recognize traffic signs based on colour information in image sequences. The studies show that SVM gives good segmentation and classification results and can perform well in high dimensional spaces and the algorithm is very versatile and effective in cases where the number of dimensions is greater than the number of samples.

5) Linear Regression (LR): An analytical technique which is used in model relationships for 2 variables by setting a linear equation to observe the data. X represents an independent variable (the variable wants to explain or forecast) while Y is a dependent variable (explaining the other variables). The leverage of LR works with almost any kind of dataset and gives good information about the features. The opposite of LR is quite some assumptions than accurate decisions. Experimental use of LR in connected vehicles has been made by authors in [17] where they measure the accurate Estimated Time of Arrival (ETA). The reason researchers use this technique is to predict the clearance time after an accident occurred. The purpose of LR calculating the clearance time is to update ETA for indexed trips and giving the most accurate time. As a result, this novel predicts that using one of these methods will decrease the accident rate and give high accuracy and latency results.

6) Clustering and K-Mean Clustering: Clustering is an example of unsupervised classification technique in ML, where several similar data points, are divided into a group different from other groups of similar data points. In a CV environment, an example of the usage of clustering in [20] shows a weight-based clustering algorithm of vehicles in the same road segment to determine the primary cluster head (PCH) and secondary cluster head (SeCH). Another type of clustering is K-Mean Clustering which is a repetitive type of clustering, In [52], K-Mean Clustering was used to collect the journey time and volume data of several clusters, to identify the boundary value of each cluster. Clustering offers several solutions to CV environment implementations but would require a combination with other techniques such as Fuzzy Logic (FL) and DT to produce a reasonable outcome.

7) Artificial Neural Network (ANN): ANN is a collection or group of multiple neurons or perceptrons, where inputs are processed only in a forward-facing direction, that performs well when precise knowledge of a relationship requires some functional approximation. In a CV environment, authors in

[16], [53] implemented ANN in their experiment of studying the effect of the number of hidden layers in the performance of the wireless network. An interesting study in [54] has also implemented ANN to predict the level of severity of a vehicle's driver, at signalized intersections where an accident has occurred. ANN models and implementation have seen an uprise in the field of transportation, due to their adaptive capability and nature. Most implementations of ANN in research are to investigate its capabilities to enhance the wireless network performance, as well as other aspects such as the condition of the vehicle driver. This would prove beneficial for CV infrastructure safety applications.

8) Convolutional Neural Network (CNN): CNN is a neural network which is used productively for the classification and recognition. CNN is highly adept in areas like identification of objects and traffic signs, besides being able to generate vision on self-driving cars. In CV, application and usage of CNN can be seen in the literature [18], [19], [21], [55], where most CNN models are used for accident analysis and prevention, that are applied in some research to efficiently map crash risk, traffic conflicts and perception models for network traffic control. A software-defined network (SDN) model, SeDaTive [21] implements CNN model to provide data input to the model, where the CNN model studies the hidden patterns in data nodes, to plan the most optimal route for the model. This illustrates that using CNN for data classification helps to ensure effective network traffic control. This deep learning method can be considered for a data classifier in collision prevention model.

9) Fuzzy Logic (FL): FL is an AI method that bears a resemblance to human reasoning. The technique and process of FL emulate the way humans make decisions, involving all possibilities between the values of YES and NO. FL can be applied and used in automotive systems, including CV environment. Several studies [25], [56] discuss the use of FL techniques and algorithms, that includes, (i) using FL as a base algorithm for vehicle's decision making systems that can make decisions based on reasoning similar to human reasoning, and (ii) detection of attacks such as message injection through the classification and differentiation of injected malicious, fabricated and normal packets for the vehicle network. Based on these two authors and several other studies, it can be seen that fuzzy logic is an ML algorithm that fits well in-vehicle environment control, which would benefit the CV environment implementations.

10) ML and DL Techniques in CV implementation: Table I illustrates a summary of all works of literature with the techniques or methods of ML and DL related to the CV environment that we surveyed. Some examples of studies include; calculating the severity of injury based on accident impact, road friction estimation, collision prediction and avoidance and cluster recognition. From these studies, can be used as a reference for this study to determine the method or technique which is suitable for our model.

TABLE I. MACHINE LEARNING AND DEEP LEARNING TECHNIQUES USED IN CONNECTED VEICHILE

Studied	Techniques	Parameters
[17]	LR, NB and Distributed RF	Distance and speed
[18]	Multi-layer perceptron (MLP), CNN and DT	Speed, acceleration, deceleration, wait time and wait ratio
[19]	Regional–Convolution Neural Network (R-CNN) and Deep Neural Network (DNN)	Speed ratio, driver characteristic and road/environmental condition
[21]	CNN	Bluetooth, Zigbee, DSRC, Wave, DSA, Wi-Fi, WiMAX and Cellular
[16]	Logistic Regression, SVM and ANN	Level friction segment (error rate, sensitivity and specificity)
[22]	SVM	Image Processing (color, shape, pattern recognition)
[51]	K-Nearest Neighbor and NB	Accelerator pedal, brake pedal, engine speed, steering wheel and vehicle speed
[49]	RF and RNN	Accelerations, lane changes, braking and aggressive and normal left and right turns
[50]	RF	Background filtering, object clustering, vehicle recognition, lane identification and vehicle tracking
[20]	Clustering	Number of neighbouring node, position, speed, road id, direction,
[52]	Fuzzy Logic	Volume, journey time, speed, and weather conditions
[56]	Fuzzy Logic	Speed and acceleration
[54]	ANN	Driver info, seat belt, vehicle type, speed, area type and point of impact

Table II and Table III illustrate the allotment of the ML and DL techniques which is supervised and unsupervised according to the review conducted in Table I. The classifications are also based on a study conducted by [43]. Some examples of supervised ML and DL methods are DT, RF, CNN and RNN. Supervised learning is where the techniques can perform learning on a dataset that is labelled in which the accuracy of the training data can be evaluated provided by an answer key for the algorithm to use. Meanwhile, the unsupervised model algorithm needs to process unlabeled data and learn to understand through features and pattern extraction. Some of the ML and DL techniques which are unsupervised are Clustering and FL.

TABLE II. APPROACH FOR ML METHODS

Machine Learning	
Supervised	Unsupervised
DT	Clustering
RF	FL
NB	Classification
SVM	Classification

TABLE III.   APPROACH FOR DL METHODS

Deep Learning
Supervised
CNN
RNN

IV. CONCEPTUAL MODEL FOR SAFETY AND SECURITY IN CONNECTED VEHICLE USING BIG DATA ANALYTICS

Based on the literature that we surveyed in this study, we proposed the conceptual model for safety and security illustrated in Fig. 5. Each aspect has its variable to ensure each component functions correctly and efficiently. We propose a safety and security concept to ensure both aspects can give virtuous impact on functionality for CVs implementation. The variables and parameters listed are leaning more towards safety, which is considered the main focus. This is due to being cautious regarding the physical part of a CV, that includes the vehicle's and driver's safety, including the driver's condition, the level of driver's injury if accidents occur, as well as the driving style. Meanwhile, for security, existing technologies such as LiDAR, ADAS and others are used to secure the virtual part of the CV implementation, such as networking, CV communication and attack prevention. Parameters data and existing technologies will be analyzed using several ML and DL methods that include CNN, NB, SVM and others.

A. Conceptual Model for Safety and Security in Connected Vehicle

To address the safety and security for CVs environment, data analysis which is implemented in the application layer would be conducted in the other two layers which are both Physical and Network Layer need to be considered.
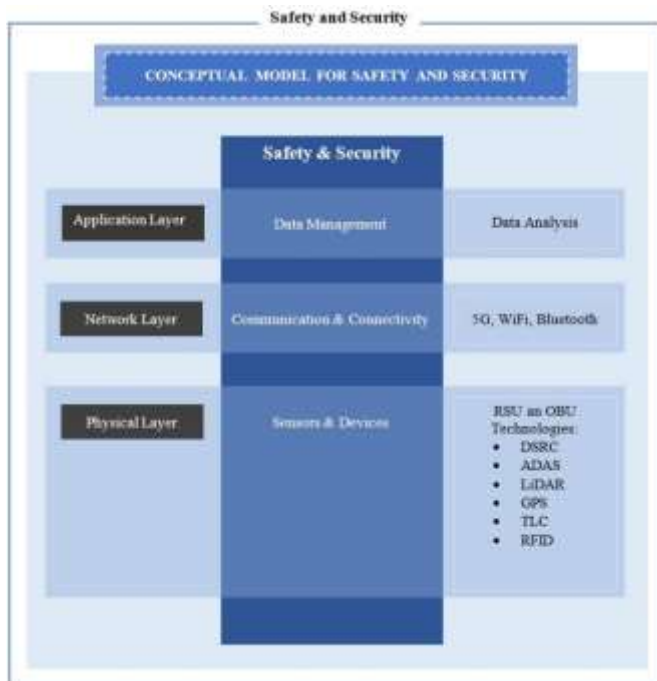

Fig. 5.   Conceptual Model for Safety and Security in Connected Vehicle.

For security, data analysis in network layer determines to be the method in the communication transmission process. In CV, communication is important for the interaction between RSU and OBU(CV) for data collection and process to produce results of the possibility of congestion, which will be sent to other CVs for action. The transmission process is essential to maintain network security from attacks. If communication is interrupted, hence the possibility of obtaining results would be hindered, that can cause chaos due to inaccurate information.

For safety, data analysis in physical layer determines to be the method of accident and collision prevention, where the model would utilize the ML methods to predict any possible collision from occurring during the congestion. Based on the ML techniques, an alert notification message would be able to be transmitted for the CVs to take action and be prepared for the incoming congestion to avoid any possible collision.

1) Safety: Data Analysis in Physical Layer using Conventional Neural Network and Naïve Bayes: For safety aspect, CNN and Naïve Bayes are considered as two main selected techniques in this paper, based on [17]–[19]. NB, are used in [17]as they are reliable, and fast for collision prediction through the sending of alerts and notifications. In the CV environment, for real-time and minimum data collection, NB is a technique that suits well as it has a fast computation time and produces the right decision quickly. CNN is also selected, as it can analyze collision risks at intersections [18], as well as to detect vehicles and lane through image processing obtained from a single front-facing camera [19]. It is believed that the combinations of NB and CNN can provide an optimum alert or notification system for collision prediction in a CV environment.

2) Security: Data Analysis in Network Layer using Decision Tree and Fuzzy Logic Decision Tree and Fuzzy Logic would be considered as the two main techniques for security, based on [23], [24], [56] as it can classify and detect connected and autonomous vehicles in cyber-attack. DT is selected as a feature selection and attack classification purposes. DT can make an efficient classification for prediction, especially on detecting which nodes are malicious and non-malicious, to prevent cyber-attack, mainly in a CV environment. Fuzzy Logic algorithms help in detecting network attacks such as message injection through the classification and differentiation of injected malicious, fabricated and normal CAN packets. Based on these two techniques, the security of the communication in CV based on the proposed model can be improved. This is important in the model, where the RSU/OBU needs to ensure constant connection and communication with all the CVs in the environment. If the network is attacked, or infiltrated with unwanted and malicious nodes, the alert or notifications of collision would not be able to be transmitted, hence increasing the risk of collision and accidents from occurring.

B. Data Analysis using Machine and Deep Learning

In this study, in regards to safety and security for CV implementation, we are proposing the application of the model

in Collision Prediction Model and Congestion Prediction Model. The scenario in a CV environment using ML and DL methods to provide an overview of how to predict collision from occurring and how traffic congestion can be estimated. Also communicated between CVs so that a notification can be sent amongst CVs to alert drivers of the upcoming traffic in front of them. By applying the model in this application, both safety and security aspects will be addressed. The application for collision and congestion prediction also alert notification, mainly to alert the driver of the upcoming or potential collision and congestion so that they can be alert and provide necessary action for their vehicle. Fig. 6 illustrates the structure for the implementation of the proposed model in a CV environment. The application approach is divided into three main sections which are Infrastructure, variables and parameters, and the Collision and Congestion Prediction Model. The infrastructure focuses on network functions so that the network is safe from any internal or external threats. Infrastructure depends on the main parameters which are; (i) Speed, (ii) Distance, (iii) Time, (iv) Position. These are the basic elements in a study of CV environment [55]. Each parameter will communicate and updated in this model, to provide real-time alert notification if a collision is potentially occurring and if traffic road is congested. This can help drivers to prevent from an accident by taking necessary action and to reduce traffic congestion in which CV can re-plan route for involving traffic jam. Safety and Security aspects are illustrated in Fig. 5. Each aspect has its parameters to ensure each component functions correctly and efficiently.

C. Applicable Simulation Scenario for Collision Prediction and Congestion Prediction

Fig. 7 and 8, is a simulated scenario of CV infrastructure in a smart city. Essential components including RSU, OBU and CV. RSU and OBU use technological sensors that are relevant for CV implementation, such as GPS, LiDAR, ADAS, TLC, DSRC and RFID. These components will collect variables data such as Time, Speed, Distance and Position of the CV to calculate and provide a prediction of the upcoming potential collision.

In this scenario, several types of communication occur, that includes V2V, V2I and V2X. All the communication would be a pathway to transfer alerts and notifications regarding collision potentially occur. Through the sent alerts and notifications, drivers would get information regarding the collision earlier and enable them to slow down the CV and give a clear passage for authorities such as ambulances, police and fire squad to arrive at the location. V2V communications are focusing on communications between vehicles, while V2I is the communication between vehicles to infrastructures. V2X is where all components interact and communicate with each other. The notifications are an outcome of the collected and analyzed variable data that will be available and sent for all types of road users.

1) Collision prediction: In Fig. 7, the prediction of collision can be performed when CV A have lost control of driving. The analysis is based on the parameters which are collected by the sensors from RSU to CV(OBU) and OBU to RSU, then the data will be processed. All parameters must

functional to be calculated are; (i) Time: To get the current time, (ii) Position: to get the current position of CV from time to time (iii) Distance: distance between CV positions, (iv) speed: to get the speed of CV. If a CV is driving at fast speed with a short distance between other CVs (based on the position and time), an alert notification will be sent to alert the CVs of the upcoming possible collision.

If a collision occurs in RSU A area, hence information in RSU A will be analyzed to calculate the level of the collision to the level set in the system. If the level of collision is similar and accurate to the level set in the system, an alert notification will be sent to RSU B and the information will be communicated to all CV in the area. This process would continue for all the available RSUs. By the alert notification sent, the other incoming CVs can take action and slow down their vehicle when there are getting near and arriving at the location of the collision. If no alert notification is being sent, the possibility for a larger scale of collision can occur. Through the alert notification, CVs can provide passage for emergency authorities to attend to the collision location and the other CVs can decide to re-route and take another road to avoid the area.

2) Congestion prediction: In Fig. 8, the prediction of congestion can be performed when RSU detecting potential congestion by the amount of CV in a one RSU area. The analyzed is based on the parameters that be taken by the sensors from RSU to OBU(CV) and OBU to RSU then the data will be processed. The parameters to be calculated are: (i) Time: To get the current time, (ii) Position: to get current position of CVs from time to time (iii) Distance: distance between CV positions, (iv) speed: to get the speed of CV. CVs that are moving slowly in a short distance between one another within an RSU area indicates that a congestion might have occurred, as the time for CV to change position is longer. When a congestion has been detected, an alert notification will be sent to CVs for the upcoming congestion for them to take necessary action such as slowing down, or reroute.

In Fig. 8, there are several components which is RSU and CV (OBU) that are important for communication. RSU A will connect with CV for their data parameters and calculates the possibility of congestion. If traffic congestion occurs in the range of RSU A, the information would be sent to RSU B, for RSU B to send notifications and alert to all CVs within its area. With the communicated information, the rate of traffic congestion can be reduced if the CVs can prepare for the congestion or change its route to another less congested route.
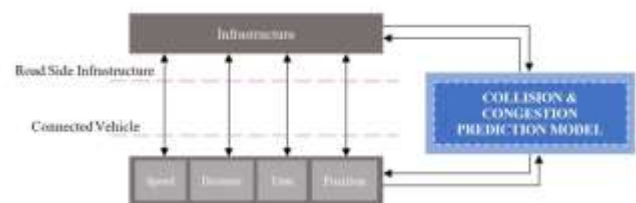

Fig. 6. Collision and Congestion Prediction Model for Safety and Security in Connected Vehicle.
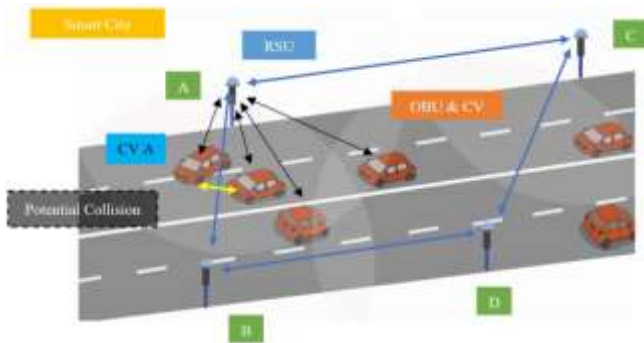
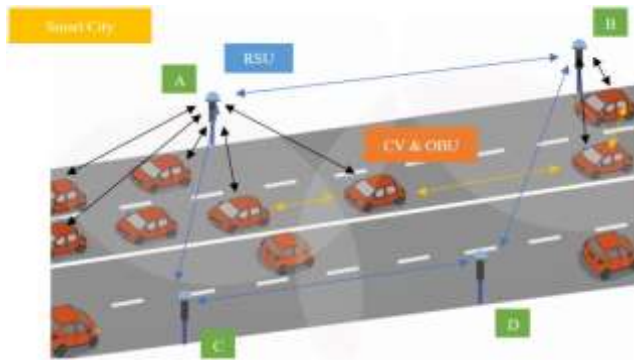Fig. 7.   Simulation Scenario for Collision Prediction.

Fig. 8.   Simulation Scenario for Congestion Prediction.

V.   CONCLUSION

In this paper, thorough research has been conducted to study the safety and security in connected vehicles through the implementation of big data analytics technology which is machine learning and deep learning. The study also includes a discussion of CV's related technologies and the techniques of machine learning and deep learning that have been applied by other researchers for various implementations. In this study, a conceptual model for both safety and security of CV has been proposed, which includes an application for collision and congestion prediction, by implementing several machine learning and deep learning techniques. The proposed model concerns all layers of CV implementation in IoT which is an application, physical and network layer. A simulation scenario has also been proposed and discussed theoretically, in which for future work, a real simulation, which collects real data as well as data analysis using necessary devices would be conducted.

ACKNOWLEDGMENT

REFERENCES

[1]   B. Lv et al., "LiDAR-Enhanced Connected Infrastructures Sensing and Broadcasting High-Resolution Traffic Information Serving Smart Cities," IEEE Access, vol. 7, pp. 79895–79907, 2019.

[2]   M. Soyturk, K. N. Muhammad, M. N. Avcil, B. Kantarci, and J. Matthews, From vehicular networks to vehicular clouds in smart cities. Elsevier Inc., 2016.

[3]   U. Z. A. Hamid, S. Z. Ishak, and F. Imaduddin, "Current Landscape of the Automotive Field in the ASEAN Region: Case Study of Singapore, Malaysia and Indonesia- A Brief Overview," Asean J. Automot. Technol., vol. 1, no. 1, pp. 21–28, 2019.

[4]   A. Sumalee and H. W. Ho, "Smarter and more connected: Future intelligent transportation system," IATSS Res., vol. 42, no. 2, pp. 67–71, 2018.

[5]   H. A. Ameen, A. K. Mahamad, S. Saon, D. M. Nor, and K. Ghazi, "A review on vehicle to vehicle communication system applications," Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 1, pp. 188–198, 2019.

[6]   Y. J. S. Yorgos, M. Golias, G. Dedes, C. Douligeris, and S. Mishra, "Challenges, Risks and Opportunities for Connected Vehicle Services in Smart Cities and Communities," IFAC-PapersOnLine, vol. 51, no. 34, pp. 139–144, 2019.

[7]   H. Kobayashi, C. Konno, M. Kayashima, and M. Nakano, "Approaches for Vehicle Information Security," Ipa, 2013.

[8]   E. Uhlemann, "Introducing connected vehicles [Connected vehicles]," IEEE Veh. Technol. Mag., vol. 10, no. 1, pp. 23–28, 2015.

[9]   A. F. Hasan, M. F. Che Husin, K. A. Rosli, M. N. Hashim, and A. F. Zainal Abidin, "Multiple Vehicle Detection and Segmentation in Malaysia Traffic Flow," IOP Conf. Ser. Mater. Sci. Eng., vol. 318, no. 1, 2018.

[10] D. Tokody, A. Albini, L. Ady, Z. Raynai, and F. Pongrácz, "Safety and Security through the Design of Autonomous Intelligent Vehicle Systems and Intelligent Infrastructure in the Smart City," Interdiscip. Descr. Complex Syst., vol. 16, no. 3, pp. 384–396, 2018.

[11] L. Ben Othmane, A. Al-Fuqaha, E. Ben Hamida, and M. Van Den Brand, "Towards extended safety in connected vehicles," IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, no. Itsc, pp. 652–657, 2013.

[12] S. Doecke, A. Grant, and R. W. G. Anderson, "The Real-World Safety Potential of Connected Vehicle Technology," Traffic Inj. Prev., vol. 16, pp. S31–S35, 2015.

[13] W. G. Najm et al., "Description of Light-Vehicle Pre-Crash Scenarios for Safety Applications Based on Vehicle-to-Vehicle Communications," no. May, p. 109p, 2013.

[14] X. Zeng, K. Balke, and P. Songchitruksa, "Potential Connected Vehicle Applications to Enhance Mobility, Safety, and Environmental Security," Southwest Reg. Univ. Transp. Cent., vol. 7, no. 2, p. No. SWUTC/12/161103-1, 2012.

[15] C. Sun et al., "An active safety control method of collision avoidance for intelligent connected vehicle based on driving risk perception," J. Intell. Manuf., 2020.

[16] G. Panahandeh, E. Ek, and N. Mohammadiha, "Road friction estimation for connected vehicles using supervised machine learning," IEEE Intell. Veh. Symp. Proc., no. Iv, pp. 1262–1267, 2017.

[17] H. Al Najada and I. Mahgoub, "Anticipation and alert system of congestion and accidents in VANET using Big Data analysis for Intelligent Transportation Systems," 2016 IEEE Symp. Ser. Comput. Intell. SSCI 2016, no. December, 2017.

[18] J. Hu, M. Huang, and X. Yu, "E ffi cient mapping of crash risk at intersections with connected vehicle data and deep learning models," Accid. Anal. Prev., vol. 144, no. June, p. 105665, 2020.

[19] N. Formosa, M. Quddus, S. Ison, M. Abdel-Aty, and J. Yuan, "Predicting real-time traffic conflicts using deep learning," Accid. Anal. Prev., vol. 136, no. December 2019, 2020.

[20] "(T) abubakar_2019_Enhanced weight-based clustering algorithm to provide reliable delivery for VANET safety applications.pdf."

[21] A. Jindal, G. S. Aujla, N. Kumar, R. Chaudhary, M. S. Obaidat, and I. You, "SeDaTiVe: SDN-Enabled Deep Learning Architecture for Network Traffic Control in Vehicular Cyber-Physical Systems," IEEE Netw., vol. 32, no. 6, pp. 66–73, 2018.

[22] C. G. Kiran, L. V. Prabhu, A. Rahiman V., and K. Rajeev, "Traffic sign detection and pattern recognition using support vector machine," Proc. 7th Int. Conf. Adv. Pattern Recognition, ICAPR 2009, pp. 87–90, 2009.

[23] Q. He, X. Meng, R. Qu, and R. Xi, "Machine Learning-Based Detection for Cyber Security Attacks on Connected and Autonomous Vehicles," pp. 1–19, 2020.

[24] M. Aloqaily, S. Otoum, I. Al Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," Ad Hoc Networks, vol. 90, no. May, 2019.

[25] F. Martinelli, F. Mercaldo, V. Nardone, and A. Santone, "Car hacking identification through fuzzy logic algorithms," IEEE Int. Conf. Fuzzy Syst., pp. 0–6, 2017.

[26] J. Barrachina, P. Garrido, M. Fogue, and F. J. Martinez, "Road Side Unit Deployment : A Density-Based Approach," vol. 5, pp. 30–39, 2013.

[27] T. J. Wu, W. Liao, and C. J. Chang, "A cost-effective strategy for road-side unit placement in vehicular networks," IEEE Trans. Commun., vol. 60, no. 8, pp. 2295–2303, 2012.

[28] Q. Yang, L. Wang, W. Xia, Y. Wu, and L. Shen, "Development of on-board unit in vehicular ad-hoc network for highways," 2014 Int. Conf. Connect. Veh. Expo, ICCVE 2014 - Proc., pp. 457–462, 2014.

[29] E. Paikari, S. Tahmasseby, and B. Far, "A simulation-based benefit analysis of deploying connected vehicles using dedicated short range communication," IEEE Intell. Veh. Symp. Proc., no. Iv, pp. 980–985, 2014.

[30] S. Dabral, S. Kamath, V. Appia, M. Mody, B. Zhang, and U. Batur, "Trends in camera based Automotive Driver Assistance Systems (ADAS)," Midwest Symp. Circuits Syst., pp. 1110–1115, 2014.

[31] Z. Zhong, S. Liu, M. Mathew, and A. Dubey, "Camera radar fusion for increased reliability in ADAS applications," IS T Int. Symp. Electron. Imaging Sci. Technol., pp. 1–4, 2018.

[32] J. Chen, S. Tian, H. Xu, R. Yue, Y. Sun, and Y. Cui, "Architecture of Vehicle Trajectories Extraction With Roadside LiDAR Serving Connected Vehicles," IEEE Access, vol. 7, pp. 100406–100415, 2019.

[33] G. H. Kulkarni and P. G. Waingankar, "Fuzzy logic based traffic light controller," ICIIS 2007 - 2nd Int. Conf. Ind. Inf. Syst. 2007, Conf. Proc., no. August, pp. 107–110, 2007.

[34] W. M. El-Medany and M. R. Hussain, "FPGA-based advanced real traffic light controller system design," 2007 4th IEEE Work. Intell. Data Acquis. Adv. Comput. Syst. Technol. Appl. IDAACS, no. September, pp. 100–105, 2007.

[35] S. Mehan, "Introduction of Traffic Light Controller with Fuzzy Control System," Int. J. Electron. Commun. Technol., vol. 2, no. 3, pp. 119–122, 2011.

[36] J. Wang, D. Ni, and K. Li, "RFID-based vehicle positioning and its applications in connected vehicles," Sensors (Switzerland), vol. 14, no. 3, pp. 4225–4238, 2014.

[37] N. Lu, S. Member, N. Cheng, S. Member, N. Zhang, and S. Member, "Connected Vehicles : Solutions and Challenges," vol. 1, no. 4, pp. 289–299, 2014.

[38] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," Transp. Res. Part C Emerg. Technol., vol. 71, pp. 143–163, 2016.

[39] B. Y. R. Khatoun and S. Zeadally, "Smart Cities : Consepts."

[40] X. Yang, J. Liu, F. Zhao, and N. H. Vaidya, "A vehicle-to-vehicle communication protocol for cooperative collision warning," Proc. MOBIQUITOUS 2004 - 1st Annu. Int. Conf. Mob. Ubiquitous Syst. Netw. Serv., pp. 114–123, 2004.

[41] "Challenges Facing Using AI in Cybersecurity | OpenMind." [Online]. Available: https://www.bbvaopenmind.com/en/technology/artificial-intelligence/challenges-facing-using-ai-in-cybersecurity/. [Accessed: 05-Jul-2020].

[42] M. H. Eiza and Q. Ni, "Driving with Sharks: Rethinking Connected Vehicles with Vehicle Cybersecurity," IEEE Veh. Technol. Mag., vol. 12, no. 2, pp. 45–51, 2017.

[43] M. A. Al-garadi, A. Mohamed, A. Al-ali, X. Du, and M. Guizani, "Surveys," Underst. Commun. Res. Methods A Theor. Pract. Approach, pp. 222–237, 2014.

[44] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," J. Urban Technol., vol. 22, no. 1, pp. 3–21, 2015.

[45] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," J. Big Data, vol. 2, no. 1, p. 21, Dec. 2015.

[46] P. Russom, "BIG DATA ANALYTICS - TDWI BEST PRACTICES REPORT Introduction to Big Data Analytics," TDWI best Pract. report, fourth Quart., vol. 19, no. 4, pp. 1–34, 2011.

[47] J. Narula, "Are we up to speed?: from big data to rich insights in CV imaging for a hyperconnected world.," JACC. Cardiovasc. Imaging, vol. 6, no. 11, pp. 1222–4, Nov. 2013.

[48] S. Amini, I. Gerostathopoulos, and C. Prehofer, "Big data analytics architecture for real-time traffic control," in 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017, no. Tum Llcm, pp. 710–715.

[49] D. Alvarez-Coello, B. Klotz, D. Wilms, S. Fejji, J. M. Gomez, and R. Troncy, "Modeling dangerous driving events based on in-vehicle data using random forest and recurrent neural network," IEEE Intell. Veh. Symp. Proc., vol. 2019-June, no. Iv, pp. 165–170, 2019.

[50] Y. Cui, H. Xu, J. Wu, Y. Sun, and J. Zhao, "Automatic Vehicle Tracking with Roadside LiDAR Data for the Connected-Vehicles System," IEEE Intell. Syst., pp. 44–51, 2019.

[51] Y. Xun, Y. Sun, and J. Liu, "An Experimental Study Towards Driver Identification for Intelligent and Connected Vehicles," IEEE Int. Conf. Commun., vol. 2019-May, 2019.

[52] L. Abberley, K. Crockett, and J. Cheng, "Modelling Road Congestion Using a Fuzzy System and Real-World Data for Connected and Autonomous Vehicles," IFIP Wirel. Days, vol. 2019-April, pp. 1–8, 2019.

[53] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," IEEE Commun. Surv. Tutorials, vol. 21, no. 4, pp. 3039–3071, 2019.

[54] H. T. Abdelwahab and M. A. Abdel-Aty, "Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections," Transp. Res. Rec., no. 1746, pp. 6–13, 2001.

[55] S. Aoki, T. Higuchi, and O. Altintas, "Cooperative Perception with Deep Reinforcement Learning for Connected Vehicles," Apr. 2020.

[56] J. Rios-Torres, A. Malikopoulos, and P. Pisu, "Online Optimal Control of Connected Vehicles for Efficient Traffic Flow at Merging Roads," IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, vol. 2015-Octob, pp. 2432–2437, 2015.