A Methodological Review on Prediction of Multi-Stage Hypovigilance Detection Systems Using Multimodal Features

QAISAR ABBAS AND **ABDULLAH ALSHEDDY**

College of Computer and Information Sciences, Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

Corresponding author: Qaisar Abbas (qaabbas@imamu.edu.sa)

ABSTRACT Several hypovigilance detection systems (HDx) were developed to avoid road-side accidents due to driver fatigue. They have suffered from several limitations. Notably many of these are focused on center-head position to define an area of interest (often referred to as PERCLOS (percentage eye closure)) without considering the face occlusion problem, light illumination, and suffer poor response time. These HDx systems mostly depend on image processing, vision-based, and multisensor-based features. To address these problems, the author utilized vision, sensors, environmental, and vehicular-based features that integrated together by fusion to predict multistage of HDx. Lately, few studies have utilized the combination of multimodal features and deep learning (DL) architectures. Those multimodal-based features (M-HDx) were feasible to predict stages of driver fatigue (multi-stage). However, there is a need to critically measure the performance of these M-HDx by carrying out a comparative analysis to recognize multi-stage of fatigue in terms of hardware-based benchmarks. Moreover, it is important to evaluate the M-HDx systems using different features-set with respect to traditional and advanced machine learning techniques. Therefore, the primary aim of this work is in algorithm and feature modeling, then compare the advantages and differences with other work. In this paper, a different study is conducted compare to state-of-the-art survey articles by statistically measuring the performance. After experiments on M-HDx systems, this paper concludes that there is still a research gap to real-time development of multistage M-HDx systems. In the end, the paper summarizes the directions, challenges, and applications in the development of HDx systems to assist other researchers for further research.

INDEX TERMS Computer vision, driver hypovigilance, driver distraction, deep learning, hybrid learning models, multimodal features, machine learning, multisensor fusion, transfer learning.

I. INTRODUCTION

Fatigue and loss of vigilance (hypovigilance) among the drivers are very common problems. This problem assumes even acute significance for long-haul logistics drivers. These risks can potentially translate into hazardous accident situations. During the last few years, researchers have developed many hardware and software-based techniques to avoid such risks [1]. In particular, an effort has been to develop techniques for automated monitoring of the drivers' activities. These systems have the ability to provide intelligent feedback

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

and generate alert messages to the drivers for recognition the uncontrollable situations. According to a report published by World Health Organization (WHO) [2], more than 1 million people lose their lives due to traffic accidents, and approximately 50 million more get injured causing severe disabilities. In particular, more than seven thousand death and thirty-eight thousand injuries are recorded due to road accidents every year in Kingdome of Saudi Arabia (KSA). The development of real-time driver fatigue detection and prediction system is a challenging task related to computer vision technologies. In such systems, detection of low vigilance and high fatigue level results in the generation of alerts and warnings to the driver about his/her poor state of driving through an alarm.

As a result, the development of this driver drowsiness system is critical for detecting the drivers' ability of perception, recognition, and vehicle control during roadside driving.

Most of the existing real-time hypovigilance detection systems (HDx) techniques have a poor response time as mentioned in the literature [3]. This calls for a need to develop such systems that respond effectively in less amount of time. Another characteristic of the desired technique should be its non-intrusive nature making it independent of any specific hardware on the driver's body. One such project to be cited is by MIT researchers named Smart Car [4]. This system however is intrusive as the driver needs to wear and use a wristband on the arm or wrist to calculate and measure the heart rate. Some other methods use the eyes and gaze movements of drivers that are installed on the helmet or by using special lenses. Though useful yet the intrusive nature of these types makes these practically unacceptable by many communities.

Hypovigilance detection systems (HDx) systems were developed in the past to detect driver drowsiness or inattention [5] under the state of drowsiness. To measure the drivers' state, in this case, several authors such as CogBeacon-ML [124], BROOK-DenseNet [126], Ford-dataset [74], and Riani-M-HDx [128] utilized different measures as shown in Figure 1. The past HDX developed systems were based on the integration of PERCLOS (percentage eye closure) and multisensor fusion approaches [6] to detect and predict driver drowsiness. Those systems were based on advanced image processing, signal processing, and machine learning techniques to define PERCLOS measures. Health measures and vitals such as EEG and/or heart rate monitoring systems can also be used to detect driver fatigue [7], [8]. Some other systems use non-visual features based on driver physiological measures and vehicle parameters. In the case of physiological parameter measurements, the researchers predict driver fatigue based on different parameters such as steering-wheel, acceleration pedal, and speed. In practice, those approaches were mostly dependent on the road-shape, way of driving, and performance of the vehicle. The author utilized electroencephalograph (EEG), electrocardiogram (ECG), electrooculography (EOG), and surface electromyogram (EMG) [9] sensors to predict driver fatigue. Yet, these techniques dependent on contactable sensors, which lessen user experience and increase hardware cost. On the other hand, it might increase the effectiveness of HDx systems.

Driver hypovigilance [10], [11] can be determined by applying a combination of the following measures. Vehicle-dependent measures are determined by monitoring the deviation of the lane direction, the rotation of the steering wheel, the acceleration pedal strength, etc. If the calculated metrics go beyond a small scale, so the hypo-surveillance risk of the driver is very high. The physiological measures are also used to determine fatigue state using medical vitals such as ECG, EMG, and EEG. Any visible distortion in these vitals can signal a change in the driver's state which should generate an alarm. Several studies have been published on
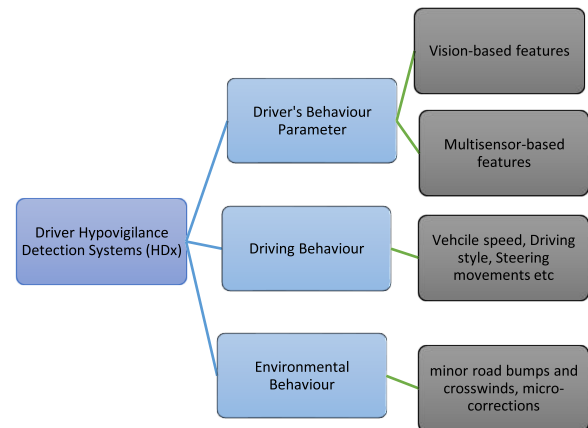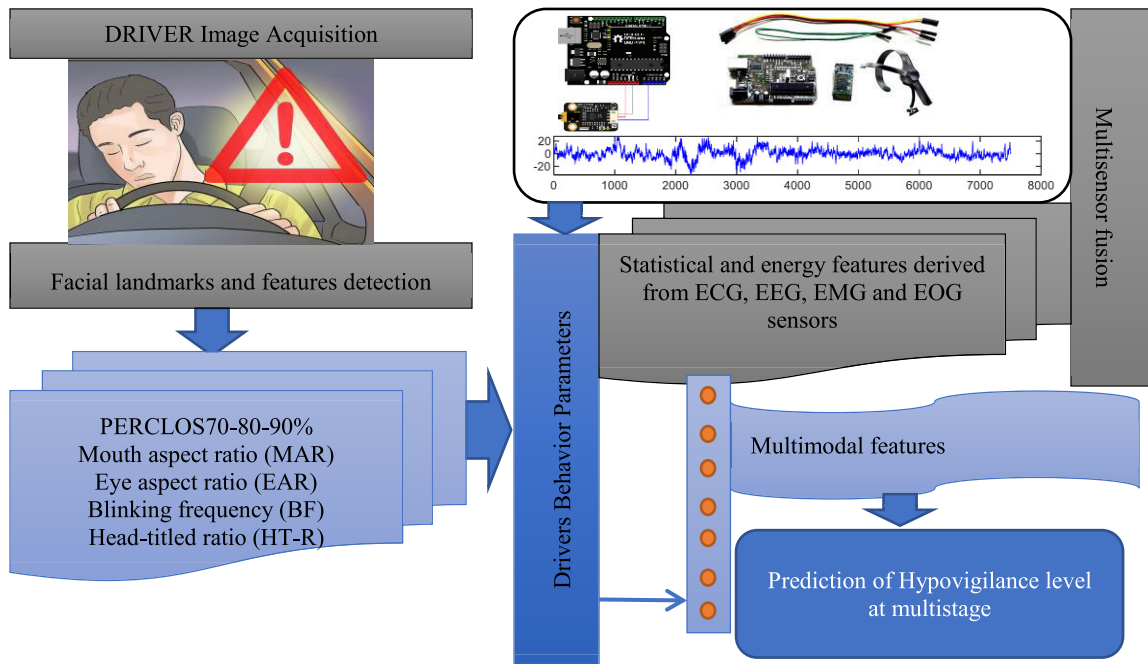

FIGURE 1. A visual example of state-of-the-art Hypovigilance prediction to measure any one of the following measures or by a combination of these measures using multimodal features.

the identification of driver drowsiness or the detection of driver inattention by a physiological signal. Driver activity such as yawning, eye closing, eye blinking, head location, etc. is tracked by a vision-based camera and the device warns the driver if hypovigilance is observed. A visual example of these different measures is displayed in Figure 1.

The development of the HDx system is very critical due to many factors such as specific machine-learning methods, inadequate parameter settings that result in a false negative and false positive. In such systems, the prediction of driver's drowsiness or fatigue level using different vision and multisensor-based features [14], [15] does not achieve absolute accuracy. Combining multiple modalities can be a possible solution to resolve some of these constraints [16]. Modality fusion can be applied to achieve that which is characterized as the incorporation and combination of various types of data, which are collected from a subject. Modality fusion [17] may be done at any level of data processing, such as sensor-level, feature-level, or decision-level fusion. In practice, many researchers have been utilizing feature-level fusion techniques to implement HDx systems in a real-time environment. In feature-level fusion, in the first step, different features from different modalities are merged into a unique vector in an early and late binding fashion. It has been shown that integrating different modalities [18] at the feature-level is an efficient way to create a model for driver HDX detection. One of the main challenges facing effective modality fusion is the criteria for combination or convergence of modalities at the stage of fusion (sensor, feature, score, or decision). So far, feature-level fusion has been commonly used to test the stress levels and fatigue of drivers. Hence, it is desirable to develop an approach that integrates different modalities together to design a more effective and robust hypovigilance prediction system in the form of multistage output. In this paper, our focus has been exactly on the same concern as shown in Figure 2.

Various methodological techniques for the development of HDx systems have been described and discussed in this

FIGURE 2. A visual example of state-of-the-art Hypovigilance prediction to measure any one of the following measures or by a combination of these measures using multimodal features.

paper. We have compared the efficiency and performance of these techniques against the technique proposed by authors in [127]. In that paper, we describe the implementation of an IoT-based architecture for detecting driver drowsiness through mobile, energy computing in 5G and cloud-based environments. The comparison carried out in this paper highlights the need to integrate modern innovations in the domain of image processing, machine learning, and intelligence algorithms for multimodal and deep-learning-based computing. Although, several survey articles [12] have been written to address the same issue. However, it is apparent that none of these studies carried out a methodological comparison in the domain of deep-learning and multimodal features-learning. In a recent survey article, the different papers were described to detect real-time driver drowsiness [13] without the comparison involving multimodal and deep learning architectures. Therefore, in this paper, we reviewed recognize methods for detection of driver fatigue based on vision-based and multisensor-based features.

A. MAJOR CONTRIBUTIONS

The paper's primary work is in algorithm and feature modeling, then compare the advantages and differences with other work. This section briefly described the main contributions of this article that are described as follows.

1) Different machine learning algorithms are evaluated on hardware-based benchmarks (CPU and GPU).
2) Methodological reviews have been carried out to highlight recent trends for the development of hypovigilance detection (HDx) systems based on

vision-based (V-HDX), sensor-based (S-HDX), and multimodal-based (M-HDX) hypovigilance detection systems.
3) Comparison of different unimodal and multimodal features by using various machine learning algorithms and state-of-the-art HDx systems have been performed.
4) Online data sources along with parameters are also provided to assist other researchers in training the network.
5) Current and future directions for the development of the HDx system are also provided in this paper to continue further research in this domain.
6) Explain the importance of multimodal-features-based driver fatigue recognition system in the deep-learning context, which is a new review article in this domain.
7) State-of-the-art comparisons have been performed on recent four multimodal-based HDX systems to further discuss challenges in this domain.

B. PAPER ORGANIZATION

The remaining organization of this paper is as follows. In Section II, the research protocol is described that consists of research questions, comparisons to other survey articles, and selection of papers for this review article. A study of recent trends in the state-of-the-art in the field of machine and deep learnings for detection of driver's fatigue based on vision-based (V-HDX), sensor-based (S-HDX), and multimodal-based (M-HDX) driver fatigue is detailed described in section III. Section IV describes different deep-learning-based models used in the past. In section V, we have performed different experiments on selected datasets and

HDx systems based on different statistical metrics. Those results highlight merits, demerits, and limitations about different parameter settings and environments to develop HDx systems in real-time. Also, this section highlights the recent datasets used in the past. In addition, the discussion about obtained results is detailed described in section VI. Besides, this section also presents the current limitation and future directions in this domain to assist other researchers. In addition, this section also describes the challenges and future directions to help other researchers so that they can develop HDx systems up-to-the-mark. Finally, this paper concludes in section VII.

II. RESEARCH PROTOCOL

The research protocol defined the detailed layout, the method that is a plan to use in the review. The protocol also defined the rules and instruction for conducting a survey on the existing work on recent state-of-the-art hypo-vigilance detection systems (HDx) which offers assistance to research beginners in this domain. And, also provide enough information regarding traditional machine learning and deep learning-based approaches in the development of HDx systems through multi-modal features as shown in Figure 3. The survey also assists the user in regards to performance comparison of traditional and deep learning-based HDx systems, the impact of deep learning and multimodal technologies, the kind of deep model are already implemented, and how the performance of the HDx systems can be improved? Moreover, the survey also provides help regarding the cost, data analysis and management, achieved result and also find the challenges related to both traditional and deep learning-based HDx systems.

A. RESEACH QUESTIONS

The fundamental core of the study of either the literature review or project is the research questions. This describes the methodological starting point of scientific study in all disciplines. Table 1 reveal the research queries linked with the HDx systems trait, the motivation and linked to a different section of the paper.

B. COMPARISONS WITH SURVEY ARTICLES

We have done extensive studies in terms of methodological comparisons of different features and machine-learning algorithms compare to state-of-the-art survey articles in this domain. Table 2 shows the difference between our study and existing surveyed articles.

C. PAPERS QUALITY EVALUATION

The visual, non-visual, and multimodal features with machine-learning and deep-learning architectures keywords are used to search papers. In general, the standard is considered after defining the research question and before the research process is performed out. Here, for instance, irrelevant papers and out of concerned papers were denied. If the large field of study relates to our subject, we would consider the research. Published research manuscripts from top
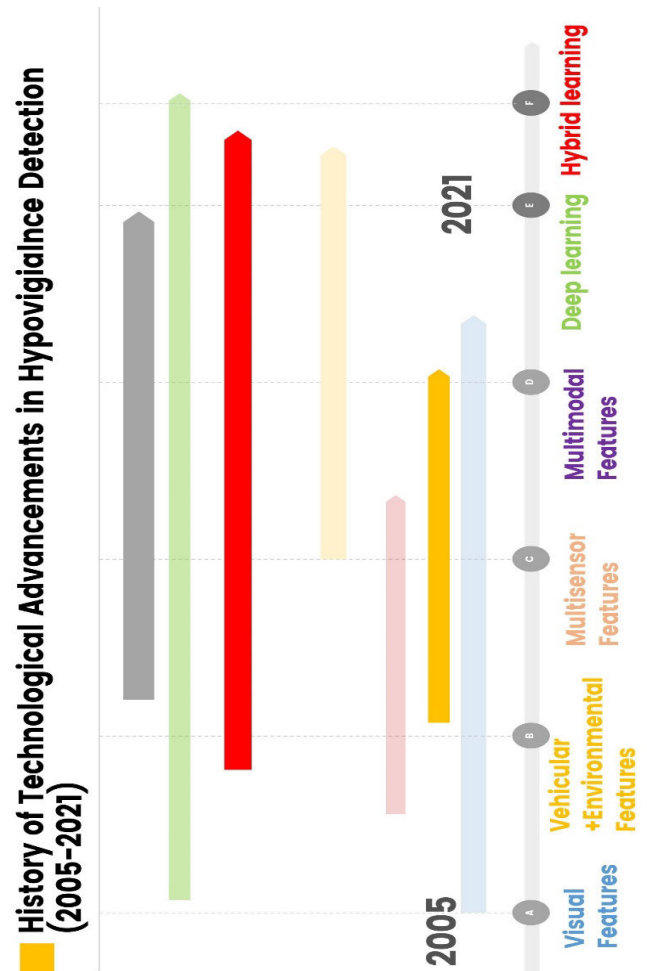
FIGURE 3. An overview of historical figure to show the technological evolution in the development of hypovigilance detection systems (HDx) related to different features and machine-learning techniques.

Journals (SCI and ESCI) and Top conferences must always be acknowledged. However, old studies and research which do not fulfill our goal must be exempted.

III. RECENT TRENDS OF HYPOVIGILANCE DETECTION SYSTEMS

Detection of driver drowsiness and fatigue has been an area of research for scientists for a while now. In the early research, the steering motion of automobiles was regarded as an indication of hypovigilance [6], [8]. More recently, techniques have used face detection and recording for this purpose. Besides, some methods require working with health vitals such as brain waves and heart rate to equate them with states of sleepiness [9]–[11]. They combined vision-based methods with multisensor-based methods to define multimodal features for the detection of hypovigilance. Some authors used physiological measurements [12] that can be extracted by using different sensors to account for heart rate, skin conductance, and respiration rate to detect hypovigilance (HDx) of drivers.

Orientation-based fatigue detection methods detect and verify the position of the eyes. The methods used different

TABLE 1. Summary of research questions and motivations involves in this study.

S. No.	Research Questions	Motivations
RQ1	What type of research is being done for recognition of single-state driver fatigue using deep learning architectures?	To know more about recent studies related to binary decision (Fatigue or normal), which are being conducted in this domain.
RQ2	What type of research is being done for recognition of multi-state driver fatigue using deep learning architectures?	To know about recent popular multi-stage driver fatigue's prediction systems.
RQ3	What are the visual features utilized to detect driver fatigue?	To know about in-depth visual-related features utilized in the past.
RQ4	What are the non-visual features utilized to detect driver fatigue?	To know about in-depth non-visual-related features utilized in the past.
RQ5	What are the multi-modal features used in the past to detect fatigue?	To understand the needs of combining visual and non-visual features to ensemble as multimodal features for the development of DFD systems.
RQ6	What are the main datasets being used by state-of-the-art studies?	To know about the availability of datasets for testing and training.
RQ7	What is the data size available to facilitate the latest machine-learning architectures?	To study and know about the size of dataset is really mattered in case of using latest approaches.
RQ8	What are traditional-machine algorithms used in the past?	To understand bout the different traditional classifier used in the past.
RQ9	What are the main performance measures used for the validating of results?	To know about the statistical metrics used in the past.
RQ10	What are the different state-of-the-art multimodal-based systems are available?	To know about existing multimodal-based DFD systems.
RQ11	What is the role of deep learning architectures?	To study various latest-machine-learning including deep-learning approaches.
RQ12	What is the importance of transfer learning (TL) in multimodal DFD system?	To know about usage of TL algorithm for the development of DFD system.
RQ13	What is the role of features involved in the detection of driver fatigue?	Try to understand existing features set used by different systems.
RQ14	What is the importance of multimodal features in deep-learning architectures?	To know about the pros and cons about the multimodal-features used in the past.
RQ15	What is the best features combination and deep-learning architectures?	To know about the latest trends in features and deep-learning models.
RQ16	What is the importance of selecting of benchmark for experiments?	To understand the importance about hardware benchmark for DFD systems.
RQ17	What are the current limitations and future directions of the detection of driver fatigue?	To identify research-gap for the better development of the DFD systems.

TABLE 2. Summary of existing survey papers on hypo-vigilance detection systems.

No.	Related Survey articles	Reviewed Up to	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6	RQ7	RQ8	RQ9	RQ10	RQ11	RQ12	RQ13	RQ14	RQ15	RQ16	RQ17
1	Kaplan et al [6]	2014	✔		✔	✔		✔											✔
2	Mathews et al [1]	2017	✔		✔	✔													
3	Ramzan et al [11]	2018	✔		✔	✔		✔	✔				✔		✔				✔
4	Khan et al [12]	2018	✔		✔	✔				✔					✔				✔
5	Sikander et al [13]	2018	✔		✔	✔		✔		✔					✔				✔
6	Rastgoo et al [14]	2018	✔		✔	✔		✔		✔									
7	Elassad et al [157]	2019	✔		✔	✔				✔	✔								✔
8	Němcová et al [16]	2019	✔		✔	✔	✔	✔	✔	✔									✔
9	Abbas et al [127]	2020	✔	✔	✔	✔	✔	✔	✔		✔		✔		✔				✔
10	Watling et al [156]	2020	✔		✔	✔				✔					✔				✔
11	**Our Review Article**	**2021**	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

techniques such as template matching and appearance-based functionality to detect the state. Nevertheless, these distinctive methods faced problems because of the errors that are generating from the signal disambiguation. These also faced problems due to lack of facial detection [14] resulting from abrupt gestures, and the discomfort due to constant hardware application. Medical signals such as electroencephalography (EEG) and electrocardiogram (ECG) [15] have also been used and analyzed to differentiate between the sleepiness and alertness states of the drivers. As a result, some authors suggested integrating vision-based methods with multisensor-based [16] techniques to extract multimodal features. Wu *et al.* [17] studied methods adopted to detect the alertness of drivers and categorized them into methods related to the condition of drivers. One category involved metrics such as eyelid motions and percentage of eye closing, whereas another category comprised of methods related to the efficiency of drivers, such as vehicle distance and lane detection and multi-modal methods that incorporated these approaches. Using a single camera or multiple cameras and several sensors were used in the past to detect the hypovigilance state of the drivers.

The major limitation of this type of system is their inability for early and accurate detection [18]. It is needed to detect early if the warning occurs at the initial stages. Literature is replete with various parameters and measures that can be used to detect driver fatigue. We need to use a combination of these measures as a single measure or different parameters may not be able to detect fatigue [19] in terms of multi-stage. If the driver used glasses or hide their face then eyes or facial features cannot be detected easily. Therefore, more than one measure or parameter can be suitable for this type of system to detect driver fatigue.

A comprehensive literature review suggested that the authors are deploying a combination of techniques instead of using a single machine learning technique to optimize the performance. To achieve this goal, some recent studies [107] have employed hybrid solutions to enhance the accuracy of the fatigue detection system. The system employed various vehicle parameters such as speed, acceleration, vehicle lane position, steering angle, braking. These features were then combined with facial features to predict driver drowsiness. Research shows that these hybrid systems are providing more accurate results with the increase of computational time. Several studies also used deep-learning models to classify

driver fatigue. The above-mentioned parameters are the main parameters of fatigue detection due to their ability to monitor variation. There can be environmental factors that may yield inaccurate data for these parameters. The lighting conditions may fluctuate because of diverse weather conditions. Similarly, sometimes it can be day or night. The background of the driver also affects these types of systems. Another factor that affects such type of system is the vibration of the car to collect visual data.

Researchers are constantly working on the development of HDx systems based on multimodal features instead of a single modality to overcome the above-mentioned limitations. By monitoring the heart rate of drivers [19] with a wearable tracker or extracting their facial expressions with an RGB camera, driving fatigue can be observed more effectively and robustly. However, such systems can present the problem of intrusiveness and result in inconvenience. The application of an RGB camera on the other hand can be affected again by light and other such factors. Also, the temporal details of fatigue characteristics and the relationship between the characteristics can't be neglected if we rely completely on such approaches. To approximate the detection of the driver's fatigue condition, these fusion methods need to be combined with transient fatigue features into the classifier's input vector. It is therefore required that we obtain features over a regular period-of-time to recognize multilevel (drowsy, alert, very alert, very drowsy or normal) driver drowsiness state. While the methods described above are encouraging, further challenges must be considered as mentioned in Table 3, Table 4, and Table 5. According to these tables, it is needed to design a new multimodal fusion-based approach to develop an HDx system so that the maximum performance level may be achieved.

The precision of fatigue detection may be significantly improved by integrating physical characteristics and vehicular features to detect fatigue. A technique that incorporates both driver characteristics and vehicle characteristics were proposed by Cheng *et al.* [20]. Parameters selected for fatigue identification included PERCLOS, blink rate, maximum time eyes are near, non-steering percentage, percentage of on-center driving, standard steering wheel angle deviation, and standard lane direction deviation. In the trial, 20 participants took part. To fuse data from both the driver and the car, the authors developed a model based on a multisource data fusion approach. A self-adaptive dynamic recognition model was introduced in a more recent study [21] that used the most powerful features for feature fusion to detect fatigue. These included physical characteristics as well as vehicle characteristics. The features were based on visual measurements of the driver as well as the actions of the car. This technique achieves an accuracy of 90.8 percent by using only the vehicle action features. The accuracy of technique with only facial features was measured at 91.6 percent accuracy. The fusion of all fatigue features produced an accuracy of 92.1 percent while accuracy of 93.8 percent was achieved when using only the most powerful features.

Apart from the detection of drowsiness, another challenge is to forecast deficiency of the operating status of a car driver [22]. In that paper, the authors investigated if it is feasible or not to use the standard sources of knowledge to detect drowsiness for a certain degree of drowsiness. The authors developed the HDx based on the behavioral, vehicular, and psychological factors to define multimodal features. For defining the behavioral features, the author used head and eyelid movements (blink duration, frequency, and PERCLOS) whereas to extract psychological features, they used variability of heart rate, respiration rate. Besides, the authors used speed, steering wheel angle, and position on the lane to measure the vehicular features. Two intelligent models were created to measure drowsiness and the time required to achieve a certain rate of drowsiness. With behavioral metrics and extra knowledge, the best result in both identification and prediction is obtained. Since they used multimodal features and obtained higher results of prediction compared to other approaches. Based on this point, we have methodological reviewed many papers in terms of vision-based, multisensor-based, and multimodal-based features. These technical details are presented in the upcoming sections. In this section, we will review state-of-the-art hypovigilance detection (HDx) systems. These HDx systems are broadly categorized into vision-based (V-HDx), multisensor-based (S-HDx), and multimodal-based (M-HDX) Systems.

A. VISION-BASED (V-HDx) SYSTEMS

Much research has gone into the development of vision-based hypovigilance detection (V-HDx) systems to diagnose drowsiness in drivers by analyzing driver behaviors. The main parameters for identifying drowsiness are the eye state, eye blinking level, mouth state, and yawning frequency of a driver [39]. The important parameter for the diagnosis of driver drowsiness is eye closure length. The V-HDx Systems that use this tool normally measure eye states and the location of the iris over a particular period to approximate the extent of eye blinking and the length of eye closure. PERCLOS (Percentage of eye closure) is a consistent and accurate measure for assessing the driver's alertness level [12], [49]–[51] for such devices. Eye condition research typically makes use of the PERCLOS measure as a drowsiness measure reaching 80% of the time for eyes closed [52], [53]. Generally, if the driver is sleepy, the time of eye closing will increase and the value of PERCLOS is greater than a driver's waking times.

In these techniques, the eye area including the pupil region needs to be separated to measure PERCLOS. Limitations of these techniques include improper lighting conditions, face occlusion, and sunglasses are main of them. To solve these issues, some V-HDx systems use IR (Infrared) cameras [24], [49], [54], [55]. The most useful V-HDx methods that focused on visual characteristics are explained in detail in the following paragraphs and compared in Table 3. In the past, the authors utilized different strategies for the detection of drowsiness based on V-HDx approaches. To classify these

TABLE 3. Summary of state-of-the-art vision-based hypovigilance (V-HDx) systems using visual features and machine learning algorithms.

Ref.	Techniques	Classifiers/ Features	Implementation	Detection	Critical Problems
[23]	Window Growing, PERCLOS	HOSVD	Pentium Core i7 i.73 GHz CPU, Two Cameras, LED	---	Environment Conditions, Sunglasses, High Complexity
[24]	Eye detecting and tracking	PCA, LDA, sparseness, and kurtosis, SVM	Intel Pentium-M processor running at 1.60 GHz with 500 MB RAM	---	Environment Parameters and Sunglasses
[25]	Haar face detector	SVM	Pavilion DV6, CORE I5 2.4 CPU, RAM 4 GB	---	Varying lighting Conditions and Sunglasses
[26]	blinking measurement and the 3Dhead pose estimation	CAMSHIFT, Haar wavelets, Hough Transform, SVM	Intel Pentium R Dual-Core CPU processor, 2 processor of 2.10 GHz, 4 GO RAM	---	Computational Expensive, Environment Parameters
[27]	Pyramid transform domain (PLBP) and Multi-Block Histogram LBP (BHLBP)	SVM and MLP	N/A	---	Varying lighting Conditions and Head Position
[28]	to detect human faces and eye regions	AdaBoost and Contour Circle	NA	91.5%	Head positions and face occlusion.
[29]	Face, head and eye regions detected	Haar-like features and adaptive boosting method	Intel Core2 Dou 2.66 GHz and 2 GB RAM	---	Computational Expensive, Environment Parameters
[30]	DWT and LBP	Decisions Tree C5.0, KNN, and SVM	Pentium 4, 1.6 Ghz with 256 MB RAM	---	Difficult to implement, Sunglasses and Environment parameters
[31]	Simultaneous eye detection and eye state estimation contextual features	cascade regression framework	NA	---	Environment Conditions, Sunglasses
[32]		three separate MCSVM classifiers	NA	---	High Computational cost
[33]	Geometric features from center of the eyes and the nose tip	manifold embedding and linear regression	2.67GHz Intel(R) Xeon(R) CPU	---	Computational high cost, Environment parameters, Sunglasses
[34]	fatigue related facial action units' (AU) identification employing a photometric stereo (PS) testbed for 3D AU reconstruction	deep neural networks	NA	95%	Environment parameters and Face occlusion
[35]	Multilayer CNN model	SoftMax classifier	NA	73.06%	Head position and Sunglasses
[36]	Viola and Jones method and corner points Shi-Tomasi detector	K-Means clustering, eye template	NA	---	Environment parameters and sunglasses
[37]	Drivers' fatigue state, like duration of eyes closure, blinking, and yawning, using video images, without equipping their bodies with any devices	Multiple Convolutional Neural Networks with Kernelized Correlation Filters (MCNN-KCF)	NA	95%	Environment parameters, head position and face occlusion
[38]	infrared imaging Gabor filter with template matching	eye-validation process and AdaBoost	CPU: i5; memory: 4GB	---	Environment parameters and sunglasses
[39]	PERCLOS Features estimation	PCA and SVM	Intel 2.80GHz CPU (single core) with 4 GB	---	Sunglasses, Head position
[40]	Multilayer CNN model	SVM	NA	73.06%	Head position and Sunglasses
[41]	PERCLOS Features	SVM	NA	Sensitivity: 95.58% and specificity: 100%	Environment parameters, sunglasses and head position
[42]	Kalman filter and Image Processing	SVM and RBF	NA		Environment parameters, sunglasses and head position
[43]	Motion-based blink detection	RNN model	NA	Accuracy: 90%	Environment parameters
[44]	Fatigue facial expression recognition	SVM with Boost-LBP features	NA	85.5%	Environment parameters and Sunglasses
[45]	recognize the face and then detecting the eye in every frame	PERCLOS measure and YawDD video dataset	---	80%	Face occlusion and Sunglasses but light illumination problem resolve
[46]	computer-vision-based driver fatigue detection systems	CNN	NA	91%	Environment parameters, sunglasses and head position
[47]	Detection of driver fatigue.	MTCNN	NA	97.06%	Environment parameters, sunglasses and head position
[48]	Hand gestures, facial expressions, behavioral features and head movements	AlexNet, VGG-FaceNet, FlowImageNet and ResNet, which use RGB videos of drivers	Intel Core i5 processor with 2.4 GHz CPU along with 8 GB RAM with python	85%	Environment parameters, sunglasses and head position
[49]	Vision-based driver fatigue recognition System	Features extraction (DCT, PCA) and classification (KNN,SVM)	NA	NA	Environment parameters, sunglasses and head position
[50]	Eyes, mouth and head motions used to detect fatigue	hierarchical temporal Deep Belief Network (HTDBN)	MATLAB implementation using a Core i5, 3.1GHz PC with 16GB RAM	84.82%	Environment parameters, sunglasses and face occlusion

[1] HOSVD: Higher order Singular value decomposition, MTCNN: multi-task cascaded convolutional neural networks, PCA: Principle component analysis, LDA: Linear discriminate analysis, SVM: Support vector machine, MLP: Multilayer Perceptron, KNN: K Nearest Neighbor MCSVM: separate multiclass support vector machine, ANNs: artificial neural networks, CNN: Convolutional neural network, RBF: Radial Basis kernel function, RNN: Recurrent Neural Network, PSO: particle swarm optimization DWT: Discrete Wavelet Transform, LBP: Local binary pattern, contextual features: continuous driving time, sleep duration time, and current time PERCLOS: Percentage eye openness tracking.

states, the author used many machine learning approaches (traditional and latest technologies such as deep learning).

In various research and engineering areas, artificial neural networks (ANNs) have been used [27]. The fuzzy inference method (FIS) [50] is another method that is used in the past for the development of V-HDx systems. The support vector machine (SVM) was used [12] which is based on the methodology of statistical learning and can be used for the classification of patterns and the inference of nonlinear interactions between variables. The SVM method's learning methodology makes it appropriate to test humans' cognitive states. SVMs can produce both linear and nonlinear models and the nonlinear models can be calculated as accurately as linear models. AdaBoost uses boosting algorithm [51] for pattern recognition. Its benefits include high performance in detection, quick process time for identification, and the potential expansion of recognition functionality. Also, Bayesian networks (BNs) classifiers are applied to the simulation of human behavior and have been used to diagnose inattention [52]. Despite these benefits, it takes considerable computing capacity and a vast volume of training data to construct a right and reliable BN model.

Another new development that has been to borrow from the field of speech processing and language technology techniques focused on hidden Markov models (HMMs) [53]. In [53], the authors built an HMM to predict route schemes using vehicle speed, steering-wheel angle, and braking force. To detect visual features from the driver's face, some authors suggested a single vision-based camera but some other authors used multiple cameras to effectively detect multiview features. Researchers in [23] propose a hybrid visual framework focused on driver's eye recognition for tracking driver's drowsiness. Using two cameras working in the visible and close infra-red spectra respectively, safe operation in-car conditions and processing in every day and night conditions were accomplished. A cascade of two classifiers conducts image recognition in both of these spectra. The exact description of the eye state in [24] is a criterion for the avoidance of car crashes due to driver sleepiness. These researches highlighted that previous classification approaches are susceptible to eye localization errors and visual obstructions.

In [25], another approach was presented combining image processing and machine learning techniques to detect driver drowsiness. The authors used Haar cascade classifier with SVM to recognize the current state of the eyes whether it is open or closed. In [26], a multilevel fusion system was developed through the measurement of eye blinking and 3D head pose estimation. They estimated the head rotation in the three directions by using only three interest points onto the driver's face. This system was evaluated by both DEAP and MiraclHB databases. Whereas in paper [27], the authors used two descriptors namely Pyramid transform domain (PLBP) and Multi-Block Histogram LBP (BHLBP) with SVM to detect only human eyes in grayscale images. In [28], an Adaboost and Contour Circle (ACC) algorithm was developed for recognizing whether eyes are in an open state or closed state.

In [29], the driver monitoring system for hypovigilance (fatigue and distraction) was developed based on the symptoms associated with the facial and eye regions. A fuzzy expert system (FIS) was used to combine the symptoms to estimate the level of driver hypovigilance. In [30], and Intelligent Drowsy Eye Detection, using an Image Mining (IDEDIM) system was developed. The authors have used a cascaded regression in [31], different visual features were used jointly to estimate the eye orientation and likelihood of openness. This paper describes three of the most powerful contextual attributes, i.e., continuous driving time, sleep length time, and current time, by using a multi-class SVM classifier to enable the real-time (online) identification of the exhaustion condition. As in [33], only three facial keypoints were used by the authors on a compilation of geometric features, namely the middle of the eyes, the corner of the mouth, and the tip of the nose.

In [34], a new V-HDx method was developed to incorporate machine vision for intelligent transportation through deep learning (DL) architecture and action units (AUs). They achieved a significant (95%) improvement in terms of accuracy compared to others. In that study, the author demonstrated an innovative driver fatigue detection method based on fatigue-related facial action units' (AU) identification. Whereas in [35], given an RGB input video of a car, a DL architecture is referred to as a deep drowsiness detection (DDD) network for learning successful features and detecting drowsiness. Experimental findings reveal that on the NTHU-drowsy driver detection benchmark dataset, DDD achieves 73.06 percent detection precision. Eye detection is a very fascinating area of study that verifies eye detection in [36]. In [37], a modern method called DriveCare was proposed which uses video clips to assess the exhaustion condition of the drivers, such as the length of eye closing, blinking, and yawning. To enhance the tracking precision, Multiple Convolutional Neural Networks with kernelized correlation filters (MCNN-KCF) were used. Different experimental results found that about 95 percent accuracy was reached by DriveCare. In [38], the authors initially use infrared imagery to capture the image of a vehicle at night and then created an algorithm to detect the face of the driver. Later, a modern eye-detection algorithm was applied to combine a Gabor filter with a similar prototype to find the location of the corners of the eye and introduce an eye-validation process thus improving the precision of the rate of detection. As the third step, to match the eyelid curve, they used a spline feature. This device has been checked with more than 200 faces on the IMM Face Database, as well as in a real-time simulation.

In [39], researchers proposed a vision-based fatigue warning system for the control of bus drivers. The system consists of head-shoulder detection units, face detection, eye detection, an estimate of eye openness, fusion, estimation of eyelid closure percentage (PERCLOS) sleepiness scale, and classification of fatigue level. To approximate the eye state based on adaptive convergence on the multimodal detections on both eyes, a fusion algorithm was integrated. The facial

markings on the observed face are pointed later in [40] and the eye aspect ratio, mouth opening ratio, and nose length ratio are consequently measured and the drowsiness length ratio is calculated based on their values. Deep learning algorithms were introduced in an offline way as well. Via SVM-based classification, a sensitivity of 95.58 percent and accuracy of 100 percent have been achieved. In [41], a template matching method for feature extraction was also extended to the kinematics of gait cycles segmented by our stepwise search-based segmentation algorithm with the SVM model for classification. The findings of fatigue identification through data from 20 recruited participants showed an accuracy of 90 percent.

The scientists implemented the first technique in [42] that senses wink completeness. These blinks vary in speed and length. However, the Recurrent Neural Network (RNN) is used as a classifier because of its suitability for sequence-based features. In [43], it was suggested that a new modular architecture method for early identification of the driver fatigue framework be proposed, taking into consideration the optimization of system output by the optimization of the particulate swarm (PSO). The findings obtained in terms of accuracy (90.4 percent), sensitivity (92.6 percent), and specificity are considered in line with the state-of-the-art (90.7%). In [44], by using the LBP technique, the authors implemented a fast and robust face detection algorithm to describe and normalize facial expression images and then used SVM to detect driver fatigue. A novel computer vision-based technique was developed in [45] to detect driver sleepiness from a video taken by a camera. The suggested approach was tested based on the YawDD public video dataset.

A V-HDx system was developed in [46] by detection of head-shoulder, face detection, eye detection, eye openness estimation, fusion, drowsiness measure percentage of eyelid closure (PERCLOS) estimation, and fatigue level classification. Using another approach in [47], the authors developed multi-task cascaded convolutional neural networks (MTCNNs) to detect driver fatigue by using multi-facial features. In [48], the authors developed a V-HDx system based on an architecture that detects the sleepiness of drivers. The authors used RGB videos of drivers as input and help in detecting drowsiness. They used mainly different transfer learning algorithms to classify these features into four classes. In [49], the authors used discrete wavelet transform (DWT) and entropy analysis to detect face features. Moreover in [50], the authors developed a V-HDx system based on the hierarchical temporal Deep Belief Network (HTDBN) method. In that research, the authors first extracted images of high-level facial and head features and then used them to identify symptoms linked to drowsiness. These are used to model and capture the interactive interactions between the gestures of the eyes, mouth, and head. In search of broad variations of driver footage, they also collected a huge detailed data set comprising different ethnicities, races, lighting conditions, and driving scenarios.
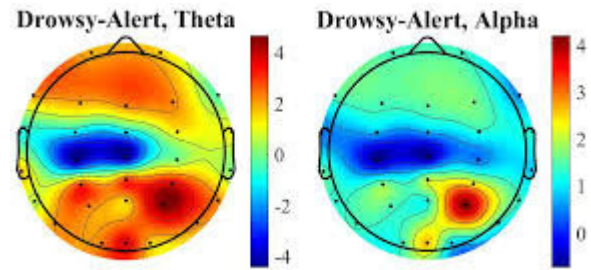
FIGURE 4. An example of electroencephalogram (EEG) signals [58] across each spectral band (delta, theta, alpha, beta, and gamma) for two mental states alert and Drowsy in all subjects with different stages.

B. MULTISENSOR-BASED (S-HDx) SYSTEMS

In comparison to V-HDx systems, the multisensor-based S-HDx approaches can help to detect driver drowsiness in multistage phases (drowsy, very drowsy, normal, and extreme drowsy). S-HDx systems use physiological or non-visual features to detect drowsiness. S-HDx is broadly classified into two main groups i.e., driver-based and vehicular-based hypovigilance detection systems. Driver-based characteristics typically apply to a driver's brain activity and heart rate, while vehicle-based characteristics include features such as brake pressure, vehicle speed variations, wheel angles, etc. In Table 4, most of the existing S-HDx systems are described in terms of different parameters.

1) DRIVE-BASED S-HDX SYSTEMS

The physiological analysis of an organism is specifically influenced by exhaustion and sleepiness leading towards dangerous situations. It was mentioned in [66] that the physiological indexes of the state of sleepiness could be differentiated from the normal state through different sensors. To assess physiological features, many authors used an individual's health measures (either individually or in combination) such as ECG, EOG, and EEG, etc. Among these methods, the most promising and feasible approach is based on calculating the EEG, as seen in Figure 4. The EEG describes the state of activity in the brain in terms of alert and drowse. It is stated that the activity of delta and theta waves is significantly increased and the activity of alpha waves is marginally increased while in a state of drowsiness [55]. The EEG is commonly recognized by researchers as a measure of the transition between various periods of sleep [67]. Many clinical [55]–[70]. These methods are intrusive because it becomes mandatory that drivers wear electrode helmets when driving to add EEG signals to driver HDx identification. However, some authors used many other sensors that are also participating in the detection of multistage of driver drowsiness.

Several studies have used EEG signals to classify driver drowsiness [55]–[70]. These methods are intrusive because it becomes mandatory that drivers wear electrode helmets when driving to add EEG signals to driver HDx identification. However, some authors used many other sensors that are also participating in the detection of multistage of driver

TABLE 4. Summary of state-of-the-art driver's fatigue detection systems using non-visual features and machine learning algorithms.

Ref.	Multisensor	Pre-processing	Analysis methods	Classification	Accuracy	Limitations
[55]	EEG	---	Drowsy or alert	DGM-SVM	91.10%	Two-stage detection and limited application in real-time processing.
[56]	EEG	DWT	Multiple entropy features	AR	98.3%	Two-stage detection
[57]	ECG	---	HRV	SAE	90%	Three-stage detection
[58]	EEG	--	Fatigue or Alert	Fisher-score	--	Four-Stage level detection
[59]	EEG	Filtering	Fatigue or Alert	AlexNet	90%	Four-Stage level detection
[60]	EEG	Filtering	Four Entropies	SVM	98.75%	Two-stage detection
[61]	EEG	--	Four Entropies	Ensemble Classifier	79.4%	Four-Stage level detection
[62]	ECG	--	Multi-index fusion of entropies	NN	-	Five-Stage level detection
[63]	ECG	--	Three-stage drowsiness	CNN	98.79%	Three-Stage level detection
[64]	EEG	Filtering	energy (α, β, θ) parameters	-	-	Three-Stage level detection
[65]	EEG	--		Hierarchical DL		Five-Stage level detection

DBN: Dynamic Bayesian network, ANN: Artificial neural network, ET: eye tribe eye tracker (ET), VnV: Visual and non-visual features, AUC: Area under-receiver operating curve, ACC: Accuracy, QS: (Quantified Self)-auto sensor, LSTM: Long short-term memory, RNN: Recurrent neural network, NA: Not applicable, PCCR: pupil center corneal reflection, CNN: Convolutional neural network, nV: non-visual features, RF: Random forest, PSO: Particle swarm optimization, ECG: electroencephalography, EEC: electrocardiography, EOG: electrooculography, SVM: support vector machine, KPCA: kernel principal component analysis

TABLE 5. State-of-the-art multimodal features based Hypovigilance detection (M-HDX) systems.

Cited. Ref.	Features and Parameters	Participants	Environment	Methods	Result
[15]	EEG, Gyroscope and Vision-based features	25	Real-time	NN	Less than 90% due to use of less dataset for training and test.
[19]	Heart rate, wearable sensors and facial expression	20	Real-time	RNN	Less than 85% on average.
[50]	Physical and Vehicular – Eye mouth, Physical and Biological	20 males and 3 females	Real-time and Simulation	DBN	Feature fusion provided best result compare to only visual or non-visual features Effect of occlusion
[72]	Vision-based, multisensor	25 participants	Real time	CNN, transfer learning and fine-tuning	--
[73]	Vision-based, mobile sensors	30	Simulator	CNN + RNN	85%
[76]	EEG, EOG and frequency of eye blinks	25	Simulator	SAE	87%
[77]	Eye blink, Cognitive physiological features	30	Real-time	CNN	90%

drowsiness. In practice, the EEG [12] based S-HDx systems have a temporal resolution of 0.001 s and a spatial resolution of 20 mm and are commonly used in the area of research into brain function. A driver's fatigue can be effectively detected using the frequency-domain features of EEG data (e.g., mean frequency, EEG continuum center of gravity, and energy contents of a, β, θ, and δ bands). Similarly, the time-domain characteristics of EEG data, such as the standard deviation, the average value, and the sum of the squares of the amplitudes, provide useful information on brain function. Due to its complicated structure, its intrusive nature, and effect on the efficiency of the driver, these techniques are still not very feasible for driving in a real environment. In studies of [67], [68], the authors used heart rate (HR) beats per minute to detect the sleepy stage. It was noted that there is a difference in HR at different times such as during long drives at night when a decrease in HR is reported. Moreover, the attention, mental behavior, and body energy of a driver also

factors that affected the HR [69], [70]. Respiration rate (RR) is also a valuable measure which is the number of exhaled and inhaled breaths in one minute.

The authors of [77] attempted to create a relation between somnolence and RR, according to which RR begins to fall with the initialization of drowsiness and sets in, and continues to fall before sleep begins. The Figure 6 shows the visual example of this five-stage based drowsiness detection system. The electrooculography (EOG) is another sensor that offers a calculation of the eye [78] monitoring. The EOG signals are changed by eye movements, such as eye action and blinking [60], [79]. In collecting EOG data, the positioning of EOG electrodes takes on special significance [85, [85]. Additionally, hypovigilance identification-ion systems focused on blink activity are highly person-dependent. Such programs do not work well with people suffering from mental disorders, since they may do more blinks in wakeful environments or their eyes can stay open even in sleepy conditions.

Electromyography (EMG) is another sensor for measuring and recording the electrical signal produced by the contraction of the muscle [86]–[88]. The HDx based on EMG data has been researched by several scholars [89]–[93]. A change in the middle-frequency portion towards the lower spectral band is observed during muscle contraction [94]–[96]. The Galvanic Skin Response (GSR) is another sensor used to measure skin conductance [92], [96]–[98]. Electro-Dermal Activity (EDA) offers a measure of skin conductance that varies due to sweat gland secretion. This technique, however, is particularly susceptible to moisture and temperature in the atmosphere. Also, some authors used Skin Temperature (ST) to measure the temperature of the surface of the skin. For example, five degrees of drowsiness are defined in [99] by measuring the temperature of the nasal skin, the temperature of the forehead, and the temperature of the muscles.

Compared to visual characteristics (V-HDx) systems, the physiological properties that are extracted by EEG, ECG, ST or EOG provide more reliability and precision [51]. A major challenge to these measures however is to design systems that overcome the intrusive nature of required hardware. Thus, the use of wireless technology such as Zigbee and Bluetooth [52] to calculate physiological signals in a non-intrusive fashion is a potential way to address this restriction but it can affect the reliability.

2) VEHICULAR-BASED S-HDX SYSTEMS

Research has been going on to use vehicular-based features to consider the state of drowsiness of the driver based on the study of car gestures, such as steering wheel rotation, lane holding, acceleration pedal movement, and braking, etc. The steering wheel rotation and the normal deviation of the lateral/lane direction are the two most widely used vehicle movement characteristics to identify the degree of driver drowsiness. The rotation of the steering wheel (SWM) [53] is measured by a steering angle sensor mounted on the steering column. Also, for environmental conditions such as minor road bumps and crosswinds, micro-corrections in steering are required. With rising drowsiness, drivers prefer to decrease the number of micro-corrections in the motions of the steering wheel. Due to its specialized technological needs, SWM-based systems have limited applicability.

Now we examine the most recent studies that utilized these driver's behavior, physiological and vehicular-based features to develop HDx systems. The EEG-based sensor was used in [55] to detect driver drowsiness. A hybrid deep generic model (DGM)-based support vector machine technique was developed for this purpose. The experimental results revealed that with 91.10% of sensitivity and 55.48% of accuracy. In another work [56], the authors used EEG data to study the multiple entropy fusion processes and compare several channel areas. A 98.3% accuracy, a 98.3% sensitivity, and a 98.2% precision were achieved. On the other hand, the authors in [57] developed a technique to predict driver fatigue using ECG as the heart rate variability (HRV) measure and deep learning model. In another work [58], the authors

used EEG signals to detect driver behavior using the Fisher score. In another recent work [59], the author developed EEG-based systems to detect hypovigilance state along with pre-processing filtering. They used AlexNet transfer learning architecture to classify signals as either normal or as a fatigue condition with an accuracy of 90%. In [60], the authors used four types of entropies to extract features from EEG signals and SVM was the classifier to detect the driver stage. The authors were reported 98.75% of detection accuracy.

The EEG-based framework was developed in [61] to detect driver drowsiness by using many different entropy metrics to evaluate EEG signals, including spectral entropy, approximate entropy, sample entropy, and fuzzy entropy. The average accuracy of the classification was found to be greater than 94% by gradient-boosted DT. Authors in [62] collected the ECG signals were and then applied multi-index fusion theory to correctly detect the level of fatigue. The ECG signals were collected in [63] and used with a convolutional neural network (CNN) model to detect fatigue-level. The highest 98.79% of detection accuracy was achieved. Likewise, in [64], the authors used energy (α, β, θ) parameters from EEG signals to study driver fatigue. Hierarchical deep learning algorithms with EEG signals were also used in [65] to detect driver fatigue on the publicly available dataset SEED-VIG. Similarly, in [66], the stack-based autoencoder (SAE) was used to detect four fatigue-level by using EEG signals. In [67], the authors used EEG sensors along with the features associated with electrocortical activities and eyeblink recognition analysis. The spectral analysis of the EEG samples immediately preceding the lane departure events showed changes in the spectral density.

In [68], the authors used optimization algorithms to detect the hypovigilance state of the drivers using EEG signals. It was a combinatorial technique where they used hierarchical extreme learning-based with particle swarm optimization known as PSO-H-ELM. To test the performance of PSO-H-ELM algorithms, they used different multisensor such as EEG, electrocardiogram (ECG), electrooculography (EOG), and surface electromyogram (sEMG). On average, the PSO-H-ELM algorithm achieved 83.12% of detection accuracy. The authors in [69] detected the drowsiness stage by using EEG signals with different entropies techniques. In contrast with these approaches, the authors in [70] suggested using multiple techniques for monitoring the drowsiness of the driver. The authors used a low-cost ECG sensor to extract data on heart rate variability (HRV) for fatigue detection. In another work [71], the authors used successive driving levels, heart rate variability (HRV) characteristics and determined fatigue-level by sample entropy (SampEn) and reported a high detect rate.

C. MULTIMODAL-BASED (M-HDx) SYSTEMS

In recent studies, the Multimodal-based (M-HDX) Systems have gained a lot of traction because of their ability to use deep-learning architectures to recognition driver's different activities and fatigue at different levels. The features
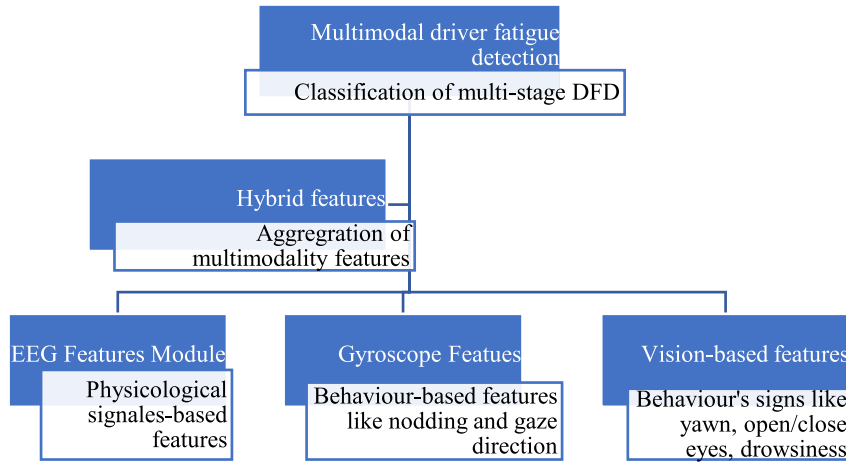
FIGURE 5. M-HDX systems are developed based on hybrid features combined by vision-based, sensors-based and vehicle-based features.
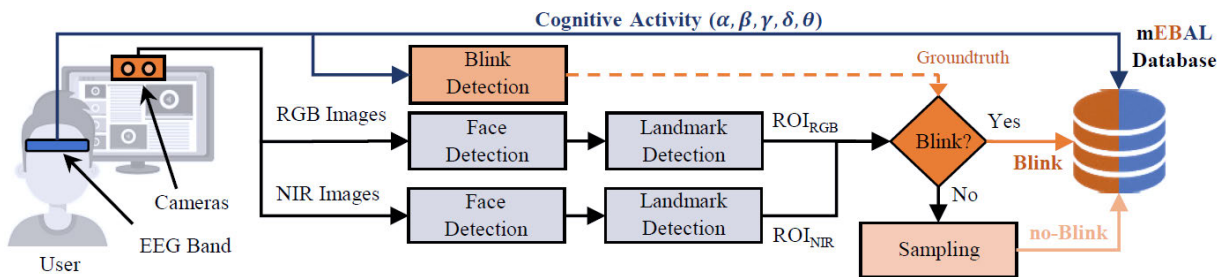
FIGURE 6. Grand-averaged scalp topographies across [77] each spectral band (delta, theta, alpha, beta, and gamma) for two mental states (a) and five drowsiness levels.

of M-HDx systems are described in Table 5 and visually presented in Figure 5. Many authors now use various forms of data [71], such as driver's physical conditions, audio and visual features, vehicle information, etc. To enhance the system's generalization ability for the development of HDx systems, the authors suggested integrating sensor data into the vision-based models. In recent years, early and late fusion techniques have been applied to combine multisensor and vision-based features into a single feature vector known as multimodal features vector. This integration of sensor data with vision-based driver detection significantly increases the overall performance of M-HDx systems. In this section, we describe some of the current state-of-the-art M-HDx systems.

A hybrid multimodal NN architecture was developed in [15] to detect driver drowsiness by integrating EEG data, Gyroscope data, and vision-based features. In another work [74], the authors suggested that a multimodal solution can have functionality that can be more effective in detecting the level of alertness of the drivers, based on developments in sensor technology. They developed a multimodal alertness dataset that comprising physiological, environmental, and vehicular features given by Ford Auto Company. Some other companies are trying to develop intelligent vehicles

that used advanced driver assistance systems (ADASs) [75]. In the paper [19], authors suggested that tracking the heart rate of drivers with a wearable tracker or extracting their facial expressions with an RGB camera are mostly used in the past. To overcome the intrusive nature of existing hardware, they used a single RGB-D camera used to resolve these issues to derive three fatigue characteristics such as pulse rate, level of eye-opening, and level of mouth opening. More significantly, this paper presented a new multimodal recurrent neural fusion network (MFRNN) that combines the three features to increase the accuracy of the detection of driver fatigue. To gain temporal information, a recurrent neural network (RNN) layer is used in the MFRNN. In particular, the authors are used different fusion strategies to get effective fatigue features. The MFRNN model was used to improve the efficiency of the drowsiness detection system.

In another work, distracted driving [72] was determined through the vehicle's driver image that includes the face, arms, and hands taken from a digital camera. Besides, in this work, the authors suggested that it is necessary to integrate a multisensor along with a vision-based camera to define multimodal features. They mentioned that it helps to improve the performance of M-HDx. In the first step, vision-based convolutional neural network (CNN) models were generated

using transfer learning and fine-tuning approaches. Afterward, the LSTM-RNN model was built in the second stage by using sensor and vision-based features together. Whereas in paper [73], the authors a mixture of camera video records and sensor data obtained on a cell phone to assess the behavior of the pilot. Image and sensor data were used together for the first three actions, and for the last three actions, only image data was used. Three separate deep learning methods were used for the classification process: CNN, CNN + RNN, and CNN + SVM. With 87%, the highest classification rate was reached by using CNN + RNN.

For the detection of driving fatigue, physiological signals such as EEG and EOG [76] have been successfully used in a single modality. A multimodal-based hypovigilance detection system was developed by integrating partial EEG and forehead EOG sensors to improve driving fatigue detection. The researchers found that the main region of the brain is an important place to detect effective EEG signals that can easily mix with forehead EOG signals. By experimental findings, they showed that when combined with forehead EOG to obtain mutual characteristics, the temporal EEG signals from six-channels provide the best output. Besides, to learn a better mutual representation, they suggested a novel multimodal fusion approach using the deep stack-based autoencoder (SAE) model. They measured useful components of the three characteristics when they already have saccade, blink, and fixation (the length of blink or saccade). Although some characteristics display an insignificant association with fatigue states, they did not use them in all. Afterward, they used EOG to reflect the forehead EOG rather than the conventional EOG. To monitor precise eye movement data, a new method was used by using eye-tracking of glasses and new PERCLOS measured is computed from Eq. (1) and Eq. (2) as follows:

$$ES_{msr} = blink + W_T CLOST.Interval(n) \quad (1)$$

$$Inval(n) = blink + saccade + fixation + PES_{msr} \quad (2)$$

This work introduces mEBAL [77] a multimodal database for the identification of eye blinks and the measurement of attention levels. The frequency of eye blink is related to cognitive function, and automated eye blink detectors have been suggested for several tasks, including estimating the level of attention, evaluating neurodegenerative disorders, identifying deceit, detecting drive exhaustion, or anti-spoofing of the face. The mEBAL dataset was created to help the other researchers effectively train the network for recognition of the hypovigilance stage based on multimodal features. The multimodal features were extracted using several experiments through vision-based cameras (both NIR and RGB) and EEG sensors. In total, the authors provided almost 6,000 samples from 38 different persons when they are engaged with e-learning tasks. This preliminary study was also included eye blink detection using CNN and persons' attention level was estimated based on their eye blink frequency. The overall system diagram of the mEBAL system is visually represented
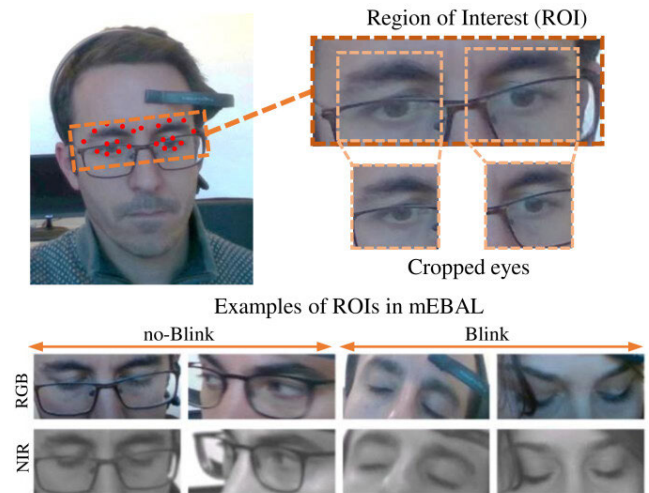
FIGURE 7. An example of mEBAL [77] dataset, which was collected in a constrained environment, but it is rich in pose, illumination changes, and other naturally-occurring factors.

in Figure 7. Surface electromyography (sEMG), electroencephalography (EEG), interface pressure of the seat, blood pressure, heart rate, and level of oxygen saturation data was collected as multimodal features in this paper [78].

Twenty male participants volunteered to conduct 60 minutes of driving on a static simulator in this study. In the back and shoulder muscle classes, findings from sEMG showed substantial physical exhaustion (r < 0.05). The EEG showed substantial (r < 0.05) increases in alpha and theta activity and significant decreases in monotonous driving beta activity. Another M-HDx system was developed in [79] by using physiological signals such as ECG, galvanic skin response, and respiration were recorded in real-world drive environments from 14 drives performed in a specified direction. Features from time, spectral and wavelet multi-fusion were widely extracted. Afterward, the prediction was performed based on sparse Bayesian learning (SBL) and principal component analysis (PCA) to look for the optimal feature sets. Average accuracy of 89% was achieved. In another study, the authors [80] used EEG, EOG, and ECG measurements for real-time mental fatigue detection. By using SVM, the authors achieved classification scores ranging from 80 ± 3 percent with a 4-s time window to 94 ± 2 percent with a 30-s time window. Similarly, in [81], the 68 electrodes of EEG/ECG/EOG and 8 channels of fNIRS data were simultaneously collected to develop an M-HDx system.

The EEG and EOG signals were used in the paper of [82] to detect the hypovigilance state of the drivers. The authors used the PERCLOS index, which is collected from eye monitoring glasses as a surveillance annotation. Then a novel electrode placement is identified for forehead EOG to enhance the feasibility and wearability. Similarly, in [83], a new multimodal architecture for the Electroencephalogram and Electrooculogram in-vehicle vigilance calculation was developed. A deep Long Short-Term Memory (LSTM) network was used to develop the HDx system. Also, in this research [84], they

proposed a Multimodal deep learning-based method (LSTM) that recognizes both visual and physiological changes in drowsiness. The combination of EEG and EOG signals was used in [85] to detect the hypovigilance state of the drivers and applied in SEED-VIG publicly available dataset.

The authors used 58 driver's multimodal data to test the HDx system in the paper of [86]. In that research, they used blink rate and posture information to form multimodal features. They achieved very impressive results to detect all stages of drowsiness (F1-score 53.6%, root mean square error 0.620). Whereas in [87], the authors used multimodal-based physiological signals to assess the alertness level of drivers. In [88], the authors used CNN and DRL models to detect driver fatigue based on EEG signals. Although in [89], the authors used 8 participants for collecting data from the motion signals (accelerometer and gyroscope), electrocardiogram (ECG), galvanic skin response, and CAN-Bus signals. Those signals are combined into a single multimodal feature vector to enhance the accuracy of the m-HDx system. To reduce the feature space, the authors have also contributed towards an optimal selection of multimodal features.

For detecting the drowsy state in humans, a multimodal method was studied in paper [90]. Video information and multisensor signals, for analysis, are two modalities considered. Visual data conveys a great deal of human alertness. For analysis and detection, the exact indicators from the video information need to be identified and captured. The multisensor signal indicating alertness of the human brain is an EEG signal. For the detection of a human being's drowsiness state, physical and mental alertness were analyzed. For drowsiness detection of humans in real-time, a framework was proposed. In [91], a methodology was developed based on the eye patterns of people, which monitored by video streams for the detection of drowsiness. In that study, the authors utilized data from the vision-based camera and sensor-based EOG signals. Whereas in [125], the authors developed Face2Multi-modal dataset based on In-vehicle multi-modal data. Especially in that paper, the authors investigated the drivers' estimates of heart rate, skin conductance, and vehicle speed to determine multimodal features. Those multimodal features are presented in the Face2Multi-modal open-source link. They believe that Face2Multi-modal offers a dataset that is already recorded the physiological status and vehicle status of drivers. This initiative serves as the building block for many current or future customized designs of drivers. More details and updates about the project are available online at Face2Multi-modal, available at https://github.com/unnc-ucc/Face2Multimodal/. To estimate the multi-modal states [126] of the driver (skin conductance and heart rates) and driving status (speed) through their facial expressions, the authors presented an in-vehicle real-time system. They used DenseNet as the model architecture for each type of data flow. Training information includes heart rates (per minute), skin conductance (uS), and velocity (km\h). The procedure for training is operated by PyTorch. Further training is work-in-progress, and we are

optimistic that the accuracy can be improved by comprehensive hyperparameter modification.

D. AVAILABLE DATA SOURCES

Different HDXs have utilized different data sources which are either freely available online or exist as private datasets. These datasets are used to extract both visual and non-visual features. Tables 4, 5, and 6 show some of the online data sets used by researchers to extract vision-based, multisensor-based, and multimodal-based features, respectively for training and testing of the machine learning classifier. The application of datasets used can be seen as Ultra-RLDD used in [112], NTHU-DDD in [50], MultiPIE in [113], 3MDAD in [114] as shown in Figure 8, MiraclHB in [26], BU-3DFE in [115] and AUC-DD in [116]. These datasets are based on computer vision technology to define visual features for driver fatigue. It can be noted from Figures 4 and 5, that we have a dataset of RGB images with 65-landmark points which can be used to train the network classifier for defining the features. To develop a robust HDXs system, it requires that those online and private data sources should be used to train the machine learning algorithms for the selection of effective visual-features. Even though, it is also required to train the classifier for recognition of driver fatigue in the smartphone or cloud computing-based platforms.

To develop hybrid HDX systems, several researchers are also using visual-features and various EEG-based biosensors to predict drowsiness. In practice, EEG signals are sometimes used to detect drowsiness, with three main building blocks. Several online data sources (see Table 8) are also available online to test and train the machine learning algorithms. A visual example of EEG spectrogram images visual with drowsiness and alert is displayed in Figure 5. In the past, most studies developed based on EEG biosensors that datasets are available online such as Min et al s' Fatigue-EEG [117], Cao et al s' Fatigue-Multi-channel [118], [119] EEG, and G. Cattan et al s' EEG-Alphawave [120].

One important dataset is mEBAL [77] contains a database of multiple types of blink detection and attention level measurement. The mEBAL dataset is used to train the network for defining effective visual and non-visual features of drivers. In this dataset, there are different parameters are measured and suggested for some functions including attention level, analysis of neuro-related diseases, detection of deception, fatigue, or coping with anti-fraud. mEBAL dataset develops preliminary data based on sensory input detection and calibration of camera sampling. In particular, three different sensors are monitored simultaneously such as Infrared Cameras (NIR) and RGB face-to-face capture and Electroencephalography (EEG) band for capturing user cognitive function and blinking events.

The mEBAL has a total of 3,000 blinking samples from both eyes received by 1 RGB camera and 2 NIR cameras. Each sample consists of 19 frames (around 600 ms.) with a total of 342,000 images ($3,000 \times 19 \times 2 \times 3$). Factors such as user status and light changes were considered at the time of

TABLE 6. Online state-of-the-art vision-based benchmark datasets to extract visual features that can be used to train and test the network.

Cited	Data Source	Features	Link URL
[112]	Ultra-RLDD dataset	Video-30 hours, 3-features: alertness, low vigilance, and drowsiness, frame rate:30fps, participant: 60	http://vlm1.uta.edu/~athitsos/projects/drowsiness/
[50]	NTHU-DDD Dataset	36 subjects, video: 9 and ½ hours, 5-different classes,	http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/
[113]	MultiPIE	different subjects, poses, illumination, occlusions, 68-landmark points	https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/
-	Kaggle-distracted drivers	22,424 images of size (480 × 680), 10 classes	https://www.kaggle.com/c/state-farm-distracted-driver-detection
[114]	3MDAD	60-subjects, 16-different actions	https://sites.google.com/site/benkhalifaanouar1/6-datasets#h.nzos3chrzmb2
[26]	MiraclHB	AVI format with a resolution of 640 × 480 and frequency 30 fps, 12: subjects	http://www.belhassen-akrout.com/
[115]	BU-3DFE	100: subjects with 2500 facial expression models	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
[116]	AUC-DD	44 (15/29)	https://heshameraqi.github.io/distraction_detection

TABLE 7. Online multisensor-based benchmark datasets that are available to train and test the network by using EEG sensors data.

Cited	Data Source	Features	Link URL	Format
[117] Min et al	Fatigue-EEG	12-subjects, 40-channel	https://figshare.com/articles/dataset/the_original_EEG_data_for_driver_fatigue_detection /5202739/1	. cnt
[118,119] Cao et al	Fatigue-Multi-channel EEG	27-subjects, 32-channels , EEGLab software	https://figshare.com/articles/Multi-channel_EEG_recordings_during_a_sustained-attention_driving_task/6427334/2	. set
[120] G. Cattan et al	EEG-Alphawave	20-subjects, 16-channels	https://zenodo.org/record/2348892#.X4bfptAzaM9	. mat

TABLE 8. Online Multimodal-based datasets are available to train and test the network by using EEG sensors data.

Cited	Data Source	Features	Link URL
[77] R Daza et al	mEBAL	3,000 blink samples from both eyes acquired with 1 RGB and 2 NIR cameras with EEG sensors	https://github.com/BiDAlab/mEBAL
[74] Abouelenien et al	Ford-dataset	Physiological, environmental, and vehicular modalities.	https://www.kaggle.com/c/stayalert/data
[82] Zheng et al	SEED-VIG	Twenty and ten subjects participated in the simulated and real-world driving environment experiments EEG and EOG during the experiments	http://bcmi.sjtu.edu.cn/~seed/seed-vig.html
[122] Massoz et al	DROZY	Face/Head Physiological	http://www.drozy.ulg.ac.be/
[123] Ortega et al	dBehaviourMD	41 hours of RGB, depth and IR videos from 3 cameras capturing face, body and hands of 37 drivers	https://dmd.vicomtech.org/
[124] Papakostas et al	CogBeacon	76 sessions collected from 19 male and female EEG functionality, facial keypoints, real-time self-reports on cognitive fatigue	https://github.com /MikeMpapa /CogBeacon-MultiModal_ Dataset_for_Cognitive_Fatigue
[125] Huang et al	Face2Multi-modal	The estimations of Heart Rate, Skin Conductance, and Vehicle Speed of the drivers used.	https://github.com/unnc-ucc/Face2Multimodal/
[126] Peng et al	BROOK-DenseNet	Heart Rate (minute), Skin Behavior (US), Speed (km\h)	https://github.com/unnc-ucc/BROOK

acquisition to mimic real-life e-learning situations. The 11 out of 38 students used glasses. A visual example of the mEBAL multimodal-based dataset is displayed in Figure 7. Several other important Multimodal-features based datasets that can be used as training models are described in Table 7. From this table, we have also visually displayed the most utilized multimodal datasets to develop M-HDX.

IV. TYPES OF DEEP LEARNING MODELS FOR HDx

In this section, we have described Deep-learning architectures (DLAs) with special emphasis on modern machine learning algorithms. For detection of driver drowsiness, several studies reported that the DLAs based algorithms achieved high accuracy [127], [128] with the reason to identify suitability for other authors. In this section, we described concepts, architectures, and techniques commonly utilized to detect driver drowsiness through visual features. Deep learning algorithms are recent techniques utilized for

FIGURE 8. An example of 3MDAD (Multimodal Multiview and Multispectral Driver Action Dataset) data set to predict and detect driver fatigue in case of real-time environment.

the detection of driver fatigue in the case of real-time [95], [97], [129]–[131]. Those DLAs methods were applied for pattern recognition and feature learning on mainly video frames. There are several variants of DLAs that have been

used in state-of-the-art systems for the detection of driver fatigue. Those DLAs systems are described and compared in Table 4. Based on the aim of DLAs, we have divided them into sub-categories for helping potential readers. Deep learning models for classification include Convolutional Neural Networks (CNN) [124], [129]–[131], Recurrent Neural Networks (RNN) [97], [95], [132], Stack-based Autoencoders (SAEs) [133], [134], Restricted Boltzmann Machine (RBM) and Deep Belief networks (DBN). For feature engineering, SAEs, RNN, and CNN are applied in the earlier driver fatigue systems but failed to classify data. However, RBM and DBN are the best deep learning algorithms for data representation but these algorithms did not test in the past driver fatigue detection systems. For our comparative analysis, we have selected CNN, RNN-LSTM, stack-based autoencoders (SAEs), and pre-train transfer learning models. Compare to these DL models, the transfer learning (TL) [135]–[141] algorithms are used.

V. METHODOLOGICAL COMPARISIONS

As a contribution of this work, the authors have compared the current leading research in the domain of driver fatigue through visual-based features techniques. Prediction of these features was performed by model-based, rule-based, supervised, and non-supervised and deep-learning algorithms (DLAs) in the past systems. In this paper, we have focused on DLAs, which were applied in the past to predict driver drowsiness instead of conventional machine learning techniques. As discussed earlier, to detect real-time driver fatigue, there is a dire need to develop an HDx system. Several traditional (MLP, SVM, or Boosting) and advanced deep learning algorithms [82], [89]–[128] (CNN, RNN, or DBN) are often used to develop a hypovigilance monitoring system. Compare to traditional-machine learning approaches, there is the latest trend toward the utilization of multilayer deep-learning algorithms (DLAs). These DLAs algorithms have reported very encouraging performance in a wide range of applications in particular detection of driver drowsiness. Moreover, the parameter setting and selection of layers are the most important factors in terms of DLAs in the case of detection of real-time driver fatigue.

To implement a real-time HDx system, the researchers have been using a set of rules or models to classify driver states during driving in different conditions. In brief, those systems are presented in Tables 3, 4, and 5 in the case of visual features, sensor-based features, and multimodal-based features, respectively. While reviewing these features, it was noticed that the multimodal and deep-learning (DLAs) in combination achieved higher performance compared to other approaches. Compare to simple DLAs, the authors are also combining different DLAs with traditional machine learning algorithms to develop a hybrid classifier. There is no trend observed in the past about transfer learning algorithms to develop HDx systems that are also considered in this paper to test the performance. The researchers have also utilized different variants of deep learning algorithms such as CNN and

RNN models. Those models are compared in Table 8 in terms of different architectures and accuracy. Though the parameter setting of CNN and RNN required a big training dataset and time complexity is the biggest thread for classification tasks.

Table 8 represents various hybrid and latest DLAs algorithms along with their detection accuracy which is not up-to-the-mark if we considered different parameters such as face occlusions, camera position, and type of sensors. In a real-time HDs system, there are certain limitations and problems of those models that are described in the upcoming sections. One of the main issues to set up parameters for DLAs is required very specialized knowledge about the model. For example, we have mentioned some parameters in Table 8 for the CNN model that is mostly utilized in the past to detect driver drowsiness as a feature selector and recognizer. It was observed that the main focus of researchers was on the design of assistant programs and aim to help drivers find their distraction time and create a warning alarm. Such hybrid or multimodal systems are providing a solid and reliable solution for predicting driver drowsiness. To increase the durability of the multimodal systems (M-HDx), it is possible to give weight to each sensor after training the mechanical separation. On these networks, the choice of machine learning algorithms is also an important task. The time and difficulty of the space always play a tradeoff for the deployment of these programs to monitor driver negligence.

Since several recent studies focused on the use of in-depth (DLAs) learning methods with multimodal features to develop an M-HDx system without understanding the practical analysis of this topic. In this paper, we have focused on overcoming this shortcoming. We have examined the impact of the construction of deep structures based on the discovery of driver fatigue. In particular, we are interested in studying the structural properties of driver fatigue in terms of visual and non-visual characteristics. Recently the detection of driver fatigue using in-depth learning algorithms has been greatly improved and in-depth by the number of layers and the number of processing units per layer to keep performance up-to-the-mark through the latest advances in computer vision systems. However, it is important to notice that the deep architectures are required to be as powerful as their calculation capabilities for real-time driver fatigue recognition. In this article, the data sets used in the past and some HDx systems are compared in different scenarios. As a contribution of this work, the authors have compared the current leading research in the domain of driver fatigue through visual-based features techniques. Prediction of these features was performed by model-based, rule-based, supervised, and non-supervised and deep-learning algorithms (DLAs) in the past systems.

In this paper, we have focused on DLAs, which were applied in the past to predict driver drowsiness instead of conventional machine learning techniques. As discussed earlier, to detect real-time driver fatigue, there is a dire need to develop an HDx system. Several traditional (MLP, SVM, or Boosting) and advanced deep learning algorithms [82], [89]–[128] (CNN, RNN, or DBN) are often

used to develop a hypovigilance monitoring system. Compare to traditional-machine learning approaches, there is the latest trend toward the utilization of multilayer deep-learning algorithms (DLAs). These DLAs algorithms have reported very encouraging performance in a wide range of applications in particular detection of driver drowsiness. Moreover, the parameter setting and selection of layers are the most important factors in terms of DLAs in the case of detection of real-time driver fatigue.

To implement a real-time HDx system, the researchers have been using a set of rules or models to classify driver states during driving in different conditions. In brief, those systems are presented in Table 3, Table 4, and Table 5 in case of visual features, sensor-based features, and multimodal-based features. While reviewing these features, it was noticed that the multimodal and deep-learning (DLAs) in combination achieved higher performance compared to other approaches. Compare to simple DLAs, the authors are also combining different DLAs with traditional machine learning algorithms to develop a hybrid classifier. There is no trend observed in the past about transfer learning algorithms to develop HDx systems that are also considered in this paper to test the performance. The researchers have also utilized different variants of deep learning algorithms such as CNN and RNN models. Those models are compared in Table 8 in terms of different architectures and accuracy. Though the parameter setting of CNN and RNN required a big training dataset and time complexity is the biggest thread for classification tasks.

A. SELECTIVE M-HDx SYSTEMS

To assess the performance of current state-of-the-art m-HDX systems, a qualitative comparative analysis is performed in this paper. Results of that comparison are presented here by using different datasets and environment scenarios. To perform statistical comparisons, we have fixed binary, ternary, and five stages of driver drowsiness in a real-time and simulator environment. In all these multistage, we considered the various parameters such as accuracy, sensitivity, specificity, reliability, F1-score, and area under the receiver operating curve (AUC). The testing environment contains face occlusion including sunglasses and night-time driving conditions. In general, visual, vehicular, and non-visual features are extracted in this comparison study to detect affective performance by using various machine learning and deep learning algorithms for predicting driver drowsiness. We have selected some recent papers to complete the state-of-the-art comparisons such as on CogBeacon-ML [124], BROOK-DenseNet [126], Ford-dataset [74], and Riani-M-HDx [128] systems on the real-time datasets. Those m-HDX systems were selected due to easy integration to a real-time system and focused on multimodal datasets. Those m-HDX systems have been briefly explained in the previous sections. The real-time processing on multisensor, vehicular, and environment parameters considered as compare to IoT-based architecture [126].

CogBeacon-ML [124] is a multimodal dataset designed to target the effects of cognitive fatigue in human performance. This dataset can be easily used to train the network for the prediction of real-time drive fatigue because the CogBeacon-ML system is developed based on multimodal features This dataset consists of 76 sessions obtained from 19 male and female participants conducting various variations of the Wisconsin Card Sorting Test (WCST), a common experimental and clinical psychological cognitive test designed to examine cognitive versatility, reasoning, and cognitive functioning aspects. During each session, the EEG functionality of the user, facial keypoints, real-time self-reports on cognitive exhaustion, and detailed success measurements obtained during the cognitive task are documented and thoroughly annotated (success rate, response time, number of errors, etc.). Also, there is an open-source machine learning analysis that can be used to predict cognitive fatigue by using multimodal features. Compare to facial points, the authors used mainly to detect cognitive fatigue by using EEG sensors at different sampling frequency rates. Afterward, they calculated the absolute band power for a given frequency range as the logarithm of the sum of the Power Spectral Density of the EEG data over that frequency range by using Eq. (3).

$$EEG(\emptyset) = \log \min \frac{1}{n} \sum_{i=1}^{n} \lceil G(f_i) \rceil^2 \qquad (3)$$

The FFT of the raw EEG signal G is where I and n are the minimum and maximum frequencies of the frequency band x, and $\emptyset$ is the five-frequency band. The relative frequency bands (R): γ, β, α, θ and δ at sampling frequency of 10 Hz are measured during each task. They also reported differences in the movement of the face to capture behavioral changes, capturing a set of 68 facial keypoints and four corners for bounding boxes, with a webcam mounted at a frame rate of 2 FPS on the top of the screen. The basic technique was applied by using Regression tree to classify facial keypoints.

The BROOK-DenseNet [126] is another platform that used M-HDx architecture by using BROOK multimodal dataset. The new version of BROOK consists of 34 passengers, driving respectively in automatic and manual modes for about 20 minutes. Their facial videos, multi-modal details, and driving status were protected by BROOK. In this BROOK dataset, the experiment included three driving conditions. Driving data was registered during the statistical analysis. The participants were asked to fill out a questionnaire on the cognitive facets of their experience during each driving session, including an appraisal of perceived confidence, relaxation, and situational awareness. For each driver, the analysis lasted roughly one hour. The BROOK dataset now consists of 11 data dimensions, including Facial Video, Vehicle Speed, Vehicle Acceleration, Vehicle Coordinate, Vehicle Ahead Size, Steering Wheel Coordinates, Throttle Status, Brake Status, Heart Rate, Eye Monitoring, and Skin Conductance. The visual example of BROOK data is displayed in Figure 9.
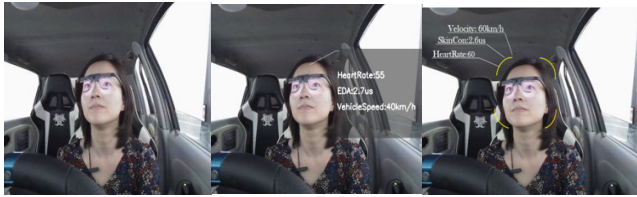
FIGURE 9. A visual example of Brook [126] data set to predict and detect driver fatigue in case of real-time environment.

FIGURE 10. An example of DROZY data set to predict and detect driver fatigue in case of real-time environment.

In BROOK-DenseNet, the authors developed DenseNet-BC-100 [126] convolutional neural network architecture for detection of driver drowsiness stage. The Brook Feature defined driving Status such as Vehicle Speed, Vehicle Accelerations, Vehicle Coordinate, Distance Ahead, Brake status, Heart rate, skin conductance, eye tracking. Those features are recorded in different sessions. Heart rates (per minute), skin conductance (uS), rpm, and training data are included (km\h) in this dataset. The procedure for training is operated by PyTorch. They also used an optimization technique to adjust hyperparameter that would enhance the accuracy of further training work-in-progress. In all, the 320,000 frames/images are used in the ready-to-train BROOK, and they have divided the training, test, and validation data sets into a ratio of 8:1:1. The BROOK-DenseNet was trained on an NVIDIA GeForce GTX 1660 GPU and specifics of the DenseNet architecture are given as follows. The author used many configurations based on DenseNet's recommendations. The model involves Depth/Layers 100 of Growth Rate, 12 of dense Blocks, Batch Size of 128, and Initial Learning Rate of 0.1 along with Testing Epochs of 50. The initial learning rate for our case is set at 0.1, which is separated by 10 at 50 percent and 75 percent of the cumulative number of training cycles.

Ford company provided another multimodal dataset [74] (Ford-dataset) for the training set, validation set, and test of different modalities such as physiological, environmental, and vehicular. They have been used a simulator to collect data from driving sessions. They used 100 drivers of different ages and genders and sequential measurements were collected every 100 milliseconds during the two minutes trial. There were 8 physiological features, 11 environmental features, and 11 vehicular features that were the feature distribution across the three modalities. For the training set, the cumulative number of instances was 604,329 (previously separate training and validation) and 120,840 for the test set, from 469 training, 31 validation, and 100 field trials. Warning and drowsy instances may be placed within the same prosecution. Visualization of this multimodal dataset with three characteristics. We made two interesting observations before processing the results. First, of all the cases in the training and test sets, there was one physiological feature and two vehicular features which had a value of zero. Thus, these three characteristics were eliminated, namely P8, V7, and V9, resulting in a final set of 7 physiological, 11 environmental, and 9 vehicular characteristics. The Ford-dataset mainly contains the "Phys + Env + Veh" features to fused different

modalities for selecting all 27 features in total. The author used Naive Bayes classifier to predict driver drowsiness based on these 27 features.

Riani-M-HDx [128] provided another dataset to use multimodal features by using action units (AUs), which were extracted from the facial deformation of drivers. They built an M-HDx system based on the OpenFace platform to identify 18 different AUs. In practice, the OpenFace framework is provided the best visual features to detect the driver's yawn. In the Riani-M-HDx dataset, the authors used four physiological sensors to extract statistical features. The final feature collection consisted of a total of 77 physiological characteristics, including 50 features of Blood Volume Pulse (BVP), 7 features of Skin Conductance (SC), 9 features of Respiration Rate (RR), 7 features of Skin Temperature (ST) and 4 features derived from the combined BVP and RR sensors, such as the mean and max-min heart rate differential, which is a measure of the variability of breath to heart rate. These all features were concatenated these new measurements into a new vector of 308 measurements after measuring the limit, mean, minimum, and standard deviations to produce four distinct vectors. They used a decision tree classifier (DT) to detect driver drowsiness.

A DROZY dataset as shown in Figure 10 is also provided to train and test the M-HDx system. However, we did not use it for comparisons because it does not come with source code to analyze the data. In contrast of this, we used both software and data sources of the same time flow are consolidated in this comparison study as displayed in Figure 11. To perform comparisons, we have used all these four data sources with comparisons to our IMSIU-DFD system [127] and results are reported in the upcoming sections. Major components and devices used to compare state-of-the-art hybrid systems by using IMAM University driver's simulator environment (IMSIU-DFD) [127] are represented in Table 9.

B. EXPERIMENTAL SETUP

To assess the performance of current state-of-the-art m-HDX systems, a qualitative comparative analysis is performed in this paper. Results of that comparison are presented here by using different datasets and environment scenarios. To perform statistical comparisons, we have fixed binary, ternary, and five stages of driver drowsiness in real-time and simulator environments. In these multistage, we considered the

TABLE 9. Performance comparisons of state-of-the-art driver fatigue detection based on visual-features and deep learning algorithms.

Cited	Deep learning models	Network Architecture	Dataset	Environment	Accuracy
[148]	MLP + CNN	3 CNN layers	NTHU	Real-time	80%
[149]	CNN	4 CNN Layers		Simulator	97%
[150]	CNN	5 CNN layers	DROZY	Real-time	--
[112]	RNN	6 CNN layers	NTHU-DDD	Simulator	--
[151]	CNN	4 CNN layers	CEW	Real-time	92.5%
[152]	CNN + Fuzzy logic	3 CNN layers	--	Simulator	--
[153]	CNN	6 CNN layers	CEW and ZJU	Real-time	94.9%
[154]	CNN + RNN-LSTM	5 CNN and 6 RNN	--	Simulator	--
[155]	CNN + RNN-LSTM	6 CNN and 4 RNN	--	Real-time	--

CNN: Convolutional neural network, RBF: Radial Basis kernel function, RNN: Recurrent
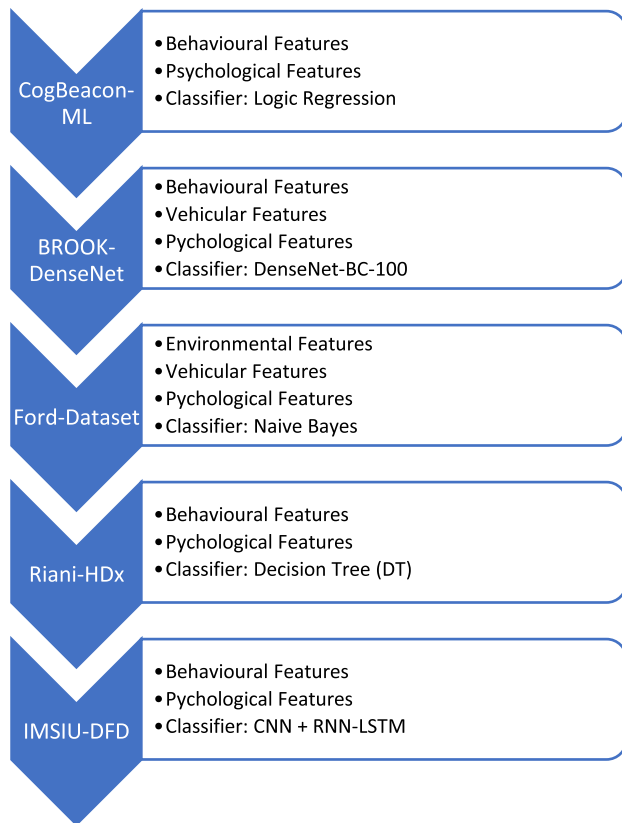
FIGURE 11. A visual example of four different datasets and software to perform comparisons, where CogBeacon-ML [124], BROOK-DenseNet [126], Ford-dataset [74], Riani-HDx [128] and IMISU-DFD [127].

various parameters such as accuracy, sensitivity, specificity, reliability, F1-score, and area under the receiver operating curve (AUC). The testing environment contains face occlusion including sunglasses and night-time driving conditions. In general, visual, vehicular, and non-visual features are extracted in this comparison study to detect affective performance by using various machine learning and deep learning algorithms for predicting driver drowsiness.

We have selected some recent papers with benchmarks to complete the state-of-the-art comparisons such as on CogBeacon-ML [124], BROOK-DenseNet [126],

TABLE 10. Major components and devices used to compare state-of-the-art hybrid systems by using IMAM university driver's simulator environment (IMSIU-DFD) [127].

Hardware	Parameters setup
CPU	Single core hyper threaded Xeon Processors @2.3Ghz (1 core, 2 threads)
Screen Resolution	1280 X 960
RAM	~12.6 GB
Network	Ethernet Network Driver
Hard Disk	100 GB
GPU	xTesla K80, having 2496 CUDA cores, 12GB GDDR5 VRAM
Arduino	Microcontroller: ATmega 328, Operating Voltage: 5V
Sensors	ECG, EEG, Skin Conductance for Arduino
EEG Sensors	Neurosky Mindwave EEG
ECG Sensors	Heart Rate variability (HRV)
Platform	Google Colab
Camera	Microsoft Kinect camera

Ford-dataset [74], and Riani-M-HDx [128] systems on the real-time datasets. Those m-HDX systems (as shown in Figure 11) were selected due to easy integration to a real-time system and focused on multimodal datasets. Those m-HDX systems have been explained in the previous sections. The real-time processing on multisensor, vehicular, and environment parameters was considered as compare to IoT-based architecture [126].

C. STATISTICAL METRICS FOR ASSESSMENTS

To evaluate the performance of HDx systems, we have used several statistical metrics to check the effectiveness of the classifier with respect to the dataset and environmental settings. In these comparisons, we have also used AUC curve to show the impact of different classification algorithms. The statistical metrics for evaluating the framework performance is measured as follows:

1) Accuracy (ACC) is the first measure that is used to determine the type of HDx system and is calculated by the following equation.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

2) Specificity (SP) is the second statistical measure that is used to precisely identify the authentic the results of m-HDx system and is measured by the following equation.

$$SP = \frac{TN}{TN + FP}$$

3) Sensitivity (SE) is the third statistical measure that is used to take into account the capacity of a classification model to correctly recognize the fatigue stage and is measured by the following equation.

$$SE = \frac{TP}{TP + FN}$$

4) Precision (PR) is the fourth statistical measure that is used to define an approximate value of the two measure values and is calculated by the following equation.

$$PR = \frac{TP}{TP + FP}$$

5) Matthews's correlation coefficient (MCC) is the fifth statistical measure that is used to define the superiority of classification technique and is calculated by the following equation.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

6) F1Score (F1) is the seventh statistical measure that is used to quantify the accuracy of a test and is formulated by the following equation.

$$F1 = 2 \times \frac{PR \times SP}{PR + SP}$$

where the approximate positive cases that are positive are TP. The expected negative case, which is negative, is TN. The expected negative cases which are currently positive are FN. These instances are referred to as errors of form. The FP parameter is the positive cases expected, which are in the true negative. These instances are referred to as error category one. Where the true positive rate (TPR) indicates the right number of decisions to identify the class. The overall accuracy of detection (ACC) for detecting HDx systems is calculated based on average. The performance of different HDx systems is evaluated by the estimators of precision (PR), sensitivity (SE), and specificity (SP). To compute these estimators, true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR) should be first measured. For the multi-class classification, we divided the comparison results into two major steps. First, we did comparisons based on the 2-class (binary), 3-class (ternary) and 5-classes based classification problems as describe in Table 8. The final estimators are calculated by taking the average results among all these experiments.

TABLE 11. Comparisons results based on binary, ternary and FIVE-STAGE multistage based HDx systems.

Classifiers	Definition of Class
Binary Class (2-classes)	DRW: Drowsy, ALT: ALERT (With/without distraction)
Ternary Classes (3-classes)	DRW: Drowsy, ALT: ALERT, Dist: Distracted (drowsiness with distraction + alertness with distraction)
Five Classes (5-classes)	V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert or drowsy, M: Moderate Drowsy, E: Extremely drowsy

D. EXPERIMENTAL ANALYSIS

Extracted from physiological cues, the feature set can contain redundant information (features that are unable to be distinguished between alertness and drowsiness), which, when added directly to the classifier, may lead to output loss. Therefore, prior to the classification stage, all the feature sets were subjected to a feature selection process. We have used paired t-test to classify only the statistically relevant traits ($p < 0.05$) that can distinguish between all participants between the warning and drowsy states. Before carrying out the t-test, we verified that the data were naturally distributed by means of the Kolmogorov-Smirnov test [41], as this is a requirement before the applying t-test. Various EEG-based systems for detecting drowsiness have been developed by using multisensor. The EEG signals are also used for the detection of drowsiness in this paper, with three key building blocks forming the HDX system. The proposed building blocks use both raw EEG signals and their corresponding spectrographs. In the first building block, while kurtosis.

The energy distribution and zero-crossing distribution properties are measured from the raw EEG signals. The EEG spectrograph images are used that are extracted spectral entropy and instantaneous frequency characteristics. To extract non-visual features, the deep feature extraction is used directly on EEG spectrograph images in the second building block using pre-trained RNN-LSTM model. In the third building block, the discrete wavelet transform (DWT) technique is used to decompose EEG signals into similar sub-bands. Instantaneous frequencies of the sub-bands are the spectrogram representations of the sub-bands and the collected statistical characteristics, such as mean and standard deviation of the instantaneous frequencies of the sub-bands. For the purposes of classification, each function group from each building block is fed to a long-short term memory (LSTM) network. Afterwards, ECG data channel was used to get time series of human heart variability to measure the movements of human body that were considered as statistical samplings. Then the distribution of those values in these samplings was analyzed by calculation of mean, standard deviation, skewness, and kurtosis. These measures very obtained through IMSIU-DFD [127] M-HDx system, which is used as base system to test and compare. Mostly, we have implemented behavioral and physiological features to develop this IMSIU-DFD system but to perform other
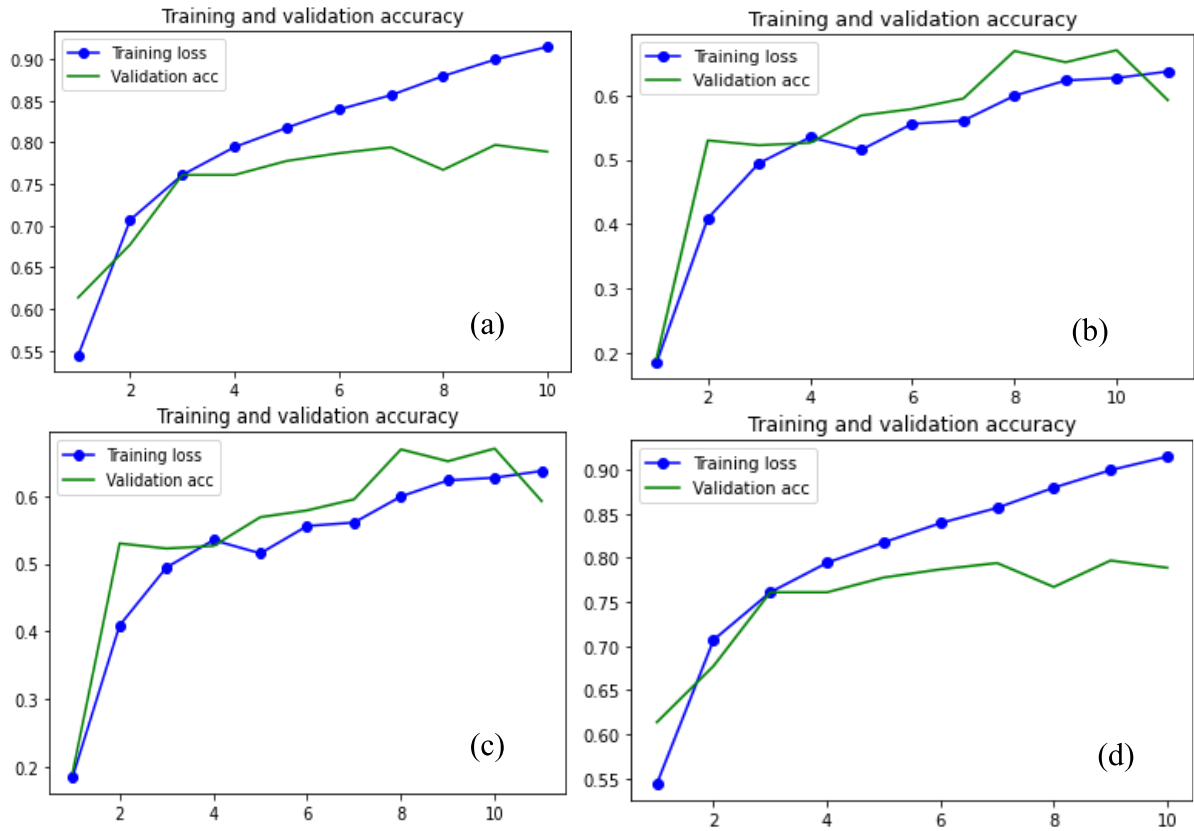
FIGURE 12. A training and validation accuracy graphs are represented where Figures (a) CogBeacon-ML [124], (b) BROOK-DenseNet [126], (c) Ford-dataset [74] and (d) Riani-HDx [128] by using 10 epochs.

comparisons based on environmental and vehicular, we have mounted a camera and utilized USB (OBD scan) to capture real-time vehicular data from Toyota RAV4, 2013 model.

By incorporating the different elements, multimodal classification was carried out. There were four different methods were used in two stages. Firstly, by concatenating the features obtained from the visual and physiological sources, we conducted an early modality fusion to create a single feature vector, which was then used for classification. The fused decision is computed by using Eq. (4) based on multistage classification problem.

$$multiStage\_HDx^{1,2,3,4,5}$$
$$= Major\_Vote(\max_{n}(Behav(.))$$
$$\cdot \min_{n}(ln(Phych)) + Environment + Vehicular) \quad (4)$$

where the parameter n is the number of features that are extracted from each different modality, is the function to indicate visual-features from the driver, and shows physiological features of the drivers. Also, the majority voting scheme is used to predict the state of driver drowsiness. Also, the environmental and vehicular features are integrated to compare the performance with other state-of-the-art M-HDx systems. To evaluate the performance, we used four different M-HDx systems such as CogBeacon-ML [124], BROOK-DenseNet [126], Ford-dataset [74], and Riani-HDx [128].

By using 10 epochs, the training and validation accuracies are also computed. A visual representation of training and validation accuracy graphs are represented in Figure 12 where Figures (a) CogBeacon-ML [124], (b) BROOK-DenseNet [126], (c) Ford-dataset [74] and (d) Riani-HDx [128] on a total of 10 epochs.

To perform experiments, we used various state-of-the-art machine learning approaches to characterize the driver's behaviors especially fatigue. These comparisons help us to make a clear difference between drowsy and fatigue parameters in different conditions and levels. The five-stages were measured such as used V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy based on SE: sensitivity, specificity: SP and AUC: area under the receiver operating curve. In all subsequent tables and paragraphs, we used notation "*/*/*" corresponding to SE/SP/AUC values. Also, the first highest performance values are displayed in terms of the bold and second one in terms of underline. We run a 10-leave-one-out cross-validation scheme for our evaluation, and the average performance is presented.

Experiment I: Table 12 represented state-of-the-art comparison results based on five-stage based hypovigilance detection (HDx) systems using driver's behaviors features and 15 different subjects in 25 minutes of recorded time.

TABLE 12. State-of-the-art comparison results based on five-stages based hypovigilance detection (HDx) systems using driver's behaviors features and 15 different subjects in 25 minutes of recorded time, where "∗/∗/∗" means SE/SP/AUC values and high values provide better results.

Classifiers	V	F	N	M	E
CogBeacon-ML [124]	45.6/47.5/0.46	40.2/43.2/42	**44.5/46.2/0.45**	37.6/40.5/0.38	41.6/43.5/0.42
BROOK-DenseNet [126]	<u>46.4/48.5/0.46</u>	<u>41.2/44.4/0.43</u>	40.5/43.4/0.42	<u>38.6/39.5/0.39</u>	<u>43.6/45.5/0.44</u>
Ford-dataset [74]	NA	NA	NA	NA	NA
Riani-M-HDx [128]	**50.6/52.5/0.51**	**43.5/44.3/0.44**	<u>40.5/45.5/0.44</u>	**40.6/41.5/0.40**	**44.6/47.5/0.45**

V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy, SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve.

TABLE 13. State-of-the-art comparison results based on five-stages based hypovigilance detection (HDx) systems using driver's physiological features and 15 different subjects in 25 min of recorded time, where "∗/∗/∗" means SE/SP/AUC values and high values provide better results.

Classifiers	V	F	N	M	E
CogBeacon-ML [124]	35.6/37.5/0.36	30.2/33.2/0.32	<u>34.5/36.2/0.32</u>	30.6/30.5/0.30	31.6/32.5/0.31
BROOK-DenseNet [126]	<u>36.4/38.5/0.36</u>	<u>31.2/34.4/0.33</u>	30.5/33.4/0.32	30.6/30.5/0.31	<u>31.6/33.5/0.32</u>
Ford-dataset [74]	30.1/32.5/0.31	28.5/24.4/0.26	30.5/34.1/0.33	<u>30.6/31.5/0.30</u>	30.1/32.5/0.31
Riani-M-HDx [128]	**46.2/48.5/0.46**	**43.5/44.3/0.42**	**40.5/45.5/0.43**	**50.6/54.5/0.51**	**60.6/63.5/0.61**

V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy, SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve.

In Ford-dataset [74], there are unavailable behavior features so we did not perform comparisons. On average, the CogBeacon-ML [124] system achieved somewhat lower results based on four-stage (V, F, M, E) compare to N-stage. However, in the case of BROOK-DenseNet [126], the performance is somewhat similar to CogBeacon-ML [124] system. It was noticed that in four-stage (V, F, M, E), both approaches achieved the same performance level when used driver's behavior features. Also, CogBeacon-ML achieved best performance in terms of N-stage than BROOK-DenseNet and Riani-M-HDx. Compare to these methods, the Riani-M-HDx [128] achieved better performance but this is also not up-to-the-mark to implement in a real-time environment. As shown in this table, the Riani-M-HDx system achieved the performance results as V-stage of 50.6/52.5/0.51, F-stage of 43.5/44.3/0.44, M-stage of 40.6/41.5/0.40, and E-stage of 44.6/47.5/0.45. This perform-ance was achieved due to the use of the best features and deep-learning (DL) architecture compare to the traditional logic regression technique used in the CogBeacon-ML system. Besides, the deep-learning architecture utilized in (BROOK-DenseNet and Riani-M-HDx) systems required many training samples, fine-tuning, and selection of layers. In contrast, if a vision-based camera is unavailable to get due to environmental or face occlusion factors then all these methods were completely failed.

Experiment II: Five-stage based hypovigilance detection (HDx) systems using physiological features and 15 different subjects in 25 minutes of recorded time are described in Table 13. On average, the Ford-dataset system achieved somewhat very lower results in terms of four-stage (V,F,N,E) compare to other M-stage. The M-stage obtained second highest result by Ford-dataset system. Where the case

of the CogBeacon-ML system achieved somewhat similar performance as the Ford-dataset approach. In addition, the BROOK-DenseNet achieved similar performance as Ford-dataset and CogBeacon-ML systems and achieved results. Compare to these methods, the Riani-M-HDx achieved better performance. The higher results are obtained due to the best utilize of psychological features by the Riani-M-HDx system compare to other systems. Also in the case of environment and face occlusion, this system obtained good performance. Due to the use of psychological features compares to behavior features, the Riani-M-HDx system achieved the best performance. In terms of behavior features, the usage of DL methods achieved higher detection accuracy but this is also not up-to-the-mark to implement in a real-time environment.

Experiment III: Table 14 represents the comparison results obtained based on five-stage based hypovigilance detection (HDx) systems using driver's vehicular features on 15 different subjects and recorded time is 25 minutes. Among four different systems, there are only two HDx systems (BROOK-DenseNet [126], Ford-dataset [74]), which used vehicular features to detect driver fatigue. The obtained results are mentioned in this table. The results show that poor performance is obtained by using only vehicular features and even used DL architectures and normal environmental conditions. On average, the Ford-dataset and BROOK-DenseNet achieved similar performance when used a simulator as comparisons. The Ford-dataset HDx systems achieved better than BROOK-DenseNet.

Experiment IV: The comparative results obtained from experiment IV are represented in table 15. There is only one Ford-dataset HDx system that used environmental features

TABLE 14. State-of-the-art comparison results based on five-stages based hypovigilance detection (HDx) systems using driver's vehicular features and 15 different subjects in 25 min of recorded time, where "∗/∗/∗" means SE/SP/AUC values and high values provide better results.

Classifiers	V	F	N	M	E
CogBeacon-ML [124]	NA	NA	NA	NA	NA
BROOK-DenseNet [126]	26.4/28.5/0.26	21.2/24.4/0.23	20.5/23.4/0.22	28.6/29.5/:0.29	**23.6/21.5/0.22**
Ford-dataset [74]	**27.1/29.5/0.26**	**22.5/24.4/:0.24**	**20.5/24.1/:0.23**	**29.6/20.5/0.27**	23.6/22.5/0.23
Riani-M-HDx [128]	NA	NA	NA	NA	NA

V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy, SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve.

TABLE 15. State-of-the-art comparison results based on five-stages based hypovigilance detection (HDx) systems using driver's environmental features and 15 different subjects in 25 min of recorded time, where "∗/∗/∗" means SE/SP/AUC values and high values provide better results.

Classifiers	V	F	N	M	E
CogBeacon-ML [126]:	NA	NA	NA	NA	NA
BROOK-DenseNet [128]	NA	NA	NA	NA	NA
Ford-dataset [74]	**27.1/29.5/0.26**	**22.5/24.4/0.24**	**20.5/24.1/0.23**	**29.6/20.5/0.27**	**23.6/22.5/0.23**
Riani-M-HDx [128]	NA	NA	NA	NA	NA

V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy, SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve.

for the detection of driver fatigue-stage. We have carried out state-of-the-art comparative findings in this table based on five-stage hypovigilance detection (HDx) systems using driver's environmental characteristics on 15 different subjects and a reported period of 25 minutes. There is only one HDx system out of four different systems (Ford-dataset [74]) that use environmental features to detect driver fatigue. In this table, the results obtained are described and we have observed very poorer efficiency. On average, the performance results for the Ford dataset in five-stages are obtained as V-stage of 27.1/29.5/0.26, F-stage of 22.5/24.4/0.24, N-stage of 20.5/24.1/0.23, M-stage of 29.6/20.5/0.27, and E-stage of 23.6/22.5/0.23. These results indicate that it should be used separately to detect driver fatigue in the 5-stage.

Experiment V: These experimental results are presented in Table 16 by using multimodal features (Behavioral + Psychological) and it has been mostly utilized in the past to develop the HDx systems. These results are based on five-stage based HDx systems using multimodal (Behavioral + Psychological) features, early fusion, and tested on 15 different subjects in 25 minutes of recorded time. On average, the Riani-M-HDx achieved better performance compare to other three systems. Similarly, the BROOK-DenseNet obtained second highest results in most of the stages to detect driver fatigue when used multimodal features.

The multimodal representations display a gradual increase in multiclass grouping relative to the use of the characteristics of individual modalities. When using early fusion, the precision exceeds 45% for the five-classes-based M-HDx systems. In comparison, by using early fusion, the per-class AUC in all multi-class methods reached the highest performance. The precision of the late fusion process, however, did not

disclose comparable findings to early fusion. What is most interesting is that despite the fact that early modality fusion outperforms all other approaches. This may suggest a close association between visual characteristics and the states identified by the classes "Alert" and "Drowsy" but also their failure to perform as well when distractions are targeted as explicit states. Physiological data tend to have somewhat more consistent findings and are likely to have a positive impact on our multimodal studies as well. The small number of available data, which do not accurately reflect the targeted groups, particularly in the 5-stage and 3-stage problems, maybe a potential explanation for this finding. The problems related to face occlusion, head position, and environmental conditions are solved by using multimodal features in DL architecture. In addition, we expect to further explore these observations.

Next, there is a dire need to determine the effect of machine learning algorithms on multimodal-based features (M-HDx) systems for the detection of different levels of driver fatigue. We have selected traditional machine learning algorithms such as ANN, SVM, Logic Regression +ANN, and hybrid DL architectures (HDLA) such as CNN +SVM, CNN +Naïve Bayes, CNN+ RNN-LSTM, and DenseNet-BC100 based on early fusion. These machine learning algorithms were selected because they have been mostly utilized in the past for performing experiments no VI to predict driver's hypovig-ilance states. These experiments are conducted based on SE, SP, ACC, PR, MCC, AUC, and F1 statistical measures on two-, and three-stages-based DFD detection systems. The notation "∗/∗/" means that drowsy/alert/distracted values and high values provide better performance results.

TABLE 16. State-of-the-art comparison results based on five-stages based hypovigilance detection (HDx) systems using multimodal (Behavioral + Psychological) features, early fusion and 15 different subjects in 25 min of recorded time, where "*/*/*" means SE/SP/AUC values and high values provide better results.

Classifiers	V	F	N	M	E
CogBeacon-ML [124]	75.6/77.5/0.76	65.2/67.2/0.66	63.5/64.6/0.64	65.6/68.0/0.66	51.6/53.5/0.52
BROOK-DenseNet [126]	87.4/89.5/0.88	90.2/94.5/0.93	85.5/87.5/0.86	87.6/89.5/0.88	85.6/88.5/0.86
Ford-dataset [75]	80.1/83.5/0.82	78.5/80.4/0.79	81.5/83.1/0.82	84.2/85.5/0.84	74.6/76.5/0.75
Riani-M-HDx [130]	**90.6/93.5/0.92**	90.5/93.3/0.91	**90.5/95.5/0.93**	**92.6/94.5/0.92**	**95.6/97.5/0.96**

V: Very alert (without distraction), F: Fairly alert (drowsiness with distraction), N: Neither alert nor drowsy, M: Moderate Drowsy, E: Extremely drowsy, SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve

TABLE 17. State-of-the-art comparison of different classifier used in CogBeacon-ML [124] based on two-stages based hypovigilance detection (HDx) systems using multimodal features (Behavioral + Psychological), early fusion, different classifiers and 20 different subjects in 35 minutes of recorded time, where "*/*" means DRW/ALT values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Binary Classification Results=DRW: Drowsy, ALT: ALERT (With/without distraction)							
ANN	45.6/47.5	47.6/48.5	45.2/46.1	47.6/49.5	0.25/0.26	0.45/0.46	0.45/0.46
SVM	45.6/47.5	47.6/48.5	46.3/45.2	49.6/47.5	0.25/0.27	0.45/0.47	0.45/0.47
Logic Regression +ANN	45.6/47.5	47.6/47.5	45.4/44.3	48.6/48.5	0.25/0.27	0.45/0.47	0.45/0.47
CNN +SVM	46.6/47.5	48.6/49.5	47.2/48.2	50.6/51.5	0.26/0.27	0.46/0.47	0.46/0.47
CNN +Naïve Bayes	48.6/49.5	50.6/50.5	49.0/50.5	53.6/52.5	0.28/0.29	0.48/0.49	0.48/0.49
CNN+ RNN-LSTM	**50.6/51.5**	**52.6/52.5**	**51.0/50.5**	**54.6/53.5**	**0.20/0.21**	**0.54/0.55**	**0.53/0.54**
DenseNet-BC100	51.0/51.1	52.1/52.1	51.2/50.2	53.2/53.3	0.21/0.22	0.51/0.52	0.51/0.52

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

TABLE 18. State-of-the-art comparison of different classifier used in BROOK-DenseNet [126] based on two-stages based hypovigilance detection (HDx) systems using multimodal features (Behavioral + Psychological), early fusion, different classifiers and 20 different subjects in 35 minutes of recorded time, where "*/*" means drw/alt values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Binary Classification Results=DRW: Drowsy, ALT: ALERT (With/without distraction)							
ANN	85.6/87.5	84.6/86.5	85.2/86.1	80.6/82.5	0.55/0.56	0.85/0.86	0.85/0.86
SVM	85.6/87.5	85.6/88.5	86.3/85.2	81.6/83.5	0.55/0.57	0.85/0.87	0.85/0.87
Logic Regression +ANN	85.6/87.5	84.6/86.5	85.4/84.3	81.6/88.5	0.55/0.57	0.85/0.87	0.85/0.87
CNN +SVM	86.6/87.5	87.6/85.5	87.2/88.2	83.6/81.5	0.56/0.57	0.86/0.87	0.86/0.87
CNN +Naïve Bayes	**88.6/89.5**	**88.6/89.5**	**89.0/80.5**	**84.6/82.5**	**0.58/0.59**	**0.88/0.89**	**0.88/0.89**
CNN+ RNN-LSTM	80.6/81.5	80.6/80.5	81.0/80.5	82.6/81.5	0.50/0.51	0.80/0.81	0.80/0.81
DenseNet-BC100	81.6/81.5	82.6/82.5	81.0/80.5	81.6/82.5	0.50/0.51	0.80/0.81	0.80/0.81

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

Experiment VI: From Table 17, the CogBeacon-ML system achieved better performance in the case of hybrid DL algorithms such as CNN + RNN-LSTM and DenseNet-BC100 compare to other ML algorithms based on two-stages of the driver such as drowsy and alert. On average, the ANN achieved very low performance. Though in the paradigm of SVM, this M-HDx system achieved similar performance. Also, in the case of CNN + Naïve Bayes and CNN + RNN-LSTM model achieved the best performance and comparable to the DenseNet-BC100 model. These results are obtained based on M-HDx features (behavioral and psychological) when measured on 20 different drives in 35 minutes of recorded time. These experiments show that the performance of CogBeacon-ML is improved if used M-HDx features with HDLA. However, it is not having better results compare to other M-HDx systems. It is also observed that if the size of training data is slightly changed then many variations in prediction results are observed. Similarly, we have done experiments on other M-HDx systems (BROOK-DenseNet, FORD-dataset, and Riani-M-HDX) based on two-stage DFD decisions. Those results are reported from Table 18 to Table 20. Based on these experiments, the BROOK-DenseNet and Riani-M-HDx achieved higher performance values compared to Ford-dataset and

TABLE 19. State-of-the-art comparison of different classifier used in Ford-dataset [126] based on two-stages based hypovigilance detection (HDx) systems using multimodal features (Vehicle + Environment + Psychological), early fusion, different classifiers and 20 different subjects in 35 minutes of recorded time, where "$*/*$" means drw/alt values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Binary Classification Results=DRW: Drowsy, ALT: ALERT (With/without distraction)							
ANN	55.6/57.5	57.6/58.5	55.2/56.1	57.6/59.5	0.35/0.36	0.55/0.56	0.55/0.56
SVM	55.6/57.5	57.6/58.5	56.3/55.2	59.6/57.5	0.35/0.37	0.55/0.57	0.55/0.57
Logic Regression +ANN	55.6/57.5	57.6/57.5	55.4/54.3	58.6/58.5	0.35/0.37	0.55/0.57	0.55/0.57
CNN +SVM	56.6/57.5	58.6/59.5	57.2/58.2	60.6/61.5	0.36/0.37	0.56/0.57	0.56/0.57
CNN +Naïve Bayes	58.6/59.5	60.6/60.5	59.0/50.5	63.6/62.5	0.37/0.39	0.58/0.59	0.65/0.67
CNN+ RNN-LSTM	**60.6/61.5**	**62.6/62.5**	**61.0/60.5**	**64.6/63.5**	**0.39/0.36**	**0.60/0.61**	**0.68/0.66**
DenseNet-BC100	60.0/61.1	62.1/62.1	60.2/60.2	64.2/63.3	0.31/0.32	0.57/0.59	0.59/0.60

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

TABLE 20. State-of-the-art comparison of different classifier used in Riani-M-HDx [128] based on two-stages based hypovigilance detection (HDx) systems using multimodal features (Behavioral + Psychological), different classifiers and 20 different subjects in 35 minutes of recorded time, where $*/*$" means drw/alt values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Binary Classification Results=DRW: Drowsy, ALT: ALERT (With/without distraction)							
ANN	88.6/90.5	82.6/84.5	80.2/82.1	77.6/79.5	0.60/0.61	0.81/0.82	0.75/0.76
SVM	88.6/90.5	84.6/86.5	81.3/82.2	78.6/80.5	0.62/0.63	0.82/0.83	0.77/0.78
Logic Regression +ANN	88.6/90.5	83.6/85.5	84.4/86.3	81.6/83.5	0.63/0.65	0.83/0.84	0.79/0.81
CNN +SVM	89.6/90.5	84.6/86.5	85.2/87.2	82.6/83.5	0.63/0.65	0.84/0.85	0.80/0.81
CNN +Naïve Bayes	**91.6/91.5**	**90.6/91.5**	**89.0/87.5**	**85.6/86.5**	**0.68/0.69**	**0.88/0.89**	**0.88/0.89**
CNN+ RNN-LSTM	88.6/86.5	85.6/86.5	86.0/85.5	83.6/81.5	0.63/0.64	0.84/0.83	0.80/0.81
DenseNet-BC100	89.6/90.5	86.6/85.5	87.0/88.5	84.6/82.5	0.65/0.64	0.87/0.88	0.85/0.86

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

TABLE 21. State-of-the-art comparison of different classifier used in BROOK-DenseNet [126] based on three-stages based hypovigilance detection (HDx) systems using multimodal features, different classifiers and 25 different subjects in 40 minutes of recorded time, where "$*/*$/" means drw/alt/dist values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Ternary Classification Results= DRW: Drowsy, ALT: ALERT, Dist: Distracted (drowsiness with distraction + alertness with distraction)							
ANN	76.6/77.5/76.5	74.6/75.5/75.5	85.2/86.1/86.5	81.6/82.5/ 81.5	0.55/0.56/0.57	0.85/0.86/0.84	0.85/0.86/0.83
SVM	77.6/77.5/76.5	76.6/77.5/76.5	86.3/85.2/86.5	82.6/81.5/ 80.5	0.55/0.57/0.57	0.85/0.87/0.85	0.85/0.87/0.82
Logic Regression +ANN	78.6/77.5/76.5	77.6/77.5/76.5	85.4/84.3/86.5	83.6/85.5/ 86.5	0.55/0.57/ 0.57	0.85/0.87/0.86	0.85/0.87/0.81
CNN +SVM	84.6/87.5/86.5	85.6/86.5/87.5	87.2/88.2/86.5	84.6/83.5/86.5	0.56/0.57/0.57	0.86/0.87/0.85	0.86/0.87/ 0.83
CNN +Naïve Bayes	**89.6/89.5/86.5**	**88.6/88.5/86.5**	**89.0/80.5/86.5**	**85.6/84.5/86.5**	**0.58/0.59/ 0.57**	**0.88/0.89/0.87**	**0.88/0.89/0.82**
CNN+ RNN-LSTM	80.6/83.5/86.5	82.6/82.5/86.5	81.0/80.5/ 86.5	82.6/83.5/ 82.5	0.50/0.51/0.57	0.80/0.81/0.84	0.80/0.81/0.85
DenseNet-BC100	83.6/81.5/86.5	82.6/82.5/86.5	81.0/80.5/86.5	83.6/81.5/82.5	0.50/0.51/0.57	0.80/0.81/0.86	0.80/0.81/0.85

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

CogBeacon-ML fatigue detection systems. However, different trends in hybrid deep-learning algorithms (HDLA) are observed. Still, the HDLA are provided the best performance compare to other machine-learning algorithms by using M-HDx features.

Experiment VII: Experiments VII results are presented in Table 21 by using multimodal features (M-HDx) and machine learning algorithms. We used the BROOK-DenseNet M-HDX system based on three-stage on 25 different subjects in 40 minutes of recorded time. In the case of CNN + Naïve Bayes and CNN + RNN-LSTM model achieved the best performance and comparable to other algorithms such as the DenseNet-BC100 model. On average, SE of (89.6/89.5/86.5), SP of (88.6/88.5/86.5), ACC of (89.0/80.5/86.5), PR of (85.6/84.5/86.5), MCC of (0.58/0.59/ 0.57), AUC of (0.88/0.89/0.87), and F1 of (0.88/0.89/0.82) are measured by using a hybrid algorithm (CNN +Naïve Bayes). This experiment shows that the BROOK-DenseNet

TABLE 22. State-of-the-art comparison of different classifier used in Riani-M-HDx [128] based on three-stages based hypovigilance detection (HDx) systems using multimodal features, different classifiers and 25 different subjects in 40 minutes of recorded time, where "∗/∗/∗" means drw/alt/dist values and high values provide better results.

Classifiers	SE	SP	ACC	PR	MCC	AUC	F1
Ternary Classification Results= DRW: Drowsy, ALT: ALERT, Dist: Distracted (drowsiness with distraction + alertness with distraction)							
ANN	76.6/77.5/76.5	78.6/78.5/76.5	85.2/86.1/86.5	80.6/82.5/83.5	0.45/0.46/0.47	0.85/0.86/0.84	0.85/0.86/0.83
SVM	76.6/77.5/76.5	80.6/81.5/82.5	86.3/85.2/86.5	82.6/81.5/83.5	0.46/0.47/0.47	0.85/0.87/0.85	0.85/0.87/0.82
Logic Regression +ANN	74.6/77.5/76.5	82.6/82.5/82.5	85.4/84.3/86.5	83.6/84.5/85.5	0.47/0.47/0.47	0.85/0.87/0.86	0.85/0.87/0.81
CNN +SVM	74.6/77.5/79.5	83.6/84.5/86.5	87.2/88.2/86.5	84.6/81.5/85.5	0.48/0.49/0.50	0.86/0.87/0.85	0.86/0.87/0.83
CNN +Naïve Bayes	80.6/83.5/87.5	84.6/85.5/86.5	89.0/80.5/86.5	85.6/83.5/86.5	0.49/0.50/0.51	0.88/0.89/0.87	0.88/0.89/0.92
CNN+ RNN-LSTM	90.6/90.5/92.5	94.6/95.5/98.5	90.6/93.5/96.5	86.6/84.5/85.5	0.52/0.51/0.54	0.90/0.91/0.94	0.90/0.91/0.95
DenseNet-BC100	**93.6/91.5/96.5**	**95.6/94.5/ 98.5**	**93.6/91.5/96.5**	**88.6/87.5/86.5**	**0.55/0.54/0.57**	**0.90/0.91/0.96**	**0.92/0.91/0.93**

SE: sensitivity, specificity: SP, AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient, PR: Precision, F1: F1-score

TABLE 23. State-of-the-art comparisons of hypovigilance detection system (m-HDx) based on multimodal features (Behavioral + Psychological) systems on five-stage in terms of hardware-based benchmarks such as (a) Graphical-processing unit (GPU) and (b) central processing unit (CPU).

Ref. No.	Cited	Accuracy	Training Time	Testing Time
(A) GPU: xTesla K80, having 2496 CUDA cores, 12GB GDDR5 VRAM				
[124]	CogBeacon-ML [124]	58.5%	6.70s	1.20s
[126]	BROOK-DenseNet [126]	85.5%	7.88s	**2.18s**
[74]	Ford-Dataset [74]	83.5%	**14.33s**	1.33s
[128]	Riani-M-HDx [128]	**91.5%**	6.35s	0.35s
(B) CPU: Intel(R) Xeon(R) CPU, RAM: 12.6 GB@2.3Ghz (1 core,2 threads)				
[126]	BROOK-DenseNet [126]	85.5%	**27.88s**	3.18s
[74]	Ford-Dataset [74]	83.5%	24.33s	2.33s
[128]	Riani-M-HDx [128]	**91.5%**	26.35s	1.35s
[124]	CogBeacon-ML [124]	58.5%	26.70s	**3.3s**

M-HDx system is improved if used CNN and Naïve Bayes classifiers to detect driver fatigue compare to DenseNet-BC 100 model. These results can be affected in case of increased training data size. Also, the Riani-M-HDx system is used to test the performance of M-HDX features in different machine learning algorithms based on a three-stage prediction of fatigue. It was noticed that the DenseNet-BC100 and CNN + RNN-LSTM model achieved the best performance and comparable to other machine learning algorithms. From table 22, it noticed that the DenseNet-BC100 classifier outperformed compare to other learning algorithms.

Experiment VIII: It is also important to measure the computational time when implemented by using the same learning strategy and different benchmarks. This experiment is presented in Table 23 based on five-stage based detection of driver fatigue. For benchmarks, we used GPU of xTesla K80, having 2496 CUDA cores with dedicated 12GB GDDR5 VRAM and CPU of Intel(R) Xeon(R) CPU, RAM: 12.6 GB@2.3Ghz (1 core,2 threads). These results are reported in Table 23 based on different M-HDx systems. Also, this complexity is calculated based on 20 minutes of data recorded by 10 persons. On average, the Riani-M-HDx system achieved the highest accuracy of 91.5% with low

training: 6.35s and testing: 0.35s times compare to other systems. It was calculated based on GPU, which is ultimately provided faster calculations due to dedicated memory as compare to shared-memory of CPU. Moreover, it is noticed that if increased the 7 cores of the processor then testing time performance is slightly improved but still, GPU is provided faster testing and training time. This point is not noticed by many M-HDx systems in the past to develop multistage's driver fatigue systems.

Experiment IX: The transfer learning (TL) algorithms are the latest trend in deep-learning models. Therefore, separate experiments are conducted to check the performance of the TL algorithm on Riani-M-HDx compared to other M-HDx systems because mostly, this system is outperformed. The TL algorithms used in this comparison are VGG-16, VGG-19, ResNet-50, DenseNet, InceptionV3, and Pretrain-CNN models. From these experiments, it was observed that the hybrid HDLA and transfer-learning (TL) algorithms are achieving high performance in terms of detection and prediction accuracy compared to standard machine learning and DL algorithms. Therefore, the TL algorithms are selected based on experiment analysis for predicting driver fatigue. To set up the TL algorithms, we describe the validation loss,

TABLE 24. Comparisons of validation loss and training time for all transfer learning models using Riani-M-HDx [128] architecture.

Model	Epoch	Validation Loss	Training Time (Epoch)	Testing Time	Convergence Time
VGG-16	97	<u>0.1925</u>	50 s	0.02020 s	426 s
VGG-19	90	0.0525	**65 s**	0.02134 s	389 s
ResNet-50	<u>98</u>	0.0496	53 s	0.02340 s	345 s
DenseNet	95	**0.4531**	56 s	**0.02550 s**	245 s
InceptionV3	**120**	0.0453	<u>60 s</u>	<u>0.02440 s</u>	**800 s**
Pretrain-CNN	**120**	0.0453	<u>60 s</u>	<u>0.02440 s</u>	<u>750 s</u>

TABLE 25. Performance measures of different transfer learning algorithms for prediction of five-stage based on Riani-M-HDx [128] architecture.

Model	Precision	Recall	F1	MCC	AUC
VGG-16	0.824	0.902	0.821	0.8501	0.908
VGG-19	0.814	0.892	0.810	0.8455	0.901
<u>ResNet-50</u>	<u>0.854</u>	<u>0.934</u>	<u>0.873</u>	<u>0.8655</u>	<u>0.933</u>
DenseNet	**0.864**	**0.945**	**0.895**	**0.8955**	**0.943**
InceptionV3	0.834	0.910	0.845	0.8550	0.914
Pretrain-CNN	0.846	0.926	0.853	0.8520	0.923

AUC: area under the receiver operating curve, MCC: Matthews correlation coefficient and F1: F1-score
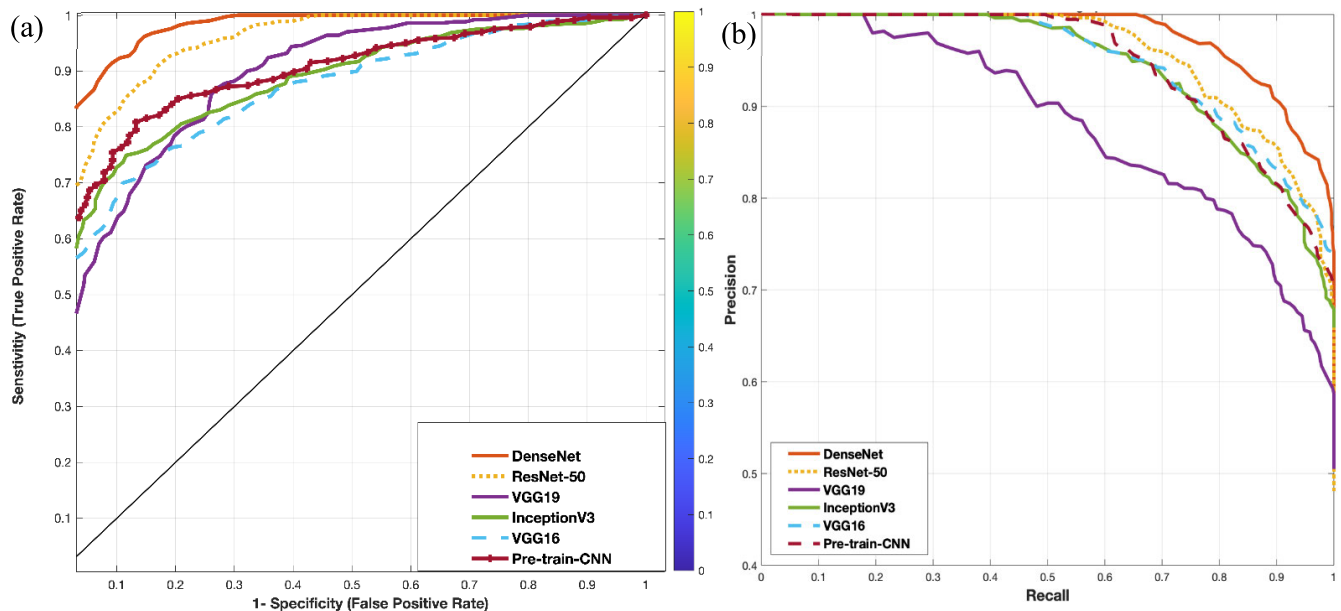
FIGURE 13. Transfer learning (TL) experiment for prediction of 5-classes based on RIANI-M-HDX [128] architecture, where (a) area under the receiver operating curve (AUC) and (b) precision-recall graphs for different multimodal-based hypovigilance detection (M-HDx) systems by using 10-fold cross validation test.

epochs, training, and testing times in Table 24 when used the RIANI-M-HDX system. Compare to VGG-16, VGG-19, InceptionV3, pre-train CNN TL, the ResNet-50, and DenseNet are achieved the highest detection accuracies with the less computational burden. These results are described in Table 25. A graphically those results are represented in terms of AUC and recall-precision (as shown in Figure 13) from Table 25. These results are obtained even in the case of 5-classes based detection of driver drowsiness. Hence, the transfer learning algorithms are provided the best

performance compared to traditional machine learning, hybrid, and standard deep learning algorithms. However, the results can be further improved to increase the training dataset and selection of deep layers. The results may be further enhanced by including different environmental conditions and position of the camera to define visual features.

VI. DISCUSSIONS

After presenting the results in detail in the previous section, we focus our attention on discussing the significance of

these results in this section. Several hypovigilance detection (HDX) systems were developed based on the integration of vision-based features, sensor-based features, and multimodal-based approaches. By using vision-based features, the drowsiness of a person reveals itself in physical and physiological changes which can either be visually observed by eyes (PERCLOS measure) and/or mouth activity and head nodding. Those visual features are extracted and then used to classify driver hypovigilance or drowsiness by the traditional machine learning (TML), deep-learning (DL), and transfer learning (TL) architectures. Also, there is another way to detect drowsiness through digitally registered via EEG and/or heart rate monitoring systems. The non-visual features are extracted in the past systems based on the driver's physiological measure and vehicle parameters. In the case of physiological parameter measurements, the authors predict driver fatigue based on different parameters such as steering-wheel, acceleration pedal, and speed. In practice, those approaches were mostly dependent on the road-shape, way of driving, and performance of the vehicle. Also, some other studies utilized electroencephalograph (EEG), electrocardiogram (ECG), electrooculography (EOG), and surface electromyogram (sEMG) sensors to predict driver fatigue. Based on the sensors, the authors were detected drowsy and alert conditions of drivers. Later on, the authors used multimodal–based features to develop M-HDx systems based on an approach to integrate vision-based, vehicular-based or environmental-based and sensor-based features to detect driver fatigue into multistage (Very alert (without distraction), Fairly alert (drowsiness with distraction), Neither alert nor drowsy, Moderate Drowsy, Extremely drowsy). According to a literature review, it noticed that the multimodal-based features provided higher accuracy compare to unimodal-based techniques. Therefore in this review paper, multimodal-based features approaches are described and compared in detail with emphasis on machine-learning algorithms (TML, DL, and TL).

Numerous survey articles were written by authors to develop the need for driver fatigue detection (DFD) systems. However, many research questions need to be addressed as mentioned in Table 1. Those survey articles are unable to address all problems as described in Table 2. A recent survey shows that several kinds of research have been going on to develop an automatic solution [127] for detecting and predicting the driver's fatigue using multimodal and deep learning (DL) architectures to recognize multistage drowsiness. An automatic solution for drivers' fatigue [4] is considered significant to improve the visual attention of drivers. In the past studies, the researchers found that road attention is very sensitive to driver fatigue as discussed in [130]. Also, it is very much important to regularly monitor the traffic environment as well as vigilance [131]. For sustained driver attention, fatigue detection in real-time is beneficial to save accidents. An alert system is also designed in the past techniques to detect upcoming hazards in their path [132]. An automatic driver fatigue detection system is always required to advance the development system for road safety.

Visual features are also provided an efficient way to detect real-time driver fatigue without using non-visual features as suggested by many authors [12], [45]–[64]. In particular [63] study, the authors observed that the light-illumination has a considerable effect on extracting features based on PERCLOS measure. Also, in [67], the authors reported that face occlusion with big-black glasses and head position might affect the measurement's reliability and accuracy. In many studies to detect driver fatigue, the PERCLOS measure was employed using a single camera. However, in real-time driver detection, if the driver's head is not centered and it is out of focus. This poses a problem as it will be very hard to measure PERCLOS. As a result, domain-expert knowledge and image processing must be robust through at-least three cameras in a three-different position to capture the driver's face. Night-vision is another limitation for the PERCLOS measure to detect visual features. To minimize these light-illumination problems, many authors are suggesting using IR cameras.

Studies [122]–[127] suggested using a single-view camera to detect visual features and it is very much difficult to detect perfect facial features through one camera. As a result, it might require multiview Online, there are several datasets available to researchers to test their M-HDx systems but with limited real-time conditions. A video database of YawDD [142] is available to test development techniques. The videos were acquired in a day and nighttime driving conditions, with drivers belonging to different gender and race. Another challenging aspect of conventional fatigue detection systems is to make them compatible with changing times and trends.

During our studies, it was observed that numerous visual and non-visual features (as described in Table 3 and Table 4) are considered in the past to detect and predict driver drowsiness levels. However, the sensors that are utilized in [76]–[79] must not interfere with the real-time driving process, the life expectancy of sensors, and even the effect on the health of drivers. Instead of using the camera, many researchers have worked with health vitals such as EEG, ECG, and EMG sensors [80]–[94] to extract biological features from the driver. Some important visual-feature-based datasets are described in Table 6. These psychological features are used to detect multistage fatigue detection, which is not possible to use only behavior features.

Our study finds that it is not possible to use a single visual or non-visual feature to be used to detect real-time driver fatigue in all environmental conditions. Some research suggests using hybrid systems that combined both visual and non-visual features to get higher accuracy. Mostly, the authors utilized PERCLOS visual features and EEG, ECG sensors for non-visual features. Those hybrid or multimodal features are feasible to detect driver fatigue. In Table 8, we presented online data resources about different multimodal-based HDx (M-HDx) systems. Those datasets can be used to test and train

M-HDx systems. The correctness of driver fatigue detection methods, in this case, depends on various factors such as real-time processing as fast and accurate results. A comparison of the fatigue detection techniques is presented in Table 5. It is concluded that the most suitable features for driver fatigue detection are visual features and machine learning algorithms especially deep-learning methods.

The latest trend is the utilization of deep learning algorithms (DLAs) [95], [97], [128]–[141] and multimodal-features to detect and predict driver drowsiness. Those DLAs-based DFD systems are presented in Table 9. In particular, a convolutional neural network (CNN) method is used in [35] to recognize driver fatigue based on minimal network structure and facial points. The authors reported that they achieved 73.06% classification accuracy. They used SoftMax and CNN algorithms together with three dropout hidden layers to do the final prediction. It was noticed that the authors did not train the model on the huge dataset and they used CNN for the prediction of driver fatigue. Since the CNN model is used to detect features from images and selection of features map are difficult steps for real-time detection of fatigue. As a result, the computational complexity is also high due to CNN methods and it is difficult to detect multistage drowsiness. Similarly, there are other deep-learning (DL) architectures that require many data augmentation and selection of multilayers to train and test the model. To evaluate the performance of traditional machine-learning, deep-learning, and transfer-learning algorithms on the detection of multistage of fatigue, we used four different M-HDx systems such as CogBeacon-ML [124], BROOK-DenseNet [126], Ford-dataset [74], and Riani-HDx [128]. By using 10 epochs, the training and validation accuracies are also computed.

Our comparative analysis based on different machine-learning algorithms and distinct features is described in Table 12 to Table 22. Based on these experiments, the application of deep learning (DL) models provided the best performance compared to traditional machine learning algorithms. In particular, the hybrid DL (HDL) algorithms are outperformed to detect driver fatigue in three-stage and five-stage level detection of drowsiness. It was noticed that many authors used unimodal or multimodal features with DL architectures to get state-of-the-art prediction results. This distribution can be seen in Figure 14 and Figure 15. This shows that the application of DL with transfer learning models is highly prevalent with both visual and non-visual feature-based fatigue detection systems. It was observed that the transfer learning (TL) algorithms were only utilized 5% in the past. However, our comparison analysis based on four M-HDx systems indicates that the TL algorithms are the best candidate for the detection of fatigue in multistage. By using TL algorithms, it is difficult to provide sufficient training samples based on visual and non-visual features. If we can provide enough training datasets then it can provide effective performance. To access computational performance, we have done another experiment and those results are mentioned in Table 23. In this table, we used the GPU of
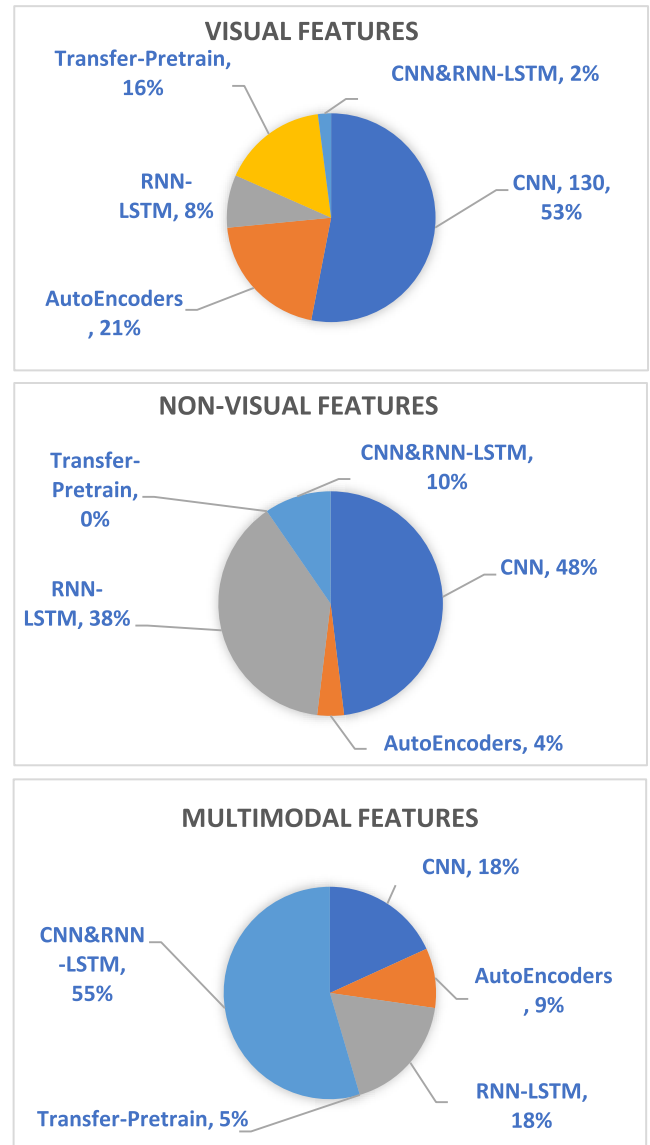

FIGURE 14. Distribution of different deep-learning models that utilized in the past for the development of hypovigilance detection systems.

xTesla K80, having 2496 CUDA cores with dedicated 12GB GDDR5 VRAM and CPU of Intel(R) Xeon(R) CPU, RAM: 12.6 GB@2.3Ghz (1 core,2 threads). On average, the Riani-M-HDx system achieved the highest accuracy of 91.5% with low training: 6.35s and testing: 0.35s times compare to other systems. It was calculated based on GPU, which is ultimately provided faster calculations due to dedicated memory as compare to shared-memory of CPU.

In this review and comparative analysis study, we incorporated a preliminary comparison to find a use of multimodal driver alertness for transfer learning algorithms. Also, the identification of multimodal datasets and available M-HDx systems are described to further conduct a pilot study to differentiate among 2-stage, 3-stage, and 5-stages of driver drowsiness. It is our first attempt to present a detailed study that uses physiological, visual, and environment modalities to track the different driver states together. Besides, three key
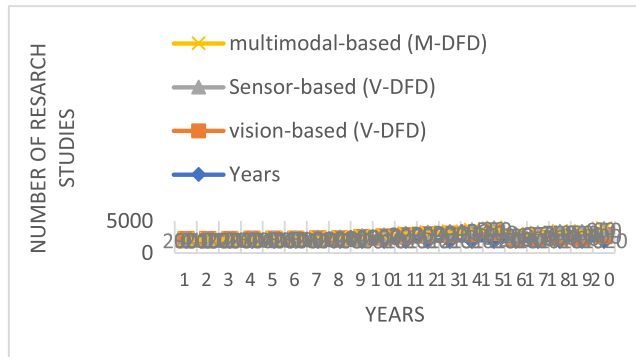
FIGURE 15. Distribution of number of multimodal-based, sensor-based and vision-based HDx systems utilized in the past.

aims were covered by this study. First, as it is a widely under-researched process, we explored the benefits of the integration of multimodal features in tracking multistage detection of fatigue. Second, our comparisons work extended the standard binary classification problem as mostly addressed in the past into a three- and five-classes problem to detect different levels of driver's alertness. Finally, we studied which methods have a greater ability to discriminate to classify alertness in drivers. Our experimental findings demonstrated the benefits provided by multimodal feature learning, highlighting substantial drawbacks for both multiclass classification systems overall individual modalities. Compared to individual modalities for multiclass classification, early modality fusion leads to improved performance, with an overall accuracy of 79.73 percent for the three-class system and 55.12 percent for the five-class approach if you used all parameter settings and different facial occlusions such as a scarf, big sunglasses. Finally, our study would not only seek the discovery of universal trends of the multiple driver states, but also the detection of personalized features associated with five-stage of driver drowsiness. This study shows that the state-of-the-art M-HDx systems are based on unimodal- or multimodal-based features to detect multi-stage of fatigue. In the subsequent sub-sections, we show the limitation of the training dataset, deep-learning methods, and future directions to assist other researchers.

A. LIMITED DATASETS FOR TRAININGS

In addition to the above technological problems in deep learning (DL) algorithms, there is an increasing need for multiple enormous datasets to undertake a successful training process. In fact, in terms of classes and instances that impact the training accuracy of deep learning algorithms, the online datasets available are small with limited variability in environments. However, we have identified the most recent datasets including vision-based, sensor-based, and multimodal-based. Those datasets can be easily integrated into any real-time applications for training the network but we did not find a dataset that covered all aspects of driver drowsy in multiple stage recognition on day and night times of driving. In the literature, these mentioned datasets along contained trained networks

that can be used as a pre-train model for the transfer learning domain. It can help to increase classification accuracy.

B. CHALLENGES OF DL ARCHITECTURES

The development of multilayer deep-learning (DL) architectures including hybrid learning (HL) and transfer learning (TL) is a recent trend in machine learning to develop hypovigilance (HDx) systems. In a large variety of applications, these DL algorithms [143]–[147] have obtained very promising results. To detect real-time multistage of driver fatigue, the DL architectures are still facing the challenges of computational cost, the complexity of network parameters, and system performance. Since the authors did not focus on computing cost and parameter complexity to develop many HDx systems when used DL algorithms. Therefore, a robust DL-based architecture is necessary to develop for feature extraction and classification tasks. Moreover, the computational complexity of the DL model is very high during the training phase because of the large number of parameters used in a network. As a result, the model reduction methods such as pruning, removal of redundant neurons, and layers are required to solve complex problems. In comparison relative to the classical DL models implemented in the CPU, the GPU-based implementation of the DL also offers high performance. Thus, this can remove the high computation cost.

The HDx architectures have recently been expanded and are becoming deeper in terms of the number of layers and the number of processing units per layer, following the new developments. Still, in terms of their computing power and training sophistication, the HDx architectures need to be more efficient. Henceforth, we are going to discuss certain recommendations to improve the computing capacity of the DL architectures in terms of the development of M-HDx systems for a real-time environment.

In this work, we have studied and compared various DL architectures based on multimodal-based HDx (M-HDx) systems to predict multistage driver fatigue. The latest techniques for M-HDx systems are DL models, where the emphasis is on extracting multimodal features fusion without defining handcrafted features. Multimodal feature fusion will mitigate the deficiencies of both visual features and non-visual features, thus enhancing HDx efficiency. While multimodal M-HDx has improved dramatically over the past few years, when integrating multiple modalities, current works fall short of optimum efficiency. Deciding at which stage of information fusion is a big difficulty such as early fusion, function level fusion, and decision/late fusion, the modalities should be fused [15]. Multimodal fusion aims to obtain complimentary knowledge from modalities to correctly perform the task of analysis. The major concern in multimodal fusion is to find the best example or stage to combine the modalities. The widely employed techniques are, based on this philosophy, data level or early fusion, function level or intermediate fusion, and judgment level or late fusion [16]. Some new fusion models have arisen from

the advent of deep learning. Since feature extraction can be performed on any layer in deep learning models, particularly CNN, we can now have early and late feature level fusions.

The most commonly used technique for incorporating knowledge in deep learning models is feature-level fusion. The biggest advantage of early fusion is that at an early stage it exploits the similarity between the modalities. Besides, to execute a task, only one classifier is required, rendering the training process less repetitive. However, time synchronization is a noteworthy constraint of the function level fusion, as the data is collected at various rates and formats in different modalities [17]. Decision stage fusion is the other broadly employed fusion tactic. The major advantage of this strategy is that it helps one to analyze each modality directly, thus significantly minimizing the probability of superiority of one modality over the other. Most current works either feature level fusion or decision fusion, losing the chance to fuse rich representations of mid-level features available in a CNN-based architecture. Hence, new deep architecture with fusion frameworks for HDx needs to discuss to resolve the aforementioned shortcomings and to exploit various fusion strategies. The major drawback in current HDx deep learning-based fusion approaches using deep and inertial sensors is that the fusion is carried out at a single level or point without providing us semantic knowledge from the features to the classifier.

C. FUTURE DIRECTIONS AND IMPROVEMENTS
After doing extensive experiments on different features and classification algorithms, we conclude that there is still a big research gap to implement DL architecture for the recognition of the different states of the drivers in terms of drowsiness. The below points are described to improve the architecture of DL models with future improvements and given as follows.

1) Create deeper DL architectures to increase performance and override the problem of degradation by adding more layers.
2) Selection of multimodal features and fusion approach will be important concerns to develop accurate model for prediction of multistage fatigue.
3) Increasing training datasets with more than 50 epochs to build the transfer-learning (TL) model.
4) Parameters for DL methods should be decreased and there will be extensive optimization techniques to compare and test the performance.
5) Designing of an appropriate loss-function is an important factor for DL architecture to improve the discriminating power of the network.
6) Feature extraction and fine-tuning are important steps for the accurate development of the deep learning model.
7) Learning using hierarchical functions, i.e., learning variant layers with variant characteristics.
8) To make use of the complimentary benefits of these DL models, the development of hybrid models will be

considered as best instead of using single DL model for features extraction or classification tasks.
9) A different kinds of DL architecture should be investigated such as few-short learning (FSL) and semantic-based learning (SBL). These kind of DL model are significantly improved identification accuracy with fewer training samples.
10) To compare the performance of vision transformers (VTs) with DL model is also required for the development of M-HDx systems.
11) Instead of using generalize DL architecture, it is also required to develop DL model specific of HDx-related problem.
12) To detect driver fatigue, many researchers [10]–[30] used visual features that are defined through vision-based technique. The authors used single view or multi-view camera to define PERCLOS measure but unfortunately, it is very much difficult to detect driver's features in case of face occlusion, light illumination and head is not center-aligned.
13) Few authors used combine approach by using vision-based and sensors-based devices together to detect driver fatigue to avoid face occlusion or light illumination problem.
14) Few current models collect both temporal function data and temporal interaction information between driver fatigue detection features. A greater degree of mouth opening, for instance, does not generally reflect fatigue, which may be reassessed over time by shifts in heart rate and level of eye opening.
15) In order to build an integrated framework that proactively identifies driver tension levels, it is important to capture, pass, pre-process, minimize, incorporate data and use it automatically to make the final decision.
16) Implement TL algorithms on GPU-based platform with cloud-computing platform.

A comprehensive literature review has taught that each approach has its drawback. Therefore, the use of many methods will be best be used. The considered features are derived from non-intrusive sensors that are related to the changes in driving behavior and visual facial expressions. To get enhanced visual facial features, three cameras can be deployed at different angles. Afterward, it should be trained on big multimodal datasets with transfer learning algorithms that can better assist automatic drowsiness detection in a real-time environment. However, time is also an important factor in a real-time environment. Hence, if transfer learning will be implemented in GPU then 35% on average training and testing time will be enhanced compared to CPU based processing.

VII. CONCLUSION
This article shows the state-of-art current research and development efforts on recognition of driver drowsiness through vision-based, sensor-based, and multimodal-based features

techniques. Literature has shown that the prediction of required features was performed by standard and latest deep learning (DL) algorithms in the past. In this paper, we have focused on deep-learning with transfer learning algorithms that have been utilized in the past to predict driver drowsiness instead of conventional machine learning techniques. A critical review is also presented to show the different driving influence factors, traditional and advance deep-learning algorithms to help the researchers identify the research gap. We conclude that many multimodal-based M-HDx systems were developed and tested on different hardware benchmarks such as (CPU and GPU). However, there are several review articles in this domain but according to our limited knowledge, none of them focused on comparisons about multimodal-based M-HDx systems in different benchmark settings. In the future, we will focus on more comparison studies related to hybrid systems for the implementation of the fatigue recognition system in a multistage. Moreover, we have assessed and described the latest methodologies used in this domain such as deep-learning algorithms. According to our knowledge, several review articles have been written to address driver fatigue detection problems but none of them focused on the domain of deep-learning multi-layer architectures. The future trends will implement the combination of supervised and unsupervised algorithms to enhance driver adaptability and cognitive performance without giving more details. With the development of hardware, the internet of things (IoT), 5G network, energy computing, and more transfer-learning based algorithms will be tested in the future trend.

REFERENCES

[1] G. Matthews, C. Neubauer, D. J. Saxby, R. W. Wohleber, and J. Lin, "Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle," *Accident Anal. Prevention*, vol. 126, pp. 85–94, May 2019.

[2] A. A. Hyder, N. Paichadze, T. Toroyan, and M. M. Peden, "Monitoring the decade of action for global road safety 2011–2020: An update," *Global Public Health*, vol. 12, no. 12, pp. 1492–1505, 20117.

[3] G. Sasikala and V. R. Kumar, "Development of advanced driver assistance system using intelligent surveillance," in *Proc. Int. Conf. Comput. Netw. Commun. Technol.* Singapore: Springer, 2019, pp. 991–1003.

[4] J. Healey and R. Picard, "SmartCar: Detecting driver stress," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4. Barcelona, Spain, 2000, pp. 218–221.

[5] D. P. Spangler, K. R. Cox, J. F. Thayer, J. R. Brooks, and B. H. Friedman, "Interplay between state anxiety, heart rate variability, and cognition: An ex-Gaussian analysis of response times," *Int. J. Psychophysiol.*, vol. 159, pp. 60–70, Jan. 2021.

[6] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.

[7] C. Diaz-Piedra, M. V. Sebastián, and L. L. Di Stasi, "EEG theta power activity reflects workload among army combat drivers: An experimental study," *Brain Sci.*, vol. 10, no. 4, p. 199, Mar. 2020.

[8] V. Parekh, D. Shah, and M. Shah, "Fatigue detection using artificial intelligence framework," *Augmented Human Res.*, vol. 5, no. 1, p. 5, Dec. 2020.

[9] F. Zhou, A. Alsaid, M. Blommer, R. Curry, R. Swaminathan, D. Kochhar, W. Talamonti, L. Tijerina, and B. Lei, "Driver fatigue transition prediction in highly automated driving using physiological features," *Expert Syst. Appl.*, vol. 147, Jun. 2020, Art. no. 113204.

[10] K. Li, Y. Gong, and Z. Ren, "A fatigue driving detection algorithm based on facial multi-feature fusion," *IEEE Access*, vol. 8, pp. 101244–101259, 2020.

[11] 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. 61904–61919, 2019.

[12] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," *Sensors*, vol. 19, no. 11, p. 2574, Jun. 2019.

[13] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.

[14] M. N. Rastgoo, B. Nakisa, A. Rakotonirainy, V. Chandran, and D. Tjondronegoro, "A critical review of proactive detection of driver stress levels based on multimodal measurements," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–35, Jan. 2019.

[15] N. S. Karuppusamy and B.-Y. Kang, "Multimodal system to detect driver fatigue using EEG, gyroscope, and image processing," *IEEE Access*, vol. 8, pp. 129645–129667, 2020.

[16] A. Nemcova, M. Seitl, T. Dominik, M. Semela, M. Sucha, R. Kolar, V. Svozilova, K. Bucsuhazy, R. Smisek, M. Mezl, B. Hesko, M. Belak, M. Bilik, and P. Maxera, "Multimodal features for detection of driver stress and fatigue: Review," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 10, 2020, doi: 10.1109/TITS.2020.2977762.

[17] W. Wu, W. Sun, Q. M. J. Wu, Y. Yang, H. Zhang, W.-L. Zheng, and B.-L. Lu, "Multimodal vigilance estimation using deep learning," *IEEE Trans. Cybern.*, early access, Oct. 7, 2020, doi: 10.1109/TCYB.2020.3022647.

[18] R. Li, H. Brand, A. Gopinath, S. Kamarajugadda, L. Yang, W. Wang, and B. Li, "Driver drowsiness behavior detection and analysis using vision-based multimodal features for driving safety," SAE Tech. Paper 2020-01-1211, 2020.

[19] G. Du, T. Li, C. Li, P. X. Liu, and D. Li, "Vision-based fatigue driving recognition method integrating heart rate and facial features," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 16, 2020, doi: 10.1109/TITS.2020.2979527.

[20] B. Cheng, W. Zhang, Y. Lin, R. Feng, and X. Zhang, "Driver drowsiness detection based on multisource information," *Hum. Factors Ergonom. Manuf. Service Industries*, vol. 22, no. 5, pp. 450–467, Sep. 2012.

[21] W. Sun, X. Zhang, S. Peeta, X. He, Y. Li, and S. Zhu, "A self-adaptive dynamic recognition model for fatigue driving based on multi-source information and two levels of fusion," *Sensors*, vol. 15, no. 9, pp. 24191–24213, Sep. 2015.

[22] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accident Anal. Prevention*, vol. 126, pp. 95–104, May 2019.

[23] B. Cyganek and S. Gruszczyński, "Hybrid computer vision system for drivers' eye recognition and fatigue monitoring," *Neurocomputing*, vol. 126, pp. 78–94, Feb. 2014.

[24] J. Jo, S. J. Lee, K. R. Park, I.-J. Kim, and J. Kim, "Detecting driver drowsiness using feature-level fusion and user-specific classification," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1139–1152, Mar. 2014.

[25] G. J. AL-Anizy, M. J. Nordin, and M. M. Razooq, "Automatic driver drowsiness detection using Haar algorithm and support vector machine techniques," *Asian J. Appl. Sci.*, vol. 8, no. 2, pp. 149–157, Mar. 2015.

[26] B. Akrout and W. Mahdi, "Spatio-temporal features for the automatic control of driver drowsiness state and lack of concentration," *Mach. Vis. Appl.*, vol. 26, no. 1, pp. 1–13, Jan. 2015.

[27] D. E. Benrachou, F. N. dos Santos, B. Boulebtateche, and S. Bensaoula, "Automatic eye localization; multi-block LBP vs. pyramidal LBP three-levels image decomposition for eye visual appearance description," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Cham, Switzerland: Springer, Jun. 2015, pp. 718–726.

[28] M. Wang, L. Guo, and W.-Y. Chen, "Blink detection using AdaBoost and contour circle for fatigue recognition," *Comput. Electr. Eng.*, vol. 58, pp. 502–512, Feb. 2017.

[29] M.-H. Sigari, M. Fathy, and M. Soryani, "A driver face monitoring system for fatigue and distraction detection," *Int. J. Veh. Technol.*, vol. 2013, pp. 1–11, Jan. 2013.

[30] A. Emam, "Intelligent drowsy eye detection using image mining," *Inf. Syst. Frontiers*, vol. 17, no. 4, pp. 947–960, Aug. 2015.

[31] C. Gou, Y. Wu, K. Wang, K. Wang, F.-Y. Wang, and Q. Ji, "A joint cascaded framework for simultaneous eye detection and eye state estimation," *Pattern Recognit.*, vol. 67, pp. 23–31, Jul. 2017.

[32] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3408–3420, Dec. 2017.

[33] K. Diaz-Chito, A. Hernández-Sabaté, and A. M. López, "A reduced feature set for driver head pose estimation," *Appl. Soft Comput.*, vol. 45, pp. 98–107, Aug. 2016.

[34] G. Sikander and S. Anwar, "A novel machine vision-based 3D facial action unit identification for fatigue detection," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 25, 2020, doi: 10.1109/TITS.2020.2974263.

[35] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, Nov. 2016, pp. 154–164.

[36] S. El Kaddouhi, A. Saaidi, and M. Abarkan, "Eye detection based on the Viola-Jones method and corners points," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 23077–23097, Nov. 2017.

[37] U. Gopikrishnan and R. Jose, "DriveCare: A real-time vision based driver drowsiness detection using multiple convolutional neural networks with kernelized correlation filters (MCNN-KCF)," in *Proc. IEEE 5th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Oct. 2020, pp. 408–413.

[38] F. You, Y. H. Li, L. Huang, K. Chen, R. H. Zhang, and J. M. Xu, "Monitoring drivers' sleepy status at night based on machine vision," *Multimedia Tools Appl.*, vol. 76, no. 13, pp. 14869–14886, 2017.

[39] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[40] A. Kumar and R. Patra, "Driver drowsiness monitoring system using visual behaviour and machine learning," in *Proc. IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Apr. 2018, pp. 339–344.

[41] A. Baghdadi, F. M. Megahed, E. T. Esfahani, and L. A. Cavuoto, "A machine learning approach to detect changes in gait parameters following a fatiguing occupational task," *Ergonomics*, vol. 61, no. 8, pp. 1116–1129, 2018.

[42] A. Fogelton and W. Benesova, "Eye blink completeness detection," *Comput. Vis. Image Understand.*, vols. 176–177, pp. 78–85, Nov. 2018.

[43] H. Al-libawy, A. Al-Ataby, W. Al-Nuaimy, and M. A. Al-Taee, "Modular design of fatigue detection in naturalistic driving environments," *Accident Anal. Prevention*, vol. 120, pp. 188–194, Nov. 2018.

[44] Y. Zhang and C. Hua, "Driver fatigue recognition based on facial expression analysis using local binary patterns," *Optik*, vol. 126, no. 23, pp. 4501–4505, Dec. 2015.

[45] S. Junaedi and H. Akbar, "Driver drowsiness detection based on face feature and PERCLOS," in *Proc. J. Phys., Conf.*, vol. 1090, Sep. 2018, Art. no. 012037.

[46] X. Ma, L.-P. Chau, and K.-H. Yap, "Depth video-based two-stream convolutional neural networks for driver fatigue detection," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Dec. 2017, pp. 155–158.

[47] 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, May 2019.

[48] M. Dua, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection," *Neural Comput. Appl.*, vol. 33, pp. 3155–3168, Jul. 2020.

[49] S. A. Khan, S. Hussain, S. Xiaoming, and S. Yang, "An effective framework for driver fatigue recognition based on intelligent facial expressions analysis," *IEEE Access*, vol. 6, pp. 67459–67468, 2018.

[50] C. H. Weng, Y. H. Lai, and S. H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, Nov. 2016, pp. 117–133.

[51] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012.

[52] K. Malhi, S. C. Mukhopadhyay, J. Schnepper, M. Haefke, and H. Ewald, "A zigbee-based wearable physiological parameters monitoring system," *IEEE Sensors J.*, vol. 12, no. 3, pp. 423–430, Mar. 2012.

[53] S. Lawoyin, D.-Y. Fei, and O. Bai, "Accelerometer-based steering-wheel movement monitoring for drowsy-driving detection," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 229, no. 2, pp. 163–173, Feb. 2015.

[54] P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig, and H. Nguyen, "EEG-based driver fatigue detection using hybrid deep generic model," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 800–803.

[55] J.-H. Jeong, B.-W. Yu, D.-H. Lee, and S.-W. Lee, "Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals," *Brain Sci.*, vol. 9, no. 12, p. 348, Nov. 2019.

[56] R. Bhardwaj, P. Natrajan, and V. Balasubramanian, "Study to determine the effectiveness of deep learning classifiers for ECG based driver fatigue classification," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2018, pp. 98–102.

[57] L. Cao, J. Li, Y. Xu, H. Zhu, and C. Jiang, "A hybrid vigilance monitoring study for mental fatigue and its neural activities," *Cognit. Comput.*, vol. 8, no. 2, pp. 228–236, Apr. 2016.

[58] W. M. Shalash, "Driver fatigue detection with single EEG channel using transfer learning," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Dec. 2019, pp. 1–6.

[59] Z. Mu, J. Hu, and J. Min, "Driver fatigue detection system using electroencephalography signals based on combined entropy features," *Appl. Sci.*, vol. 7, no. 2, p. 150, Feb. 2017.

[60] P. Wang, J. Min, and J. Hu, "Ensemble classifier for driver's fatigue detection based on a single EEG channel," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1322–1328, Dec. 2018.

[61] J. Tan, X. Y. Shi, and W. Guo, "Research on driving fatigue level using ecg signal from smart bracelet," in *Green, Smart and Connected Transportation Systems*. Singapore: Springer, 2020, pp. 799–810.

[62] J. Huang, X. Luo, and X. Peng, "A novel classification method for a driver's cognitive stress level by transferring interbeat intervals of the ECG signal to pictures," *Sensors*, vol. 20, no. 5, p. 1340, 2020.

[63] D. Jing, D. Liu, S. Zhang, and Z. Guo, "Fatigue driving detection method based on EEG analysis in low-voltage and hypoxia plateau environment," *Int. J. Transp. Sci. Technol.*, vol. 9, no. 4, pp. 366–376, Dec. 2020.

[64] W. Ko, K. Oh, E. Jeon, and H.-I. Suk, "VIGNet: A deep convolutional neural network for EEG-based driver vigilance estimation," in *Proc. 8th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2020, pp. 1–3.

[65] E. Q. Wu, P.-Y. Deng, X.-Y. Qu, Z. Tang, W.-M. Zhang, L.-M. Zhu, H. Ren, G.-R. Zhou, and R. S. F. Sheng, "Detecting fatigue status of pilots based on deep learning network using EEG signals," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jan. 1, 2020, doi: 10.1109/TCDS.2019.2963476.

[66] L.-W. Ko, O. Komarov, W.-K. Lai, W.-G. Liang, and T.-P. Jung, "Eye-blink recognition improves fatigue prediction from single-channel forehead EEG in a realistic sustained attention task," *J. Neural Eng.*, vol. 17, no. 3, Jun. 2020, Art. no. 036015.

[67] Y. Ma, S. Zhang, D. Qi, Z. Luo, R. Li, T. Potter, and Y. Zhang, "Driving drowsiness detection with EEG using a modified hierarchical extreme learning machine algorithm with particle swarm optimization: A pilot study," *Electronics*, vol. 9, no. 5, p. 775, May 2020.

[68] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, Mar. 2009.

[69] M. Gromer, D. Salb, T. Walzer, N. M. Madrid, and R. Seepold, "ECG sensor for detection of driver's drowsiness," *Procedia Comput. Sci.*, vol. 159, pp. 1938–1946, 2019.

[70] F. Wang, H. Wang, and R. Fu, "Real-time ECG-based detection of fatigue driving using sample entropy," *Entropy*, vol. 20, no. 3, p. 196, Mar. 2018.

[71] Z. Ahmad and N. Khan, "Human action recognition using deep multilevel multimodal (M^2) fusion of depth and inertial sensors," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1445–1455, Feb. 2020.

[72] F. Omerustaoglu, C. O. Sakar, and G. Kar, "Distracted driver detection by combining in-vehicle and image data using deep learning," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106657.

[73] C. Streiffer, R. Raghavendra, T. Benson, and M. Srivatsa, "DarNet: A deep learning solution for distracted driving detection," in *Proc. 18th ACM/IFIP/USENIX Middleware Conf., Ind. Track*, Dec. 2017, pp. 22–28.

[74] M. Abouelenien, M. Burzo, and R. Mihalcea, "Cascaded multimodal analysis of alertness related features for drivers safety applications," in *Proc. 8th ACM Int. Conf. Pervas. Technol. Rel. Assistive Environments*, Jul. 2015, pp. 1–8.

[75] J.-L. Yin, B.-H. Chen, K.-H.-R. Lai, and Y. Li, "Automatic dangerous driving intensity analysis for advanced driver assistance systems from multimodal driving signals," *IEEE Sensors J.*, vol. 18, no. 12, pp. 4785–4794, Jun. 2018.

[76] L.-H. Du, W. Liu, W.-L. Zheng, and B.-L. Lu, "Detecting driving fatigue with multimodal deep learning," in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 74–77.

[77] R. Daza, A. Morales, J. Fierrez, and R. Tolosana, "MEBAL: A multimodal database for eye blink detection and attention level estimation," 2020, *arXiv:2006.05327*. [Online]. Available: http://arxiv.org/abs/2006.05327

[78] M. Jagannath and V. Balasubramanian, "Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator," *Appl. Ergonom.*, vol. 45, no. 4, pp. 1140–1147, Jul. 2014.

[79] L.-L. Chen, Y. Zhao, P.-F. Ye, J. Zhang, and J.-Z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Syst. Appl.*, vol. 85, pp. 279–291, Nov. 2017.

[80] F. Laurent, M. Valderrama, M. Besserve, M. Guillard, J.-P. Lachaux, J. Martinerie, and G. Florence, "Multimodal information improves the rapid detection of mental fatigue," *Biomed. Signal Process. Control*, vol. 8, no. 4, pp. 400–408, Jul. 2013.

[81] S. Ahn, T. Nguyen, H. Jang, J. G. Kim, and S. C. Jun, "Exploring neuro-physiological correlates of Drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data," *Frontiers Hum. Neurosci.*, vol. 10, p. 219, May 2016.

[82] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *J. Neural Eng.*, vol. 14, no. 2, Apr. 2017, Art. no. 026017.

[83] G. Zhang and A. Etemad, "Capsule attention for multimodal EEG and EOG spatiotemporal representation learning with application to driver vigilance estimation," 2019, *arXiv:1912.07812*. [Online]. Available: http://arxiv.org/abs/1912.07812

[84] H.-T. Choi, M.-K. Back, and K.-C. Lee, "Driver drowsiness detection based on multimodal using fusion of visual-feature and bio-signal," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 1249–1251.

[85] H. Li, W.-L. Zheng, and B.-L. Lu, "Multimodal vigilance estimation with adversarial domain adaptation networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–6.

[86] M. Sunagawa, S.-I. Shikii, W. Nakai, M. Mochizuki, K. Kusukame, and H. Kitajima, "Comprehensive drowsiness level detection model combining multimodal information," *IEEE Sensors J.*, vol. 20, no. 7, pp. 3709–3717, Apr. 2020.

[87] A. Sengupta, A. Dasgupta, A. Chaudhuri, A. George, A. Routray, and R. Guha, "A multimodal system for assessing alertness levels due to cognitive loading," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 1037–1046, Jul. 2017.

[88] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cognit. Neurodynamics*, vol. 12, no. 6, pp. 597–606, Dec. 2018.

[89] O. Dehzangi, V. Sahu, M. Taherisadr, and S. Galster, "Multi-modal system to detect on-the-road driver distraction," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2191–2196.

[90] C. Anitha, "Detection and analysis of drowsiness in human beings using multimodal signals," in *Digital Business*. Cham, Switzerland: Springer, 2019, pp. 157–174.

[91] C. B. S. Maior, M. J. D. C. Moura, J. M. M. Santana, and I. D. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113505.

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

[93] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

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

[95] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, "A comprehensive review of recent advances on deep vision systems," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 39–76, Jun. 2019.

[96] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, early access, Mar. 17, 2020, doi: 10.1109/TKDE.2020.2981333.

[97] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, "Video scene analysis: An overview and challenges on deep learning algorithms," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 20415–20453, Aug. 2018.

[98] Q. Abbas, I. Fondon, A. Sarmiento, S. Jiménez, and P. Alemany, "Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features," *Med. Biol. Eng. Comput.*, vol. 55, no. 11, pp. 1959–1974, Nov. 2017.

[99] Q. Shi, Z. Zhang, T. He, Z. Sun, B. Wang, Y. Feng, X. Shan, B. Salam, and C. Lee, "Deep learning enabled smart mats as a scalable floor monitoring system," *Nature Commun.*, vol. 11, no. 1, pp. 1–11, Dec. 2020.

[100] X. Liao, K. Li, X. Zhu, and K. J. R. Liu, "Robust detection of image operator chain with two-stream convolutional neural network," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 955–968, Aug. 2020.

[101] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.

[102] A. Drif, H. E. Zerrad, and H. Cherifi, "EnsVAE: Ensemble variational autoencoders for recommendations," *IEEE Access*, vol. 8, pp. 188335–188351, 2020.

[103] Q. Abbas and M. E. Celebi, "DermoDeep–A classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network," *Multimedia Tools Appl.*, vol. 78, no. 16, pp. 23559–23580, Aug. 2019.

[104] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, and D.-S. Huang, "Deep learning-based methods for person re-identification: A comprehensive review," *Neurocomputing*, vol. 337, pp. 354–371, Apr. 2019.

[105] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.

[106] B. K. Reddy and D. Delen, "Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology," *Comput. Biol. Med.*, vol. 101, pp. 199–209, Oct. 2018.

[107] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[108] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, Oct. 2018, pp. 270–279.

[109] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019.

[110] P. Du, E. Li, J. Xia, A. Samat, and X. Bai, "Feature and model level fusion of pretrained CNN for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2600–2611, Aug. 2019.

[111] L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2798–2810, May 2018.

[112] Y. Ed-Doughmi, N. Idrissi, and Y. Hbali, "Real-time system for driver fatigue detection based on a recurrent neuronal network," *J. Imag.*, vol. 6, no. 3, p. 8, Mar. 2020.

[113] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Portland, OR, USA, Jun. 2013, pp. 896–903.

[114] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD," *Signal Process., Image Commun.*, vol. 88, Oct. 2020, Art. no. 115960.

[115] H. Ujir, M. Spann, and I. H. M. Hipiny, "3D facial expression classification using 3D facial surface normals," in *Proc. 8th Int. Conf. Robotic, Vis., Signal Process. Power Appl.*, in Lecture Notes in Electrical Engineering, vol. 291, H. M. Sakim and M. Mustaffa, Eds. Singapore: Springer, 2014, doi: 10.1007/978-981-4585-42-2_29.

[116] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, "Driver distraction identification with an ensemble of convolutional neural networks," *J. Adv. Transp.*, vol. 2019, pp. 1–12, Feb. 2019.

[117] J. Min, P. Wang, and J. Hu, (2017). *The Original EEG Data for Driver Fatigue Detection. Figshare. Dataset.* Accessed: Feb. 1, 2019. [Online]. Available: https://doi.org/10.6084/m9.figshare.5202739.v1

[118] Z. Cao, C. K. Michael, J. T. Lin, and C. Teng (2018). *Multi-Channel EEG Recordings During a Sustained-Attention Driving Task*. figshare. Fileset. [Online]. Available: https://doi.org/10.6084/m9.figshare.6427334.v2

[119] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Dec. 2019.

[120] C. Grégoire, P. L. C. Rodrigues, and M. Congedo. *EEG Alpha Waves Dataset.* Accessed: Feb. 20, 2020. [Online]. Available: https://zenodo.org/record/2348892#.X4bNidAzaM8

[121] C. Spencer. (Jun. 12, 2020). *A Comparison of Unimodal and Multimodal Measurements of Driver Stress in Real-World Driving Conditions.* [Online]. Available: https://psyarxiv.com/en5r3

[122] Q. Massoz, T. Langohr, C. Francois, and J. G. Verly, "The ULg multi-modality drowsiness database (called DROZY) and examples of use," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–7.

[123] J. Diego Ortega, N. Kose, P. Cañas, M.-A. Chao, A. Unnervik, M. Nieto, O. Otaegui, and L. Salgado, "DMD: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis," 2020, *arXiv:2008.12085*. [Online]. Available: http://arxiv.org/abs/2008.12085

[124] M. Papakostas, A. Rajavenkatanarayanan, and F. Makedon, "CogBeacon: A multi-modal dataset and data-collection platform for modeling cognitive fatigue," *Technologies*, vol. 7, no. 2, p. 46, Jun. 2019.

[125] Z. Huang, R. Li, W. Jin, Z. Song, Y. Zhang, X. Peng, and X. Sun, "Face2Multi-modal: In-vehicle multi-modal predictors via facial expressions," in *Proc. 12th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Sep. 2020, pp. 30–33.

[126] X. Peng, Z. Huang, and X. Sun, "Building BROOK: A multi-modal and facial video database for human-vehicle interaction research," 2020, *arXiv:2005.08637*. [Online]. Available: http://arxiv.org/abs/2005.08637

[127] Q. Abbas and A. Alsheddy, "Driver fatigue detection systems using multi-sensors, smartphone, and cloud-based computing platforms: A comparative analysis," *Sensors*, vol. 21, no. 1, p. 56, Dec. 2020.

[128] K. Riani, M. Papakostas, H. Kokash, M. Abouelenien, M. Burzo, and R. Mihalcea, "Towards detecting levels of alertness in drivers using multiple modalities," in *Proc. 13th ACM Int. Conf. Pervas. Technol. Rel. Assistive Environments*, Jun. 2020, pp. 1–9.

[129] Q. Abbas and M. E. A. Ibrahim, "DenseHyper: An automatic recognition system for detection of hypertensive retinopathy using dense features transform and deep-residual learning," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31595–31623, Nov. 2020.

[130] S. Chen, Z. Wang, and W. Chen, "Driver drowsiness estimation based on factorized bilinear feature fusion and a long-short-term recurrent convolutional network," *Information*, vol. 12, no. 1, p. 3, Dec. 2020.

[131] M. Papakostas, K. Das, M. Abouelenien, R. Mihalcea, and M. Burzo, "Distracted and drowsy driving modeling using deep physiological representations and multitask learning," *Appl. Sci.*, vol. 11, no. 1, p. 88, Dec. 2020.

[132] L. He, J. C.-W. Chan, and Z. Wang, "Automatic depression recognition using CNN with attention mechanism from videos," *Neurocomputing*, vol. 422, pp. 165–175, Jan. 2021.

[133] Y. Zhang, Y. Chen, and C. Gao, "Deep unsupervised multi-modal fusion network for detecting driver distraction," *Neurocomputing*, vol. 421, pp. 26–38, Jan. 2021.

[134] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, "CorrNet: Fine-grained emotion recognition for video watching using wearable physiological sensors," *Sensors*, vol. 21, no. 1, p. 52, Dec. 2020.

[135] S. Sony, K. Dunphy, A. Sadhu, and M. Capretz, "A systematic review of convolutional neural network-based structural condition assessment techniques," *Eng. Struct.*, vol. 226, Jan. 2021, Art. no. 111347.

[136] S. S. Poorna, A. Deep, K. Hariharan, R. R. Jain, and S. Krishnan, "A transfer learning approach for drowsiness detection from EEG signals," in *Innovations in Computational Intelligence and Computer Vision*. Singapore: Springer, 2021, pp. 369–375.

[137] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021.

[138] A. Alghamdi, M. Hammad, H. Ugail, A. Abdel-Raheem, K. Muhammad, H. S. Khalifa, and A. A. Abd El-Latif, "Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities," *Multimedia Tools Appl.*, pp. 1–22, Mar. 2020.

[139] Y. Gu, H. Liu, T. Wang, S. Li, and G. Gao, "Deep feature extraction and motion representation for satellite video scene classification," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 140307.

[140] R. Hou, Y. Zhao, Y. Hu, and H. Liu, "No-reference video quality evaluation by a deep transfer CNN architecture," *Signal Process., Image Commun.*, vol. 83, Apr. 2020, Art. no. 115782.

[141] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet," *IEEE Access*, vol. 8, pp. 24697–24712, 2020.

[142] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD: A yawning detection dataset," in *Proc. 5th ACM Multimedia Syst. Conf.*, Mar. 2014, pp. 24–28.

[143] Y. Shi, D. Feng, Y. Cheng, and S. Biswas, "A natural language-inspired multilabel video streaming source identification method based on deep neural networks," *Signal, Image Video Process.*, pp. 1–8, Jan. 2021, doi: 10.1007/s11760-020-01844-8.

[144] A. Mohite, P. Rege, and D. Chakravarty, "Human activity recognition using positioning sensor and deep learning technique," in *Advances in Signal and Data Processing*. Singapore: Springer, 2021, pp. 473–489.

[145] A. S. Parihar, D. Varshney, K. Pandya, and A. Aggarwal, "A comprehensive survey on video frame interpolation techniques," *Vis. Comput.*, pp. 1–25, Jan. 2021, doi: 10.1007/s00371-020-02016-y.

[146] A. Mahmoud and A. Mohammed, "A survey on deep learning for time-series forecasting," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*. Cham, Switzerland: Springer, 2021, pp. 365–392.

[147] K. Pawar and V. Attar, "Automated surveillance model for video-based anomalous activity detection using deep learning architecture," in *Innovations in Computational Intelligence and Computer Vision*. Singapore: Springer, 2021, pp. 327–334.

[148] L. Celona, L. Mammana, S. Bianco, and R. Schettini, "A multi-task CNN framework for driver face monitoring," in *Proc. IEEE 8th Int. Conf. Consum. Electron.-Berlin (ICCE-Berlin)*, Sep. 2018, pp. 1–4.

[149] K. W. Lee, H. S. Yoon, J. M. Song, and K. R. Park, "Convolutional neural network-based classification of driver's emotion during aggressive and smooth driving using multi-modal camera sensors," *Sensors*, vol. 18, no. 4, p. 957, 2018.

[150] M. Hashemi, A. Mirrashid, and A. Beheshti Shirazi, "CNN-based driver drowsiness detection," 2020, *arXiv:2001.05137*. [Online]. Available: http://arxiv.org/abs/2001.05137

[151] V. Tumen, O. Yildirim, and E. Ergen, "Detection of driver drowsiness in driving environment using deep learning methods," in *Proc. Electr. Electron., Comput. Sci., Biomed. Engineerings' Meeting (EBBT)*, Apr. 2018, pp. 1–5.

[152] M. Hashemi, A. Mirrashid, and A. Beheshti Shirazi, "CNN-based driver drowsiness detection," 2020, *arXiv:2001.05137*. [Online]. Available: http://arxiv.org/abs/2001.05137

[153] M. M. Rahman, M. S. Islam, M. K. Ara Jannat, M. H. Rahman, M. Arifuzzaman, R. Sassi, and M. Aktaruzzaman, "EyeNet: An improved eye states classification system using convolutional neural network," in *Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2020, pp. 84–90.

[154] J.-M. Guo and H. Markoni, "Driver drowsiness detection using hybrid convolutional neural network and long short-term memory," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 29059–29087, Oct. 2019.

[155] S. Arefnezhad, S. Samiee, A. Eichberger, M. Frühwirth, C. Kaufmann, and E. Klotz, "Applying deep neural networks for multi-level classification of driver drowsiness using vehicle-based measures," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113778.

[156] C. N. Watling, M. Mahmudul Hasan, and G. S. Larue, "Sensitivity and specificity of the driver sleepiness detection methods using physiological signals: A systematic review," *Accident Anal. Prevention*, vol. 150, Feb. 2021, Art. no. 105900.

[157] Z. Elamrani Abou Elassad, H. Mousannif, H. Al Moatassime, and A. Karkouch, "The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103312.

QAISAR ABBAS is currently working as an Associate Professor with the College of Computer and Information Sciences, Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia. He has numerous research articles in both journals and conferences papers. His research interests include artificial intelligence, computer vision, and deep learning.

ABDULLAH ALSHEDDY is currently working as an Associate Professor with the College of Computer and Information Sciences, Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia. He has numerous research articles in both journals and conferences papers. His research interests include artificial intelligence, optimization, and soft computing.

• • •