A review of driver distraction detection while driving based on convolutional neural networks

Ghady Alhamad, Mohamad-Bassam Kurdy
Department of Information Technology, Syrian Virtual University, Damascus, Syria

Article Info

Article history:

Received Sep 1, 2024
Revised Jul 5, 2025
Accepted Oct 16, 2025

Keywords:

Convolutional neural network
Cognitive distraction
Distraction
Driver behavior
Facial features

ABSTRACT

Driver distraction represents a major cause of traffic accidents, posing a serious threat to human life. In this review, we present the latest research findings of driver distraction detection based on convolutional neural networks (CNNs). In general, the analysis of driver behavior while driving is represented by either detecting driver drowsiness or attention diversion from driving by other activities, all of which fall under the definition of driver distraction. Facial features are often the basis for detecting driver drowsiness. In most papers, it is typically done by eye blinking, yawning, and head movement. As for the driver attention diversion, it is through the position of the hand and face. It involves many activities, text messages, making phone calls, adjusting the radio, consuming beverages, reaching for objects behind the driver, applying makeup, interacting with passengers, and other similar distractions. However, suggesting new methodologies in driver distraction detection and choosing appropriate CNN-based techniques is a big challenge given the wide variety experiments and studies in this field. Therefore, previous papers should be revisited to produce new methods by taking advantage of the techniques used. As a result, this paper reviews research approaches and reveals the effectiveness of CNN in detecting driver distraction. Finally, the article lists techniques that can be used as benchmarks in this context.

Corresponding Author:

Ghady Alhamad
Department of Information Technology, Syrian Virtual University
Damascus, Syria
Email: ghady_213452@svuonline.org

1. INTRODUCTION

Distracted driving has become a dominant cause of traffic accidents [1]. According to the National Highway Traffic Safety Administration (NHTSA), distracted driving refers to any action that diverts a driver's focus away from safely operating the vehicle, and includes anything that shifts attention from the task of driving [2]. Based on its report from the United States, between 2011 and 2020, approximately 32,483 people lost their lives in crashes influenced by driver distraction. In 2020 alone, distraction-related fatalities reached 3,142 nationwide, accounting for 8% of all motor vehicle deaths, marking an increase of 23 compared to 2019. Crashes involving distracted driving represented 14% of injury crashes and 13% of all police-reported traffic accidents that year. Among drivers aged 15 to 20 involved in fatal crashes, 7% were reported as distracted, making this age group the most affected by distraction during deadly accidents. Additionally, 396 fatalities were linked specifically to cell phone-related distractions, comprising 13% of all deaths involving a distracted driver. In 2020, distracted driving was responsible for the deaths of 587 non-occupants, including pedestrians, cyclists, and other individuals [3].

There are many reasons for distraction that can lead to impaired driving and accidents. The use of mobile phones is one of the main causes of distraction. In fact, using mobile devices is not the only cause of distraction, but according to NHTSA, it also includes, talking or texting on your phone, eating and drinking, talking to people in your vehicle, fiddling with stereo, entertainment or navigation system [2]. The centers for disease control and prevention (CDC) identifies three main categories of distracted driving: cognitive, visual, and manual. Cognitive distractions occur when a driver's mind wanders away from the task of driving, meaning that even if their body remains in a proper driving posture, their focus is mentally diverted. Visual distractions happen when a driver's eyes are taken off the road due to fatigue, drowsiness, inattention, or the use of smartphones. Manual distractions involve temporarily taking the hands off the steering wheel to perform tasks such as using a phone, eating or drinking, grooming, or interacting with passengers [4]. Previous work has aimed to address the issues related to detecting driver fatigue and drowsiness [5]–[7] as visual distraction, and detecting attention diversion from driving by other activities [8]–[10] as manual distraction. However, the studies still lack attention to cognitive distraction. Even the definitions of cognitive distraction are not completely agreed upon in the field of driving safety. Some researchers have defined cognitive distraction in ways that overlap with visual distraction [11], or with the concept of driver mental workload [12]. Other studies define it as shifting attention to secondary tasks that are not related to the driving task [1]. In fact, from my point of view, current studies have not paid enough attention to the driver's feelings and psychological state as cognitive distraction, for example. Although Chan and Singhal [13] in old study analysed the relationship between emotional side and cognitive distraction. To achieve this goal, a driving simulator was used, and the emotional words were divided into three categories: neutral, negative, and positive. In addition, Chand and Karthikeyan [14] proposed a model composed of two main components: detecting driver fatigue and analyzing the driver's emotional state to prevent reckless driving. They combined fatigue assessment with emotion analysis, observing that driver behavior can vary across multiple states, including normal, fatigued, aggressive, disturbed, and under the influence of alcohol.

Recently, a variety of methods have been proposed for driver distraction detection. Euclidean aspect ratio (EAR) [15], [16], percentage of eyelid closure (PERCLOS), frequency of open mouth (FOM) [17], support vector machine (SVM) [14], [17]–[19] model, long short-term memory (LSTM) [20], [21], convolutional neural networks (CNN) [14], [19]–[23]. EAR which measures the distance between vertical and horizontal eye landmark points by using an Euclidean distance method to detect the eye state [15]. PERCLOS represents the proportion of frames in which the eyes are closed relative to the total frames within a given time period, while FOM refers to the proportion of frames in which the mouth is open compared to the total frames over the same time interval [17]. The SVM model is especially suitable for classification tasks involving small sample sizes [20]. LSTM is a specialized variant of the recurrent neural network (RNN), typically used with CNN, and can effectively capture the time information of the input image sequence [24]. RNNs are well-suited for analyzing time series data; however, they are generally not regarded as effective for image processing tasks. However, CNN is the most promising way in computer vision-based. It is the most established algorithm among various deep learning models [25]. CNNs resemble standard neural networks in that they are composed of neurons with learnable weights and biases [26]. It is very widely used to perform image classification, object detection, image recognition, face recognition and several other tasks related to image processing [27].

Since the solution is based on CNN, this will lead researchers to select the best among a set of techniques to achieve their goal, such as preprocessing techniques, model architecture and CNN's approaches, datasets, and methodologies. The dealing of each of these technologies may sometimes differ depending on the type of distraction. Detecting driver distraction requires real-time driving monitoring. Therefore, the video is captured by either a smartphone camera or an attached camera. Birrell and Fowkes [28] recorded videos with four cameras mounted inside the vehicle. The first camera, attached to a smartphone, recorded high-definition video focused on the driver's face. The other three cameras recorded in standard definition: two monitored the forward and rearward driving scenes, while the third captured the activity on the smartphone. As for Mali *et al.* [29] used the front camera of a smartphone to capture images of the driver, and then feed the images to the smartphone for image processing. Furthermore, many other things should be considered for those recordings. It will be discussed in section 2. A typical CNN architecture consists of three main layers: a convolutional layer, a pooling layer, and a fully connected layer [26]. The arrangement and configuration of these layers define the model, which is subsequently designed and trained on a dataset to address a specific problem. However, a pre-trained model refers to a network created and trained by others on a large dataset, intended to address a problem similar to the one at hand. Most studies reuse well-known pre-trained models, including VGG-16 and VGG-19, which are adopted by [30], [31], ResNet by [32], [33], and GoogleNet (Inception v1) by [34]. Training the models from scratch would spend a serious amount of time as well as consume huge data. Several references have shown reviews on pre-trained models, such as [35], [36]. In fact, the pre-trained model is often used in combination with one of the

"Transfer learning" or "Fine tuning" approaches. The major challenge identified that each of the reviewed studies used different datasets to achieve their goals. In addition, these datasets are often limited, and some of them are not public. ImageNet is one of the most popular public datasets. The ImageNet project is a large visual database designed for use in visual object recognition software research. More than 14 million images have been hand-annotated by the project to identify the objects present in each image, and for over one million images, bounding boxes are also included. ImageNet encompasses more than 20,000 categories [37]. There are also datasets that are more relevant to a specific topic, such as StateFarm for distracted driver detection that is used by Omerustaoglu *et al.* [38]. Due to the limited datasets, some researchers tend to train their models on their own dataset such as [39].

This paper will provide a review of driver distraction detection, considering the reliance on CNNs at all stages, leveraging appropriate techniques, and proposing new approaches in this field. This review will discuss driver distraction based on CNN with focusing on network input, pre-trained models, datasets, and methodologies. Finally, recommendations will be presented in conclusion to derive new methods for detecting driver distraction using CNN techniques.

2. NETWORK INPUT

CNN is specifically designed to process input images. These images are extracted from video frames of drivers while driving. Videos are captured by cameras mounted in the car. The researchers typically prepare the input images during a pre-processing step, then feed them into the network. They use various techniques to achieve that. Preprocessing the input images occurs more or less in the following set of common stages, as shown in Figure 1.

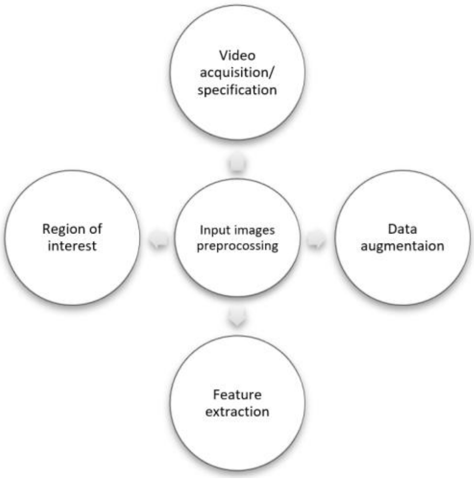


Figure 1. Input images pre-processing

2.1. Video acquisition/specification

During this stage, the video obtained from a fixed camera or smartphone is divided into individual frames for further analysis [40]. Video is captured in two ways. The first is suitable for visual distractions such as drowsiness detection systems, where only the driver's face is captured. The second is suitable for other distraction detection systems, which can use multiple cameras in different directions. The target is mostly the hands and head. Tran *et al.* [41] proposes a deep learning approach to detecting distracted driving behaviors. The images captured using two cameras, by which the body movements and face of the driver are monitored respectively. Chen *et al.* [24] presented a driver drowsiness estimation system. They use a mobile phone camera positioned in front of the driver's face to capture various shots of facial expressions, including eyes and mouth.

The video specification is related to video speed (frames per second (FPS)) number of frames, infrared camera (IR), grayscale or RGB, and image resolution. All of them could affect directly the result of model. According to Magán *et al.* [42], it is crucial to determine the appropriate frame rate at which the camera communicates with the system. A high frame rate may increase system load due to the large number of FPS that require processing, whereas a low FPS can adversely affect system performance. For instance, given that the average blink duration ranges from 100 to 400 ms [42], a frame rate of 10 FPS may suffice to detect blinks while preventing system overload. The total number of frames depends on both the video duration and the FPS. Magán *et al.* [42] evaluate 600 frames each time a new frame is captured by the camera. They evaluated the driver's drowsiness level at a given moment using data collected from the preceding 60 seconds. Some researchers tend to use an infrared camera as an IR camera allows one to construct vision-based systems that can work in different light conditions, for example during the night or during bad weather events [43]. Jabbar *et al.* [44] record with infrared camera to obtain night-time video. The resulting content is 9.5 hours of videos that have a resolution of 640×480 at 30 FPS. While some rely on RGB cameras. In addition, others convert frames to grayscale, since most libraries that used to detect objects take grayscale images, such as OpenCV. Image resolutions are often determined on the basis of models and proposed methods. Table 1 shows how the papers prepocessed the input images.

Table 1. Preprocessing the input images-video specification

Authors	Camera	FPS	Frames number	Image resolution	Image color	Behavior	Distraction type	Accuracy (%)
Florez *et al.* [45]	-	25	7,500-3,000	112×112	Gray-scale	Fatigue	Visual	99.71
Jabbar *et al.* [44]	Mobile camera/ on dashboard	30	600,000	640×480	Infrared camera	Fatigue	Visual	88
Ghazal *et al.* [46]	Dashboard camera	14.9	-	100×100	Gray-scale	Fatigue	Visual	95
Tran *et al.* [41]	Dual cameras:in front of the driver/right side of the driver	8	-	64×64	RGB	Fatigue, texting, phone, radio, reaching backward, adjusting hair or makeup, and interacting with passengers	Visual Manual	96.7
Magán *et al.* [42]	Camera mounted under the front mirror	10	600	64×64	-	Fatigue	Visual	Training data: 60 Testing data: 60 Fuzzy: 93

2.2. Data augmentation

Data augmentation refers to the process of artificially generating new data from existing training datasets. Common techniques include resizing, flipping, rotating, cropping, and padding. This approach helps mitigate issues such as overfitting and limited data availability, while enhancing model robustness and overall performance [47]. For example, Jabbar *et al.* [44] utilized CodeBox to create additional images by applying a series of augmentation operations to those extracted from video frames, as illustrated in Figure 2 in addition, Table 2 shows other data augmentation articles examples.

Figure 2. Data augmentation [44]

Table 2. Preprocessing the input image-data augmentation

Authors	Data augmentation	Behavior	Distraction type	Accuracy
Florez *et al.* [45]	Rotation range: 20% Horizontal: true Fill mode: nearest Result: 5 images of each image	Fatigue	Visual	99.71%
Chand and Karthikeyan [14]	Brightness range: 75% Rotation interval: ±2 degree Sheer range: ±2% Zoom transformation interval: ±2%	Fatigue, drunken, reckless	Visual Cognitive	93%
Chen *et al.* [4]	Spatial stream: cropping, rotating, horizontal flipping, and shifting. Temporal stream: random cropping and horizontal flipping.	Texting, phone, radio, reaching, reaching backward, adjusting hair or makeup, and interacting with passengers	Manual	Enhance the accuracy by approximately 30%
Aytekin and Mençik [30]	Random rotation	Fatigue	Visual	91%

2.3. Feature extraction

It is a method of CNN that recognizes key patterns in an image for classification purposes. Feature extraction in drowsiness detection systems depends mainly on extracting the coordinates of facial landmarks from the image, such as eyes and mouth. This aims to classify driver fatigue through yawning and eyes blinking, such as [24], [42], [46]. As for other driver distraction detection systems, they are not limited to extracting the facial landmarks. Zhang *et al.* [48] extracted facial, mouth, and hand features from images captured by a camera mounted on the vehicle's dashboard. Tran *et al.* [41] extracted facial and body features using two cameras, one mounted in front of the driver and the other on the driver's right side. Figure 3 shows face and eyes feature extraction and Figure 4 shows body feature extraction examples. Studies use various techniques to extract these features, Yan *et al.* [21] investigate the contextual cues that most significantly influence specific driver actions by analyzing skin-like regions. Skin regions are initially extracted using a Gaussian mixture model (GMM) trained on skin images, and subsequently processed by R*CNN to classify the driver's actions.

Figure 3. Face or eyes feature extraction [46]

Figure 4. Body feature extraction [20]

Seshadri *et al*. [49] employed a histogram of oriented gradients (HOG) in combination with an AdaBoost classifier, trained separately for each side of the facial regions, to categorize mobile phone usage as right hand, left hand, or none. Das *et al.* [50] introduced a video-based hand detection dataset in an automotive environment and utilized an aggregate channel features (ACF) object detector. Ghazal *et al.* [46] employed a learning-based approach for face detection that relies on the Haar-like features proposed by Viola and Jones, combined with cascade classifiers. To optimize processing efficiency, an integral image is utilized to minimize the computational cost of initial image processing and to enable fast calculation of rectangular features. The value of the integral image at any point (x, y) is computed in a single pass over the original image and represents the sum of all pixel intensities located above and to the left of (x, y), as expressed in (1).

$$I(x,y) = i(x,y) + I(x-1,y) + I(x,y-1) + I(x-1,y-1), \tag{1}$$

Where the integral image is $\sum_{x' \le x, y' \le y} i(x', y')$, and i (x, y), is the original image.

To enable rapid and accurate face detection, the authors applied the AdaBoost algorithm to construct a lightweight yet effective face detector derived from the calculated image feature values obtained through the integral image method. Haar-like features and a cascade of classifiers were utilized to identify both facial regions and specific subregions such as the eyes and frontal face. This section provides a brief overview of several feature recognition techniques, with additional examples summarized in Table 3.

Table 3. Preprocessing the input image-feature extraction

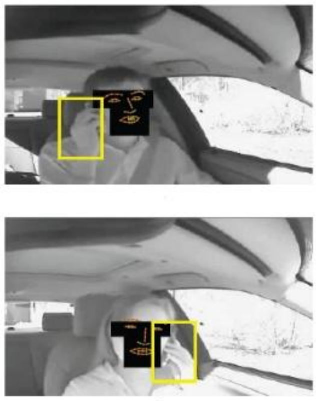
Authors	Measure	Behavior	Distraction type	Accuracy
Florez *et al.* [45]	Eye	Fatigue	Visual	99.71%
Chand and Karthikeyan [14]	Eye under conditions like: fatigue, drunkenness and aggression	Fatigue, drunken, reckless	Visual Cognitive	93%
Chen *et al.* [4]	Distraction behavior	Texting, phone, radio, reaching backward, adjusting hair or makeup, and interacting with passengers	Manual	Increase the accuracy rate by nearly 30%
Jabbar *et al.* [44]	Yawning, reduced blink rate, head nodding, conversations, and drowsy eyes	Fatigue	Visual	88%

2.4. Region of interest

The region of interest (ROI) typically refers to the significant and relevant portions of an image. By focusing on the ROI, unnecessary processing of non-essential image areas can be avoided, thereby improving computational efficiency [51]. Figure 5 illustrates examples of ROI extraction used in different studies. Florez *et al.* [45] proposed a method for detecting driver drowsiness by concentrating on the eye region, as illustrated in Figure 5(a). Out of the 468 landmarks identified during the facial landmark detection phase, only four points were selected to define the ROI. In the MediaPipe face mesh model, the selected points were 63, 117, 293, and 346, which, when connected, form an irregular quadrilateral representing the ROI. Then, they proposed a method for ROI correction to correct the selected irregular area. Seshadri *et al.* [49] extracted another ROI, with different method.

The Viola and Jones face detection algorithm was employed to identify the subject's face in the initial frame of a video. The facial landmarks localized in this first frame were then used as initialization points for tracking in subsequent frames. A total of 49 facial landmarks were detected by the algorithm. During the training phase, regions of interest were cropped for both positive and negative class samples based on the facial alignment results. As illustrated in Figure 5(b), rectangular crops of size 50×80 were generated, using landmark 18 as the top-right corner and landmark 23 as the top-left corner of the cropped region. These regions were utilized to create positive and negative class samples corresponding to cases in which subjects were either holding or not holding a cell phone in their left or right hand. Table 4 presents a list of papers with regions of interest and facial method used.

(a)



(b)

Figure 5. ROI extraction of (a) eyes open and eyes closed and (b) right hand and left hand [45], [49]

Table 4. Preprocessing the input image-ROI extraction

Authors	Facial method	ROI	Behavior	Distraction type	Accuracy (%)
Chen *et al.* [24]	CNN-LSTM	Face/eyes and mouth	Fatigue	Visual	75
Chirra *et al.* [52]	Haar cascade	Eyes	Fatigue	Visual	96.42
Park *et al.* [53]	VGG-FaceNet	Face	Fatigue	Visual	73.06
Phan *et al.* [54]	Dlib	Face/eyes and mouth	Fatigue	Visual	97
Hashemi *et al.* [31]	Viola and Jones	Face/eyes	Fatigue	Visual	Avg. 96.42
Seshadri *et al.* [49]	Supervised descent method (SDM) algorithm	Hands	Using cell phone	Manual	-

3. PRE-TRAINED MODELS

Pre-trained CNNs refer to models that have been previously trained on large-scale datasets (e.g., ImageNet) and subsequently adapted to perform diverse tasks such as image classification, object detection, and facial recognition [55]. For example, The VGG-16 architecture is a pre-trained CNN designed for image classification. Trained on a large-scale and diverse dataset, it can be efficiently fine-tuned to perform accurately on domain-specific image classification problems [55], [56]. Pre-trained models can serve as a foundation for various tasks, even when these tasks differ significantly from the original training objective. This approach is referred to as transfer learning. Leveraging pre-trained models allows for considerable savings in both time and computational resources that would otherwise be needed to train a neural network from scratch. Moreover, by fine-tuning the pre-trained model on a specific dataset, it can be adapted to a particular problem, often requiring substantially less data [55]. Table 5 presents comparison between the most famous pre-trained models. The studies reused pre-trained models in their purposed methods. They sometimes tend to use more than one model at the same time in comparing the accuracy of their results. Some examples are presented in Table 6.

Table 5. Comparison of pre-trained models [56]

Model/Paper	Size (MB)	Parameters	Top-1 accuracy (%)	Top-5 accuracy (%)	Depth	Default image size
Xception/Chollet *et al.* [57]	88	22,910,480	79	94.5	126	229×229
VGG-16/Simonyan *et al.* [58]	528	138,357,544	71.3	90.1	23	224×224
VGG-19/Simonyan and Zisserman [58]	549	143,667,240	71.3	90	26	224×224
ResNet-50/He *et al.* [59]	98	25,636,712	74.9	92.1	-	224×224
Inception-V3/Szegedy *et al.* [60]	92	23,851,784	77.9	93.7	159	229×229
MobileNet-V2/Sandler *et al.* [61]	14	3,538,984	71.3	90.1	88	224×224

Table 6. Pre–trained models used in studies

Authors	Measure	Pre-trained models	Behavior	Distraction type	Accuracy (%)
Gu *et al.* [17]	Eye/mouth	AlexNet, ResNet, MSP-NET (developed new)	Fatigue	Visual	97.12/98.6, 97.8/98.5, 98.1/98.9
Florez *et al.* [45]	Eye	ResNet-50-V2, Inception-V3, VGG-16	Fatigue	Visual	99.7, 99.3, 99.4
Chen *et al.* [4]	Distraction behavior	VGG-16	Texting, phone, radio, reaching backward, adjusting hair or makeup, and interacting with passengers	Manual	Increase the accuracy rate by nearly 30%
Aytekin *et al.* [30]	Eye and mouth	VGG-16	Fatigue	Visual	91

4. DATASETS

One of the main challenges in this process is identifying a sufficiently large public dataset that adequately represents the expected outcomes for such systems. Therefore, this section provides a list of some public datasets in this field with detailed information about each which can be used as a benchmark.

i) National Tsing Hua University (NTHU) dataset: this dataset comprises 22 subjects from diverse ethnic backgrounds, recorded under both daytime and nighttime conditions. Simulated driving scenarios included behaviors such as Yawning, reduced blink rate, head nodding, conversations, and drowsy eyes. Videos were captured using an infrared camera at a resolution of 640×480 pixels and 30 FPS [62].
ii) Southest University Driving-posture (SEU) driving-posture dataset: developed by Zhao *et al.* [63]. this dataset consists of videos captured using a side-mounted Logitech C905 CCD camera under daylight conditions, with a resolution of 640×480 pixels. A total of 20 drivers in the dataset, comprising ten males and ten females.
iii) ZJU eyeblink dataset: consists of 80 video clips from 20 individuals, with each individual contributing four clips: frontal view without glasses, frontal view with black-frame glasses, and upward view without glasses. Eye images are classified as open or closed and separated into training and testing sets. The dataset contains 7,000 open-eye images (5,770 for training, 1,230 for testing) and 5,570 closed-eye images (1,574 for training and 410 for testing), with each image sized 24×24 pixels [64].
iv) Closed eyes in the wild (CEW) dataset: this dataset contains online images of approximately 2,423 participants from multiple racial groups, including Asians and light-skinned non-Asians. Among these, 1,192 images feature closed eyes and 1,231 show open eyes. The images were selected from the labeled faces in the wild (LFW) database [65].
v) DROZY dataset (ULg multimodality drowsiness database): includes 14 participants (3 males and 11 females), each contributing videos roughly 10 minutes long, accompanied by psychomotor vigilance test (PVT) scores measuring drowsiness. Time-synchronized Karolinska sleepiness scale (KSS) ratings are provided for each participant [65].
vi) Eye and mouth detection (EMD) dataset: comprises 36,764 eye samples and 15,185 mouth samples from 21 volunteers. The dataset covers real-world driving conditions, including participants with or without glasses, frontal and lateral views, as well as day and night environments [17].

vii) Yawn detection dataset (YawnDD) dataset: contains recordings of 322 drivers in real-car conditions, captured with in-car cameras. The dataset is divided into four classes—yawn, no-yawn, open-eye, and closed-eye—totaling 2,900 samples: 726 closed-eye, 726 open-eye, 725 no-yawn, and 723 yawn images [30].

5. METHOD

In this section, proposed approaches and mechanisms adopted in 4 papers to detect driver distraction will be listed, in completion of the techniques and methods presented in the previous sections.

5.1. Chen *et al.* [24]/2020

They introduced a driver drowsiness detection model that integrates factorized bilinear feature fusion with an LSTM-based recurrent convolutional network to accurately identify signs of driver sleepiness, as illustrated in Figure 6. The primary contributions of their research include:

i) Design of a novel multilevel driver drowsiness estimation system composed of the following main components: extraction of deep feature representations associated with the driver's eyes and mouth from the dataset; fusion of features related to fatigue indicators; and temporal modeling of fatigue features through a long short-term recurrent convolutional network (LSTM).
ii) Regarding fatigue feature fusion, they proposed a new factorized bilinear feature fusion model suitable for multi-model feature input and performed bilinear fusion of the extracted deep feature representations of eyes and mouth to solve the limitations of the feature linear fusion process.

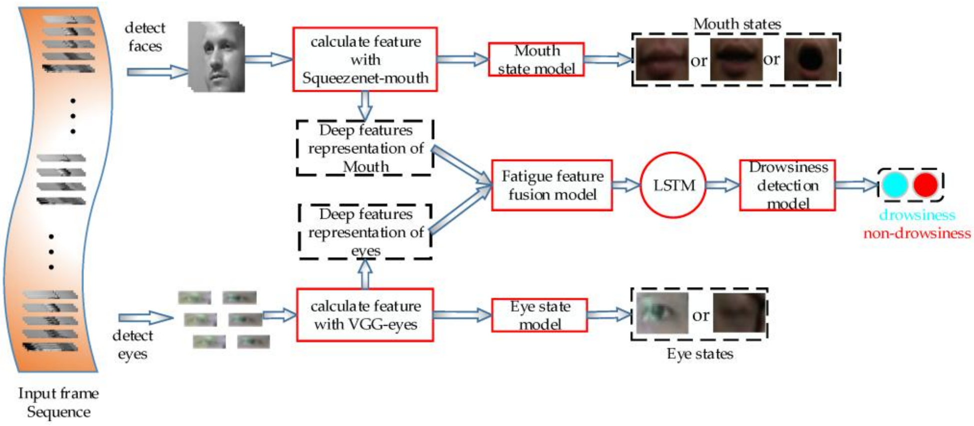


Figure 6. Overview of the proposed framework architecture [24]

5.2. Magán *et al.* [42]/2022

The aim of this work is to develop a system capable of estimating driver fatigue using sequences of images in which the subject's face is clearly visible. The system includes the following components:

i) They perform fatigue detection tasks at a given moment based on the analysis of a sequence of images for the last 60 s.
ii) In this study, two alternative solutions are presented, focusing on minimizing false positives.
iii) The first approach employs a combination of a RNN and CNN.
iv) The second approach utilizes deep learning techniques to extract numerical features from the images, which are then processed by a fuzzy logic-based system.
v) A Gaussian blur was applied to the original image to minimize noise and soften edges, ensuring that the main content and structure of the image remained largely unaffected.
vi) The DLIB library was used to detect the facial region within each image.
vii) Face coordinates were determined using HOG features in conjunction with a linear SVM.

5.3. Chen *et al.* [4]/2020

They developed a two-stream CNN architecture for distraction detection, as depicted in Figure 7. The architecture operates as follows:

i) The model comprises three main sub-networks: a spatial stream convolutional network, a temporal stream convolutional network, and a fusion network.
ii) The spatial and temporal streams are responsible for extracting spatial and temporal features, respectively, which are subsequently combined within the fusion network.

iii) Leveraging transfer learning, the spatial stream network is constructed based on the well-known VGG-16 architecture, allowing the use of pre-trained weights from the ImageNet dataset.
iv) Initially, average pooling across the temporal dimension is applied to 10 consecutive RGB frames, and the output is then forwarded to the subsequent layers of the spatial stream network.
v) Finally, the fusion network comprising two convolutional layers and two fully connected layers-merges the spatial and temporal representations to classify ten distinct types of distracted driving behaviors.

5.4. Ghazal *et al.* [46]/2018

A low-cost and real-time embedded system for fatigue detection is proposed in this study using CNN. The paper includes many ideas, as follows:
i) Pre-trained model weights were loaded to leverage transfer learning and reduce reliance on large-scale datasets.
ii) Face detection was applied to define the ROI.
iii) The study focused on solutions suitable for resource-constrained devices, aiming for a scalable, cost-effective, and robust implementation. For this purpose, a Raspberry Pi 3 Model B was employed due to its affordability.

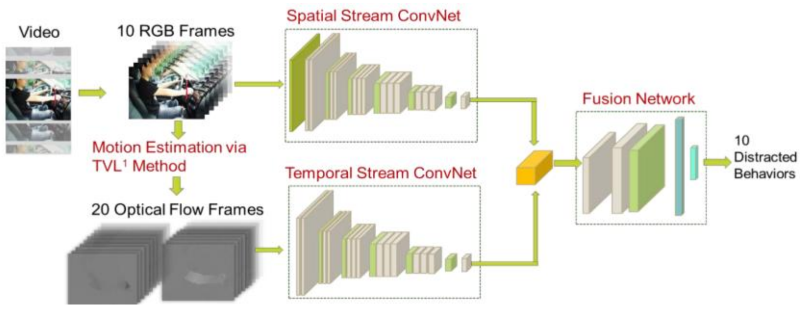


Figure 7. Driver distraction system flowchart [4]

6. CONCLUSION

The driver distraction detection systems play an assistant role in protecting the driver from traffic accidents. Deep learning models have the potential to become a mainstream solution due to their high accuracy and stable robustness, provided that sufficiently large datasets are available for training. This paper reviews a selection of case studies conducted between 2014 and 2024 that focus on deep learning and CNN-based approaches. Many successful studies have demonstrated the effectiveness of CNNs in detecting driver distraction, sometimes combined with other techniques, have high accuracy, and have advantages and disadvantages. The methodologies proposed so far in all the papers are effective and can integrate together to launch new creative findings and ideas. There are three types that can be used to measure level of distraction (cognitive, visual, and manual). It can be noted that the papers reviewed focused more on visual and manual distraction than on cognitive distraction, because researchers did not pay sufficient attention to cognitive distraction. Perhaps the largest problem however is that the definitions typically refer to distraction as a reduction of attention, but never define attention. Actually, cognitive distraction has a negative impact on driving and can lead to serious accidents. As a result, we recommend digging more into cognitive distraction by taking advantage of CNN and the techniques reviewed. It is also recommended to build large public dataset covering this aspect to support broader studies.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ghady Alhamad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Mohamad-Bassam Kurdy	✓	✓								✓		✓	✓	

C	:	**C**onceptualization	I	:	**I**nvestigation	Vi	:	**Vi**sualization
M	:	**M**ethodology	R	:	**R**esources	Su	:	**Su**pervision
So	:	**So**ftware	D	:	**D**ata Curation	P	:	**P**roject administration
Va	:	**Va**lidation	O	:	Writing - **O**riginal Draft	Fu	:	**Fu**nding acquisition
Fo	:	**Fo**rmal analysis	E	:	Writing - Review & **E**diting			

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

DATA AVAILABILITY

This paper presents seven datasets that have been used and are available at:

– NTHU dataset at https://www.kaggle.com/datasets/banudeep/nthuddd2.
– SEU dataset is not available public.
– ZJU dataset at https://github.com/elnino9ykl/ZJU-Dataset.
– CEW dataset at https://parnec.nuaa.edu.cn/_upload/tpl/02/db/731/template731/pages/xtan/ClosedEyeDatabases.html.
– DROZY dataset at http://www.drozy.ulg.ac.be/.
– EMD dataset at https://www.kaggle.com/datasets/umapriyasr/eye-mouth-detection-dataset.
– YawnDD dataset at https://qualinet.github.io/databases/video/yawdd_a_yawning_detection_dataset/.

REFERENCES

[1] Q. Hua, L. Jin, Y. Jiang, M. Gao, and B. Guo, "Cognitive distraction state recognition of drivers at a nonsignalized intersection in a mixed traffic environment," *Advances in Civil Engineering,* vol. 2021, pp. 16, 2021, doi: 10.1155/2021/6676807

[2] NHTSA, "Distracted driving," *National Highway Traffic Safety Administration*, U.S., 2022. Accessed: Dec. 31, 2023. [Online]. Available: https://www.nhtsa.gov/risky-driving/distracted-driving

[3] Traffic Safety Marketing, "2023 distraction fact sheet," *Traffic Safety Marketing*, 2023. Accessed: Dec. 31, 2023. [Online]. Available: https://www.trafficsafetymarketing.gov/sites/tsm.gov/files/2023-02/15776a-Distraction%20Fact%20Sheet_%20020323v2-tag_0.docx

[4] J.-C. Chen, C.-Y. Lee, P.-Y. Huang, and C.-R. Lin, "Driver behavior analysis via two-stream deep convolutional neural network," *Applied Sciences,* vol. 10, no. 6, 2020, doi: 10.3390/app10061908.

[5] A. Dasgupta, A. George, S. L. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Transactions on Intelligent Transportation Systems,* vol. 14, no. 4, pp. 1825-1838, 2013, doi: 10.1109/TITS.2013.2271052.

[6] J.-J. Yan, H.-H. Kuo, Y.-F. Lin, and T.-L. Liao, "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing," in *2016 International Symposium on Computer, Consumer and Control*, 2016, pp. 243-246, doi: 10.1109/IS3C.2016.72.

[7] W. Zhang, B. Cheng, and Y. Lin, "Driver drowsiness recognition based on computer vision technology," *Science and Technology,* vol. 17, no. 3, pp. 354-362, 2012, doi: 10.1109/TST.2012.6216768.

[8] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," *Expert Systems with Applications,* vol. 149, 2020, doi: 10.1016/j.eswa.2020.113240.

[9] L. Li, B. Zhong, C. H. Jr, Y. Liang, W. J. Horrey, and X. Xu, "Detection of driver manual distraction via image-based hand and ear recognition," *Accident Analysis & Prevention,* vol. 137, 2020, doi: 10.1016/j.aap.2020.105432.

[10] C. Huang, X. Wang, J. Cao, S. Wang, and Y. Zhang, "HCF: a hybrid CNN framework for behavior detection of distracted drivers," *IEEE access,* vol. 8, pp. 109335-109349, 2020, doi: 10.1109/ACCESS.2020.3001159.

[11] E. Muhrer and M. Vollrath, "The effect of visual and cognitive distraction on driver's anticipation in a simulated car following scenario," *Transportation research part F: traffic psychology and behaviour,* vol. 14, no. 6, pp. 555-566, 2011, doi: 10.1016/j.trf.2011.06.003.

[12] Y. Yang, Z. Ye, S. M. Easa, Y. Feng, and X. Zheng, "Effect of driving distractions on driver mental workload in work zone's warning area," *Transportation research part F: traffic psychology and behaviour,* vol. 95, pp. 112-128, 2023, doi: 10.1016/j.trf.2023.03.018.

[13] M. Chan and A. Singhal, "The emotional side of cognitive distraction: Implications for road safety," *Accident Analysis & Prevention,* vol. 50, pp. 147-154, 2013, doi: 10.1016/j.aap.2012.04.004.

[14] H. V. Chand and J. Karthikeyan, "CNN based driver drowsiness detection system using emotion analysis," *Intelligent Automation & Soft Computing,* vol. 31, no. 2, 2022, doi: 10.32604/iasc.2022.020008.

[15] A. K. Biswal, D. Singh, B. K. Pattanayak, D. Samanta, and M.-H. Yang, "IoT-based smart alert system for drowsy driver detection," *Wireless communications and mobile computing,* vol. 2021, doi: 10.1155/2021/6627217.

[16] S. Mehta, S. Dadhich, S. Gumber, and A. J. Bhatt, "Real-time driver drowsiness detection system using eye aspect ratio and eye closure ratio," in *Proceedings of international conference on sustainable computing in science, technology and management, Amity University Rajasthan*, 2019, pp. 1333-1339, doi: 10.2139/ssrn.3356401.

[17] W. H. Gu, Y. Zhu, X. D. Chen, L. F. He, and B. B. Zheng, "Hierarchical CNN-based real-time fatigue detection system by visual-based technologies using MSP model," *IET Image Processing,* vol. 12, no. 12, pp. 2319-2329, 2018, doi: 10.1049/iet-ipr.2018.5245.

[18] S. Zepf, T. Stracke, A. Schmitt, F. v. d. Camp, and J. Beyerer, "Towards real-time detection and mitigation of driver frustration using SVM," in *2019 18th IEEE International Conference on Machine Learning and Applications*, Boca Raton, Florida, 2019, doi: 10.1109/ICMLA.2019.00039.

[19] J. F.-Monroy, M. N.-Miyatake, G. S.-Perez, and H. P.-Meana, "Visual-based real time driver drowsiness detection system using CNN," in *2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control*, Mexico City, Mexico, 2021, doi: 10.1109/CCE53527.2021.9633082.

[20] H. Chen and L. Chen, "Support vector machine classification of drunk driving behaviour," *International journal of environmental research and public health,* vol. 14, no. 1, 2017, doi: 10.3390/ijerph14010108.

[21] S. Yan, Y. Teng, J. S. Smith, and B. Zhang, "Driver behavior recognition based on deep convolutional neural networks," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, Changsha, China, 2016, doi: 10.1109/FSKD.2016.7603248.

[22] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: a deep learning approach," *IEEE transactions on Vehicular Technology,* vol. 68, no. 6, pp. 5379-5390, 2019, doi: 10.1109/TVT.2019.2908425.

[23] M. E. V.-Recalde, J. Pérez, and J. Echanobe, "Driver monitoring system based on cnn models: an approach for attention level detection," in *Intelligent Data Engineering and Automated Learning–IDEAL 2020: 21st International Conference,* Guimaraes, Portugal, 2020, doi: 10.1007/978-3-030-62365-4_56.

[24] 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, 2020, doi: 10.3390/info12010003.

[25] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging,* vol. 9, pp. 611-629, 2018, doi: 10.1007/s13244-018-0639-9.

[26] Stanford University, "CS231n: convolutional neural networks for visual recognition," *Spring 2020.* Accessed: Dec. 31, 2023. [Online]. Available: https://cs231n.stanford.edu/

[27] V. Kaushish, N. Singh, K. Maggo, P. Nagrath, and R. Jain, "Driver drowsiness detection system with OpenCV and Keras," in *Proceedings of the International Conference on Innovative Computing & Communication*, 2021, doi: 10.2139/ssrn.3884784.

[28] S. A. Birrell and M. Fowkes, "Glance behaviours when using an in-vehicle smart driving aid: a real-world, on-road driving study," *Transportation research part F: traffic psychology and behaviour,* vol. 22, no. 2014, pp. 113-125, 2014, doi: 10.1016/j.trf.2013.11.003.

[29] S. Mali, K. Bangale, R. Walekar, and P. P. Shitole, "Driver drowsiness detection using smartphone," *International Journal for Research in Applied Science and Engineering Technology,* vol. 10, no. 7, 2022, doi: 10.22214/ijraset.2022.45003.

[30] A. Aytekin and V. Mençik, "Detection of driver dynamics with VGG16 Model," *Applied Computer Systems,* vol. 27, no. 1, pp. 83-88, 2022, doi: 10.2478/acss-2022-0009.

[31] M. Hashemi, A. Mirrashid, and A. B. Shirazi, "Driver safety development: real-time driver drowsiness detection system based on convolutional neural network," *SN Computer Science,* vol. 1, no. 5, pp. 289, 2020, doi: 10.1007/s42979-020-00306-9.

[32] S. Ma, T. Huang, X. Sun, and Y. Wei, "Driver drowsiness detection based on ResNet-18 and transfer learning," in *2021 33rd Chinese Control and Decision Conference*, Kunming, China, 2021, pp. 2390-2394, doi: 10.1109/CCDC52312.2021.9601668.

[33] A. Jeyasekar and V. R. Iyengar, "Driver's drowsiness detection based on behavioural changes using resnet," *International Journal of Recent Technology and Engineering,* vol. 8, no. 3, pp. 5708-5712, 2019, doi: 10.35940/ijrte.B2494.098319.

[34] P. S. Chaitanya, P. B. Lakshmi, B. Suchita, S. R. Kowsar, and K. J. Rani, "Distracted driver detection using inception V1," in *2023 4th International Conference on Electronics and Sustainable Communication Systems*, Coimbatore, India, 2023, pp. 215-221, doi: 10.1109/ICESC57686.2023.10193551.

[35] A. Stančić, V. Vyroubal, and V. Slijepčević, "Classification efficiency of pre-trained deep CNN models on camera trap images," *Journal of Imaging,* vol. 8, no. 2, 2022, doi: 10.3390/jimaging8020020.

[36] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data,* vol. 8, no. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[37] D. O. Durusoy, "Open-source datasets for image processing and artificial intelligence research: a comparison of imageNet and MS COCO datasets," *International Journal of Sciences and Innovation Engineering*, vol. 2, no. 5, pp. 639–653, 2025, doi: 10.70849/IJSCI0205202575.

[38] F. Omerustaoglu, C. O. Sakar, and G. Kar, "Distracted driver detection by combining in-vehicle and image data using deep learning," *Applied Soft Computing,* vol. 96, 2020, doi: 10.1016/j.asoc.2020.106657.

[39] F. Faraji, F. Lotfi, J. Khorramdel, A. Najafi, and A. Ghaffari, "Drowsiness detection based on driver temporal behavior using a new developed dataset," *arXiv:2104.00125,* 2021.

[40] M. Ngxande, J.-R. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: a review of state-of-art techniques," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics*, Bloemfontein, South Africa, 2017, pp. 156-161, doi: 10.1109/RoboMech.2017.8261140.

[41] D. Tran, H. M. Do, J. Lu, and W. Sheng, "Real-time detection of distracted driving using dual cameras," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, USA, 2020, pp. 2014-2019, doi: 10.1109/IROS45743.2020.9340921.

[42] E. Magán, M. P. Sesmero, J. M. A-Weber, and A. Sanchis, "Driver drowsiness detection by applying deep learning techniques to sequences of images," *Applied Sciences,* vol. 12, no. 3, 2022, doi: 10.3390/app12031145.

[43] G. Borghi, E. Frigieri, R. Vezzani, and R. Cucchiara, "Hands on the wheel: a dataset for driver hand detection and tracking," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition*, Xi'an, China, 2018, pp. 564-570, doi: 10.1109/FG.2018.00090.

[44] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver drowsiness detection model using convolutional neural networks techniques for android application," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies*, Doha, Qatar, 2020, pp. 237-242, doi: 10.1109/ICIoT48696.2020.9089484.

[45] R. Florez, F. P.-Quispe, R. J. C.-Castillo, J. C. H.-Levano, T. Paixão, and A. B. Alvarez, "A CNN-based approach for driver drowsiness detection by real-time eye state identification," *Applied Sciences,* vol. 13, no. 13, 2023, doi: 10.3390/app13137849.

[46] M. Ghazal, Y. A. Haeyeh, A. Abed, and S. Ghazal, "Embedded fatigue detection using convolutional neural networks with mobile integration," in *2018 6th International Conference on Future Internet of Things and Cloud Workshops*, Barcelona, Spain, 2018, pp. 129-133, doi: 10.1109/W-FiCloud.2018.00026.

[47] M. Johnson, "What is data augmentation in a CNN? Python examples," *NNART*, 2023. Accessed: Feb. 5, 2024. [Online]. Available: https://nnart.org/what-is-data-augmentation-in-a-cnn/

[48] X. Zhang, N. Zheng, F. Wang, and Y. He, "Visual recognition of driver hand-held cell phone use based on hidden CRF," in *Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety*, Beijing, China, 2011, pp. 248-251, doi: 10.1109/ICVES.2011.5983823.

[49] K. Seshadri, F. J. -Xu, D. K. Pal, M. Savvides, and C. P. Thor, "Driver cell phone usage detection on strategic highway research program (SHRP2) face view videos," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Boston, MA, USA, 2015, pp. 35-43, doi: 10.1109/CVPRW.2015.7301397.

[50] N. Das, E. O.-Bar, and M. M. Trivedi, "On performance evaluation of driver hand detection algorithms: challenges, dataset, and metrics," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Gran Canaria, Spain, 2015, doi: 10.1109/ITSC.2015.473.
[51] Q. Zhang and H. Xiao, "Extracting regions of interest in biomedical images," in *2008 International Seminar on Future BioMedical Information Engineering*, Wuhan, China, 2008, pp. 3-6, doi: 10.1109/FBIE.2008.8.
[52] V. R. R. Chirra, S. R. Uyyala, and V. K. K. Kolli, "Deep CNN: a machine learning approach for driver drowsiness detection based on eye state," *Revue d'Intelligence Artificielle,* vol. 33, no. 6, pp. 461-466, 2019, doi: 10.18280/ria.330609.
[53] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks," in *Computer Vision – ACCV 2016 Workshops*, 2016, pp. 154-164, doi: 10.1007/978-3-319-54526-4_12.
[54] A.-C. Phan, N.-H.-Q. Nguyen, T.-N. Trieu, and T.-C. Phan, "An efficient approach for detecting driver drowsiness based on deep learning," *Applied Sciences,* vol. 11, no. 18, 2021, doi: 10.3390/app11188441.
[55] D. Masood, "Pre-trained CNN architectures design, performance analysis and comparison," *Medium*, 2023. Accessed: Feb. 20, 2024. [Online]. Available: https://medium.com/@daniyalmasoodai/pre-train-cnn-architectures-designs-performance-analysis-and-comparison-802228a5ce92
[56] Keras, "Keras applications," *Keras.* Accessed: Mar. 5, 2024. [Online]. Available: https://keras.io/api/applications/
[57] F. Chollet, "Xception: deep learning with depthwise separable convolutions," In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, doi: 10.1109/CVPR.2017.195.
[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, doi: 10.1109/CVPR.2016.90.
[60] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
[61] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. -C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
[62] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Computer Vision–ACCV 2016 Workshops*, Taiwan, Taipei, 2016, pp. 117–133, doi: 10.1007/978-3-319-54526-4_9.
[63] C. H. Zhao, B. Zhang, J. He, and J. Lian, "Recognition of driving postures by contourlet transform and random forests," *IET Intelligent Transport Systems,* vol. 6, no. 2, pp. 161-168, 2012, doi: 10.1049/iet-its.2011.0116.
[64] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognition,* vol. 47, no. 9, pp. 2825-2838, 2014, doi: 10.1016/j.patcog.2014.03.024.
[65] M. Ngxande, J.-R. Tapamo, and M. Burke, "Bias remediation in driver drowsiness detection systems using generative adversarial networks," *IEEE Access,* vol. 8, pp. 55592-55601, 2020, doi: 10.1109/ACCESS.2020.2981912.

BIOGRAPHIES OF AUTHORS

Ghady Alhamad obtained bachelor degree in Information Systems Engineering from Syrian Virtual University, Syria, December, 2012, Master in Web Science from Syrian Virtual University, Syria, Feb, 2012. She worked remotely through 2013-2024 in various promising companies in different countries. E-commerce freelancer, web/mobile applications developer in UAE/Fujairah, mobile applications developer KSA/Riyadh, react front developer Canada, web/desktop applications developer KSA/Jeddah and more. She is supervisor of Graduation Projects in Department of Information Technology, Master in Web Science, Syrian Virtual University since 2015. She did some distinctive projects: Java compiler, e-commerce sites, web applications, messengers, Tas-heel mobile application for happiness measurement, chatbot for restaurants, and view art in room mobile applications. She can be contacted at email: eng.ghadyalhamad@gmail.com.

Mohamad-Bassam Kurdy obtained master degree in Information Systems Engineering from INPG, France 1986, Ph.D. in Mathematical Morphology from Mines ParisTech, France 1990, He worked at HIAST 1991-2013. He was Head of Computer Sciences at HIAST between 1997-2003. Country Manager for Syria of EC project EUMEDIS - Medforist. Actually, Professor at SVU, ESC Dijon and ESC Rennes teaching: advanced data mining, big data, information retrieval, content-based image retrieval (cbIR) and supervising many masters student projects (postgraduate). He can be contacted at email: bassamkurdy7@gmail.com.