This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

Detecting Human Driver Inattentive and Aggressive Driving Behavior using Deep Learning: Recent Advances, Requirements and Open Challenges

MONAGI H. ALKINANI[1], WAZIR ZADA KHAN[2] (Senior Member, IEEE) QURATULAIN ARSHAD[3]

[1]College of Computer and Engineering, Department of CS & AI, University of Jeddah, Saudi Arabia (e-mail: malkinani@uj.edu.sa)
[2]Farasan Networking Research Laboratory, Faculty of CS & IT, Jazan University, Saudi Arabia, 45142 (e-mail: wazirzadakhan@jazanu.edu.sa)
[3]Department of CS, Comsats University, Wah Campus, Pakistan. (e-mail: brightsuccess_12@yahoo.com)

Corresponding authors: Wazir Zada Khan, Monagi H. Alkinani (e-mail: wazirzadakhan@jazanu.edu.sa, malkinani@uj.edu.sa).

ABSTRACT Human drivers have different driving styles, experiences, and emotions due to unique driving characteristics, exhibiting their own driving behaviors and habits. Various research efforts have approached the problem of detecting abnormal human driver behavior with the aid of capturing and analyzing the face of driver and vehicle dynamics via image and video processing but the traditional methods are not capable of capturing complex temporal features of driving behaviors. However, with the advent of deep learning algorithms, a significant amount of research has also been conducted to predict and analyze driver's behavior or action related information using neural network algorithms. In this paper, we contribute to first classify and discuss Human Driver Inattentive Driving Behavior (HIDB) into two major categories, Driver Distraction (DD), Driver Fatigue (DF), or Drowsiness (DFD). Then we discuss the causes and effects of another human risky driving behavior called Aggressive Driving behavior (ADB). Aggressive driving Behavior (ADB) is a broad group of dangerous and aggressive driving styles that lead to severe accidents. Human abnormal driving behaviors DD, DFD, and ADB are affected by various factors including driver experience/inexperience of driving, age, and gender or illness. The study of the effects of these factors that may lead to deterioration in the driving skills and performance of a human driver is out of the scope of this paper. After describing the background of deep learning and its algorithms, we present an in-depth investigation of most recent deep learning-based systems, algorithms, and techniques for the detection of Distraction, Fatigue/Drowsiness, and Aggressiveness of a human driver. We attempt to achieve a comprehensive understanding of HIADB detection by presenting a detailed comparative analysis of all the recent techniques. Moreover, we highlight the fundamental requirements. Finally, we present and discuss some significant and essential open research challenges as future directions.

INDEX TERMS Deep Learning, Human Inattentive Driving Behavior, Connected Vehicles, Road Accident Avoidance, Abnormal Behavior Detection, Distraction or Aggressiveness Detection, Fatigue or Drowsiness Detection.

I. INTRODUCTION

THE massive development in automobile industry has enhanced vehicle technology to ensure safe and secure travel. But still many accidents occur due to the unsafe, careless and dangerous driving of the driver. Hence, even today, the problem of traffic safety concerns the worldwide. Road accident is undoubtedly a global tragedy that has become an ever rising trend. Road traffic accidents and injuries impose serious problems on society by causing considerable economic losses to individuals, their families and to nations as a whole [1]. World Health Organization (WHO) announced nearly 1.25 million fatalities each year and on average 3,287 deaths a day [1], [2]. It is estimated that road traffic deaths continue to climb, reaching 1.35 million in 2016 [2]. It is thus, very crucial to understand the various factors and reasons associated with these road accidents.

The vision of highly automated vehicles is accident free driving in the future. In theory, the technology of self-driving cars is established with six levels of autonomy ranging from 0 to 5. In order to detect potentially dangerous driving situations Advanced Driving Assistance Systems (ADAS) or Advanced Driving Systems (ADS) are suggested to be capable of reducing human errors while driving. Although these systems offer many advantages and can handle some specific dangerous situations but in reality, they become incompetent to detect dangerous driving menacing situations. This is due to the fact that these systems depend on sensors and in adverse situations, the accuracy of these sensors diminishes significantly [3]. For example, lane departure warning systems use sensors to register lane markings or the road edge, which may be problematic on roads that aren't well marked or are covered with snow. Some sensors may not function well in low light or inclement weather. Moreover, some systems only work at certain speeds. There have been recent news reports of fatal Tesla crashes that occurred when automation failed to detect obstacles during a period when the driver was not monitoring the automation [4]. Hence, a human driver needs to be alert and ready for taking control of his car (vehicle) in emergency situations [5], [6].

The semi-autonomous vehicles or partial driving automation (SAE Level 2) [7], [8] are driver assistance systems that are becoming more prevalent these days and are increasingly available in passenger vehicles with level SAE Level 2. For example, the automation systems that are on the road are launched from companies such as Tesla, Mercedes, GM, and Volvo, are Level 2. These cars can control steering and speed on a well-marked highway, but a driver still has to supervise. Comparatively, a Honda vehicle is a Level 1 which is equipped with its "Sensing" suite of technologies, that include emergency braking detection, adaptive cruise control and lane keeping assistance. It is often confused with highly automated driving, however, in fact the semi-autonomous vehicles require human driver to monitor automation [9]. Furthermore, under some safety conditions, safety considerations may require automation to shut it-self off to protect an inattentive driver [10]. Thus, for both levels of 2 and 3 ADAS systems, human drivers must be well prepared to intervene, whenever system failures or limitations occur. According to [11], increasing levels of vehicle automation has shifted the human driver's active role to a supervisory role for monitoring automation. It is thus important for the human driver to be more attentive and simultaneously aware of multiple vehicle's status in order to respond quickly in the event of failure or malfunction.

The automation in vehicles is ultimately degrading the attentiveness of a human driver. The study in [12] pointed out that human drivers fail to monitor the automation and thus cannot detect critical signals related to systems functionality due to misunderstanding or over trust in these systems. A multitude of research concerning human performance issues including trust in automation, situation awareness etc., is covered by [13] and evolving driving roles are specifically covered in [14]. According to [15], automotive manufacturers might avoid this supervisory role altogether and only 5% of the studies have addressed this area, thus it is still opened to debate. In this regard, Banks, Stanton and Harvey indicated in [16] that automation in driving and additional assistance subsystems would increase cognitive loads on a human driver rather than reducing his workload and they have analyzed in [17] the videos of the participants operating a *Tesla Model S* in Autopilot mode, finding that drivers were not properly supported in adhering to their new monitoring responsibilities, and were showing signs of complacency and over-trust. Moreover, they suggested that certain levels of driving automation need not be implemented even if they are feasible from a technical point of view, and while considering human factors or human supervisory role in automated driving, only two levels driver driving (DD) and driver not driving (DND) should be preferred. Another important issue regarding the automated driving is the concept of trust calibration which means that human drivers while supervising the automation in driving intervene only when they believe that their own decisions are superior to the automation system's decisions. In this regard, [13] suggested to convey the the capabilities and limitations of the automation to the operator whenever feasible.

Intensive efforts have been made in understanding, detecting, identifying and predicting the human driving styles and behavior since driving is an essential daily activity for many people [18]. Various research efforts have approached the problem of detecting abnormal human driver behavior with the aid of capturing and analyzing face of driver and vehicle dynamics via image and video processing but the traditional methods are not capable of capturing complex temporal features of driving behaviors. The studies [19]–[23] that utilized accelerometers and gyro-sensors built into smart phones for detecting driving behavior, fail to provide accurate results if the vehicle is operated in a GPS-deprived area, or if the GPS receiver demonstrated poor performance. Thus, aggressive driving remained difficult to detect. Moreover, such systems fail to recognize aggressive driving in the form of quick and irregular turns and frequent braking while the vehicle is used on a mountain road and other winding-type roads. In Bio-signal-based methods [24], [25], the problem is that expensive sensors are used and the attachment of sensors may cause discomfort. In the single camera based solutions [26], measurement becomes difficult at night or in tunnels. Driver fatigue can be detected over a wide range by leveraging dual Near-infrared (NIR) cameras [27], however, physical characteristics that can be observed by a naked eye cannot be detected. Aggressive driving emotion detection-based convolution neural networks (CNN) method [28] has used both NIR and thermal cameras where NIR cameras can detect facial feature points and measure their changes and Thermal cameras can measure temperature changes in a driver's body, which cannot be checked by the naked eye. In

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

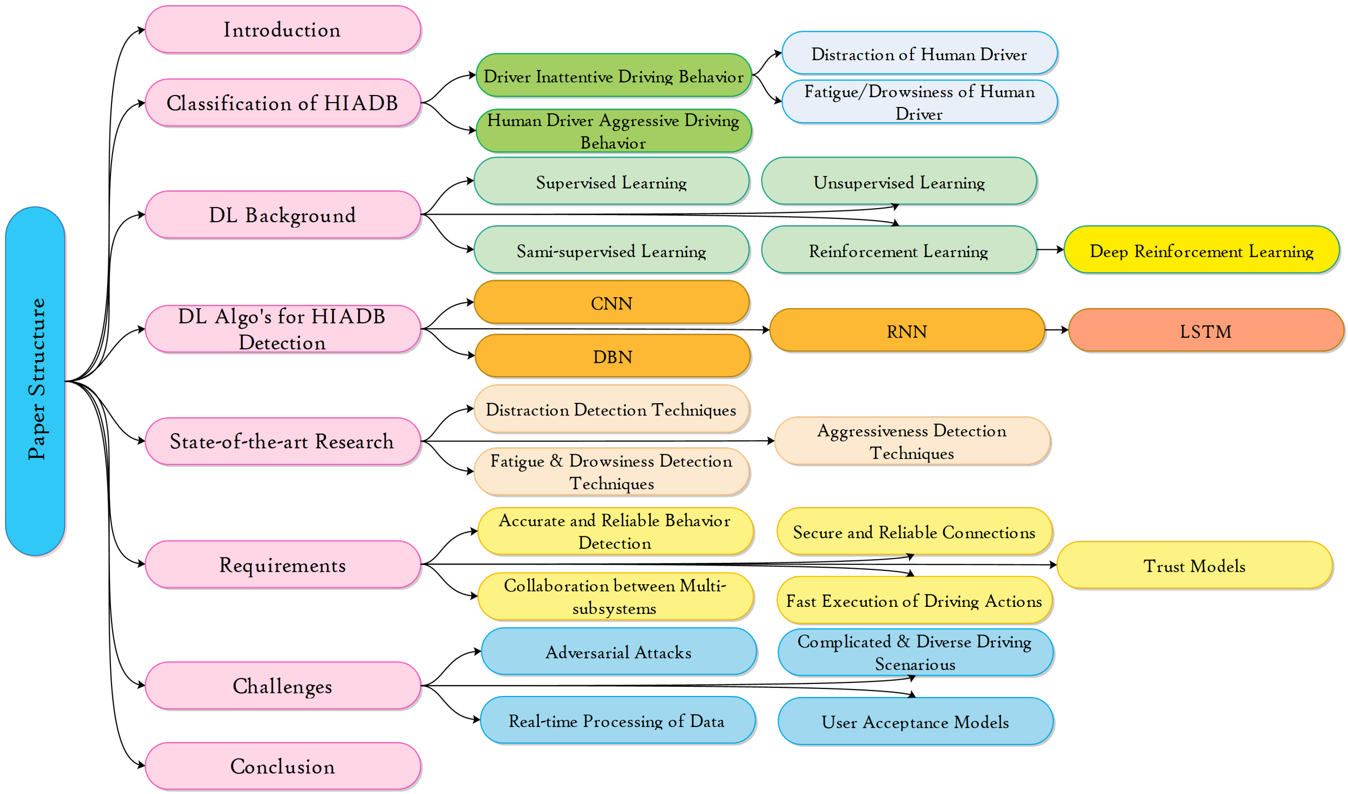

FIGURE 1. Structure of the Paper.

this method, an intensively trained CNN is robust in various environmental and driver conditions. However, the use of two cameras increases algorithm complexity and processing time.

The topic of driver behavior detection is intuitively so wide and several surveys and studies exist in the literature on driver behavior detection [29]–[32], but what appears to be lacking in these attempts is that none of them presented a fuller and more holistic picture of human driving behavior (HDB) detection techniques based on deep learning. The survey [29] covers only the analysis of the existing image processing methods of driver drowsiness detection and recent deep learning techniques and methods for drowsiness detection of a driver are still missing. In another survey [30], the authors focus on the utilization of deep learning models for various traffic state prediction like traffic speed and flow in Intelligent Transportation Systems (ITS). Another recent comprehensive survey [31] focused on utilization of deep learning methods for various ITS problems including traffic flow prediction, traffic signal control, travel time estimation etc. The authors in [32] investigated ML based ITS techniques, applications and services, such as co-operative driving and road hazard warning and expanded the concepts of various ITS tasks but lacks human driving behavior recognition techniques. Motivated by the fact that monitoring and correcting driver intention in time is critical to ADAS systems in order to avoid making conflicting decisions with the driver, more recently the authors in [33] provide an overview of the ego-vehicle driver intention inference (DII) systems. They mainly focus on driver's lane change intention on highways and classified human driver intention inference mechanisms into three major categories; strategy intention,operation intention and tactical intention.

This paper focuses on the most recent deep learning based systems, algorithms and techniques for the detection of Human Driver Inattentive and Aggressive Driving Behavior (HIADB) by classifying human Inattentive driving behavior (HIDB) into two major categories; Distraction and Fatigue/Drowsiness. We also discuss in detail causes and effects of human driver Aggressive driving behavior and its detection through deep learning methods. Aggressive driving behaviors often result in property as well as bodily injury damages. A number of factors including driver's experience/inexperience in driving, decline (change) in driving skills due to natural aging or illness/diseases, gender, different driving characteristics as well as the other environmental conditions like road adhesion, traffic conditions, and weather conditions affect the attentiveness of a human driver while driving either negatively or positively. But the study of these factors on human driving behavior are out of the scope of this paper. The challenges and requirements regarding the detection of HIADB are also highlighted. In an attempt to achieve a comprehensive understanding of human driving behavior, a detailed comparative analysis of all the recent techniques is also performed, presenting in the form of tables and figures and thus, this review serves as trove of information which is not only helpful in informing the specialists and the students about the current state of the art HIADB detection but also assist them with design choices.

The major contributions of this study are as follows:

1) The human driver inattentive driving behavior (HIDB) is classified into two major categories; distraction and fatigue/drowsiness. The detection of driver distraction and driver fatigue/drowsiness is classified according to the features selected for the detection of human driving behavior in the literature.
2) A more risky human driving behavior called Aggressive driving behavior is also explained and discussed, highlighting the causes and effects of different aggressive driving styles on human safe driving. The detection of human driver aggressive driving behavior (HADB) is classified according to the aggressive driving styles adopted by an aggressive driver.
3) The most recent deep learning based solutions for human driver Inattentive and Aggressive Driving Behavior (HIADB) detection are reviewed systematically and comprehensive comparative analysis is performed, highlighting their detection accuracies.
4) The fundamental requirements for detecting human abnormal driving behavior from developing trust of ADAS systems to including trust, security and privacy of such automated systems are presented and discussed.
5) The imperative open research challenges in the field of HIADB detection are identified. These challenges are discussed as future research directions that need to be addressed for the accurate detection of abnormal human driving behavior so that risk of accidents can be reduced.

The structure of our paper which is illustrated in Figure 1 is as follows: In Section 2, we first classify and discuss the Human driver Inattentive Driving Behavior (HIDB) detection and then discuss Human driver Aggressive Driving Behavior (HADB) and classify its detection measures. In Section 3, we present some background about machine learning (ML), deep learning (DL). In Section 4, we give an overview of deep learning algorithms that are particularly employed for detecting HIADB. In Section 5, we comprehensively review deep learning based methods devised in the literature for detecting HIADB and a comparative analysis of these approaches is performed according to the classification in Section 2. In Section 6, we highlight some of the fundamental requirements for HIADB detection. Section 7 presents open research challenges in detecting human abnormal driving behavior. In Section 8, we finally conclude the paper.

II. HUMAN DRIVER INATTENTIVE AND AGGRESSIVE DRIVING BEHAVIOR (HIADB)

A. HUMAN DRIVER INATTENTIVE DRIVING BEHAVIOR (HIDB)

Driving is a dynamic process which is comprised of three key components including, driver, vehicle and the driving environment [34]. For the sake of life and property safety, a driver is responsible to make appropriate decisions and perform actions accordingly by staying aware and attentive to the environment and the current situation. A study sponsored by NHTSA, investigated 723 crashes and showed that driver behavioral error caused or contributed to 99% of these crashes [35]. It is estimated that 90% of road accidents are caused by wrong driving behavior. In this section, we categorize the HIDB that contribute to serious road accidents into two main domains; Driver Distraction (DD) and Driver Fatigue or Drowsiness (DFD).

1) Distraction of a Human Driver

Distraction of a human driver refers to loosing concentration behind the wheels while doing another event or activity and when an object, or person outside or inside of the vehicle takes away the driver's attention from the driving task. Distraction of human driver can degrade his driving performance resulting in unplanned speed changes, hiccups in vehicle control, and drifting outside the lane edges, which ultimately increases the chance of a motor vehicle crash. Distraction can be either driver initiated (where the driver starts carrying out a distracting activity) or non-driver initiated (the unpredictable actions of something or someone else) [36]. Human driving distractions can vary greatly in the form and severity. Some of the examples of human driver distraction include sending or texting a message on mobile phone, calling, listening or reaching a mobile phone [37], looking at off-road persons and events, using a navigation system, eating or drinking, operating in-vehicle-technology etc. Every year automobile accidents happen and hundreds of thousands of people are injured due to distracted human drivers and this number is continuously rising. It is reported that when divers try to multitask, road accidents happen more often and thus, more than half of the total accidents are caused by distraction of a driver [37].

Human driver distractions are of six kinds including *visual*, *cognitive*, *manual*, *auditory*, *olfactory* and *gustatory* distraction. Visual distraction is when a driver takes his/her eyes off the road. Cognitive distraction is when a driver takes his/her mind off of driving and thinking about something which is not related to vehicle driving. Manual distraction is when the driver takes his/her hands off the steering wheel and manipulating a device. Auditory distraction is when a human driver hears some sound like ringtone or music which is not related to driving and thus, distraction is caused when different sounds impede the driver from making the best use of his/her hearing, making him less observant towards controlling the vehicle safely. While driving, mood or mind set of a driver can be changed when he listens to music or having a conversation, thus, effecting the human driving behavior. Olfactory distraction is when a driver is distracted by the smell of something outside or inside of the vehicle. Gustatory distraction is when a driver is distracted by the taste of something while eating or drinking.

Driving is a multitask activity (manual/visual) or a process that demands human driver attentional resources associated with his visual features (visual perceptions), cognitive fea-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

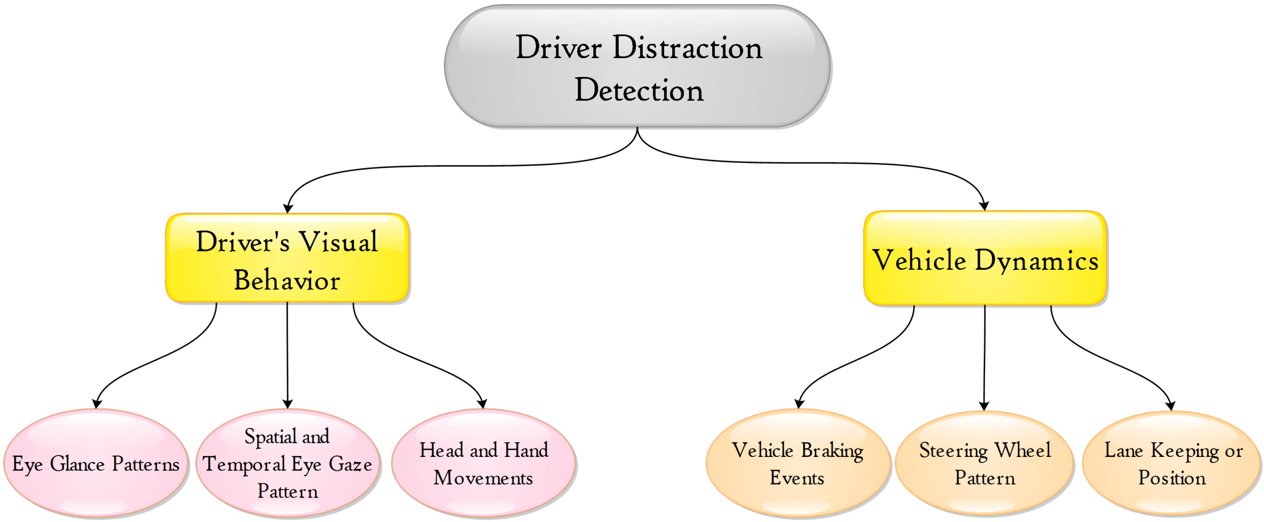

FIGURE 2. Taxonomy of driver distraction detection.

tures (spatially working memory) and manual features (motor responses). However, other human resources like auditory or verbal responses, senses like olfactory and gustatory are substantially not needed (or have minor demand) for driving tasks. Visual, cognitive and manual distractions highly degrade the driving performance and disrupt the driver's attention allocated to the driving scene and for processing information while driving whereas the other three distractions (auditory, olfactory, gustatory) have minimal or no affect on the driving performance or attentiveness of a driver. According to the estimation of the National Safety Council, 26% of all car crashes involve cell phones because it is a combination of visual, manual and cognitive distractions [38].

We categorize detection of Human driver distraction into two types of measures, human driver's visual behavior and vehicle's dynamics or vehicle related features (driving performance metrics). This categorization is shown in Figure 2. Visual, cognitive or manual human driver distraction can be detected by measuring human's visual features (eye glance patterns, spatial and temporal eye gaze distribution or pattern, head and hand movements) or vehicle related features (steering wheel pattern, lane keeping or position, vehicle braking events). When attempting to determine whether cognitive load has truly diverted the driver's attention, it is a bigger challenge to identify and assess what is and what is not cognitive load as it is difficult to understand precisely what a driver may be thinking. A number of studies exist in the literature for the detection of human driver visual, cognitive and manual distraction. According to [39], increased visual distraction resulted due to frequent glances at an object far away and off-road, the concentration of gaze. The three parameters standard deviation of lane position, eyes-off-road glance time and glance duration of head-off-road are crucial indicators of visual distraction. The authors in [40]–[42] study and analyze the parameters like gaze pattern and head movements to discriminate between focused and distracted driving and measure the cognitive distraction. A study in [43] has shown that distraction, cognitive workload, and features of eye movement (saccade, smooth pursuit, and fixation) are interlinked. Saccades are quick actions that occur when visual attention transfers from one point to another and thus, can be used for mental workload measurements.

Recently, for the non-intrusive and real-time detection of visual distraction, a method was developed in [44] which detects driver's visual distraction by measuring vehicle dynamics data and environmental data, without using eye-tracker information. A more recent attempt [45] presented a novel algorithm for detection of driver's manual distraction consisting of two modules. The first module was responsible for the prediction of the bounding boxes of the driver's right hand and right ear from RGB images. The second module on the other hand took the bounding boxes as input and predicted the type of distraction. A research study in [46] developed a distraction detection algorithm using kinematic signals from the vehicle Controller Area Network (CAN) bus. The authors continuously recorded vehicle kinematic data from naturalistic driving study program on everyday driven vehicles without specified base lines. Cabin video data were captured and manually inspected to identify attentive or distracted cases. Moreover, a Nonlinear Autoregressive Exogenous (NARX) model was developed for vehicle speed prediction. On the basis of which the mean of the absolute prediction error was proposed as a new metric for distraction detection. The authors then exploited a support vector machine for distraction detection with the identified features reporting promising detection performance results.

The three distractions including auditory, olfactory and gustatory are still in their infancy and thus, there are only a few attempts in the literature to detect them. A study in [47] proposes a procedure to quantitatively estimate auditory distractions that are not accompanied by any visual diversions to clarify the influence of non-visual distractions on driving. An auditory distraction has negative impact on the driver's driving behavior [48] whereas in other studies [49], [50], it is shown that listening to music can help human drivers to stay alert, and influences mood and physiological states of a driver without necessarily impairing his driving performance. The only two studies we found [51], [52] focused on olfactory distraction and showed that smell distracts a human driver.

2) Fatigue/Drowsiness of a Human Driver

Fatigue of a human driver refers to a state when he/she is too tired to remain alert. Human Driver's Fatigue results due to his physical or mental exertion, prolonged driving, inadequate, fragmented or interrupted sleep, strenuous work or non-work activities or a combination of other factors. Insufficient sleep and fatigue of a driver can reduce vigilance, diminish his alertness and concentration so that he is not capable enough to recognize oncoming hazards which ultimately leads to deterioration of driving performance, affect the quality of decision making, driver's significant loss of control, unpredictable vehicle trajectory and no braking response. Moreover, fatigue leads to slower reaction time, diminished steering performance and makes the driver less capable to keep distance to the car in front.

The terms "fatigue", "sleepiness" and "drowsiness" are often used interchangeably in the driving context but they differ significantly. Sleepiness can be defined as the neurobiological need to sleep [53]. Fatigue is linked to feeling tired, exhausted or low in energy, but often does not result in sleep, whereas drowsiness is a state in between sleep and wakefulness or right before sleep. Inactivity reduces physical fatigue but increases drowsiness. Although the causes of fatigue and sleepiness may be different, their effects are very much the same, namely a decrease in mental and physical performance capacity. The U.S. National Highway Traffic Safety Administration (NHTSA) reports that drowsy driving is related to at least 100,000 motor-vehicle crashes and more than 1,500 deaths per year. Since, drowsy and fatigue driving is a serious problem that is responsible for thousands of accidents and numerous fatalities every year, thus, it is a major transportation safety concern. Several studies [54], [55] reveal that 25-35% of driving mishaps are related to fatigue, making it the second major reason for road accidents.

We categorize detection of Human driver fatigue or drowsiness into two major types of measures; Driver-based measures or Vehicle-based measures. Driver based measures, derived from human driver, are further categorized into Physiological features (intrusive) and Visual features (non-intrusive). Physiological measures of a human driver involve intrusive fatigue detection by analyzing the psychological state of the driver through electroencephalographic (EEG) and electro-oculographic (EOG) information features. Measuring human visual features involve non-intrusive detection by providing fatigue warning based on facial features including Head and Eye Movement etc. Vehicle based features include those which are derived from the vehicle including Lateral displacement, Speed, Acceleration etc. Fatigue/Drowsiness detection categorization is also shown in the Figure 3.

A number of methods have been proposed in the literature in the past. The researchers have focused on utilizing certain physiological and physical phenomena and variations that undergo in the body of a fatigued driver for the detection of human driver fatigue. Recently, a driver fatigue detection system was presented [56] in which SVM was trained on driver performance data, lane position, lane heading, and lateral distance. The authors have proposed and applied a two stage parameter learning algorithm called Distributed Learning and Searching (DL&S) to select optimal parameters and used them to train a SVM for driver fatigue detection. at the first stage distributed learning was performed on a subset of training data, and at the second stage fine-tuning was performed to obtain the most optimal parameters. The data set was generated from real-world driving trips taken by 20 drivers. Their experimental results showed that the proposed DL&S algorithm consumed only 7.5% of the computational cost needed by grid search.

Driver's drowsiness can be detected by identifying initial signs of fatigue before a critical situation arises. The most recent attempt in [57] has utilized eye tracking data acquired from 53 subjects in a simulated driving experiment and simultaneously recorded multichannel electroencephalogram (EEG) signals to detect drowsiness of a driver. They employed random forest (RF) and a non-linear support vector machine (SVM) for binary classification of the state of vigilance. Another attempt in [58] presented an algorithm for drowsiness detection in drivers of several vehicles based on eye-shape. They located the face in real-time video, and feature point detection on face region for delimiting the ocular area by using a combination of HOG Linear SVM. Moreover, the authors in [59] detect driver's drowsiness by classifying surface electromyography signal features. They measured the surface electromyography signal from the upper arm and shoulder muscles including mid deltoid, clavicular portion of the pectoralis major, and triceps and biceps long heads. They have applied six classifiers for the prediction of drowsiness and concluded that the k-nearest neighbor classifier predicts drowsiness by 90% accuracy, 82% precision, 77% sensitivity, and 92% specificity. A recent study [60] leverages the data of steering wheel angles for monitoring driver fatigue level under real driving conditions by presenting an online drowsiness detection system. Another study [61] performed a comparative analysis of K-nearest neighbor, support vector machine, and artificial neural network classifiers for driver drowsiness detection by analyzing different road geometries (straight segments and curve segments) based on a driving simulator [62].

B. HUMAN DRIVER AGGRESSIVE DRIVING BEHAVIOR (HADB)

In addition to distraction and fatigue, aggressive driving behavior of a human driver is another major reason for the road

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

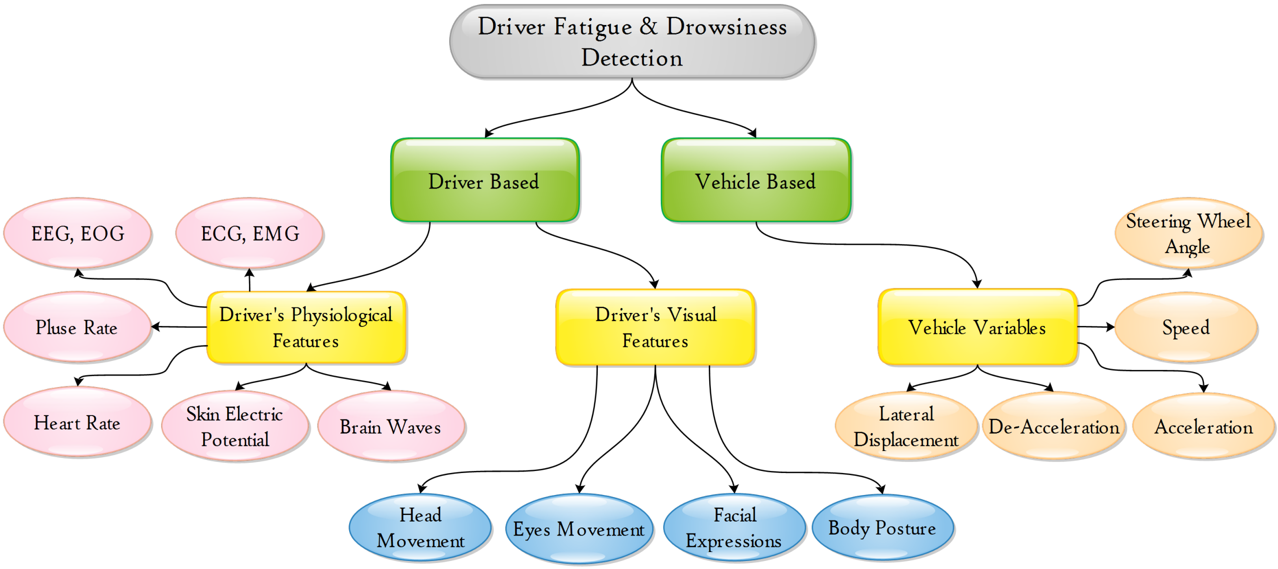

FIGURE 3. Taxonomy of driver fatigue and drowsiness detection.

accidents. National Highway Traffic Safety Administration (NHTSA) defines aggressive driving behavior as an action "when individuals commit a combination of moving traffic offences so as to endanger other persons or property" [63]. According to the British Automobile Association, flashing, obscene gestures, deliberate obstruction of equipment, verbal abuse or strike are labeled as aggressive driving behaviors or road rage [64]. Some authors even report that aggressive driving and road rage are used as synonyms but, in order to make a distinction between them, the authors in [65] said that Aggressive driving is a traffic offense and road rage is a criminal offense and thus, road rage is associated with criminal behaviors punishable by law.

We categorize Aggressiveness of a human driver into two types; Habitual Aggressiveness (Intentional) and Situational or Occasional Aggressiveness (non-intentional). Habitual aggressiveness of a human driver is an intentional risk-taking behavior that refers to unsafe and hostile driving behavior while operating a vehicle for instance, making frequent or unsafe lane changes (weaving), excessive speeding, tailgating etc. The driving in such an aggressive manner may also sometimes result in harm for other individuals (motorists) on the road and for the society, thus, risk the lives of other people. This kind of aggressive behavior may include violent behaviors such as failing to signal or yield the right of way, ignoring or disregarding traffic controls, shortcut maneuvers etc. The authors in [66] describe the perception of what people consider an aggressive behavior, and their perception of which are the most aggressive acts performed when driving. They categorized three behavioral elements of aggressive driving as intentional acts of physical, verbal, or gestured aggression; negative emotions (e.g., anger, frustration) while driving; and risk-taking. It was found that when the nuance of intentionality of a driver is added to the debate of aggressiveness, aggressive driving is not the result of simple lapses and errors while driving but a clear distinction can be made between violent but non-aggressive behaviors, and aggressive, but non-violent behaviors.

Situational or Occasional aggressive driving behavior of a human driver is a non-intentional risk taking behavior while operating a vehicle that refers to reckless driving without any intention to harm other road users and on certain occasions it can be useful for survival purposes. sometimes the road situation or environment makes a human driver to act aggressively and adopt aggressive driving styles. For instance; A simulated and on road experimental study [67] was conducted and focused on driving anger induction and detection. Three scenarios including waiting for the red light frequently, traffic congestion, and the surrounding vehicle interference were used for inducing anger and then a detection algorithm was built for which a Hidden Naive Bayes classifier was employed to detect angry driving during the on-road driving experiments. Many drivers are over confident and unaware of their bad driving habits, thus in order to increase awareness of driving habits of drivers it is important to classify aggressive and normal driving behavior, which can assist them to avoid potential accidents [3]. In this regard, in order to raise people's awareness and help them understand their driving behaviors, the authors in [68] presented and trained a population model using instances for each unique routine label (e.g., one model for aggressive and one for non-aggressive driving). They also conducted a user study in which participants drove on a predefined route and reviewed their driving behaviors using their tool hypothesizing that showing participants their aggressive behaviors and non-aggressive alternatives will change their rating of their driving expertise and quality. Their systems can automatically model, detect, generate, and reason about people's routines by enabling artificial intelligent systems. Such systems can help people reflect on and understand their behaviors, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

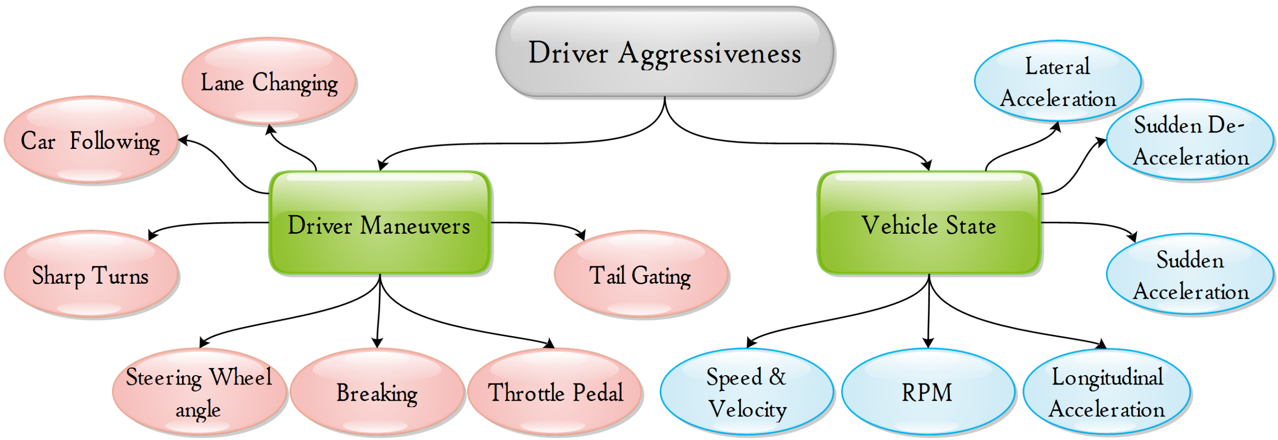

FIGURE 4. Taxonomy of driver aggressiveness.

is a step towards technologies that aid people in behavior change.

We categorize the detection of human driver aggressiveness into two kinds of monitoring the measures; measures related to driver maneuvers and measures related to vehicle state. Human driver's maneuvers or driving styles used to detect aggressive driving include Lane Changing, Car Following, Sharp Turns, Steering Wheel angle, Braking, Throttle Pedal, Tailgating etc. Aggressive behavior of human driver can also be detected by assessing the vehicle's state including Lateral Acceleration, Sudden Acceleration and Deacceleration, Longitudinal Acceleration, Speed and Velocity, RPM etc. This categorization is shown in Figure 4.

Detection of driver aggressiveness can significantly aid in reducing the number of traffic accidents leading to safe driving. There are certain behaviors associated with aggressive driving which are utilized by the researchers to identify aggressive driving behavior. Several methods have been proposed for detecting aggressive driving behaviors based on metrics from vehicle sensor data such as excessive speed, hard braking, heavy acceleration, and aggressive turns. Most of the studies have used common vehicle kinematics such as speed, longitudinal and lateral acceleration to measure driving aggressiveness [69]. An attempt in [70] has investigated that whether vehicle longitudinal jerk could be potentially used to identify aggressive drivers. The authors hypothesized that the vehicle jerk indicates how smoothly a driver accelerates and decelerates the vehicle, and aggressive drivers may use large jerk more often by operating the gas and brake pedal compared to normal drivers. In this regard they developed two jerk-based metrics; the frequency of using large positive jerk when pressing the gas pedal and the frequency of using large negative jerk when pressing the brake pedal. Speeding, Tailgating, driver's association with crash or near-crash were used to identify the aggressive drivers and results showed that aggressive drivers had significantly higher values of the two jerk-based metrics. Most recently the authors in [71] presented a two-stage clustering approach for the detection of unsafe driving styles by utilizing driving data and information on mobile usage, harsh events occurrence, speeding and acceleration profile with increasing importance with respect to safety. In this way, trips have been categorized into six distinct groups (Aggressive trips include Aggressive trips, Distracted trips and Risky trips). Non-Aggressive Trips on the other hand include Safe trips, Distracted Trips and Risky Trips. An initial clustering was performed for separating aggressive from non-aggressive trips whereas a second level clustering was performed to distinguish normal trips from unsafe trips. By grouping the drivers in relation to the trips, the authors have analyzed that drivers cannot maintain a stable driving profile through time, but exhibit a strong volatile behavior per-trip.

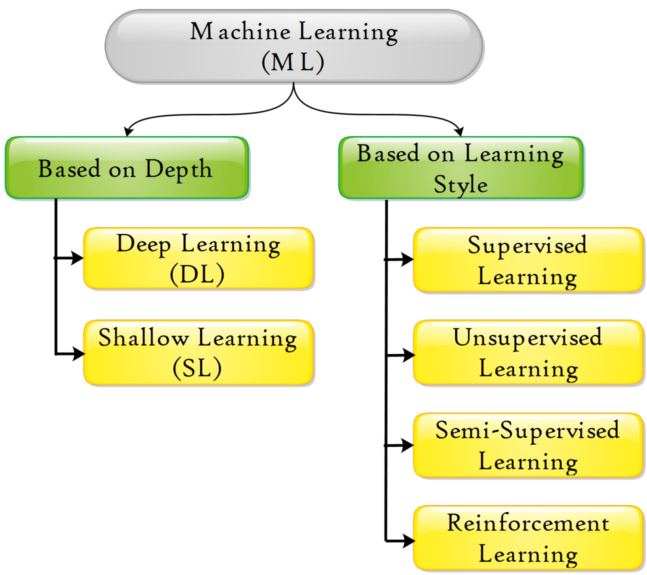

FIGURE 5. Classification of Machine Learning Algorithm.

III. BACKGROUND OF DEEP LEARNING

This section presents the background of Machine learning and deep learning and its categories. Machine Learning (ML) is a small subset of Artificial Intelligence (AI), since

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

the 1950's. ML models can be categorized into two major categories; models based on the depth and models based on learning styles; Models that are based on learning styles are of four kinds in which deep learning models are trained with data; supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Models that are based on depth are further categorized into two types; Shallow learning and Deep Learning. This categorization is clearly demonstrated in Figure 5.

A. MODELS BASED ON LEARNING STYLES

In this section, we present ML models based on learning styles which include Supervised learning, Unsupervised Learning, Semi-supervised Learning and Reinforcement Learning.

1) Supervised Learning

In supervised learning, a deep learning model is handed a full set of labeled data including both the input data and the desired output (correct answer), which a machine learning engineer or data scientist has added while training an algorithm to guide it to understand which features are important to the problem at hand. Thus, the algorithm learns on a labeled dataset, and evaluates its accuracy on the training data using an answer key. For example, when an image of flower is given to a supervised learning model it can compare it to the training examples, a labeled dataset of flower images that would tell the model which photos were of roses, daisies and daffodils and so in this way it can predict the correct label accurately and can correctly classify new images of other flowers. Deep learning supervised learning models include Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN) etc.

2) Unsupervised Learning

In unsupervised learning, a deep learning model is handed an unlabeled dataset including a collection of examples without a specific desired outcome or correct answer. The training dataset has no explicit instructions on what to do with it and thus, the algorithm learns to inherent structure from the input data to make sense by extracting features and patterns on its own. For example, if a collection of bird photos is given to an unsupervised learning model, it will look at them and separate them roughly by species and group them together, by relying on training data and cues like feather color, size or beak shape. Deep learning unsupervised learning models include Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM), etc.

3) Semi-supervised Learning

In semi-supervised learning, a learning model takes the middle ground because it is handed with a training data set in which some data is labeled but most of it is unlabeled. For example, if the semi-supervised learning model has to identify and then specify specific groups of web pages it uses some of the labeled data to identify that there are specific groups of web page types present in the data and what they might be. The algorithm is then trained on unlabeled data so that it is capable enough to define the boundaries of those web page types and may even identify new types of web pages that were unspecified in the existing data handed to the algorithm. The main disadvantage of supervised learning is the expense of labeling the training data the process of labeling massive amounts of data is time consuming and expensive whereas, unsupervised learning often doesn't work very well due to unlabeled data. Semi-supervised learning is thus, useful because it is a win-win in many problems like web page classification, speech recognition, genetic sequence and many more. Some of the semi-supervised deep learning models include Generative Adversarial Networks (GAN), Variational Autoencoders (VAE), Ladder Net etc.

4) Reinforcement Learning

Reinforcement learning is a goal oriented algorithm that falls in between supervised and unsupervised learning. It learns by trial and error and predicts the action that will yield the best result. In reinforcement learning, no supervisor is involved and the algorithm works sequentially in a unknown environment where only a reward signal is used for an agent to determine if they are doing well or not. The agent needs to find the "right" actions to take in different situations for which it is either rewarded or punished for the actions they take. Each action of an agent affects the next data it receives. The general framework of RL can be applied to any problem due to its generality and dynamic nature.

The components of RL algorithm are agent, environment, action, state, reward and policy. An *agent* is an entity that executes actions e.g. a trader deciding what to buy or sell or a robot deciding to walk on a path. The agent makes decisions without losing too much reward for which it needs to learn from its experiences in the environment. A *state* is a situation in which the agent finds itself containing the set of actions, tools, dangers, rewards or the information available to an agent which is used to determine what happens next. A *reward* is a discount which is a scalar feedback signal that represents the result of the agent's action. The reward indicates how well an agent is doing at step time t and can be immediate (short term) or delayed (long term). The reward function should be carefully designed in order to apply RL to various real world problems. An *Environment*, which can be fully observable or partially observable, is a function that transforms the action taken in the previous step into a reward and a new set of actions. A *policy* is an agent's behavior function (Deterministic or Stochastic), or the strategy that agent uses to determine the next action, based on the current state and previous rewards.

Deep reinforcement learning (DRL) is a combination of Reinforcement learning and Deep Learning in which deep learning determines the action taken at every stage by creating a sequential reinforcement learning process [72]. DRL can be further categorized into two classes; Value-Based and Policy-Based Learning. In Value based methods (Q-

Learning), an agent calculates a Q-value of each possible action. It performs back-propagation to find the most accurate Q-Values and selects the best action [73]. In Policy-based methods, an agent does not calculate a value function for each action but it learns the policy function directly, e.g. Policy Gradient. Since, there are major gaps between simulated and real environments that make it difficult to train models, DRL works very well in closed environments like video games, but it is difficult to apply to real-world environments [74]–[76].

B. MODELS BASED ON DEPTH

In this section, we present ML models based on Depth which include Shallow learning models and Deep learning models.

1) Shallow learning

Shallow learning [77] is just a buzz-word for the traditional machine learning pipeline. Shallow Learning relies on hand-crafted features based upon heuristics of the target problem. Supervised Shallow learning models include Support Vector Machines (SVM), Decision tress (DT), K-Nearest Neighbor (k-NN), Random Forest (RF), Linear Regression (LR), Logistic Regression, Naive Bayes etc. Unsupervised Shallow Learning models include Clustering, Hierarchical Clustering, K-means, Auto-encoders etc. Shallow learning models are out of the scope of this paper.

2) Deep Learning

Neural Networks (NN) is a subfield of ML that has spawned Deep Learning (DL). Deep Learning methods derive their own features directly from data (feature learning). Shallow neural Networks consists of single hidden layer whereas Deep Neural Networks consists of more hidden layers and has a large number of neurons in each layer. Since the inception of DL, it has garnered tremendous success in almost every application domain [78]. Although the history of deep learning can be traced back to the mid 1960's [79], modern-day deep learning is still a relatively new development and it developed largely from 2006 onward. It has been applied to a number of fields including computer vision, image processing, speech recognition, medical imaging, bioinformatics, robotics and control, natural language processing, cybersecurity, and many others.

Deep learning describes a family of learning algorithms rather than a single method that can be used to learn complex prediction models, e.g., multi-layer neural networks with many hidden units [80]. A procedure of learning estimates the model parameters so that the learned model (algorithm) can perform a task. These tasks include Classification, Clustering, Regression, generative modeling, dimentionality reduction, association Rule Learning etc. Deep learning consists of several layers in between the input and output layer which allows for many stages of non-linear information processing units with hierarchical architectures to be present that are exploited for feature learning and pattern classification [79], [80]. The traditional ML methods faltered but deep learning excels in solving the problems of learning from raw audio signal, the raw pixel values of images, or mapping between sentences of arbitrary lengths and their counterparts in foreign languages [81]. It is thus capable of addressing low-level perceptual data in a way that ML and previous tools could not because they are affected by the curse of dimensionality. Deep learning models include Convolution Neural Networks (CNN), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM), Autoencoders (AEs), Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) etc. Only those Deep learning models that are employed for driver behavior detection are described in detail in Section IV and the rest of the models are out of the scope of this paper.

IV. DEEP LEARNING ALGORITHMS FOR HUMAN DRIVER INATTENTIVE AND AGGRESSIVE DRIVING BEHAVIOR DETECTION

The deep learning models that are employed for the detection of Human driver Inattentive and Aggressive Driving Behavior (HIADB) are of two kinds; *Generative* models and *Discriminative* models. *Generative* model is a branch of unsupervised deep learning that learns any kind of data distribution and captures the joint probability. It describes how a data set is generated, in terms of a probabilistic model, thus, new data instances can be generated by sampling form this model. The Generative deep learning models include Auto-encoders (AE), Variational Autoencoders (VAE), Restricted Boltzmann Machine (RBM), Deep Belief Networks (DBN) and Generative Adversarial Networks (GAN). The aim of VAE is to maximize the lower bound of the data log-likelihood whereas the aim of GAN is to achieve an equilibrium between Generator and Discriminator. Some of the generative classifiers include Naive Bayes, Bayesian networks, Markov random fields, Hidden Markov Models (HMM). Generative model can also be applied to a labeled dataset to learn how to generate observations from each distinct class.

Discriminative model is a branch of supervised learning and discriminates between different kinds of data instances. It learns a function that maps an input to an output using a labeled dataset and captures the conditional probability. In these models the input data is classified into known categories by learning discriminative driving features adaptively. These algorithms learn distinctive features through non-linear transformation and classification based on probabilistic prediction. Their basic functions include feature extraction and classification. Some of the discriminative deep learning models include Convolution Neural Networks (CNN), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM) which are the focus of this paper. CNN, RNN and LSTM are the most common deep learning methodologies applied to HIADB detection.

A. CONVOLUTION NEURAL NETWORKS (CNN)

Kunihiko Fukushima was the first who introduced Convolutional Neural Networks (CNN) by designing neural networks

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

with multiple pooling and convolutional layers. In 1979, he developed an artificial neural network, called Neocognitron, which used a hierarchical, multilayered design. This design allowed the computer to recognize visual patterns by learning about the shapes of objects [82]. CNN achieved great success not only in powering vision in robots and self-driving cars but also in identifying objects, faces, traffic signs [83] etc. It has been widely adopted in the applications for image classification, speech recognition, video classification, action recognition, and sentence classification.

CNNs, often called ConvNet, have two main components; the feature extraction part and the classification part [83], [84]. In the feature extraction part, features are detected by performing a series of convolutions and pooling operations using two hidden layers; Convolution and Pooling layers. In convolution layers, filters are applied to the original image, or to other feature maps in a deep CNN. A series of filters known as convolutional kernels are present in each convolutional layer. The filter is a matrix of integers that are used on a subset of the input pixel values, the same size as the kernel. The important parameters involve in this layer are the number of kernels and the size of the kernels. Pooling layers on the other hand simply the information output from convolution layers and reduces the spatial size of the representation by decreasing the required amount of computation and weights. Thus, pooling reduces the dimensionality of the network by performing a specific function such as max pooling, which takes the maximum value in a certain filter region, or average pooling, which takes the average value in a filter region. In the classification part, fully connected layers are first organized in three dimensions; width, height and depth and then they serve as a classifier on top of these extracted features. It takes the output of the previous layers, "flattens" them and turns them into a single vector that can be an input for the next stage. They assign a probability for the object on the image and the final output is reduced to a single vector of probability scores, organized along the depth dimension. The final probabilities for each label are given by the Fully connected output layer.

B. RECURRENT NEURAL NETWORKS (RNN)

Recurrent neural networks (RNN) were based on David Rumelhart's work in 1986. John Hopfield in 1982 discovered called Hopfield networks. Recurrent Neural Networks (RNN) are neural networks with time varying behavior including the notion of dynamic change over time. RNNs are called recurrent or recursive because they perform the same task for every element of a sequence and the output depends upon the previous computations.

RNNs support processing of sequential data by using a looping mechanism allowing the information to persist and flow from one step to the next. There are three layers the input layer, the hidden layer and the output layer. The input layer takes in a sequential input loops through the input values and gives output. The Hidden layer contains the memory from previous iterations which is utilized for prediction or classification resulting through an output layer. Different kinds of RNNs include Deep Transition (DT) RNN, DT-RNN with shortcut connections, Deep Transition-Deep Output (DOT) RNN and Stacked RNN etc. [85], quasi-recurrent neural networks (QRNN) [86], hierarchical multiscale recurrent neural network (HM-RNN). RNNs are particularly useful in time dependent data where data must be handled in a sequential manner or processes change over time for example in Speech recognition, Machine translation, Image recognition, Music composition, Handwriting recognition, Stock predictions, Grammar learning and natural language processing [87]–[89]. There is also a recent attempt in [90] to employ Recurrent Neural Networks for Classifying Crash-related events and proposed a model for the detection of crash and near-crash events based on a large set of time-series data collected from naturalistic driving behavior.

1) Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) was proposed in 1997 by German researchers Sepp Hochreiter and Jurgen Schmidhuber as a solution to the vanishing gradient problem in RNN algorithm. LSTM is an artificial recurrent neural network (RNN) architecture that has feedback connections [91]. It can process single data points like images as well as the entire sequences of data) sequentially and keep its hidden state through time. A common LSTM unit architecture is composed of different memory blocks called cells (the memory part of the LSTM unit), and three "regulators", called gates (an input gate, an output gate and a forget gate). The cell remembers values over arbitrary time intervals. Similar to a computer's memory, Information can be stored in, written to, or read from a cell. The cell makes decisions about what to store, and when to allow reads, writes and erasures, via gates that open and close. There are two states that are being transferred to the next cell; the cell state and the hidden state. The three gates are responsible for regulating the flow of information inside the LSTM unit and outside the cell. The input gate is responsible for the addition of information to the cell state. The output gate is responsible for selecting useful information from the current cell state and showing it out as an output.The forget gate is responsible for removing information from the cell state that is no longer required for the LSTM to understand things or the information that is of less importance via multiplication of a filter. This process helps in optimizing the performance of the LSTM network.

Some variations of the LSTM unit do not have one or more of these gates or maybe have other gates. Some of these variations or types of LSTM include Gated Recurrent Units (GRUs), Multiplicative LSTM (mLSTM), Bidirectional LSTM (BiLSTM) with Attention mechanism etc.

Gated Recurrent Unit (GRUs) is a variant of LSTM with forget gate that was introduced in 2014 by K. Cho et al [92]. It lacks an output gate and has fewer parameters than classic LSTM. Therefore, GRU fully writes the contents from its memory cell to the larger net at each time step. GRUs only have one hidden state responsible for holding both the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

long-term and short-term dependencies at the same time due to the gating mechanisms and computations that the hidden state and input data go through. It has two gates, a reset gate and update gate to solve the vanishing gradient problem. The update gate controls information that flows into memory, and the reset gate controls the information that flows out of memory. These gates are trained to selectively filter out any irrelevant information while keeping what's useful. For proposing a new hidden state, the Reset gate decides which portions of the previous hidden state are to be combined with the current input. The Update gate determines how much of the previous hidden state is to be retained and what portion of the new proposed hidden state that is derived from the Reset gate) is to be added to the final hidden state.

Multiplicative LSTM (mLSTM) is a variant of LSTM that was introduced by Krause et al in 2016. It combines the long short-term memory (LSTM) and multiplicative recurrent neural network architectures; the input, output and forget gates that provide continuous analogues of write, read and reset operations for the cells. It is demonstrated in [93] that mLSTM outperforms standard LSTM and its deep variants for a range of character level language modelling tasks.

Bidirectional LSTM (BiLSTM) involves duplicating the first recurrent layer in the network so that there are now two layers side-by-side, then providing the input sequence as-is as input to the first layer and providing a reversed copy of the input sequence to the second. BiLSTM has two networks, one access past information in forward direction and another access future in the reverse direction.

Another most influential idea in the Deep Learning community is called Attention mechanism, which is used in various problems like neural machine translation, human action recognition [94] and so on. The attention mechanism can focus on discriminative features in a longer sequence, which can be used in many difficult tasks. Bidirectional LSTM with attention mechanism combines birdirectional LSTM with attention network [95].

C. DEEP BELIEF NETWORK (DBN)

Deep belief network (DBN) is a multilayered probabilistic generative graphical model is composed of multiple layers of stochastic, latent variables and can learn to probabilistically reconstruct its inputs. The latent variables are often called hidden units or feature detectors that typically have binary values. The top two layers that have undirected, symmetric connections between them. The lower layers receive top-down, directed connections from the layer above. The states of the units in the lowest layer represent a data vector. After learning the top-down through layer by layer procedure, the values of the latent variables in every layer can be inferred by a single, bottom-up pass that starts with an observed data vector in the bottom layer and uses the generative weights in the reverse direction.

When binary stochastic neurons are connected in a directed acyclic graph it is called a Sigmoid Belief Net (Radford Neal 1992). When binary stochastic neurons are connected by using symmetric connections, it is called a Boltzmann Machine (Hinton & Sejnowski, 1983). RBM is a two-layer stochastic network including visible layer V and hidden layer h. The hidden or invisible layers are not connected to each other and are conditionally independent. DBNs integrate Restricted Boltzmann Machines (RBMs) with Deep Feed Forward Neural Networks (D-FFNN).

RBMs (Restricted Boltzmann Machine) are stacked together to form a deep belief network and training is provided using the greedy layer wise method. The two phases of DBN training are Pre-training Phase and Fine-tuning Phase. For the first phase, an unsupervised learning approach is used. In the first step, a layer of properties is trained by inputting the original data and fixing up the parameters and the input signals are obtained from the pixels directly. The output is then used as the input of the second RBM and the rest is done in the same manner. At last a DBN with several layers is created whose parameters are suitable to extract the features of this kinds of data. For the second phase, a supervised learning approach is used. In the second step, a suitable classifier is added (such as back propagation algorithm) to the end of the DBN to to provide fine tuning of whole neural network [96], [97].

V. DETECTING HUMAN DRIVER INATTENTIVE AND AGGRESSIVE DRIVING BEHAVIOR USING DEEP LEARNING

In this section, we comprehensively review deep learning based methods and techniques in the literature for detecting Human Driver inattentive driving behavior including Distraction, Fatigue/Drowsiness and Aggressiveness according to our driver inattentive driving behavior (HIADB) classification done in Section 2.

A. COMPARATIVE STUDY OF DRIVER DISTRACTION DETECTION TECHNIQUES (DDDT) BASED ON DEEP LEARNING

The focus of this subsection is the review of the latest research attempts in detecting human driver distraction using deep learning algorithms. The comparative analysis of detection techniques for human driver distraction is presented in Table 1.

A. Behera et al. [98] have proposed a novel Multi-Stream Long Short-Term Memory (M-LSTM) net-work for recognizing fine grained driver distracted activities that are difficult to distinguish. M-LSTM integrates the concepts of LSTM and CNN for recognizing driver distraction activities like texting, talking over phone, eating and drinking. M-LSTM network is evaluated from one stream up-to four streams. The proposed architecture has three components Transferable deep CNN features, contextual cues involving body pose and body-object interaction and the proposed Multi-stream LSTM (M-LSTM) for sequence modeling and activity recognition. The authors have proposed Contextual descriptors to represent high-level knowledge involving human pose and human-object interactions and relationships between body parts and objects (cup, bottle, phone). Body-pose descriptor translates the body parts configuration to a feature vector by encoding relationships between various body parts. Body-object descriptor captures the pairwise relationship between the body joints and involved objects, encoding the relative position of an object with respect to a given joint in a scene. 70% of the dataset is used for training and the rest 30% for validation. The inputs are per-frame appearance and contextual features which are based on transfer Learning (TL) i.e. CNN features and object and body parts detectors are not trained/fine-tuned on the target dataset. The two LSTM layers have been used for capturing the sequential information in the M-LSTM. The first layer is in individual stream and the second layer is after the fusion. The authors have used unseen drivers for testing in their experiments. The proposed network is flexible to accommodate more input streams depending on the target application. Moreover, it is light-weight and can be trained using CPU.

H. Eraqi et al. [99] have proposed a reliable deep learning-based solution for the detection and identification of driver distraction. They have obtained RGB images from a camera mounted above the dashboard and trained multiple convolution neural network architectures on raw images, skin-segmented images, face images, hands images, and face + hands images. Finally they have evaluated a weighted sum of all network's outputs, and final class distribution is done using a genetic algorithm.

Munif Alotaibia and Bandar Alotaibi [100] have developed a system for the detection of distracted driver posture and they have enhanced the performance of detecting the distracted behaviors of drivers by using a combination of deep learning modules, the inception module with a residual block and a hierarchical recurrent neural network (HRNN). They have used 10several percentages to divide the data into testing and training sets like they have used 10 They have calculated the overall accuracy as the number of all images that are classified correctly divided by the total number of all samples in the test set. The proposed method is also applied to our method to the AUC distracted driver data-set in which the authors have used 75% of the data set for training and the remainder for testing. The authors have compared their proposed method with ResNet whose accuracy was 95.31% and they have concluded that the larger architectures such as Xception, Inception, VGG and ResNet50 can not be optimized easily on the State Farm Distracted Driver data set.

J. Celaya-Padilla et al. [101] have proposed a novel ubiquitous oriented methodology to detect distracted drivers who are using cellphones. The authors have mounted a wide-angle camera on the roof to compile a video of the driver and then each video is is split into 24 pictures. The final images are then used to feed a CNN algorithm in order to train it to accurately detect driver's distraction. The authors have implemented the pre-trained Google CNN architecture Inception v3 in this research work. In this work, the authors have chosen Inception CNN was chosen because CNN can be exported for low cost hardware such as Raspberry Pi and can be deployed in Android platforms. Moreover, the CNN architecture involves multiple convolution filters, edges are detected by the first layer, and the over all design is tackled by the second layer. After applying the filters on the data set, the results were then concatenated and passed forward.

Yang Xing et al. [33] have designed a driver activities recognition system based on the deep convolution neural networks (CNN) to detect whether the driver is distracted or not. The authors have identified seven common driving activities among them three of them are distracted driver activities. They have collected raw RGB images by using the low cost Kinect camera that enables the collection of multi-modal signals, such as the color image, depth image, and audio signals. For the naturalistic data collection, ten drivers are involved. Then, the images are cropped and Gaussian mixture model (GMM) algorithm is used for the segmentation that extracts the driver body from the background. Finally, the segmented images are passed to the CNN models for training and testing. Three pre-trained CNN models namely AlexNet, GoogLeNet, and ResNet50 were adopted and evaluated. For reducing the cost, transfer learning method is applied to fine tune the pre-trained CNN models. In this study no temporal information is considered, and each image is processed individually.

W. Huang et al. [102] have proposed three video-based behavior detection techniques for abnormal driving using three deep learning-based fusion model. The proposed models WGD (wide group densely network), WGRD (wide group residual densely network) and AWGRD (alternative wide group residual densely network) are motivated by DenseNet which is based on the densely connected convolutional network. To enhance the width, cardinality and generalization efficiency in WGD model without increasing the number of parameters, the group, and wide convolution replace the DenseNet's conventional convolution. To improve the learning efficiency of WGRD over WGD and DenseNet models, additional complex features, superposition of previous layers along with residual networks are employed. In AWGRD model, the training efficiency of WGRD is improved by considering the superposition of $(l-1)$ previous layers. On the other hand, in WGRD superposition of all previous layers are considered.

C. Zhang et al. [103] have designed a dedicated Interwoven Deep Convolutional Neural Network (InterCNN) architecture that uses all four-dimensional (time, height, width, RGB channels) multi-stream inputs including side video streams, side optical flows, front video streams, and front optical flows for accurate classification of driver behaviors in real-time. The authors have build a mock-up environment to emulate self-driving car conditions and in order to record the body movements and facial expressions side and front facing cameras were deployed. a total of 50 drivers (72 male and 38 female,) participated in the experiments. The hierarchical InterCNN has two components involving two simpler architectures, first is the plain CNN, which uses only the side video stream as input an second is the two-stream

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

CNN (TS-CNN) which takes the side video stream and the side optical flow as input. The entire dataset is divided into a training set (30 videos), a validation set (10 videos) and a test set (10 videos).

B. COMPARATIVE STUDY OF DRIVER FATIGUE OR DROWSINESS DETECTION TECHNIQUES (DFDDT) BASED ON DEEP LEARNING

The focus of this subsection is the review of the latest research attempts in detecting human driver fatigue or drowsiness using deep learning algorithms. The comparative analysis of detection techniques for human driver fatigue or drowsiness is presented in Table 2.

Y. Ed-doughmi et al. [104] have proposed an RNN (recurrent neural network) based driver drowsiness detection technique for road safety. The LSTM (long short-term memory) algorithm using a video public dataset is employed for the training and validation of the proposed driver drowsiness technique. The dataset is split into 7s short scenes for optimal learning and the detection of drowsiness using supervised auto-learning calculations. After data preparations, the LSTM algorithm is applied for training and prediction of the driver's drowsiness. The experimental results show that the proposed technique achieved 92.71% accuracy. However, the proposed technique requires a change in the posture of the driver after a period of time as it classifies the actions of the drivers based upon its different postures. If the diver remains still or unable to change his/her posture over a long period of time, then the proposed technique will fail to detect the drowsiness.

Z. Xiao et al. [105] have proposed driver fatigue detection method using driver's eyes blinking duration and sequences. In the proposed detection method, for training and testing of driver's eyes sequence through images the CNN with LSTM is employed. The eyes regions extracted from videos using multi-task framework based on deep cascading and using deep CNN and LSTM spatial features of driver's eyes are learned. Finally, the driver fatigue is detected using the duration and sequences of his/her eyes. The experimental results shows that the proposed technique achieved high detection rate for the detection of driver's fatigue. However, the proposed method is evaluated using author's private dataset which limits the applicability and achieving claimed high accuracy results of the proposed methods with other available dataset (captured under various real driving condition and driving scenarios) and real-time captured videos.

W. Liu et al. [106] have proposed a driver fatigue detection algorithm using two-stream network models with multi-facial features. The proposed algorithm is comprised of four parts. In the first part, mouth and eye Positioning is done with multi-task cascaded convolution neural networks (MTCNNs). In the second part, the static features are extracted from a partial facial image. in the third part, the dynamic features are extracted from a partial facial optical flow. in the fourth part, both static and dynamic features are combined using a two-stream neural network to make the classification. MTCNN consists of three network architectures, Proposal Network (P-Net), Refine Network (R-Net), and Output network (O-Net). P-Net structure was responsible for obtaining the regression vector of the candidate window and bounding box in the face area. R-Net structure was responsible for removing the false positive region by employing bounding box regression and non maximum suppression. O-net structure was used to make result of processing finer by using one more convolution layer than in R-Net. O-Net has similar working as R-Net but it supervised the face area and five obtained coordinates including the left eye, the right eye, the nose, the left part of the lip, and the right part of the lip. O-Net was responsible for face classification and facial landmark localization. The proposed algorithm first performed face detection of the driver and then intercepted the left eye area and the mouth area into the fatigue detection network, combined with the optical flow image of the left eye and mouth. The drivers were detected whether they were in a normal, speaking, yawning or dozing state.

M. Tanveer et al. [107] have proposed a driver-drowsiness detection model using deep learning techniques for BCI (brain-computer interface). The proposed detection model is evaluated through fNIRS (functional near-infrared spectroscopy). The driver's passive brain signals were captured using a car simulator and the fNIRS system is used to measure the brain signal in a continuous wave. For the classification of drowsiness and active state of the driver, DNN (Deep neural networks) were utilized and CNN (convolutional neural networks) were used for training and testing of the proposed model. The experimental results show that for differentiate between drowsiness and active state of the driver the proposed model achieved 99.3% accuracy.

C. COMPARATIVE STUDY OF DRIVER AGGRESSIVENESS DETECTION TECHNIQUES (DADT) BASED ON DEEP LEARNING

The focus of this subsection is the review of the latest research attempts in detecting human driver aggressiveness using deep learning algorithms. The comparative analysis of driver aggressiveness is presented in Table 3.

Y. Moukafih et al. [108] have proposed a driver aggressive behavior detection scheme using two deep learning algorithms LTSM (Long Short Term Memory) and FCN (Fully Convolutional Network). For the classification of driver's behavior, testing and validation of the proposed technique, a public dataset "UAH-DriveSet" is utilized. Various vehicles and environmental features (e.g., vehicle position on the road, distance from the vehicle ahead, acceleration and speed) are used for the detection of aggressive behavior of the driver. The results show that the proposed technique performs well as compared to other existing techniques in terms of processing time window size. However, for larger window size (i.e., $>$ 5 minutes), the performance of the proposed technique decreases.

M. Matousek et al. [109] have proposed a neural network-based aggressive driver behavior detection technique. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

TABLE 1. Comparative study of Driver Distraction Detection Techniques (DDDT)

Study	Category of Distraction	Detected Driver Behavior	Measure	Tools and Architectures	Classifiers	Plateform	Dataset	Efficiency
[98]	Manual	Texting, talking over phone, eating and drinking	Face, hands and skin detection	VGG16, Linux PC (Intel i7-5930K, 12 cores, 3.5 GHZ) with NVIDIA Quadro P6000 24GB GPU	Multi-stream LSTM and CNN	Simulations	COCO dataset State Farm dataset Distracted Driver dataset. Part Affinity Fields (PAFs)	91.25%
[99]	Visual and Manual	Using mobile phone	Face, hands and skin detection	ASUS ZenFone smartphone (Model ZD551KL) rear camera, DS325 Sony DepthSense camera 5 AlexNet and 5 InceptionV3)	CNN	-	Self-created dataset	90%
[100]	Manual	texting, talking on the phone, operating the radio, drinking, reaching behind, fixing hair and makeup, and talking to the passenger	Posture	a computer with an Intel core 16.0 GB of RAM and a 64-bit Windows 10 OS. ResNet	Combination of HRNN and LSTM	-	State Farm, AUC	96.23%, 92.36%
[101]	Manual	Texting on mobile phone	Hand movements	GoPro Hero 5 camera, Inception V3, Python , Google TensorFlow	CNN	-	a self -created blind dataset	92.8%
[33]	Manual	using in-vehicle radio device, texting, and answering the mobile phone	Hand movement	Kinect camera, Intel Core i7 2.5GHzCPU, C++ based on the Windows Kinect SDK and OpenCV. NVIDIA MX150 2GB GPU	CNN	Real Vehicle	Self -collected dataset	91.4%
[102]	Mnaual	Texting (using right, left hand), talking on the phone (using right, left hand), operating the radio, drinking, reaching behind, hair and makeup, talking to passenger	-	a workstation equipped with Intel Xeon Silver 4110 CPU, 128G RAM, CentOS 7 and PyTorch 1.0.0, Nvidia Titan V GPU card	Fusion models based -Wide group densely (WGD) network, Wide group residual densely (WGRD) network, Alternative wide group residual densely (AWGRD) network	Real vehicle	Kaggle state farm distracted driver detection database	-
[103]	Manual	Texting, eating, drinking, searching, talking, watching, gaming, preparing	-	OpenCV vision library, Python, 1-2 NVIDIA TITAN X and Tesla K40M GPUs	a plain CNN, two-stream CNN (TS-CNN)	Mock-up car cockpit environment	Self-collected dataset	81.66%

proposed technique employed LSTM and RNNs (auto-encoding Replicator Neural Networks) for driver behavior detection using pedals, steering wheels and most recent history of the vehicle. SUMO (Simulation of Urban Mobility) is used for dataset generation. After the detection of driver aggressive behavior, the diver assistant tries to reduce the aggressiveness of the driver and using C2X communication alert the surrounding vehicles for the possible incident. The simulation results show that the proposed technique performed better than the previously proposed techniques. However, to achieve high and reliable detection of driver behavior the proposed technique heavily dependent upon the post-processing of the recent history of the vehicle.

Y. Xing et al. [110] have proposed a joint time-series modeling approach called personalized joint time series modeling (JTSM) for predicting leading vehicle trajectory while considering different driving styles. JTSM is based on the Long Short-Term Memory (LSTM) Recurrent Neural Network model (RNN) which contains a common LSTM layer and different fully connected regression layers for three different driving styles, namely moderate, conservative and aggressive. The proposed method has following three parts, Next Generation Simulation (NGSIM) data processing, Gaussian Mixture Model (GMM)-based driving style recognition, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

TABLE 2. Comparative study of Driver Fatigue or Drowsiness Detection Techniques (DFDDT)

Study	Category	Detected Driver Behavior	Measure	Tools and Architectures	Classifiers	Dataset	Efficiency
[104]	Fatigue	Driver visual based features	Yawning, nodding, slow blink rate	PC with Alienware R17, Ubuntu 16.04 LTC, 8G GPU, and 16G RAM. Tensor-Flow	LSTM	-	92.7%
[105]	Fatigue	Driver visual based features	spatial-temporal feature of eyes	Infrared camera, MTCNN	CNN + LSTM	specialized dataset named TJPU-FDD	95.83%
[106]	Fatigue	Driver visual based features	Optical flow of mouth and the left eye area	Keras framework, MTCNN, GTX 1080 Ti	CNN	National Tsing Hua University Driver Drowsiness Detection dataset (NTHU-DDD)	97.06%
[107]	Drowsiness	Driver's physiological features	Brain waves	Optodes, wave-imaging system (DYNOT, NIRx, Medical Technologies	DNN and CNN	-	99.3%

representative feature evaluation for the driving styles with the Maximal Information Coefficient (MIC) method. In the first part, NGSIM vehicle trajectory data were collected from a different region at a different time, which can reflect congested and moderate traffic conditions, and then a data preprocessing is applied to clean the raw data set. In the second part, GMM unsupervised clustering method was applied to generate the most distinctive driving styles for the connected vehicles. A specific driving style is generated by GMM for each vehicle based on the speed, acceleration, jerk, time, and space headway features of the leading vehicle. In the third part, the MIC algorithm is used to analyze the mutual dependence between the GMM learned driving styles and the hand-crafted feature vector. The JTSM model is trained based on the following-leading pairs extracted from the NGSIM data set. The proposed approach JTSM was tested for making predictions one to five seconds ahead that the results of evaluation indicated significant advantage of JTSM over the baseline algorithms, Constant Kalman Filter (CKF), LSTM, and Multiple LSTM and achieved more precise prediction. The authors in [111] have estimated future energy consumption and the speed of a vehicle by proposing a personalized energy consumption analysis and prediction framework. They have deigned a personalized joint time series modeling system based on the long short-term memory by considering different driving styles. They have calculated the energy consumption for highway vehicles based on the energy required at the wheel function. Then they have classified the different energy consumption levels for the car and truck users based on the energy consumption map. Finally, they have found that heavy energy consumption users have distinctive driving behaviors that involve a higher speed, larger acceleration, more headway space, and less time headway.

C. Lv et al. [112] have proposed a CPS-based co-design optimization framework for an automated electrical vehicle (EV) considering different driving styles based on Platform Based Design (PBD) methodology. The authors have developed a driving style recognition algorithm using unsupervised learning method. The authors have synthesized vehicle control algorithms for typical driving styles, moderate, aggressive and conservative, with different protocol selections. The requirements for vehicle design and control involve dynamical performance, energy efficiency, and ride comfort. In order to capture driver's behavior, Maximum speed and acceleration time were used as proxies for indicating dynamic performance of the driver. The comfort level of a vehicle called drivability, was assessed by vehicle's jerk j, which is the second derivative of the vehicle's longitudinal velocity. The energy efficiency of a vehicle was represented by the energy consumed during a certain trip. The simulations were implemented iteratively with developed models under defined driving events at each operating point for the three driving styles.

VI. REQUIREMENTS FOR THE DETECTION OF HUMAN DRIVER INATTENTIVE BEHAVIOR

In this section we have highlighted the enabling requirements for the Detection of Human Driver Inattentive Driving Behavior (HIADB), including driver Distraction, Fatigue/Drowsiness and Aggressiveness.

A. ACCURATE AND RELIABLE DETECTION OF HUMAN BEHAVIOR

The successful avoidance of road accidents is highly dependant upon the accurate and reliable detection and prediction of human driver behavior. So it is a fundamental requirement to accurately and reliably detect and predict human abnormal driving behaviors like distraction, fatigue, drowsiness and aggressiveness. The human driver behavior detection system is comprised of two parts, first, the equipment or hardware part which includes video cameras, and various vehicle sensors (GPS, accelerometer, Gyroscope etc) for monitoring, braking, location etc., and second, the software part which includes the feature extraction algorithms and classifiers for classifying human driver behavior like distraction, fatigue, drowsiness and aggressiveness. In order to accurately and reliably detect and predict such human behaviors while driving which may lead to severe accidents on the road, both

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

TABLE 3. Comparative study of Driver Aggressiveness Detection Techniques (DADT)

Study	Type of Parameter	Measure	Classifiers	Dataset	Efficiency
[108]	Vehicle state and environment	Speed, Acceleration, Car position relative to lane center, time of impact to ahead vehicle, Car angle relative to lane curvature, Road width	FCN-LSTM	UAH-DriveSet	95.8% (with 5 min window size)
[109]	Vehicle state	Average speed, Speeding, Acceleration, Number of lane changes, Minimum gap	Auto-Encoding Replicator Neural Networks (RNNs) and (LSTM)	Luxembourg SUMO Traffic (LuST) scenario	93%
[110]	Vehicle state	Speed, Lateral and longitudinal position, Velocity, Acceleration, Jerk, Space headway, Time headway	LSTM-	I-80 and US-101 freeway datasets	-
[112]	Vehicle state	Throttle pedal position, Brake light switch, Longitudinal and lateral accelerations, Steering wheel angle and Vehicle speed	GMM	Self collected data using Sedan/ SUV type vehicles	Improve efficiency by 10% as compared to other related studies.

the integral parts of the detection system are required to be working accurately and reliably.

B. EFFICIENT COLLABORATION BETWEEN MULTI-SUBSYSTEMS

In order to generate a complete human driver behavior profile, for accurate detection of human driver distraction, drowsiness, Fatigue and aggressiveness, multiple sub systems work together. For drowsiness/fatigue detection of human driver, driver's eye state, facial expressions, driving posture, and vehicle related features like speed, acceleration, steering wheel angle, brake pedal etc., are required to analyze. For driver distraction, driver's manual distraction like talking on the phone, eating and drinking and other distractions like cognitive or auditory etc., are required to be analyzed. Similarly, for detecting aggressiveness detection of a human driver, different vehicle related features like risky speedy, sudden acceleration and de-acceleration, throttle pedal, steering wheel angle and driver's maneuvers like, traffic light violence, lane changing, car following etc., are needed to be analyzed. The effective coordination between these subsystems is fundamentally required, so that human abnormal driving behaviour could be efficiently and accurately detected.

C. SECURE AND RELIABLE CONNECTIONS

Autonomous, connected driving and driver assistance systems utilize the power of internet for getting real-time information about road conditions, accidents, traffic, and current weather. The data that the connected vehicle can gather regarding the objects on the road, speed bumps, and other vehicles can benefit the human driver in avoiding traffic jams and possible accidents making controlling the vehicle and parking much easier and even contact the emergency services in the event of an accident. Besides all these benefits, connected cars are vulnerable to cyber-security attacks and leakage of personal data due to security breaches can result in vehicle theft. Thus, vehicle-to-vehicle communication, vehicle-to-infrastructure communication and cloud data are needed to be more secure in internet connected vehicles. Additional protection including Advanced hardware firewalls, and incorporated network-level security elements are also required to tackle the security and privacy issues of interconnected vehicle systems that are capable enough to detect human abnormal driving behaviour like fatigue and drowsiness.

D. FAST AND REAL-TIME EXECUTION OF DRIVING ACTIONS

The human driver's actions (Distraction, Aggressive Driving) and Behavior (Fatigue and drowsiness) which are the major cause of road accidents must be accurately detected. After detection, the systems like advanced driver assistant systems (ADAS) assist the driver according to their level of automation regarding controlling the vehicle, obstacle detection, location finding, collision avoidance etc. The level 1 ADAS systems acquire information from the sensors, present the relevant information to the driver and aid the drivers by enhancing their perception about the driving environment but do not provide warning alerts. The level 2 ADAS systems assess the criticality of hazards after acquiring the information and then provide passive mitigation of hazardous through warnings. The level 3 ADAS systems are a bit more advanced and complex systems. They provide assistance in controlling the vehicle and actively mitigate the hazards through intervening on their own by using evasive measures e.g., adjusting the speed by applying the brakes for the headway within a certain threshold. Besides all these technology advancements and important enhancements in such systems, there is much room for the improvement and low cost camera based solutions that can be fully integrated with all the vehicle modules are needed which are capable enough to more accurately predict all the possible human behavior like aggressive driving and various situations like when a collision between two cars appears imminent.

E. TRUST MODELS FOR INTELLIGENT VEHICLE TECHNOLOGIES

In order to utilize the benefits of intelligent vehicle technologies like efficiency, safety, and flexibility, the exchange of information and data between different road users and elements of the infrastructure is a necessity. This represents the higher

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

possibilities of security and privacy breaches (combination of physical and digital threats) due to lack of security updates and inadequate privacy protection which degrades the user (human driver) trust upon intelligent vehicles and thus, it is the major obstacle in the public acceptance. Most recently [113], [114] the researchers identified trust as most influential factor for the acceptance of any new technology. Hence, it is required to comprehensively address the factors (diversity) through which the adoption of intelligent vehicular technology can be promoted and impeded. In order to increase driver acceptance to the output methods, the researchers need to find ways to provide assistance to drivers (elderly or young) in a way that is appropriate and that would not annoy them. Moreover, it is essential for the developers to investigate the technology acceptance during the development phase. The researchers and developers need to understand the perceptions of human drivers about intelligent vehicular technologies that can detect their abnormal driving behaviours.

VII. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Detection of Human driver Inattentive Driving Behavior (HIADB) provides useful and effective insights for preventing road accidents and saving precious human lives. However, there are several challenges that need to be addressed in order to detect and predict the HIADB successfully. In this section, we discuss some major challenges faced by HIADB techniques.

A. ADVERSARIAL ATTACKS

The core purpose of HIADB techniques is an accurate classification between normal and abnormal driving behaviors. Most of the HIADB techniques are based on images or videos collected using a camera that is usually mounted on the car dashboard. These images and videos are used as input by HIADB deep learning algorithm for training, testing and finally detection/ prediction purposes. However, deep learning-based HIADB techniques are vulnerable to the most sophisticated and dangerous attacks known as adversarial attacks. In a survey [115] the authors focused on the adversarial examples of deep learning models and reviewed current research efforts on attacking various deep neural networks in different applications. Another attempt in [116] presented a comprehensive survey on adversarial attacks on deep learning in Computer Vision. Another most recent survey [117] conducted a comprehensive review for adversarial attacks on textual deep neural models and proposed different classification schemes to organize the reviewed literature.

In adversarial attacks, even a minor alteration into the input images lead to completely wrong results with a high accuracy rate in deep learning-based algorithms. In the presence of adversarial attacks, accurate and reliable detection or prediction of human driving behavior is a challenging task. Therefore, practical solutions and methods are needed to successfully tackle the adversarial attacks so that HIADB detection can be accurately done using a deep learning algorithms.

B. REAL-TIME PROCESSING OF HETEROGENEOUS DATA

The creation of a complete human driving behavior profile through deep learning algorithms using real-time processing of heterogeneous data is a challenging task because of a large set of output classes and large number of free parameters. Processing of heterogeneous data is very important to avoid accidents and it entails the processing of huge amount of various types of real-time data generated by different subsystems. The driving data can be unstructured, semi structured or structured data. For example, for the detection of various driving behaviors (fatigue, drowsiness, distraction and aggressiveness) data from video cameras, various sensors, other near by vehicles and road side infrastructures (RSUs) are required for processing and after successful detection of driving behavior, a fast-appropriate action (e.g., generating alert for the human driver, sending alerts to other nearby cars, slowing down the car automatically, call emergency response agencies in case of critical error or accident etc.) is also needed. Such a great varieties of data require a deep learning algorithm to identify hidden relationships among heterogeneous data. Also training a deep learning network with centralized architectures is tedious, so distributed architectures are favorable.

The processing of various types of heterogeneous data (specially images and video data captured under different lighting conditions or images of drivers wearing sunglasses) and involvement of many subsystem inputs (location, speed, breaking data under various types of roads or mountains), traffic, locations (in tunnels, covered and under bridges), weather conditions might result in processing delays. Thus, efficiency can be achieved without compromising accuracy by a collection CPUs and GPUs that can enhance the speed of training. Another common challenge in this regard is that the deep learning process cannot resolve conflicts of information [118]. Deep learning solutions for the detection of driver distraction, fatigue/drowsiness and aggressiveness frequently process data coming from a single sensor modality, most of times cameras, and in some cases process data from various modalities, for example LiDAR and camera. However, there is still a lot of uncertainty on how to process and combine data from heterogeneous sensors in the best way to detect different kinds of human driving behaviors [119].

C. COMPLICATED AND DIVERSE REAL-LIFE DRIVING SCENARIOS

Detection of various human driving behaviors depends upon the driving styles of the driver and real-time driving scenarios. Modeling a generic driving behavior using deep learning algorithms is a very complicated and challenging task as every human has different driving styles and driving behaviors under various driving scenarios. In real-life there are many factors which can affect or influence the human behavior while driving. These factors include multi-agent environment (many participants), diver age (senior/young and experienced/inexperienced drivers), road condition, traffic conditions (congestion, peak and low traffic hours), driving under trauma (emotionally disturbance due to sad or happy incidents), driving styles (Left hand and right-hand driving), lack of awareness with driving rules and finally people with different fiscal appearance (e.g., small or big eyes, habitual fast eyes blinking and yawning etc. Moreover, the development of a one-for-all solution to detect and predict human driving behavior through only deep learning algorithm alone may not be appropriate or effective solution. Therefore, more research efforts are required to overcome the challenges of complicated, diverse real-life driving styles, behaviors and scenarios effectively through integration and collaboration of various subsystems available for monitoring and predicting the driveling behavior, vehicle functioning and performance.

D. ACCEPTANCE OF INTELLIGENT BEHAVIOUR TECHNOLOGIES BY DRIVERS

A complex socio-technical innovation of Autonomous driving has profound potential impact on our society and economy, having massive benefits like it gives access to mobility to people who are disabled, drastically reduces the number of accidents, deaths and injuries; reduces traffic congestion and pollution and boosts the economy. Besides all these advantages, a key barrier to its adoption is the public trust in driver-less cars and unfortunately, due to lack of trust, the general public doesn't appear to be ready to consume Autonomous Vehicle (AV) technology [120]. According to AAA's survey [121], there is an increase from 63% (2017) to 71% (2019) Americans who claim to be afraid to ride in a self-driving car. One major reason about such a negative trend of people towards AV is that the AV industry has had its setbacks; For example, the most recent one of Uber's fatal self-driving car accident in March 2018 that has damaged the public opinion about AVs.

The usage and success of new technologies is dependent upon the acceptance of its consumers. Many factors affect and influence the consumer behaviour, which makes the consumer acceptance complicated and challenging for new technologies. The detection of HIADB is a part of an intelligent vehicle system or advance driver assisted system (ADAS) which is also facing the challenge of acceptance by its users due-to many factors. These factors include the reliability of driver-less cars, the security of ADAS (from hackers), deep learning based HIADB detection systems, sensed data, hardware, communication, privacy of human driver and its credentials (location tracking and surveillance), and finally, the accuracy of deep learning algorithms under adversarial attacks. Lack of strong security and user privacy in such critical systems may lead to lack of trust upon these systems and ultimately results in low acceptance from its users and hampering its development. Moreover, acceptance models for HIADB detection are not available and so far no research study is conducted for the exploration of user acceptance. Therefore, for the successful usage and acceptance of HIADB detection systems more research efforts are required which is a major challenge for the researchers and developers.

VIII. CONCLUSION

Road accidents are a global scourge in which Human driving behavior is an important factor, affecting road safety that ultimately leads to loss of human lives. Safe driving behavior requires human driver to be alert and attentive while making fast cognitive decisions in a dynamically changing road environment. In order to improve road safety and avoid severe road accidents, there is a need for monitoring, detecting and early prediction of human abnormal, Inattentive and aggressive driving behavior. In this study, we classified and discussed the human driver's Inattentive driving behavior (HIDB) into two major categories; Distraction and Fatigue/Drowsiness. The detection of driver distraction and driver fatigue/drowsiness was classified according to the features selected for the detection of human driving behavior in the literature. Aggressive driving behavior being another more risky human driving behavior was also explained and discussed, high-lighting the causes and effects of different aggressive driving styles on human safe driving. The detection of human driver aggressive driving behavior (HADB) was classified according to the aggressive driving styles adopted by aggressive drivers. The most recent deep learning based solutions for human driver Inattentive and Aggressive Driving Behavior (HIADB) detection were reviewed systematically and comprehensive comparative analysis (quantitatively and qualitatively) was performed, highlighting their detection accuracies. The enabling and fundamental requirements for detecting human abnormal and inattentive driving behavior from developing trust of ADAS systems to including trust, security and privacy of such automated systems were presented and discussed. In the end, the imperative open research challenges in the field of HIADB detection were identified. These challenges are discussed as future research directions that need to be addressed for the accurate detection of abnormal human Aggressiveness.

We conclude that HIDB and HIADB can be efficiently detected and accurately assessed by using multiple sources of information. So far, deep learning innovative technologies have shown the potential to extend advanced strategies and methods in detecting and predicting abnormal, inattentive and aggressive driving behavior of a human driver and thus many attempts have been made to detect, predict and classify these behaviors while driving. However, low cost solutions are still research targets and high performance can be achieved by utilizing and employing other deep learning models and strategies like deep reinforcement learning and Q-learning. Moreover, considering the fact that every driver drives in a risky manner, either more often or rarely, we suggest that there is a need to provide drivers with incentives for improving their abnormal, inattentive and aggressive driving behaviors.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Saudi Arabia, under grant No. UJ-04-18-ICP. The authors, therefore, acknowledge with thanks the University technical and financial supports.

REFERENCES

[1] ASIRT, "Road safety facts," 2019. [Online]. Available: https://www.asirt.org/safe-travel/road-safety-facts/
[2] WHO, "Global status report on road safety 2018," 2018. [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
[3] S. M. Kouchak and A. Gaffar, "Estimating the driver status using long short term memory," in International Cross-Domain Conference for Machine Learning and Knowledge Extraction, pp. 67–77. Springer, 2019.
[4] R. Ferris and L. Koldny, "Feds to investigate tesla crash driver blamed on autopilot," 2018. [Online]. Available: https://www.cnbc.com/2018/01/23/tesla-on-autopilot-crashes-into-fire-truck-on-california-freeway.html
[5] I. Barabás, A. Todoruţ, N. Cordoş, and A. Molea, "Current challenges in autonomous driving," in IOP Conference Series: Materials Science and Engineering, vol. 252, no. 1, p. 012096. IOP Publishing, 2017.
[6] A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. R. Weippl, Machine Learning and Knowledge Extraction: Third IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2019, Canterbury, UK, August 26-29, 2019, Proceedings, vol. 11713. Springer Nature, 2019.
[7] S. O.-R. A. V. S. Committee et al., "Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," SAE Standard J, vol. 3016, pp. 1–16, 2014.
[8] S. O.-R. A. V. S. Committee et al., "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International: Warrendale, PA, USA, 2018.
[9] A. Eriksson and N. A. Stanton, "Takeover time in highly automated vehicles: noncritical transitions to and from manual control," Human factors, vol. 59, no. 4, pp. 689–705, 2017.
[10] G. P. Band, G. Borghini, K. Brookhuis, and B. Mehler, "Psychophysiological contributions to traffic safety," Frontiers in human neuroscience, vol. 13, 2019.
[11] A. P. Van den Beukel, M. C. van der Voort, and A. O. Eger, "Supporting the changing driverâs task: Exploration of interface designs for supervision and intervention in automated driving," Transportation research part F: traffic psychology and behaviour, vol. 43, pp. 279–301, 2016.
[12] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," Human factors, vol. 52, no. 3, pp. 381–410, 2010.
[13] J. Y. Chen, M. J. Barnes, and M. Harper-Sciarini, "Supervisory control of multiple robots: Human-performance issues and user-interface design," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 4, pp. 435–454, 2010.
[14] N. Merat and J. D. Lee, "Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind," Human factors, vol. 54, no. 5, pp. 681–686, 2012.
[15] C. D. Cabrall, A. Eriksson, F. Dreger, R. Happee, and J. de Winter, "How to keep drivers engaged while supervising driving automation? a literature survey and categorisation of six solution areas," Theoretical issues in ergonomics science, vol. 20, no. 3, pp. 332–365, 2019.
[16] V. A. Banks, N. A. Stanton, and C. Harvey, "Sub-systems on the road to vehicle automation: Hands and feet free but not âmindâ free driving," Safety science, vol. 62, pp. 505–514, 2014.
[17] V. A. Banks, A. Eriksson, J. O'Donoghue, and N. A. Stanton, "Is partially automated driving a bad idea? observations from an on-road study," Applied ergonomics, vol. 68, pp. 138–145, 2018.
[18] T. K. Chan, C. S. Chin, H. Chen, and X. Zhong, "A comprehensive review of driver behavior analysis utilizing smartphones," IEEE Transactions on Intelligent Transportation Systems, DOI 10.1109/TITS.2019.2940481, pp. 1–32, 2019.
[19] Z. Chen, J. Yu, Y. Zhu, Y. Chen, and M. Li, "D 3: Abnormal driving behaviors detection and identification using smartphone sensors," in 2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 524–532. IEEE, 2015.
[20] T.-H. Nguyen, D.-N. Lu, D.-N. Nguyen, and H.-N. Nguyen, "Dynamic basic activity sequence matching method in abnormal driving pattern detection using smartphone sensors," Electronics, vol. 9, no. 2, p. 217, 2020.
[21] D.-W. Koh and H.-B. Kang, "Smartphone-based modeling and detection of aggressiveness reactions in senior drivers," in 2015 IEEE Intelligent Vehicles Symposium (IV), pp. 12–17. IEEE, 2015.
[22] R. Chhabra, S. Verma, and C. R. Krishna, "Detecting aggressive driving behavior using mobile smartphone," in Proceedings of 2nd International Conference on Communication, Computing and Networking, pp. 513–521. Springer, 2019.
[23] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 402–427, 2013.
[24] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 38, no. 3, pp. 502–512, 2008.
[25] W.-L. Zheng, K. Gao, G. Li, W. Liu, C. Liu, J.-Q. Liu, G. Wang, and B.-L. Lu, "Vigilance estimation using a wearable eog device in real driving environment," IEEE Transactions on Intelligent Transportation Systems, 2019.
[26] B. Hariri, S. Abtahi, S. Shirmohammadi, and L. Martel, "Vision based smart in-car camera system for driver yawning detection," in 2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras, pp. 1–2. IEEE, 2011.
[27] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," IEEE transactions on vehicular technology, vol. 53, no. 4, pp. 1052–1068, 2004.
[28] K. W. Lee, H. S. Yoon, J. M. Song, and K. R. Park, "Convolutional neural network-based classification of driverâs emotion during aggressive and smooth driving using multi-modal camera sensors," Sensors, vol. 18, no. 4, p. 957, 2018.
[29] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," IEEE Access, vol. 7, pp. 61 904–61 919, 2019.
[30] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee, "Enhancing transportation systems via deep learning: A survey," Transportation research part C: emerging technologies, 2018.
[31] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," IEEE Transactions on Intelligent Transportation Systems, 2019.
[32] T. Yuan, W. B. da Rocha Neto, C. Rothenberg, K. Obraczka, C. Barakat, and T. Turletti, "Harnessing machine learning for next-generation intelligent transportation systems: A survey," 2019.
[33] Y. Xing, C. Lv, H. Wang, H. Wang, Y. Ai, D. Cao, E. Velenis, and F.-Y. Wang, "Driver lane change intention inference for intelligent vehicles: framework, survey, and challenges," IEEE Transactions on Vehicular Technology, vol. 68, no. 5, pp. 4377–4390, 2019.
[34] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," Sensors, vol. 19, no. 11, p. 2574, 2019.
[35] D. L. Hendricks, M. Freedman, J. C. Fell et al., "The relative frequency of unsafe driving acts in serious traffic crashes," United States. National Highway Traffic Safety Administration, Tech. Rep., 2001.
[36] ROSPA, "The royal society for the prevention of accidents, road safety factsheet, driver distraction fact sheet," 2017. [Online]. Available: https://www.rospa.com/rospaweb/docs/advice-services/road-safety/drivers/driver-distraction.pdf
[37] A. Khandakar, M. E. Chowdhury, R. Ahmed, A. Dhib, M. Mohammed, N. A. Al-Emadi, and D. Michelson, "Portable system for monitoring and controlling driver behavior and the use of a mobile phone while driving," Sensors, vol. 19, no. 7, p. 1563, 2019.
[38] G. M. Fitch, S. A. Soccolich, F. Guo, J. McClafferty, Y. Fang, R. L. Olson, M. A. Perez, R. J. Hanowski, J. M. Hankey, and T. A. Dingus, "The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk," Tech. Rep., 2013.
[39] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: Less than the sum of its parts," Accident Analysis & Prevention, vol. 42, no. 3, pp. 881–890, 2010.
[40] M. N. Husen, S. Lee, and M. Q. Khan, "Syntactic pattern recognition of car driving behavior detection," in Proceedings of the 11th international conference on ubiquitous information management and communication, p. 77. ACM, 2017.

[41] Y. Liao, S. E. Li, W. Wang, Y. Wang, G. Li, and B. Cheng, "Detection of driver cognitive distraction: A comparison study of stop-controlled intersection and speed-limited highway," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, pp. 1628–1637, 2016.

[42] Y. Liao, G. Li, S. E. Li, B. Cheng, and P. Green, "Understanding driver response patterns to mental workload increase in typical driving scenarios," IEEE Access, vol. 6, pp. 35 890–35 900, 2018.

[43] J. L. Harbluk, Y. I. Noy, and M. Eizenman, "The impact of cognitive distraction on driver visual behaviour and vehicle control," Tech. Rep., 2002.

[44] M. Botta, R. Cancelliere, L. Ghignone, F. Tango, P. Gallinari, and C. Luison, "Real-time detection of driver distraction: random projections for pseudo-inversion-based neural training," Knowledge and Information Systems, vol. 60, no. 3, pp. 1549–1564, 2019.

[45] L. Li, B. Zhong, C. Hutmacher Jr, Y. Liang, W. J. Horrey, and X. Xu, "Detection of driver manual distraction via image-based hand and ear recognition," Accident Analysis & Prevention, vol. 137, p. 105432, 2020.

[46] Z. Li, S. Bao, I. V. Kolmanovsky, and X. Yin, "Visual-manual distraction detection using driving performance indicators with naturalistic driving data," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp. 2528–2535, 2017.

[47] H. Uno and K. Hiramatsu, "Effects of auditory distractions on driving behavior during lane change course negotiation: estimation of spare mental capacity as a index of attention distraction," JSAE review, vol. 21, no. 2, pp. 219–224, 2000.

[48] L. Zhang, B. Cui, M. Yang, F. Guo, and J. Wang, "Effect of using mobile phones on driverâs control behavior based on naturalistic driving data," International journal of environmental research and public health, vol. 16, no. 8, p. 1464, 2019.

[49] T. Oron-Gilad, A. Ronen, and D. Shinar, "Alertness maintaining tasks (amts) while driving," Accident Analysis & Prevention, vol. 40, no. 3, pp. 851–860, 2008.

[50] M. D. Zwaag, Music directs your mood. University Library Groningen][Host], 2012.

[51] D. Dmitrenko, E. Maggioni, and M. Obrist, "I smell trouble: using multiple scents to convey driving-relevant information," in Proceedings of the 2018 on International Conference on Multimodal Interaction, pp. 234–238. ACM, 2018.

[52] D. Dmitrenko, E. Maggioni, C. T. Vi, and M. Obrist, "What did i sniff?: Mapping scents onto driving-related messages," in Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 154–163. ACM, 2017.

[53] K. Strohl, S. Merritt, J. Blatt, A. Pack, F. Council, and S. Rogus, "Drowsy driving and automobile crashes. nccdr/nhtsa expert panel on driver fatigue and sleepiness," Washington, DC: National Highway Traffic Safety Administration, 1998.

[54] D. Fatigue, "Road accidents: A literature review and position paper," The Royal Society for the Prevention of Accidents, 2001.

[55] L. National Heart, B. Institute et al., "Drowsy driving and automobile crashes," United States. National Highway Traffic Safety Administration, Tech. Rep., 1998.

[56] Y. Xie, C. Bian, Y. L. Murphey, and D. S. Kochhar, "An svm parameter learning algorithm scalable on large data size for driver fatigue detection," in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–8. IEEE, 2017.

[57] A. S. Zandi, A. Quddus, L. Prest, and F. J. Comeau, "Non-intrusive detection of drowsy driving based on eye tracking data," Transportation Research Record, p. 0361198119847985, 2019.

[58] R. Galindo, W. G. Aguilar, and R. P. R. Ch, "Landmark based eye ratio estimation for driver fatigue detection," in International Conference on Intelligent Robotics and Applications, pp. 565–576. Springer, 2019.

[59] M. Mahmoodi and A. Nahvi, "Driver drowsiness detection based on classification of surface electromyography features in a driving simulator," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 233, no. 4, pp. 395–406, 2019.

[60] Z. Li, S. Li, R. Li, B. Cheng, and J. Shi, "Online detection of driver fatigue using steering wheel angles for real driving conditions," Sensors, vol. 17, no. 3, p. 495, 2017.

[61] Z. Li, Q. Zhang, and X. Zhao, "Performance analysis of k-nearest neighbor, support vector machine, and artificial neural network classifiers for driver drowsiness detection with different road geometries," International Journal of Distributed Sensor Networks, vol. 13, no. 9, p. 1550147717733391, 2017.

[62] A. N. Assuncao, A. L. Aquino, C. d. M. Santos, C. Ricardo, R. L. Guimaraes, and R. A. Oliveira, "Vehicle driver monitoring through the statistical process control," Sensors, vol. 19, no. 14, p. 3059, 2019.

[63] A. D. Enforcement, "Strategies for implementing best practices," Technical Report, US Dept of Transportation National Highway Traffic Safety â, Tech. Rep., 2000.

[64] K. Johnson, "Frustration drives road rage," Traffic Safety (Chicago), vol. 97, no. 4, pp. 8–11, 1997.

[65] D. Shinar, "Aggressive driving: the contribution of the drivers and the situation," Transportation Research Part F: traffic psychology and behaviour, vol. 1, no. 2, pp. 137–160, 1998.

[66] F. Alonso, C. Esteban, L. Montoro, and A. Serge, "Conceptualization of aggressive driving behaviors through a perception of aggressive driving scale (pad)," Transportation research part F: traffic psychology and behaviour, vol. 60, pp. 415–426, 2019.

[67] L. Yan, P. Wan, L. Qin, and D. Zhu, "The induction and detection method of angry driving: Evidences from eeg and physiological signals," Discrete Dynamics in Nature and Society, vol. 2018, 2018.

[68] N. Banovic, A. Wang, Y. Jin, C. Chang, J. Ramos, A. Dey, and J. Mankoff, "Leveraging human routine models to detect and generate human behaviors," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 6683–6694, 2017.

[69] M. Kamrani, R. Arvin, and A. J. Khattak, "Extracting useful information from basic safety message data: an empirical study of driving volatility measures and crash frequency at intersections," Transportation Research Record, vol. 2672, no. 38, pp. 290–301, 2018.

[70] F. Feng, S. Bao, J. R. Sayer, C. Flannagan, M. Manser, and R. Wunderlich, "Can vehicle longitudinal jerk be used to identify aggressive drivers? an examination using naturalistic driving data," Accident Analysis & Prevention, vol. 104, pp. 125–136, 2017.

[71] E. G. Mantouka, E. N. Barmpounakis, and E. I. Vlahogianni, "Identifying driving safety profiles from smartphone data using unsupervised learning," Safety Science, vol. 119, pp. 84–90, 2019.

[72] Y. Li, "Deep reinforcement learning: An overview," arXiv preprint arXiv:1701.07274, 2017.

[73] S. Mirjalili, H. Faris, and I. Aljarah, "Introduction to evolutionary machine learning techniques," in Evolutionary Machine Learning Techniques, pp. 1–7. Springer, 2020.

[74] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26–38, 2017.

[75] V. Talpaert, I. Sobh, B. R. Kiran, P. Mannion, S. Yogamani, A. El-Sallab, and P. Perez, "Exploring applications of deep reinforcement learning for real-world autonomous driving systems," arXiv preprint arXiv:1901.01536, 2019.

[76] M. Bouton, A. Nakhaei, K. Fujimura, and M. J. Kochenderfer, "Cooperation-aware reinforcement learning for merging in dense traffic," arXiv preprint arXiv:1906.11021, 2019.

[77] O. Ibitoye, R. Abou-Khamis, A. Matrawy, and M. O. Shafiq, "The threat of adversarial attacks on machine learning in network security–a survey," arXiv preprint arXiv:1911.02621, 2019.

[78] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," Electronics, vol. 8, no. 3, p. 292, 2019.

[79] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.

[81] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," May, vol. 19, p. 2019, 2019.

[82] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," Pattern recognition, vol. 15, no. 6, pp. 455–469, 1982.

[83] K. Fukushima, "Recent advances in the deep cnn neocognitron," Nonlinear Theory and Its Applications, IEICE, vol. 10, no. 4, pp. 304–321, 2019.

[84] S. Indolia, A. K. Goswami, S. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network-a deep learning approach," Procedia computer science, vol. 132, pp. 679–688, 2018.

[85] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," arXiv preprint arXiv:1312.6026, 2013.

[86] J. Bradbury, S. Merity, C. Xiong, and R. Socher, "Quasi-recurrent neural networks," arXiv preprint arXiv:1611.01576, 2016.

[87] S. K. Mahata, D. Das, and S. Bandyopadhyay, “Mtil2017: Machine translation using recurrent neural network on statistical machine translation,” Journal of Intelligent Systems, vol. 28, no. 3, pp. 447–453, 2019.
[88] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Advances in neural information processing systems, pp. 3104–3112, 2014.
[89] W. Wang, S. Tulyakov, and N. Sebe, “Recurrent convolutional shape regression,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 11, pp. 2569–2582, 2018.
[90] S. Park, Y. Seonwoo, J. Kim, J. Kim, and A. Oh, “Denoising recurrent neural networks for classifying crash-related events,” IEEE Transactions on Intelligent Transportation Systems, 2019.
[91] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), DOI 10.3115/v1/D14-1179, pp. 1724–1734. Doha, Qatar: Association for Computational Linguistics, Oct. 2014. [Online]. Available: https://www.aclweb.org/anthology/D14-1179
[92] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder–decoder for statistical machine translation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, 2014.
[93] B. Krause, L. Lu, I. Murray, and S. Renals, “Multiplicative lstm for sequence modelling,” arXiv preprint arXiv:1609.07959, 2016.
[94] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, “An end-to-end spatio-temporal attention model for human action recognition from skeleton data,” in Thirty-first AAAI conference on artificial intelligence, 2017.
[95] Y. Cao, F. Yang, Q. Tang, and X. Lu, “An attention enhanced bidirectional lstm for early forest fire smoke recognition,” IEEE Access, vol. 7, pp. 154 732–154 742, 2019.
[96] Y. Hua, J. Guo, and H. Zhao, “Deep belief networks and deep learning,” in Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, pp. 1–4. IEEE, 2015.
[97] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, “An introductory review of deep learning for prediction models with big data. front,” Artif. Intell, vol. 3, no. 4, 2020.
[98] A. Behera, A. Keidel, and B. Debnath, “Context-driven multi-stream lstm (m-lstm) for recognizing fine-grained activity of drivers,” in German Conference on Pattern Recognition, pp. 298–314. Springer, 2018.
[99] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, “Driver distraction identification with an ensemble of convolutional neural networks,” Journal of Advanced Transportation, vol. 2019, 2019.
[100] M. Alotaibi and B. Alotaibi, “Distracted driver classification using deep learning,” Signal, Image and Video Processing, pp. 1–8, 2019.
[101] J. M. Celaya-Padilla, C. E. Galván-Tejada, J. S. A. Lozano-Aguilar, L. A. Zanella-Calzada, H. Luna-García, J. I. Galván-Tejada, N. K. Gamboa-Rosales, A. Velez Rodriguez, and H. Gamboa-Rosales, “âtexting & drivingâ detection using deep convolutional neural networks,” Applied Sciences, vol. 9, no. 15, p. 2962, 2019.
[102] W. Huang, X. Liu, M. Luo, P. Zhang, W. Wang, and J. Wang, “Video-based abnormal driving behavior detection via deep learning fusions,” IEEE Access, vol. 7, pp. 64 571–64 582, 2019.
[103] C. Zhang, R. Li, W. Kim, D. Yoon, and P. Patras, “Driver behavior recognition via interwoven deep convolutional neural nets with multi-stream inputs,” arXiv preprint arXiv:1811.09128, 2018.
[104] Y. Ed-doughmi and N. Idrissi, “Driver fatigue detection using recurrent neural networks,” in Proceedings of the 2nd International Conference on Networking, Information Systems & Security, p. 44. ACM, 2019.
[105] Z. Xiao, Z. Hu, L. Geng, F. Zhang, J. Wu, and Y. Li, “Fatigue driving recognition network: fatigue driving recognition via convolutional neural network and long short-term memory units,” IET Intelligent Transport Systems, 2019.
[106] W. Liu, J. Qian, Z. Yao, X. Jiao, and J. Pan, “Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection,” Future Internet, vol. 11, no. 5, p. 115, 2019.
[107] M. A. Tanveer, M. J. Khan, M. J. Qureshi, N. Naseer, and K.-S. Hong, “Enhanced drowsiness detection using deep learning: An fnirs study,” IEEE Access, vol. 7, pp. 137 920–137 929, 2019.
[108] Y. Moukafih, H. Hafidi, and M. Ghogho, “Aggressive driving detection using deep learning-based time series classification,” in 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1–5. IEEE, 2019.
[109] M. Matousek, E.-Z. Mohamed, F. Kargl, C. Bösch et al., “Detecting anomalous driving behavior using neural networks,” in 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 2229–2235. IEEE, 2019.
[110] Y. Xing, C. Lv, and D. Cao, “Personalized vehicle trajectory prediction based on joint time series modeling for connected vehicles,” IEEE Transactions on Vehicular Technology, 2019.
[111] Y. Xing, C. Lv, D. Cao, and C. Lu, “Energy oriented driving behavior analysis and personalized prediction of vehicle states with joint time series modeling,” Applied Energy, vol. 261, p. 114471, 2020.
[112] C. Lv, X. Hu, A. Sangiovanni-Vincentelli, Y. Li, C. M. Martinez, and D. Cao, “Driving-style-based codesign optimization of an automated electric vehicle: a cyber-physical system approach,” IEEE Transactions on Industrial Electronics, vol. 66, no. 4, pp. 2965–2975, 2018.
[113] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, and Q. Arshad, “Data and privacy: Getting consumers to trust products enabled by the internet of things,” IEEE Consumer Electronics Magazine, vol. 8, no. 2, pp. 35–38, 2019.
[114] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, and Q. Arshad, “Enabling consumer trust upon acceptance of iot technologies through security and privacy model,” in Advanced Multimedia and Ubiquitous Engineering, pp. 111–117. Springer, 2016.
[115] H. Xu, Y. Ma, H. Liu, D. Deb, H. Liu, J. Tang, and A. Jain, “Adversarial attacks and defenses in images, graphs and text: A review,” International Journal of Automation and Computing, vol. 17, no. 2, pp. 151–178, 2020.
[116] N. Akhtar and A. Mian, “Threat of adversarial attacks on deep learning in computer vision: A survey,” IEEE Access, vol. 6, pp. 14 410–14 430, 2018.
[117] W. E. Zhang, Q. Z. Sheng, A. Alhazmi, and C. Li, “Adversarial attacks on deep-learning models in natural language processing: A survey,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 11, no. 3, pp. 1–41, 2020.
[118] S. Karthik, A. Paul, and N. Karthikeyan, Deep learning innovations and their convergence with big data. IGI Global, 2017.
[119] N. Aranjuelo, L. Unzueta, I. Arganda-Carreras, and O. Otaegui, “Multimodal deep learning for advanced driving systems,” in International Conference on Articulated Motion and Deformable Objects, pp. 95–105. Springer, 2018.
[120] N. C. Louis Stewart, Moh Musa, “Look no hands: self-driving vehicle’s public trust problem,” 2019. [Online]. Available: https://www.weforum.org/agenda/2019/08/self-driving-vehicles-public-trust/
[121] E. Edmonds, “Three in four americans remain afraid of fully self-driving vehicles,” 2019. [Online]. Available: https://newsroom.aaa.com/2019/03/americans-fear-self-driving-cars-survey/

MONAGI H. ALKINANI received his PhD in Computer Science in 2017 from Western University, London, Canada. In 2018, he joined the Deanship of Scientific Research at University of Jeddah, Saudi Arabia, where he served as the vice-dean of research. He holds an assistance professor position at the Department of Computer Science and Artificial Intelligence and is a member of the Jeddah Computer Vision Team, where he supervises research activities and teaches image processing and artificial intelligence to bachelor students as well as signal processing to master students in computer science. At the Deanship he supervised research in the field of computer vision. He has been involved in many collaborative research projects financed by various instances including the Ministry of Education and the University of Jeddah. He currently serves as coordinator of the RD program for the Ministry of Education (https://rdo.moe.gov.sa/en). He is also officer of international collaboration of University of Jeddah under Deputyship of the International Collaboration at the Ministry of Education.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2999829, IEEE Access

WAZIR ZADA KHAN (M'16-SM'17) received his bachelor's and master's degrees in computer science from COMSATS University Islamabad, Wah Campus in 2004 and 2007, respectively. He received his PhD from the Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, Malaysia in 2015. Dr. Khan is currently working with Farasan Networking Research Laboratory, Faculty of CS & IT, Jazan University, Saudi Arabia. Dr. Khan is also serving as a Researcher at the 'Global Foundation for Cyber Studies and Research' (https://www.gfcyber.org), which is an independent, non-profit, and non-partisan cybersecurity think-tank based in Washington D.C. Dr. Khan has published over 75 research papers in the journals and conferences of international repute. He is the serving as a reviewer of many reputed journals and also a member of the technical program committee for many international conferences. He has more than ten years of teaching/professional experience in Pakistan and Saudi Arabia. His current research interests include wireless sensor networks, security and privacy, blockchain, IoT, IIoT and reinforcement learning. He is a Senior Member of the IEEE.

PLACE
PHOTO
HERE

QURATULAIN ARSHAD received her BS in Computer Science from Comsats Institute of Information Technology, Pakistan. Her research interests include network security, trust and reputation in sensor networks and ad hoc networks. Her subjects of interest include Network Security, Machine Learning, Deep Learning, Digital Image Processing and Computer Vision.