Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles

Article *in* IEEE Transactions on Intelligent Transportation Systems · April 2019

DOI: 10.1109/TITS.2019.2906038

4 authors:

Franco van Wyk
University of Tennessee
14 PUBLICATIONS **82** CITATIONS
SEE PROFILE

Yiyang Wang
University of Michigan
8 PUBLICATIONS **36** CITATIONS
SEE PROFILE

Anahita Khojandi
University of Tennessee
52 PUBLICATIONS **181** CITATIONS
SEE PROFILE

Neda Masoud
University of Michigan
45 PUBLICATIONS **364** CITATIONS
SEE PROFILE

Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles

Franco van Wyk, *Member, IEEE,* Yiyang Wang, Anahita Khojandi, *Member, IEEE,* and Neda Masoud

***Abstract*—Connected and automated vehicles (CAVs) are expected to revolutionize the transportation industry, mainly through allowing for a real-time and seamless exchange of information between vehicles and roadside infrastructure. Although connectivity and automation are projected to bring about a vast number of benefits, they can give rise to new challenges in terms of safety, security, and privacy. To navigate roadways, CAVs need to heavily rely on their sensor readings and the information received from other vehicles and roadside units. Hence, anomalous sensor readings caused by either malicious cyber attacks or faulty vehicle sensors can result in disruptive consequences, and possibly lead to fatal crashes. As a result, before the mass implementation of CAVs, it is important to develop methodologies that can detect anomalies and identify their sources seamlessly and in real-time. In this paper, we develop an anomaly detection approach through combining a deep learning method, namely convolutional neural network (CNN), with a well-established anomaly detection method, Kalman filtering with a χ^2-detector, to detect and identify anomalous behavior in CAVs. Our numerical experiments demonstrate that the developed approach can detect anomalies and identify their sources with high accuracy, sensitivity, and F1 score. In addition, this developed approach outperforms the anomaly detection and identification capabilities of both CNNs and Kalman filtering with a χ^2-detector methods alone. It is envisioned that this research will contribute to the development of safer and more resilient CAV systems that implement a holistic view towards intelligent transportation system (ITS) concepts.**

***Index Terms*—Cyber-physical systems, Fault diagnosis, Intelligent vehicles, Intrusion detection, Vehicle safety**

I. Introduction

OUR current transportation system is on the brink of transforming into a highly connected, automated, and intelligent system as a result of the rapid emergence of connected and automated vehicles (CAVs) [1]. CAVs, with various degrees of connectedness and automation, are expected to play an integral role in the next phase of the transportation revolution, leading to more accessible, more efficient, safer, more environmentally friendly, and hence sustainable, transportation options [2], [3]. CAVs use wireless technology to facilitate communication between vehicles, with roadside units (RSUs), and with personal mobile devices. This will allow them to continuously transmit and share information such as speed, position, acceleration, and braking, enabling CAVs to warn their surrounding vehicles of potentially unsafe circumstances. These vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication technologies will provide unprecedented efficiency, safety, and mobility advancements. For instance, CAV technologies are expected to decrease fatal traffic accidents by as much as 80%, reducing their corresponding $870 billion cost, while also improving traffic flow, cutting into the approximate 7 billion hours American motorists spend in traffic annually [4].

Although the ever-increasing use of CAV technologies in vehicles are expected to have numerous advantages, the potential drawbacks are not negligible. CAVs use a variety of sensors to build a virtual map of their surroundings in order to drive in the correct lane within the speed limit, avoid collisions, and detect obstacles in their immediate physical environment. Hence, anomalous sensor values caused by either malicious cyber attacks or faulty vehicle sensors can result in disruptive consequences, and possibly lead to fatal crashes. The increase in connectivity and automation has led to the scrutiny of in-vehicle network architectures used by automotive manufacturers and evaluation of vulnerabilities in their resiliency against such anomalous behavior [5], [6], [7]. For instance, the dedicated short-range communication (DSRC) technology is currently used to facilitate communication within connected networks. DSRC has a range of approximately 300 meters, hence it can protect vehicles from long-range cyber attacks, e.g., a stationary attacker would only have a short time-window to attack moving vehicles when they are in close proximity [8]. Although such a technology can prove useful, it is not comprehensive enough and leaves CAVs vulnerable to non-stationary attackers, among others. Hence, it is necessary to better understand CAVs' vulnerabilities and develop holistic, real-time methodologies that can mitigate them.

Various internal and external cyber attack surfaces exist in CAV systems, i.e., the entry point of the attack, which may enable hackers to access and compromise the safety and integrity of CAVs [6], [7], [9], [10], [16], [17]. Typical *internal* attack surfaces include in-vehicle devices, GPS system, on-board diagnostics (OBD) system, vehicle sensors including the controller area network (CAN) bus, and other sensors required for CAV operation. For instance, [9] demonstrated through an OBD port attack, it is possible to disable the brakes, turn-off head-lights, and take over steering for cars equipped with a low level of autonomy. Typical *external* attack surfaces include information from RSUs, machine vision, data from other vehicles, security system breaches of the vehicle, and navigation interference [10], [15], [19]. For instance, [10] demonstrated how an attacker could gain access to the visual recognition software used in autonomous vehicles and manipulate it by creating a simple alteration to RSUs that would cause the car to misinterpret them, possibly putting vehicle

Franco van Wyk and Anahita Khojandi are with the University of Tennessee, Knoxville, TN 37996 USA. (e-mail: fvanwyk@vols.utk.edu, khojandi@utk.edu)

Yiyang Wang and Neda Masoud are with the University of Michigan, Ann Arbor, MI 48109 USA (e-mail: yiyangw@umich.edu, nmasoud@umich.edu)

occupants at risk. Similarly, several teams have hacked traffic light controller systems, highway signs, and traffic surveillance cameras [52]. For instance, in 2017, approximately 70% of the storage devices that record data from Washington D.C. police surveillance cameras were infected with ransomware by hackers [53]. Injection of fake information and map database poisoning is considered to be one of the most dangerous cyber attacks on CAVs [16]. Future CAVs are expected to have even more attack surfaces than what has currently been investigated. Possible reasons for cyber attacks on CAVs include financial gain, collecting private information, and gaining priority access to infrastructure.

Not all anomalous sensor behavior are due to malicious attacks. Sensor readings may be influenced for a variety of reasons prompting the transmission of faulty information [9], [22], [23]. For instance, sensors in CAVs may be blinded by magnetic interference, signal outage, poor weather conditions, and other environmental circumstances. Furthermore, as sensors age, inherent errors are introduced which may result in sensor failure, therefore, affecting data availability. Faulty sensors, therefore, pose significant risks to the operation of CAVs since their safe operation depends on information obtained from sensors.

Sensor redundancy is a measure that can be implemented to protect CAVs against anomalous sensor behavior. The majority of CAV manufacturers are expected to incorporate multiple sensors that measure the same parameter [11]. For instance, CAVs may utilize various sensor systems such as cameras, radio detection and ranging (RADAR), light detection and ranging (LIDAR), and ultrasonic sensors for lane keeping purposes as well as GPS to assist navigation. To illustrate the importance of sensor redundancy, consider a short-term loss of a GPS signal in a CAV as it passes through a tunnel. In such a scenario, information from redundant sensors such as the inertial measurement unit (IMU), RADAR, and LIDAR can be used to approximate vehicle location until all systems are online again. In this example, sensor redundancy proves useful since sensors collecting the same data are not affected similarly by environmental factors; e.g., the factors that lead to a lack of access to GPS signal do not affect RADAR. Additionally, different types of sensors collecting the same data may have various degrees of precision as well as various levels of vulnerability to different cyber attack types. For instance, if a vehicle's machine vision, used for obstacle detection through video image processing, is attacked by using a high-brightness infrared LED, the redundant obstacle detection sensors including RADAR and LIDAR would be unaffected. Hence, sensor redundancy can lead to improved, dynamic sensor fusion in which anomalous sensor readings, due to either faults or attacks, can be discarded while the normal data is being fused to increase the reliability of the fused data.

A large body of work exists that examines various methods to detect anomalies, if they occur, and/or identify their source; however, only a limited number of these studies are focused on CAVs and ITS. Table VIII in the appendix summarizes some of the important works related to anomaly detection and identification in CAVs. Several studies employ distance-related metrics such as the Mahalanobis distance, affinity propagation clustering, and graph theory to detect in-vehicular network intrusions and faulty sensors [20], [21], [24], [25]. Also, Kalman filtering [14] has been used in numerous applications for fault detection [29], [30]. In addition, some recent studies have employed deep learning techniques such as convolutional neural network (CNN) [33], recurrent neural network (RNN) [48], and multilayer perceptron (MLP) models to detect anomalies in autonomous agriculture equipment vision and malicious CAN packets in modern vehicles [26], [27]. Deep learning techniques can be implemented on raw data and therefore do not require data abstraction. Lastly, [28] and [31] employed Bayesian networks and signal entropy to detect anomalies in in-vehicular networks and autonomous robotic vehicles. However, these methods require synchronized data sources and high-volume attacks to perform well.

Several gaps are apparent in the literature. First, to the best of our knowledge, there is a lack of deep learning implementations in anomaly detection and identification for CAVs. Data are becoming more readily available and deep learning models are renowned for their performance using large datasets, therefore such models may be able to outperform traditional anomaly detection techniques such as Kalman filtering with failure detectors. In addition, there is a lack of comprehensive frameworks combining different anomaly detection and identification methods that incorporate the strengths and negate the weaknesses of the individual anomaly detection methods in CAVs. Lastly, applications focusing on real-time detection and identification of anomalous information (such as cyber attacks and/or faulty sensors) are limited.

Before mass implementation of CAVs into the transportation system, we need to ensure that the design of CAVs is resilient to cyber attacks and faulty equipment. This study assumes that the participating vehicles are of levels 4 and 5 automation, as defined by the National Highway Traffic Safety Administration (NHTSA) [32]. Our main objective in this study is to detect anomalous sensor behavior and identify the source anomalous sensor in real-time for CAVs to assure the high reliability of fused data. Our framework is generic in that a 'sensor' may refer to any of the on-board sensors in a CAV, or another connected vehicle or roadside unit (RSU) that is communicating with the CAV. Specifically, we develop a holistic and generic framework by combining a deep learning technique, i.e., CNN, and Kalman filtering with a χ^2-detector, and investigate their ability to detect and identify various types of anomalous behavior in real-time. In addition, we perform various experiments to investigate the effects of anomaly type, magnitude, and duration.

Anomalous sensor readings, caused by attacks or failures, can present themselves in different ways. Several network attack taxonomies are available in the literature. Bhuyan et al. [38] summarize the taxonomy of intrusions or attacks in computer network systems, which encompass CAVs. Also, several faulty sensor behaviors are discussed in [43]. In this paper, we consider the anomalous sensor behavior resulting from both false injection attacks and sensor failures. According to the literature, anomalous sensor behavior can be represented by the five main following types:

1) Instant: A sharp, unexplained change in the observed data between two successive sensor readings.
2) Constant: A temporarily constant observation that is different from the "normal" sensor readings and is uncorrelated to the underlying physical phenomena.
3) Gradual drift: A small and gradual drift in observed data during a time period. It can result in a large discrepancy between the observed data and the true state of the system in time.
4) Bias: A temporarily constant offset from the sensor readings.
5) Miss: Lack of available data during a time period.

In this paper, consistent with literature [38], [43], we focus on detection and identification of anomalous behavior, caused by either cyber attacks or faulty sensors, resulting in 'instant,' 'constant,' 'gradual drift,' and 'bias.' These types of anomalies are some of the most dangerous for CAVs [16], [41]. In this paper, we do not explicitly account for 'miss,' which can result from DoS attacks preventing the exchange of information. However, note that 'miss,' depending on its duration, can be viewed as 'instant' or 'constant' behaviors, where the sensor reading is non-existent instead of showing a wrong value. Hence, it can partially be addressed using the same methods for detecting 'instant' or 'constant' behaviors. Regardless, we acknowledge that the considered anomaly types may not encapsulate *all* possible types of anomalies expected to occur in CAVs.

In this paper, we develop an anomaly detection approach through combining a deep learning method, namely convolutional neural network (CNN), with a well-established anomaly detection method, Kalman filtering with a χ^2-detector, to detect and identify anomalous behavior in CAVs. Our main contributions are as follows: (1) We develop an anomaly detection and identification approach based on convolutional neural networks (CNN), applied to time-series data obtained from multiple sensors. Our use of the CNN for anomaly detection in time-series data is novel, where we generate 'images' from a continuous feed of real-time raw sensor data from a fixed-width sliding window and classify these images as anomalous or normal; (2) We develop a new generic anomaly detection and identification approach through combining CNN with a well-established anomaly detection method, i.e., Kalman filtering with a χ^2-detector. The resulting CNN-empowered KF (CNN-KF) framework can effectively detect and identify sensor anomalies.

II. Methods

In this section, we first discuss the two models that form the building blocks of our framework, namely the CNN model and the Kalman filter with a χ^2-detector model (referred to as the KF model throughout), developed independently to detect anomalies caused by cyber attacks and/or faulty equipment, in a CAV trip. Next, we develop a framework that combines the two methods in order to improve detection and identification capabilities by relying on their respective individual strengths. CNN was mainly selected due to its ability to capture temporal patterns, relationship between various sensors, and its capability to automatically extract features while weight sharing. All these make CNN particularly suitable as large amounts of data are becoming available. However, there is always a risk of unknown/unseen patterns going undetected when relying only on CNNs. Hence, using an architecture that combines CNN and KF can provide an additional level of reliability for the task at hand, as CNN and KF are complementary in their ability to detect anomalies. It is worth noting that we also explored recurrent neural networks (RNNs) with long short-term memory (LSTM) units; however, CNN consistently outperformed RNN in all preliminary experiments. This is partly because in this particular application, there are many normal values between consecutive anomalous values, which generally makes it hard for RNN to distinguish between anomalous and normal values in an extended sequence of data [54].

In general, the inputs to the model are the data collected from sensors, reading the same or highly correlated physical quantities. Based on the input data, at every time step, e.g., a few milliseconds (ms), outputs are generated as to whether anomalies are present (detection) and if so, which sensor reading(s) are erroneous (identification). Consequently, erroneous data can be excluded and normal data can be seamlessly fused to support CAV operation. Please note that all notation used are summarized in Table IX in the appendix.

A. Kalman Filter

As discussed, Kalman filter combined with a failure detector is a well-established, widely used method for fault detection and identification in time-series data. In order to detect and identify anomalous sensor readings, we use an adaptive Kalman filter with a χ^2-detector to filter out process and measurement noise. Specifically, we assume our physical system is a discrete-time linear time-invariant system in the following form:

$$x(k) = Ax(k-1) + w(k-1) \quad (1)$$

where $x(k) \in \mathbb{R}^m$ is the vector of state variables at time k, $w(k) \in \mathbb{R}^m$ is the process noise at time k, and $A \in \mathbb{R}^{m\times m}$ is the state-transition matrix.

As discussed, we consider redundant sensor in this study. Let n denote the number of these sensors. That is, we consider n local subsystems, corresponding to the redundant sensors, with measurement matrices $H(k) \in \mathbb{R}^{p\times m}$ and sensing model:

$$z_i(k) = H(k)x(k) + v_i(k), \ i = 1, 2, ..., n \quad (2)$$

where $v_i(k) \in \mathbb{R}^p$ is zero mean Gaussian white noise sequences associated with the process and the measurement. The covariance matrices of $v_i(k) \in \mathbb{R}^p$ and $w(k)$ are $R_i(k)$ and $Q(k)$, respectively. We assume $v_i(k)$ and $w(k)$ are independent. In equation (2), $z_i(k) = [z_{i,1}(k), z_{i,2}(k), ..., z_{i,p}(k)]^T \in \mathbb{R}^p$ is a vector of sensor measurements for subsystem $i \in \{1, ..., n\}$. We assume all n subsystems have the same measurement matrix $H(k)$ and the readings across all subsystems are synchronized.

For each subsystem, a Kalman filter is used to estimate the state vector $x(k)$ from sensor reading $z_i(k)$. Kalman filter consists of two phases, i.e. prediction and update. The prediction phase advances the state estimate before the next measurement, and the update phase corrects the state estimate based on the measurement.

Let $\hat{z}(k|k-1)$ denote the predicted value of measurement at time k, $P(k|k-1)$ denotes the error covariance matrix of predicted state, and $\nu(k)$ denote the innovation, i.e., the difference between the measurement $z(k)$ and the predicted value of measurement at time k,

$$\nu(k) = z(k) - \hat{z}(k|k-1). \tag{3}$$

Also let $S(k)$ denote covariance matrix of innovation. In practice, covariance matrices R and Q are generally unknown a priori. Thus, we apply an adaptive Kalman filter to approximate these matrices [42]. Specifically, we use a moving estimation window of size M to adaptively estimate R and Q matrices according to the innovation sequence within the time window, i.e.,

$$\begin{aligned} \hat{R}(k) &= \hat{C}_\nu(k) - H(k)P(k|k-1)H(k)^T, \\ \hat{Q}(k) &= K(k)\hat{C}_\nu(k)K(k)^T, \end{aligned} \tag{4}$$

where

$$\hat{C}_\nu(k) = \frac{1}{M}\sum_{j=k-M}^{k-1} \nu(j)\nu(j)^T. \tag{5}$$

We use a χ^2-detector to construct χ^2 test statistics, to determine whether the new measurement falls into the gate region with the probability determined by the gate threshold γ, defined as

$$V_\gamma(k) = \{z : (z-\hat{z}(k|k-1))^T S(k)^{-1}(z-\hat{z}(k|k-1)) \leq \gamma\}. \tag{6}$$

The χ^2 test statistics for each local subsystem is defined as

$$t(k) = \nu(k)^T S(k)^{-1}\nu(k) \tag{7}$$

It is easy to show that under Gaussian assumption the test statistic has a χ^2 distribution with p degrees of freedom, where p is the number of components of the measurement vector. However, in practice, usually w and v_i do not follow Gaussian distributions, and parameter γ has to be selected empirically. Hence, the threshold γ for $t(k)$ and the time window size M can be tuned.

Selection of the threshold γ is a trade-off between sensitivity of the trained model (i.e., the proportion of correctly identified anomalies), and the false-alarm rate. For each experiment in our experiment section, we select γ using a grid search within the range $\{1, 2, ..., 50\}$, and select the value of γ resulting in the highest F1 score.

The window size M is a parameter that allows for the control of smoothing short-term fluctuations in the detector. To select the best value for M, we performed a grid search. Specifically, for various values of M in our grid search, we computed the Area Under the ROC Curve (AUC) on a validation dataset. The results suggested that $M \in \{10, 15, 20\}$ provides robust results and high AUC (≈ 0.96). The AUC values within this range are similar. Hence, consistent with previously published work (e.g., [44]), we selected the window size $M = 15$ epochs for our experiments.

Once a faulty measurement stream has been identified, in order to ensure the future estimate is reliable, the measurement stream should be rejected at once and not fused with other measurements to ensure that the information generated through data fusion is not contaminated.

B. CNN

To apply CNN for real-time detection and identification of anomalous sensor behavior, we use a fixed-width sliding window on input data from all sensors measuring the same quantity, either directly or indirectly, where conversions or combining with other sensors may be required to infer the value of the quantity. At every epoch, as new observations are collected from sensors, the sliding window shifts to include these latest observations. Hence, the input to CNN is a series of 'images' from the continuous feed of raw sensor data during a CAV trip. For instance, consider three sensors (e.g., GPS, accelerometer, and transmission vehicle speed sensor) measuring vehicle speed at the sampling interval of 0.1 seconds. Hence, an image of size 3×10 would include the data collected from the three sensors during the last one second of the trip. The CNN therefore utilizes the data from multiple sensors simultaneously for detection and identification of anomalous values.

CNN models are trained to evaluate these images to detect and identify anomalies in real-time. Specifically, a sliding window is used on retrospectively collected data from sensors to produce images for training and testing. Because the goal is both to detect and to identify anomalies, for each sensor we train a separate model using labeled images, i.e., supervised learning. That is, if an anomaly is present in an image constructed with the data from the sensor of interest, the response variable is set to 1; otherwise, it is set equal to 0. Once separate models are trained to identify anomalies for each sensor, a logical OR operator on the outcomes of the all such models determines whether anomalous readings are detected across all sensors.

In our experiments, for each CNN model, a popular image recognition architecture from the literature is adopted [36]. The parameter values for this architecture are then selected based on a number of experiments performed to maximize anomaly detection and identification performance on a validation set. In short, we used three convolution layers with max-pooling, followed by two fully connected layers with random dropout between the layers. A 1×2 pool size is used and 40, 60, and 60 filters are used for convolutional layers one to three, respectively. Also, a random dropout rate of 0.1 and batch size of 128 is used to train the CNN models. Furthermore, rectified linear unit (ReLU) activation functions are used and the Adam optimizer for Tensorflow in Python is implemented to minimize binary cross-entropy [35]. The following parameters were used for the Adam optimization algorithm: learning rate, $\alpha = 0.001$, exponential decay rates, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a fuzz factor, $\epsilon = 10^{-8}$.

It is widely acknowledged that deep learning models such as CNNs are often subject to 'overfitting' during the training process [36], [37]. To reduce the risk of overfitting, in addition to random dropouts, we use early stopping to monitor the accuracy of the validation set with a patience of 200 epochs. Therefore, when training a CNN model, starting from any training epoch, if the validation accuracy during the following 200 epochs does not increase, training is terminated and the model corresponding to 200 training epochs ago, which resulted in the highest validation accuracy, is selected.

C. CNN-Empowered Kalman Filter (CNN-KF)

In order to further improve upon the detection and identification performances of the individual KF and CNN models, we develop a new framework that relies on both CNN and Kalman filter as shown in Figure 1. In this framework, first CNN models process the images of raw data from all sensors, which are obtained by using a sliding window over all sensor readings, to identify whether the readings from each sensor are normal or anomalous. Consequently, the sensors with anomalous behavior are excluded and the readings from normal sensors are separately fed into adaptive Kalman filters with failure detectors for further examination and anomaly detection. If Kalman filter detects anomalies that are missed by CNN, the readings from the corresponding sensors are excluded and the remaining normal data are fused in order to achieve a higher degree of reliability. As time passes and the vehicle continues its trip, if the sensors that presented anomalies go back to normal behavior, as verified by both CNN models and Kalman filters, the exclusion is no longer necessary and hence, the readings of the previously excluded sensors would be used in fusion again. In this study, we use a CNN-empowered Kalman Filter (CNN-KF), as opposed to a Kalman Filter empowered CNN (KF-CNN), mainly because having the Kalman filter in the last layer of learning allows for the reliable fusion of the non-anomalous sensor values [33], which is important from a practical perspective for the application considered. In addition, based on our preliminary experiments, CNN-KF generally outperforms KF-CNN. Hence, we opted to present the results for CNN-KF only.

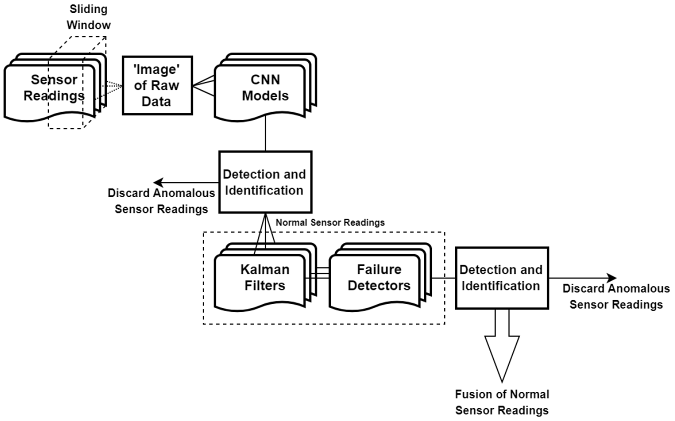
Sliding Window
Sensor Readings
'Image' of Raw Data
CNN Models
Detection and Identification
Discard Anomalous Sensor Readings
Normal Sensor Readings
Kalman Filters
Failure Detectors
Detection and Identification
Discard Anomalous Sensor Readings
Fusion of Normal Sensor Readings

Fig. 1. Overview of the CNN-KF Framework. At each time epoch, the sensor readings collected within the past several time epochs are used as an input 'image' to the CNN-KF Framework. The data are first processed with a CNN to detect and identify sensors with anomalous readings. Next, the CNN-verified non-anomalous data are fed to the KF model to further detect and identify anomalies to increase reliability.

III. Data

The data for this study is obtained from the research data exchange (RDE) database for the Safety Pilot Model Deployment (SPMD) program [34]. This program was conducted with the primary objective of demonstrating CAVs, with the emphasis on connectivity technologies such as V2V and V2I communications in real-world conditions. The program recorded detailed and high-frequency (collected every 100ms) data for more than 2,500 vehicles over a period of two years. The data features extracted from the SPMD dataset used in this study include the in-vehicle speed (denoted as sensor 1), GPS speed (sensor 2), and in-vehicle acceleration (sensor 3) for one of the test vehicles with a trip length of 2,980 seconds. Note that the in-vehicle speed and GPS speed sensors observe the same quantity, namely, speed, whereas the acceleration sensor observes the vehicle's acceleration which can be used to infer speed.

Since there are no publicly available datasets for CAVs that include anomalies in sensor measurements, due to either attacks or faults, and the ground truths, we used simulation to generate datasets for our experiments. Specifically, we accounted for the four major anomaly types including instant, constant, gradual drift, and bias. We inject all four types of anomaly into each of the three speed-related (redundant) sensors. We assume the *onset* of anomalous values in sensors, due to either attacks or faults, occur independently. That is, we do not explicitly train models on datasets containing interdependent sensor failures or systemic cyber attacks on vehicle sensors. Additionally, we assume that no more than one anomaly can *start* in every time epoch, which is indeed very unlikely considering that sensors are generally reliable, and attacks/faults to sensors occur independently. However, dependent on the types, onset times, and durations of anomalies, multiple sensors may be anomalous at the same time.

There is existing work that illustrates the sensors considered in our numerical study, i.e., speed and acceleration sensors, are vulnerable to cyber attacks or faults (e.g., see [16], [50], [51]). For in-vehicle speed and acceleration sensors, an injection attack through the CAN bus or the on-board diagnostics (OBD) system, could give rise to the four types of anomalies considered in this paper. Also, for the in-vehicle acceleration sensor, an acoustic injection attack could result in anomalous sensor values. Lastly, for the speed measurement from the GPS, both the operating environment of the vehicle and GPS spoofing/jamming attacks may result in anomalous sensor values.

We generate various datasets for our experiments at 1% or 5% rates of anomalies, denoted by α. We simulate the anomalies to occur at randomly selected onset times (discretized into 100ms) to randomly selected sensors. To simulate the corresponding attacks/faults, these anomalies are then added to each affected sensor's 'base value,' i.e., the normal sensor readings in the original dataset indicating the traveling speed of the CAV at the time that the anomaly was introduced. Algorithm 1 presents the pseudo code used for simulating the anomalies. Note that vectors V_i and V_i', $i \in \{1, 2, \ldots, n\}$, denote non-anomalous and anomalous readings for sensor i, respectively. To facilitate thorough experiments, we vary the simulated anomalies in type, magnitude and/or duration when generating the datasets. In addition, dependent on the experiment, we generate datasets where we randomly sample from a set of one or all anomaly types. The exact types of anomalies considered as well as their magnitude (and duration) will be discussed in detail for each experiment in Section IV.

Algorithm 1 Anomaly generation process

```
$\alpha \leftarrow$ anomaly rate; $n \leftarrow$ number of sensors
for time epoch $t \in T$ do
    if $\mathcal{U}(0,1) \leq \alpha$ then
        $\zeta \leftarrow \mathcal{U}(0,1)$
        for $i \in \{1, 2, \ldots, n\}$ do
            if $\zeta \geq \frac{i-1}{n}$ and $\zeta < \frac{i}{n}$ then
                Generate anomaly; $V_i' \leftarrow V_i +$ anomaly
            end if
        end for
    else
        $V_i' \leftarrow V_i, \ \ i \in \{1, 2, \ldots, n\}$
    end if
end for
```

IV. Results

In this section, we perform various analyses to investigate the anomaly detection and identification performance of the three models discussed in Section II. Specifically, we present the results obtained when using Kalman filter (KF) and CNN alone, and compare their performance to highlight their respective capabilities. In addition, we present the results obtained using CNN-KF framework and compare and contrast its performance with those of KF and CNN alone. We first investigate the detection performance of all three models when trained and tested for a single anomaly type in Section IV-A-A. We then investigate the detection and identification performance of the three models when trained and tested in the presence of all anomaly types in Section IV-B.

In our analyses, we use various datasets in which we simulate various types of anomalies of different durations and magnitudes to draw insights from the use of CNN, KF, and the CNN-KF models to detect/identify anomalies in real-time. Note that we need to select the parameter γ for KF models and also extensively train CNN models and tune their many parameters. Hence, for any given dataset, we use a training/validation/testing split of 60%/20%/20%. We use the training and validation sets to tune the model parameters. Next, to objectively evaluate their performance levels, we use the separate test sets for testing.

We evaluate the performance of the models in terms of accuracy, sensitivity, precision, and F1 score. Accuracy measures the overall proportion of correct predictions for normal and anomalous sensor values. Sensitivity assesses the proportion of correctly identified anomalous sensor values from the total number of anomalous sensor values. Precision measures the proportion of anomalous sensor values among those predicted as anomalous. Lastly, F1 score is the harmonic mean of sensitivity and precision. These metrics are particularly chosen as they measure the ability of the models to correctly differentiate between normal and anomalous sensor behavior. Note that these metrics are commonly used to evaluate the performance of classification models. Here, for consistency and to enable comparison of all models, we use the same metrics to evaluate the performance of CNN, KF, and CNN-KF models.

A. *Models Under a Single Anomaly Type*

In this section we compare the *detection* performance of the three models for the specific types of anomalies, as discussed in Section I, namely, instant, constant, gradual drift, and bias. We generate various datasets, each with a specific type of anomaly, with anomaly rate $\alpha = 5\%$. For each dataset, we train and test CNN models to measure their performance in detecting the specific anomaly type. Because each sensor reading in our experiment is one dimensional, the state transition matrix A and measurement matrix H for KF are simply single values.

1) Instant: The instant anomaly type is simulated as a random Gaussian variable with mean and variance of zero and 0.01, respectively, that is scaled by a scalar $c \in \{25, 100, 500, 1{,}000, 10{,}000\}$, i.e., $c \times \mathcal{N}(0, 0.01)$, to capture various magnitudes. The resulting simulated value is added to the base value of sensor measurement for one epoch, i.e., 100 ms.

Table I illustrates the anomaly detection performance of the KF, CNN, and CNN-KF models. Considering the detection performance of KF and CNN models, the following is noted. In general the detection performance of the KF and CNN models increases across all metrics in the magnitude of the instant anomaly type, which is consistent with the intuition. For small magnitudes of anomalous sensor values, i.e., the first two rows in Table I, the detection performance of the models are poor. However, in these cases the difference between the anomalous sensor values and non-anomalous sensor values are generally too small to pose any substantial risks to the operation of the vehicle. For magnitudes that may pose significant risk to the operation of the vehicle, i.e., rows 3–5, the models are able to detect anomalous behavior with high performance. As seen in the table, KF and CNN models have similar performance for instant anomalies with large magnitudes. The CNN models generally outperform KF models in terms of sensitivity, precision, and F1 score. Particularly, sensitivity is higher in CNN models, which can directly impact the reliability and safety of fused data in CAVs. A reason for this higher performance of CNN models compared to KF models is that when the anomalies are small enough, the attack will fall into the region of gating for the KF model, therefore it cannot be detected by the chi-square test, which results in a low sensitivity measure. Additionally, unlike the KF models that use the readings by a single sensor over time to detect any potential anomalies on that sensor, the CNN models use the readings from *all* sensors within a time window to detect anomalous behavior by each individual sensor. This redundancy in information improves the CNN performance when the anomaly magnitudes are small and therefore harder to detect.

Considering the anomaly detection performance of the CNN-KF model for the instant anomaly type, the following is noted. Similar to the results for KF and CNN models, the detection performance across all metrics increases in anomaly magnitude. Furthermore, it is seen that, in general, the CNN-KF model improves upon the detection performance of both

KF and CNN models as reported in Table I. For instance, for instant anomalies of magnitude $25 \times \mathcal{N}(0, 0.01)$ added to the base value (row 1), it is seen that sensitivity and F1 score of the CNN-KF model respectively increase by 13.1% and 18.9% over the KF model, and by 1.6% and 1.4% over the CNN model. Note that the F1 scores of CNN-KF are larger that those of CNN in rows 4–5; these numbers only appear to be the same as the numbers in the table are rounded to one decimal place. Lastly, note that for anomalies with very large magnitudes (row 5), high performance is found in all models, especially for KF.

2) Constant: The constant anomaly type is simulated as a temporarily constant observation that is different compared to the "normal" sensor readings. Similar to the previous case, we simulate the magnitude of the anomaly that is added to the base value at the onset of the anomaly using a random variable to capture various magnitudes in any given experiment. Specifically, the magnitude of a given anomaly is sampled from a uniform distribution $\mathcal{U}(0, c)$, where $c \in \{1, 3, 5\}$. In addition, here we account for various durations of the anomalous behavior. Let d denote the number of epochs during which the anomalous behavior is present, where we use $d \in \{3, 5, 10\}$.

Table II shows the results of the constant anomaly type for the KF, CNN, and CNN-KF models. As seen in rows 1–3, the performance of the KF and CNN models generally increases in anomaly duration, given that anomaly magnitudes are drawn from the same random variable. In these cases, because the anomaly magnitude can be somewhat large, KF generally outperforms CNN. Similarly, as seen in rows 3–5, given a fixed anomaly duration (with $d = 10$), the performance of both KF and CNN models are typically better when the anomaly magnitudes are generally larger. In general, similar to the detection performance of the instant anomaly type, the KF model slightly underperforms compared to the CNN model in the case of low magnitude anomalies (rows 4–5) and slightly outperforms the CNN model in detecting anomalies with stochastically larger magnitudes (rows 1–3). In addition, the CNN model generally illustrates a more consistent detection performance across various anomaly magnitudes and durations compared to KF.

Considering the anomaly detection performance of the CNN-KF model for the constant anomaly type, the following is noted. The results illustrate that the CNN-KF model outperforms the KF model when the magnitude of anomalies is relatively small, i.e., in rows 4–5. In addition, the CNN-KF model clearly outperforms the CNN model with respect to accuracy, sensitivity, and F1-score across all experiments. Note that the magnitude of gain in performance is larger when comparing CNN-KF with KF, as opposed to when comparing CNN-KF with CNN. For instance, in row 5, using the CNN-KF model, as opposed to the KF model, increases the sensitivity and F1 score by up to 9% and 7.6%, respectively. Compare these numbers, respectively, with the observed increases of up to 2.6% and 1.1% when using the CNN-KF model, as opposed to the CNN model. Note that the improved performance of CNN-KF model over the CNN model is mainly due to the ability of the Kalman filtering aspect of the CNN-KF model to detect the onset of anomalous behavior faster than that of the CNN model, especially for larger anomaly magnitudes. Lastly, similar to the results in Table I, it is seen that KF outperforms CNN-KF for anomalies with large magnitudes.

3) Gradual drift: The gradual drift anomaly type is simulated by adding a linearly increasing set of values to the base values of the sensors. Specifically, we use a vector of linearly increasing values from 0 to $c \in \{2, 4\}$, corresponding to 2 m/s and 4 m/s, respectively, denoted by the function *linspace*$(0, c)$. In addition, here again we account for various durations of the anomalous behavior, namely, $d \in \{10, 20\}$. For instance, when $c = 4$ and $d = 20$, a linearly increasing speed of up to 4 m/s (i.e., with an intercept of 0 and a slope of $4/19$) is added to the base speed of the CAV over the next 20 epochs (i.e., 2 seconds).

Table III shows the results of the gradual drift anomaly type for the KF, CNN, and CNN-KF models. Gradual drift is one of the most difficult anomalies to detect since it increases sensor values gradually, therefore making it challenging to detect the onset of anomalous sensor behavior and differentiating it from normal behavior. Nevertheless, KF and CNN perform reasonably well across various values of magnitude and duration considered. Similar to the constant anomaly type, the detection performance of the CNN model increases in duration and magnitude of the anomalous sensor behavior. Furthermore, as seen in the table, CNN models consistently outperform KF models across all magnitudes and durations for the gradual drift anomaly type. That is mainly because the sliding window implementation of CNN provides additional opportunities to detect gradual drift anomalies compared to KF. This generally leads to an improved detection performance, even if the anomaly onset is missed by CNN, which is often the case for gradual drift anomalies.

Considering the anomaly detection performance of the CNN-KF model for the gradual drift anomaly type, the following is noted. As seen in Table III, this model outperforms both KF and CNN models across all experiments. For instance, in row 4, using the CNN-KF model, the sensitivity and F1 score respectively increase by 6.7% and 9.2% compared to the KF model, and by 1.2% and 0.6% compared to the CNN model.

4) Bias: The bias anomaly type is simulated as a temporarily constant offset from the baseline sensor readings. Similar to the previous cases, we simulate the magnitude of the anomaly using a random variable to capture various magnitudes in any given experiment. Specifically, the magnitude of a given anomaly is sampled from a uniform distribution $\mathcal{U}(0, c)$, where $c \in \{1, 3, 5\}$. In addition, here we account for various durations of the anomalous behavior, where the sampled magnitude is added to all true sensor readings during the specified duration to generate the anomalous readings. Let d denote the number of epochs during which the anomalous behavior is present, where we use $d \in \{3, 5, 10\}$.

Table IV shows the results of the bias anomaly type for the KF, CNN, and CNN-KF models. As seen in rows 1–3, the performance of the KF and CNN models generally increases in anomaly duration, given that anomaly magnitudes are drawn from the same random variable. Similarly, as seen in rows 3–5, given a fixed anomaly duration (with $d = 10$), the performance

TABLE I
DETECTION PERFORMANCE OF INSTANT ANOMALY TYPE FOR THE KF, CNN AND CNN-KF MODELS.

	KF (%)				CNN (%)				CNN-KF (%)			
Anomaly Magnitude	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
base value + 25 × $\mathcal{N}(0, 0.01)$	95.7	38.4	65.8	48.5	79.1	49.9	97.7	66.0	80.0	51.5	97.6	**67.4**
base value + 100 × $\mathcal{N}(0, 0.01)$	98.6	78.4	93.6	85.3	93.5	85.7	98.1	91.5	93.6	86.2	97.9	**91.7**
base value + 500 × $\mathcal{N}(0, 0.01)$	99.7	95.6	99.0	97.3	98.2	95.8	99.8	97.8	98.3	96.0	99.7	**97.8**
base value + 1,000 × $\mathcal{N}(0, 0.01)$	99.8	96.2	100	98.1	98.7	97.1	99.8	98.4	98.8	97.1	99.8	**98.4**
base value + 10,000 × $\mathcal{N}(0, 0.01)$	99.9	100	99.7	**99.8**	99.6	99.1	100	99.5	99.7	99.2	99.8	99.5

TABLE II
DETECTION PERFORMANCE OF CONSTANT ANOMALY TYPE FOR KF, CNN AND CNN-KF MODELS.

		KF (%)				CNN (%)				CNN-KF (%)			
Anomaly Magnitude	Duration, d	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
base value + $\mathcal{U}(0, 5)$	3	98.5	91.4	98.8	**95.0**	94.5	89.1	99.8	94.1	94.9	89.9	99.6	94.5
base value + $\mathcal{U}(0, 5)$	5	98.5	94.9	98.5	**96.7**	94.6	90.7	99.2	94.8	95.1	91.7	99.0	95.2
base value + $\mathcal{U}(0, 5)$	10	97.8	96.0	98.5	**97.3**	95.5	93.7	99.2	96.4	96.2	94.9	99.1	97.0
base value + $\mathcal{U}(0, 3)$	10	95.7	92.5	96.9	94.6	94.8	92.9	98.8	95.8	95.3	93.9	98.7	**96.2**
base value + $\mathcal{U}(0, 1)$	10	88.8	78.8	92.4	85.1	90.1	85.2	99.1	91.6	91.2	87.8	98.6	**92.7**

TABLE III
DETECTION PERFORMANCE OF GRADUAL DRIFT ANOMALY TYPE FOR THE KF, CNN AND CNN-KF MODELS.

		KF (%)				CNN (%)				CNN-KF (%)			
Anomaly Magnitude	Duration, d	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
base value + $linspace(0, 4)$	10	94.7	91.4	95.3	93.4	94.4	92.0	99.3	95.5	94.7	93.1	99.2	**96.1**
base value + $linspace(0, 4)$	20	92.2	93.0	94.7	93.8	95.7	95.1	99.4	97.2	96.0	95.6	99.3	**97.4**
base value + $linspace(0, 2)$	10	90.3	86.5	89.3	87.9	92.7	89.1	99.4	94.0	93.0	89.6	99.3	**94.2**
base value + $linspace(0, 2)$	20	83.1	86.5	86.9	86.7	92.8	92.0	98.7	95.3	93.7	93.2	98.7	**95.9**

TABLE IV
DETECTION PERFORMANCE OF BIAS ANOMALY TYPE FOR KF, CNN AND CNN-KF MODELS.

		KF (%)				CNN (%)				CNN-KF (%)			
Anomaly Magnitude	Duration, d	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1	Acc	Sens	Prec	F1
base value + $\mathcal{U}(0, 5)$	3	98.7	93.4	98.1	**95.7**	94.0	87.8	99.8	93.4	94.6	89.3	99.5	94.2
base value + $\mathcal{U}(0, 5)$	5	98.2	94.6	97.5	**96.0**	94.2	89.2	99.9	94.3	94.8	90.6	99.6	94.9
base value + $\mathcal{U}(0, 5)$	10	97.3	95.3	98.0	96.6	94.4	91.3	99.8	95.4	95.9	94.4	99.1	**96.7**
base value + $\mathcal{U}(0, 3)$	10	95.9	93.2	96.5	94.8	92.4	88.4	99.6	93.7	94.4	92.6	98.6	**95.5**
base value + $\mathcal{U}(0, 1)$	10	90.1	81.0	93.8	86.9	85.0	77.6	98.6	86.8	88.0	84.8	95.9	**90.0**

of both KF and CNN models generally decrease when the magnitude of anomalies stochastically decrease. Interestingly, different from previous cases, in this case the KF model consistently outperforms the CNN model in all cases examined.

Considering the anomaly detection performance of the CNN-KF model for the bias anomaly type, the following is noted. The results illustrate that the CNN-KF model outperforms the KF model when the magnitude of anomalies is relatively small and their duration is relatively long, i.e., in rows 3–5. In addition, the CNN-KF model clearly outperforms the CNN model with respect to sensitivity and F1-score across all experiments. The improved performance of CNN-KF model over the CNN model is mainly due to the ability of the Kalman filtering aspect of the CNN-KF model to detect the onset of anomalous behavior faster than that of the CNN model, especially for larger anomaly magnitudes.

In conclusion, as seen in Table I, CNN-KF generally outperforms both CNN and KF in detecting the instant anomaly type. KF only outperforms CNN-KF when the magnitude of the anomaly is very large. As seen in Table II, the CNN and CNN-KF models perform better for anomalies with smaller magnitudes. Similar to Table I, KF performs better for anomalies with large magnitudes but less so for anomalies with small magnitudes. This is due to the fact that when the anomalies are small enough, they fall into the gating region of the KF model; therefore, they cannot be detected by the chi-square test. Also, as seen in Table III, the CNN-KF model outperforms both KF and CNN models across all magnitudes and durations considered for the gradual drift anomaly type. This is mainly because CNN-KF combines the strength of CNN where the sliding window implementation provides additional opportunities for detection and the ability of KF in detecting the first few epochs of these anomalies. Lastly, as seen in Table IV for the bias anomaly type, consistent with the results in Tables I–III, the CNN-KF models perform well for anomalies with small magnitudes and long duration whereas KF performs better for anomalies with large magnitudes and a small duration.

B. Models Under Mixed Anomaly Types

In this section we investigate the performance of the models when applied in *detection* and *identification* of various types of anomalies as opposed to the single anomaly types considered in Section IV-A-A. First, we investigate the generalizability of the models presented in Section IV-A-A with respect to unseen anomaly types to motivate the need to develop CNN-based models under mixed anomaly types. Next we develop new models, trained and tested in the presence of all four anomaly types, where the simulation is run multiple times to provide confidence intervals (CIs). Specifically, we present the mean performance along with the 95% CIs, and perform statistical tests to establish whether or not the observed improvements across models are significant. In addition, we analyze the effect of the rate at which anomalous sensor values may occur (at $\alpha = 5\%$ and $\alpha = 1\%$) on the performance of the models.

As seen in Section IV-A-A, CNN and CNN-KF models generally outperform KF models. However, note that in contrast to CNN models, KF models do not require much effort for training and they generalize well; only parameters γ and M need to be calibrated for KF models. Additionally, as we will demonstrate in this section, for CNN models to generalize well and correctly classify previously unseen observations, they require to be trained on representative training sets. However, in practice, CAV anomaly detection systems may encounter various instances of anomalies for which the models are not explicitly trained. This is particularly important for CAVs since these vehicles will be faced with numerous unfamiliar circumstances. Here we present the results of using the trained CNN models from Section IV-A-A and testing them on datasets where all types of anomalous sensor values are present. We break down the results to present the performance of the models in terms of anomalous sensor values identification. The main objective of this analysis is to investigate the degree to which each of these models can generalize to detect/identify unseen anomalies.

Table V presents the performance of training CNN models on one type of anomaly and using them to identify anomalous sensor values where all anomaly types are present. Each case presented across three rows provides the performance of a trained model on a particular training set with the given anomaly type. In the test dataset, the instant, constant, gradual drift, and bias anomalies are all present and are modeled using 1,000×$\mathcal{N}(0, 0.01)$, $\mathcal{U}(0, 5)$ with $d = 10$ epochs, $linspace(0, 4)$ with $d = 20$ epochs, and $\mathcal{U}(0, 5)$ with $d = 10$ epochs, respectively. For instance, for the first case illustrated in Table V, we train the CNN with the instant anomaly type where anomalies are sampled from 1,000×$\mathcal{N}(0, 0.01)$ and we test the model on the test set where all four anomaly types are present.

As seen in Table V, training the CNN model using only instant anomalous sensor values results in poor performance, and particularity low sensitivity. This is expected, since the constant, gradual drift, and bias anomaly types are very different from the instant anomaly type in both magnitude and duration. Also, note that the identification performance generally varies across sensors. Specifically, for sensor 3 (i.e., in-vehicle acceleration), across all experiments, it is seen that the performance metrics are worse than those for the other two speed sensors. This is partly due to the large variability in consecutive acceleration measurements compared to the other two speed sensors that tends to report much smoother readings over time. Lastly, as seen in the table, using the CNN model that is trained on either constant, gradual drift or bias anomaly type results in reasonable performance across all sensors with an F1 score of up to 90.1%.

TABLE V

IDENTIFICATION PERFORMANCE OF TRAINING CNN MODELS ON ONE TYPE OF ANOMALOUS SENSOR VALUES AND TESTING THEM TO IDENTIFY ANOMALOUS SENSOR VALUES WHERE ALL ANOMALY TYPES ARE PRESENT. THE REPORTED VALUES ARE IN PERCENTAGES.

Anomaly Type Used in Training	Sensor	Acc	Sens	Prec	F1
Instant, 1,000×$\mathcal{N}(0, 0.01)$	1	91.1	64.1	99.3	77.9
	2	88.1	59.0	95.0	72.8
	3	85.4	44.1	99.9	61.2
Constant, $\mathcal{U}(0, 5)$, $d = 10$	1	95.3	87.0	93.4	90.1
	2	90.9	70.8	94.0	80.7
	3	89.2	65.2	91.1	76.0
Gradual drift, $linspace(0, 4)$, $d = 20$	1	92.5	76.4	91.5	83.3
	2	90.6	70.2	93.5	80.2
	3	88.2	66.7	85.0	74.7
Bias, $\mathcal{U}(0, 5)$, $d = 10$	1	94.8	88.3	90.4	89.3
	2	91.7	70.6	98.1	82.1
	3	89.3	61.7	96.2	75.1

In the next set of experiments, we train the models using datasets in which all types of anomalous sensor values are present, to estimate the models' anomaly identification performance in practice. Table VI illustrates the identification performance and the 95% CIs across 10 simulation runs, when the anomaly rate is set to 5% (α = 5%) and all types of anomalies are present. Similar to the results obtained for the specialized models in Section IV-A-A, as seen in Table VI, the CNN model generally outperforms the KF model, especially with respect to sensitivity and F1 score. Furthermore, the CNN-KF model improves upon the performance of both KF and CNN models, especially with regard to F1 score as highlighted in the table. We perform paired t-tests and one-way ANOVA tests to investigate the statistical significance between the identification performance with respect to F1 score between all pairs of models, i.e., KF and CNN models, KF and CNN-KF models, and CNN and CNN-KF models. The p-values obtained indicate statistical significance at 5% level (p-value $<$ 0.05) across all tests performed.

Figure 2 presents the in-vehicle speed sensor readings, with and without superimposed anomalous sensor values, during a 3-second time window of a CAV trip in one of the test sets, and illustrates the instances at which the anomalous sensor values were detected by each of the three models. In this example, the anomaly is of type constant, has a relatively small magnitude of 0.17 m/s (i.e., results in anomalous sensor readings of 18.66 m/s compared to the base speed of 18.49 m/s at the anomaly onset), and lasts for a period of $d = 10$ epochs or 1 second, i.e., from 980 to 989 time epochs, into the CAV trip. As seen in the figure, the KF and CNN models both fail to identify all anomalous sensor readings during this 1 second period. In contrast, CNN-KF model identifies all anomaly instances. Note that, in this case, the anomalous sensor values caused by the constant anomaly type are very small compared to the base speed value, illustrating the effectiveness of the CNN-KF framework in detecting/identifying the anomalous sensor values.

TABLE VI
IDENTIFICATION PERFORMANCE AND THE 95% CIs ACROSS 10 DIFFERENT EXECUTIONS FOR ALL THREE MODELS, AT THE ANOMALY RATE OF α = 5% AND IN THE PRESENCE OF ALL TYPES OF ANOMALIES. P-VALUES INDICATE STATISTICAL SIGNIFICANCE AT 5% LEVEL USING PAIRED t-TEST AND ONE-WAY ANOVA TESTS, BETWEEN THE IDENTIFICATION PERFORMANCE OF ALL PAIRS OF MODELS. THE REPORTED VALUES ARE IN PERCENTAGES.

Model	Sensor	Acc	Sens	Prec	F1
KF	1	97.8 ±0.2	90.5 ±1.4	95.9 ±1.1	93.1 ±0.7
	2	98.0 ±0.2	90.6 ±1.1	96.0 ±1.1	93.2 ±0.8
	3	96.1 ±0.3	85.1 ±2.6	89.0 ±1.8	86.9 ±1.4
CNN	1	98.0 ±0.3	93.3 ±0.7	99.5 ±0.2	96.3 ±0.4
	2	97.0 ±0.2	90.4 ±0.6	98.7 ±0.3	94.4 ±0.4
	3	94.7 ±0.5	83.3 ±1.3	97.6 ±1.0	89.8 ±0.8
CNN-KF	1	98.1 ±0.2	94.2 ±0.5	99.4 ±0.2	**96.7** **±0.3**
	2	97.6 ±0.2	92.9 ±0.5	98.4 ±0.4	**95.5** **±0.4**
	3	95.7 ±0.4	87.3 ±1.0	97.3 ±0.7	**92.0** **±0.8**

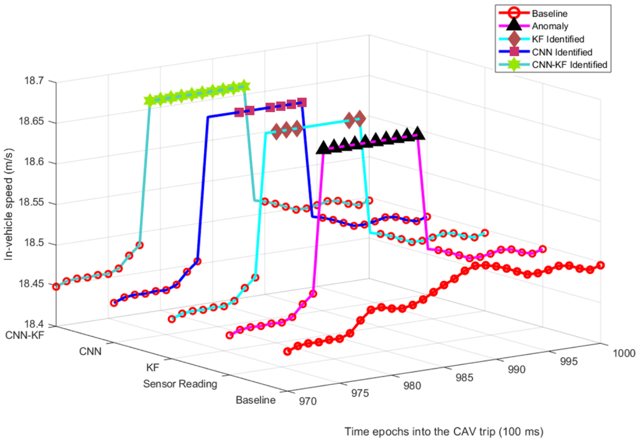

Fig. 2. An example illustrating the performance of KF, CNN, and CNN-KF models in identifying anomalous sensor values of in-vehicle speed sensor readings during a 3-second time window of a CAV trip.

Lastly, we investigate the effect of the anomaly rate α on the identification performance of the trained models. Table VII illustrates the identification performance and the 95% CIs when the previously trained models are tested across 20 different test sets in which the anomaly rate is set to 1% (α = 1%) and all anomaly types are present. In this experiment we use a larger number of test sets compared to the previous experiment due to the low anomaly rate. Similar to the α = 5% case illustrated in Table VI, the CNN model outperforms the KF model, and the CNN-KF model outperforms both the KF and CNN models. Specifically, the p-values obtained using paired t-test and one-way ANOVA tests indicate statistical significance at 5% level between the identification performance of all pairs of models with respect to F1 score. As seen in Tables VI and VII, model accuracy increases when the rate of anomalous sensor values is at α = 1%, compared to α = 5%, as there are fewer anomalous instances to misclassify; however, F1 score generally decreases at α = 1%, compared to α = 5%. Note that these results are in general better than those obtained if the training sets used were at α = 1% anomaly rate, particularly for CNN and CNN-KF models. This is mainly because if the training set is highly unbalanced, classification models tend to favor the more representative class (in our case 'normal' readings) and hence, do not generalize well to detect anomalous sensor values. This is why when training a model on an unbalanced set, either oversampling or undersampling is utilized [45], [46].

TABLE VII
IDENTIFICATION PERFORMANCE AND THE 95% CIs ACROSS 20 DIFFERENT EXECUTIONS FOR ALL THREE MODELS, AT THE ANOMALY RATE OF α = 1% AND IN THE PRESENCE OF ALL TYPES OF ANOMALIES. P-VALUES INDICATE STATISTICAL SIGNIFICANCE AT 5% LEVEL USING PAIRED t-TEST AND ONE-WAY ANOVA TESTS, BETWEEN THE IDENTIFICATION PERFORMANCE OF ALL PAIRS OF MODELS. THE REPORTED VALUES ARE IN PERCENTAGES.

Model	Sensor	Acc	Sens	Prec	F1
KF	1	99.3 ±0.1	84.5 ±2.8	92.1 ±1.3	88.0 ±1.7
	2	99.3 ±0.1	84.8 ±3.1	92.7 ±1.8	88.5 ±2.3
	3	98.7 ±0.2	70.6 ±3.6	86.9 ±3.4	77.3 ±1.9
CNN	1	99.3 ±0.1	90.7 ±1.0	97.1 ±1.1	93.8 ±0.9
	2	99.1 ±0.1	89.8 ±1.3	95.0 ±1.5	92.2 ±0.9
	3	98.2 ±0.4	79.1 ±1.9	94.8 ±2.0	86.1 ±1.1
CNN-KF	1	99.3 ±0.1	91.3 ±0.9	96.7 ±1.1	**93.9** **±0.8**
	2	99.1 ±0.1	90.9 ±1.0	94.8 ±1.4	**92.8** **±0.9**
	3	98.5 ±0.3	83.2 ±1.9	95.0 ±1.9	**88.6** **±1.1**

V. DISCUSSION

The main objective of this study is to improve safety of CAVs and robustness of their decisions in the presence of faulty sensors or cyber attacks, which may pose significant risks to transportation network users. CAVs are expected to use a number of redundant sensors measuring the same parameters, e.g. speed or location, which can be used to detect and identify anomalous sensor values by comparing the data collected from the corresponding sensors. In addition, CAVs collect large volumes of data, often at 100ms or finer intervals, enabling the use of deep learning techniques to learn complex, nonlinear patterns in the data to improve overall detection/identification performance. We take advantage of these two important features of CAVs by combining a deep learning technique, CNN, and a traditional anomaly detection technique, KF with a failure detector, to address the real-time anomalous sensor value detection and identification for CAVs. Our approach improves upon both CNN and KF models, when they are used separately, as it integrates their strengths.

Note that the developed models are able to perform detection and identification, not prediction. That is, they do not have the capability to preventively detect anomalous sensor values. However, the models are expected to be used in an online fashion, on high-frequency data (with the sampling frequency of approximately 100ms) to detect and identify anomalous sensor values. Therefore, if a sensor starts to transmit anomalous data, it can be identified and actions may be implemented within a fraction of a second to mitigate its effects.

Introducing sliding windows in CNN and KF-CNN models allows for evaluating sensor readings holistically and more than once, hence providing additional opportunities for detection and identification of anomalous sensor values. For instance, all models are at risk of missing bias and gradual drift anomaly types, particularly if the anomaly magnitudes are small. However, in CNN and CNN-KF, even if the onset of anomaly is initially missed, the anomaly is detected/identified after only a few time epochs, e.g., approximately in 300ms.

VI. Conclusion and Future Work

The goal of this study is to develop an approach to detect/identify anomalous behaviors in CAVs to improve their safety. Our results show that the use of deep learning models such as CNN to detect/identify anomalous sensor values in CAV systems in real-time is a viable path and it can improve upon the well-established methods such as Kalman filtering with failure detectors. In addition, our results show that because of the inherent differences between CNN and KF, combining these approaches can build upon their individual strengths and hence result in an improved performance. More specifically, we show that by using a CNN-empowered KF on raw sensor data, it is possible to detect and identify anomalous sensor values in real-time with high accuracy, sensitivity, precision, and F1 score. This research contributes to the field of ITS safety as a whole since the success of ITS operations is heavily dependent on the safe operation of all its respective elements. In addition, CAV manufacturers and policy-makers may benefit from the observations in this study with regards to the value of having redundant information for a specific parameter such as the speed of a vehicle. As a result, more redundant sensors and information gathering devices may be implemented and considered in CAVs to increase their resiliency against anomalous sensor values. It is also expected that the CNN-KF framework presented in this study will be applied to various other sources of information in CAVs to increase their safety.

The study is subject to limitations. First, the anomalous sensor values used in the experiments, consistent with previous studies in the literature, are simulated, mainly because this type of data are not yet readily available. In addition, due to paucity of data on connected vehicles, the experiments are limited to on-board sensors. Although the framework is generic and can address anomaly detection/identification regardless of the data source, additional testing is needed as real data becomes readily available. To that end, it should be noted that the reported performance of the proposed CNN-based methods are only valid for those attack types whose effects can be reflected by the four anomaly types considered in this study.

Furthermore, in this study, we assume the *onset* of anomalous values in sensors, due to either attacks or faults, occur independently. Although we do not explicitly train models on datasets containing interdependent sensor failures or systemic cyber attacks on a vehicle's sensors, the framework is expected to still be able to detect such faults/attacks. To improve model performance under such scenarios, however, additional training using such data is required. Indeed, it is envisioned that real-world data would be available in the near future as a result of various ongoing pilot studies (Michigan, Wyoming, New York City, and Tampa) as well as the introduction of CAVs into society by companies such as Waymo, Uber, Audi, and Tesla.

In future work, it may be useful to distinguish between anomalous and malicious information, since this will influence the action taken to mitigate their effects. Also, it may be beneficial to also identify the type of anomaly occurring. This, in turn, will enable the development of certain real-time actions to minimize the impact of cyber attacks and/or faulty sensors, therefore contributing to the development of effective counter-measures.

References

[1] Ran, B. and Boyce, D., 2012. Dynamic urban transportation network models: theory and implications for intelligent vehicle-highway systems (Vol. 417). Springer Science & Business Media.

[2] Meyer, G. and Beiker, S. eds., 2014. Road Vehicle Automation (p. 154). Heidelberg: Springer.

[3] Litman, T., 2014. Autonomous vehicle implementation predictions. Victoria Transport Policy Institute, 28.

[4] DoT, U.S., 2016. United States Department of Transportation. Connected vehicles and cyber security. Available at https://www.its.dot.gov/factsheets/pdf/cv_%20cybersecurity.pdf. Accessed 09-03-2017.

[5] Greenberg, A., 2015. Hackers remotely kill a jeep on the highway – With me in it. Wired, 7, p.21.

[6] Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H. and Savage, S., 2010, May. Experimental security analysis of a modern automobile. In Security and Privacy (SP), 2010 IEEE Symposium on (pp. 447-462). IEEE.

[7] Weimerskirch, A. and Gaynier, R., 2015, October. An Overview of Automotive Cybersecurity: Challenges and Solution Approaches. In TrustED@ CCS (p. 53).

[8] Bai, F., Stancil, D.D. and Krishnan, H., 2010, September. Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers. In Proceedings of the sixteenth annual international conference on Mobile computing and networking (pp. 329-340). ACM.

[9] Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S., Koscher, K., Czeskis, A., Roesner, F. and Kohno, T., 2011, August. Comprehensive Experimental Analyses of Automotive Attack Surfaces. In USENIX Security Symposium.

[10] Field, M., 2017, The Telegraph, Graffiti on stop signs could trick driverless cars into driving dangerously, Available at: http://www.telegraph.co.uk/technology/2017/08/07/graffiti-road-signs-could-trick-driverless-cars-driving-dangerously. Accessed 09-01-2017.

[11] Darms, M., Rybski, P. and Urmson, C., 2008, June. Classification and tracking of dynamic objects with multiple sensors for autonomous driving in urban environments. In Intelligent Vehicles Symposium, 2008 IEEE (pp. 1197-1202). IEEE.

[12] Varghese, J.Z. and Boone, R.G., 2015, September. Overview of Autonomous Vehicle Sensors and Systems. In Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering.

[13] Khaleghi, B., Khamis, A., Karray, F.O. and Razavi, S.N., 2013. Multisensor data fusion: A review of the state-of-the-art. Information Fusion, 14(1), pp.28-44.

[14] Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1), pp.35-45.

[15] Jo, M., Park, J., Baek, Y., Ivanov, R., Weimer, J., Son, S.H. and Lee, I., 2016, October. Adaptive Transient Fault Model for Sensor Attack Detection. In Cyber-Physical Systems, Networks, and Applications (CPSNA), 2016 IEEE 4th International Conference on (pp. 59-65). IEEE.

[16] Petit, J. and Shladover, S.E., 2015. Potential cyberattacks on automated vehicles. IEEE Transactions on Intelligent Transportation Systems, 16(2), pp.546-556.

[17] Yan, C., Wenyuan, X. and Liu, J., 2016. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle?, DEF CON.

[18] Hoppe, T., Kiltz, S. and Dittmann, J., 2008. Security threats to automotive CAN networks – practical examples and selected short-term countermeasures. Computer Safety, Reliability, and Security, pp.235-248.

[19] Truong,K. N., Patel, S. N., Summet, J. W., and Abowd, G. D., Preventing camera recording by designing a capture-resistant environment, in Proc.7th Int. Conf. UbiComp, 2005, pp. 73–86.

[20] Yang, S., Liu, Z., Li, J., Wang, S. and Yang, F., 2016. Anomaly Detection for Internet of Vehicles: A Trust Management Scheme with Affinity Propagation. Mobile Information Systems, 2016.

[21] Lin, R., Khalastchi, E. and Kaminka, G.A., 2010, May. Detecting anomalies in unmanned vehicles using the mahalanobis distance. In Robotics and Automation (ICRA), 2010 IEEE International Conference on (pp. 3038-3044). IEEE.

[22] Realpe, M., Vintimilla, B. and Vlacic, L., 2015, January. Sensor Fault Detection and Diagnosis for autonomous vehicles. In MATEC Web of Conferences (Vol. 30). EDP Sciences.

[23] Pous, N., Gingras, D. and Gruyer, D., 2017. Intelligent Vehicle Embedded Sensors Fault Detection and Isolation Using Analytical Redundancy and Nonlinear Transformations. Journal of Control Science and Engineering, 2017.

[24] Khalastchi, E., Kaminka, G.A., Kalech, M. and Lin, R., 2011, May. Online anomaly detection in unmanned vehicles. In The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1 (pp. 115-122). International Foundation for Autonomous Agents and Multiagent Systems.

[25] Park, J., Ivanov, R., Weimer, J., Pajic, M. and Lee, I., 2015, April. Sensor attack detection in the presence of transient faults. In Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems (pp. 1-10). ACM.

[26] Christiansen, P., Nielsen, L.N., Steen, K.A., Jorgensen, R.N. and Karstoft, H., 2016. DeepAnomaly: Combining Background Subtraction and Deep Learning for Detecting Obstacles and Anomalies in an Agricultural Field. Sensors, 16(11), p.1904.

[27] Kang, M.J. and Kang, J.W., 2016. Intrusion detection system using deep neural network for in-vehicle network security. PloS one, 11(6), p.e0155781.

[28] Bezemskij, A., Loukas, G., Gan, D. and Anthony, R., 2017. Detecting cyber-physical threats in an autonomous robotic vehicle using Bayesian Networks.

[29] Foo, G.H.B., Zhang, X. and Vilathgamuwa, D.M., 2013. A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended Kalman filter. IEEE Transactions on Industrial Electronics, 60(8), pp.3485-3495.

[30] Wei, X., Verhaegen, M. and van Engelen, T., 2010. Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques. International Journal of Adaptive Control and Signal Processing, 24(8), pp.687-707.

[31] Muter, M. and Asaj, N., 2011, June. Entropy-based anomaly detection for in-vehicle networks. In Intelligent Vehicles Symposium (IV), 2011 IEEE (pp. 1110-1115). IEEE.

[32] NHTSA, 2013. Preliminary Statement of Policy Concerning Automated Vehicles, National Highway Traffic Safety Administration, Available at https://www.nhtsa.gov, Accessed 11-03-2017.

[33] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks, 61, pp.85-117.

[34] Bezzina, D. and Sayer, J., 2014. Safety pilot model deployment: Test conductor team report. Report No. DOT HS, 812, p.171.

[35] Kingma, D. P., and Ba, J., 2014. Adam: A Method for Stochastic Optimization, presented at the 3rd International Conference on Learning Representations (ICLR), San Diego, USA.

[36] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[37] Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of machine learning research, 15(1), pp.1929-1958.

[38] Bhuyan, Monowar H and Bhattacharyya, Dhruba Kumar and Kalita, Jugal K, 2014. Network anomaly detection: methods, systems and tools. IEEE communications surveys & tutorials.

[39] Brown, R Grover, 1994. GPS RAIM: Calculation of Thresholds and Protection Radius Using Chi-square Methods; a Geometric Approach. Radio Technical Commission for Aeronautics.

[40] NXP Semiconductor, F., 2014. MMA8451Q 3-Axis, 14-bit/8-bit Digital Accelerometer. Data Sheet, Document Number: MMA8453Q.

[41] Mo, Yilin, et al., 2010. False data injection attacks against state estimation in wireless sensor networks. Decision and Control (CDC), 2010 49th IEEE Conference on. IEEE, .

[42] Mohamed, A. H., and K. P. Schwarz. 1999. Adaptive Kalman filtering for INS/GPS. Journal of geodesy 73.4 : 193-203.

[43] Sharma, A.B., Golubchik, L. and Govindan, R., 2010. Sensor faults: Detection methods and prevalence in real-world datasets. ACM Transactions on Sensor Networks (TOSN), 6(3), p.23.

[44] Loebis, D., Sutton, R., Chudley, J. and Naeem, W., 2004. Adaptive tuning of a Kalman filter via fuzzy logic for an intelligent AUV navigation system. Control engineering practice, 12(12), pp.1531-1539.

[45] Chawla, N.V., Japkowicz, N. and Kotcz, A., 2004. Special issue on learning from imbalanced data sets. ACM Sigkdd Explorations Newsletter, 6(1), pp.1-6.

[46] Park, S.H., Kim, S.M. and Ha, Y.G., 2016. Highway traffic accident prediction using VDS big data analysis. The Journal of Supercomputing, 72(7), pp.2815-2831.

[47] Marchetti, M., Stabili, D., Guido, A., & Colajanni, M. 2016. Evaluation of anomaly detection for in-vehicle networks through information-theoretic algorithms. In Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016 IEEE 2nd International Forum on (pp. 1-6).

[48] Taylor, A., Leblanc, S., & Japkowicz, N. 2016. Anomaly detection in automobile control network data with long short-term memory networks. In Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on (pp. 130-139).

[49] Levi, M., Allouche, Y., & Kontorovich, A. 2018. Advanced Analytics for Connected Car Cybersecurity. In 2018 IEEE 87th Vehicular Technology Conference (VTC Spring) (pp. 1-7).

[50] Trippel, T., Weisse, O., Xu, W., Honeyman, P., & Fu, K. 2017. WALNUT: Waging doubt on the integrity of MEMS accelerometers with acoustic injection attacks. In Security and Privacy (EuroS&P), 2017 IEEE European Symposium on (pp. 3-18).

[51] Currie, R. 2015. Developments in car hacking. SANS Institute, InfoSec Reading Room, pp. 33.

[52] Huq, N., Vosseler, R. & Swimmer, M. 2017. Cyberattacks against intelligent transporta-tion systems.

[53] Williams, C. 2017. Hackers hit D.C. police closed-circuit camera network, city officials disclose. Available at:https://www.washingtonpost.com/local/public-safety/hackers-hit-dcpolice- closed-circuit-camera-network-city-officials-disclose

[54] Malhotra, P., Vig, L., Shroff, G. and Agarwal, P., 2015, April. Long short term memory networks for anomaly detection in time series. In Proceedings (p. 89). Presses universitaires de Louvain.

Franco van Wyk is a Ph.D. candidate in the Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville. He received his M.Eng degree in Engineering Management from Stellenbosch University, South Africa. His research interests include optimal sequential decision making under uncertainty and safety in connected and automated vehicle systems. He is a member of IEEE and IISE.

Yiyang Wang is an M.S. student in the Electrical Engineering and Computer Science Department at the University of Michigan Ann Arbor. He received his B.S. degree in Telecommunications Engineering from Jilin University, China. He is currently a research assistant at Next Generation Mobility Systems Lab in the Department of Civil and Environmental Engineering at University of Michigan Ann Arbor. His research interests include statistical machine learning, stochastic processes, and cybersecurity in connected and autonomous vehicle technologies.

Anahita Khojandi is an Assistant Professor in the Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville. She received her Ph.D. in Industrial Engineering from University of Pittsburgh. Her research interests include decision making under uncertainty, data mining and machine learning with applications in a wide array of complex systems. She is a member of INFORMS, IISE, IEEE, and Society for Medical Decision Making.

Neda Masoud is an Assistant Professor in the Department of Civil and Environmental Engineering at the University of Michigan Ann Arbor. She received her Ph.D. in Civil and Environmental Engineering from University of California Irvine. She holds an M.S. degree in Physics, and a B.S. degree in Industrial Engineering. Her research interests include large scale optimization and machine learning with applications in shared-use mobility services and connected and automated vehicle systems. She is a member of INFORMS and TRB.

APPENDIX

See Tables VIII – IX.

TABLE VIII
SUMMARY OF KEY LITERATURE RELATED TO ANOMALY DETECTION IN CAVS.

Author(s)	Cyber-security/anomaly aspects investigated	Key relevant findings	Study approach (method)	Type of data used
[20]	Anomaly detection for internet of vehicles, specifically, detecting abnormal vehicles operating in a platoon	Low detection failure rate below 1%, demonstrating its ability to detect and filter the abnormal vehicles	Affinity propagation framework	Simulation in TransModeler
[15]	An adaptive transient fault model for sensor attack detection for multiple operating mode systems	Improvement on existing non-transient fault models. Uses a dynamic look-up table for the applicable system parameters	Transient fault model (TFM) using graph theory	Data obtained from an unmanned ground vehicle. Simulated attacks
[9]	Analysis of external attack surfaces of a modern automobile	Remote exploitation is feasible via a broad range of attack vectors	Experimental analyses on vehicles in the sedan segment	Real experiments on vehicles
[26]	Obstacle/anomaly detection algorithm using deep learning	High accuracy, low computation time and low memory footprint	Combine background subtraction and CNN	Field experiments
[17]	Examines the security of the sensors of autonomous vehicles, and investigate the trustworthiness of the sensors	Off-the-shelf hardware were able to perform jamming and spoofing attacks which can compromise the safety of self-driving cars	Laboratory and outdoor experiments	Collected data through experiments
[21]	A model-free approach for detecting anomalies in unmanned autonomous vehicles, based on sensor readings	Works well for a limited number of attributes	Anomaly detection using Mahalanobis distance	Data from unmanned aerial vehicle
[28]	Detecting cyber-physical threats in real time in an autonomous robotic vehicle	Can determine whether an AV is under attack and also whether the attack originated from the cyber or the physical domain	Heuristic binary classifier and Bayesian network	Simulated attacks using an unmanned ground vehicle
[31], [47]	Detecting anomalies (attacks) for in-vehicle networks	Certain attacks on the CAN-bus of a vehicle were detected using the proposed methodology. Difficult to detect low-volume attacks	Signal entropy	Field experiments using a vehicle, CAN data
[6]	Experimental security analysis of a modern automobile	It is possible to bypass rudimentary network security protections such as the malicious bridging between a cars internal subnets	Experiments in laboratory and road tests	Various experiments
[24]	Online anomaly detection in unmanned vehicles	Method is able to take into account a large number of monitored sensors and internal measurements	Anomaly detection using Mahalanobis distance	Data from robot, and a high-fidelity flight simulator
[16]	State of the art in identifying potential cyber attacks on automated vehicles	Identifies risks of various importance. Identifies GNSS spoofing and injection of fake messages as most dangerous attacks on AVs	Exploratory study	Review of literature
[27]	A deep learning model to enhance in-vehicular safety by detecting malicious CAN packets	Real-time response to the attack with a significantly improved detection ratio in controller area network (CAN) bus	Multilayer perceptron (MLP) model	Simulated data using software package (OCTANE)
[25]	Addresses the problem of detection and identification of sensor attacks in the presence of transient faults	Able to detect and identify attacks using sensor fusion	Pairwise inconsistencies between sensors to detect and identify attacks	Unmanned ground vehicle
[48]	Deep learning to detect attacks on CAN bus	Detect anomalies with low false alarm rates	Long Short-Term Memory (LSTM) recurrent neural network (RNN)	Synthesize anomalies with modified CAN bus data
[49]	Machine learning approach is proposed to protect connected vehicles	Detect anomalies with a large number of features	Combine Hidden Markov Model and regression model	Simulation using software (SUMO)

TABLE IX
NOTATION TABLE

Variable	**Description**
n	Number of sensors
α	Anomaly rate
x	State variable of Kalman filter
z	Sensor measurement
w	Process noise of state-transition model
v	Sensor measurement noise
ν	Innovation between measurement and predicted value of the measurement
A	State-transition matrix
H	Sensor measurement matrix
R	Process noise covariance matrix
Q	Measurement noise covariance matrix
$\hat{R}$	Process noise covariance matrix estimate
$\hat{Q}$	Measurement noise covariance matrix estimate
S	Covariance matrix of innovation
P	Error covariance matrix, where the error is the difference between true and predicted state values
M	Window size of adaptive Kalman filter
γ	χ^2 detector parameter