

Article

SOMN_IA: Portable and Universal Device for Real-Time Detection of Driver's Drowsiness and Distraction Levels

Jonathan Flores-Monroy , Mariko Nakano-Miyatake , Enrique Escamilla-Hernandez ,
Gabriel Sanchez-Perez  and Hector Perez-Meana *

Instituto Politécnico Nacional, Av. Santa Ana 1000, Coyoacan, Mexico City CP4040, Mexico

* Correspondence: hmperezm@ipn.mx

Abstract: In this paper, we propose a portable device named SOMN_IA, to detect drowsiness and distraction in drivers. The SOMN_IA can be installed inside of any type of vehicle, and it operates in real time, alerting the dangerous state caused by drowsiness and/or distraction in the driver. The SOMN_IA contains three types of alarm: light alarm, sound alarm, and the transmission of information about the driver's dangerous state to a third party if the driver does not correct his/her dangerous state. The SOMN_IA contains a face detector and a classifier based on the convolutional neural networks (CNN), and it aids in the management of consecutive information, including isolated error correction mechanisms. All of the algorithmic parts of the SOMN_IA are analyzed and adjusted to operate in real-time in a portable device with limited computational power and memory space. The SOMN_IA requires only a buck-type converter to connect to the car battery. The SONM_IA discriminates correctly between real drowsiness and normal blinking, as well as between real dangerous distraction and a driver's normal attention to his/her right and left. Although the real performance of the SOMN_IA is superior to the CNN classification accuracy thanks to isolated error correction, we compare the CNN classification accuracy with the previous systems.

Keywords: driver's drowsiness level; driver's distraction level; hardware implementation; convolutional neural networks (CNN); face detection; real-time operation; portable device



Citation: Flores-Monroy, J.; Nakano-Miyatake, M.; Escamilla-Hernandez, E.; Sanchez-Perez, G.; Perez-Meana, H. SOMN_IA: Portable and Universal Device for Real-Time Detection of Driver's Drowsiness and Distraction Levels. *Electronics* **2022**, *11*, 2558. <https://doi.org/10.3390/electronics11162558>

Academic Editor: Dah-Jye Lee

Received: 14 July 2022

Accepted: 12 August 2022

Published: 16 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vehicle accidents are a serious problem around the world. Specifically, in Mexico, this problem is alarming because Mexico ranks seventh worldwide and third in Latin America in deaths caused by vehicle accidents, according to the Mexican National Institute of Public Health [1]. Table 1 shows the causes of traffic accidents in urban and suburban areas in Mexico from 2016 to 2022, which are reported by the National Institute of Statistics, Geographics, and Informatics in Mexico (INEGI) [2]. According to this table, more than 90% of the accidents are caused by the negligence or irresponsibility of the driver, while the rest of the accidents are caused by the pedestrian or passenger's carelessness or inappropriate behaviors, mechanical problems, failures on the road and other factors such as natural agents. Additionally, according to the report by the Mexican Security Committee (CNS) [3], the main causes related to the driver's negligence or irresponsibility are 1. driving under the influence of alcohol, medicine, or drugs; 2. doing reckless maneuvers, and omission by the driver; 3. driving at excessive speed; 4. The physical condition of the driver; and 5. driving while drowsy, tired, or sleepy. Several worldwide statistical reports [4,5] show that more than 30% of traffic accidents are caused by driver drowsiness or driver distraction (33%). Various circumstances, such as a driver's lack of sleep, a driver's physical or mental fatigue, monotonous road conditions, etc., are considered as major factors of a driver's sleepiness. On the other hand, the widespread use of cell phones while driving causes driver distraction and then a lack of alertness to the surrounding situation.

Table 1. Report by INEGI: causes of land traffic accidents in urban/suburban areas in Mexico.

Variable	2016	2017	2018	2019	2020
Full events (absolute number)	360,051	367,789	365,281	362,729	301,678
Driver	91%	91%	92%	92%	95%
Pedestrian or passenger	1%	1%	1%	1%	1%
Vehicle failure	1%	1%	1%	1%	1%
Bad road condition	3%	3%	3%	3%	2%
Other causes	4%	4%	3%	3%	1%

Considering that almost all traffic accidents are caused by drivers, and that a third of driver-driven accidents are related to a driver's drowsiness level and distraction level, until now several approaches have been proposed to detect these driver's events to avoid a lamentable accident. Basically, there are three approaches that are used: vehicle behavior-based or a driving-pattern-based approach, a driver's physiological signal-based approach, and a visual-based approach. In the first approach, the data sensed from the vehicle, such as velocity, steering angle, acceleration, and lane deviation, are used to detect a driver's abnormal condition [6,7]. The principal drawback of this approach is the variation of the detection performance depending on the road condition and the driver's driving ability. In the second approach, a driver's physiological signals [8–10], such as electrooculogram (EOG), electrocardiogram (ECG), electromyogram (EMG), and electroencephalogram (EEG), are used for this purpose. Although this approach provides better performance, it is invasive to the driver because the driver must use the wearable sensors during driving. Finally, in the visual-based approach, the driver's face image or video data are analyzed to determine the driver's drowsiness and distraction levels [11–14]. The principal advantage of this approach is that it is not invasive for drivers and the performance does not depend on the driver's driving skill, vehicle type, or the road conditions. In almost all systems based on this approach, machine learning and/or deep learning techniques are introduced as a principal process of the system. For example, the face localization is carried out using a Haar-based face detector [15] or a MediaPipe face detector [16]. The drowsiness detection and/or distraction are carried out using some convolutional neural networks (CNN) [13,14], such as VGG16, MobileNet, ResNet, and some specifically designed CNNs for this purpose [11,12]. Almost all systems using this approach perform well in a laboratory environment using the GPU-based workstation, in which there are not any limitations in computing power, memory space, or energy consumption. However, in the real implementation of these systems inside of a vehicle, several of the limitations mentioned above are faced; moreover, the portability of the system must be considered.

In this research, we propose a technological device nominated SOMN_IA, which can detect two of the most dangerous driver's states during driving: drowsiness and distraction. The proposed device, which is portable and compact, can be installed inside any vehicle regardless of the type or model. We designed a CNN-based detection algorithm to operate in real time under the limited resources of a portable device. The proposed device uses a new generation of techniques employing a graphical interface as well as passive and active alarms, which allows one to communicate to the driver if he or she is driving dangerously, to avoid a possible accident. Unlike previously proposed devices in different research papers [11–14], SOMN_IA is a universal system, which allows its implementation in any type of vehicle because the device can be connected using a plug-in BUCK-type DC-DC converter to the car battery. The SOMN_IA contains, besides a main operation unit, an input-port for a single camera with a visible and an infrared spectrum, and a graphical interface, including a sound alarm to inform the driver about his/her dangerous state. The SOMN_IA is compact enough, which allows one to collocate it in any part of the vehicle, such as the windshield or anywhere in the passenger compartment. The details of the SOMN_IA are described later.

The rest of the paper is organized as follows: Section 2 describes the related works that have the same purpose as the proposed system. Section 3 describes each process of the SOMN_IA in detail. Section 4 describes the hardware implementation of SOMN_IA, while in Section 5 some experimental results, a comparison with previously reported systems, and some demonstrations are given to show the performance of SOMN_IA in real driving conditions. Finally, we conclude our work and describe some future works in Section 5.

2. Related Works

In this section, we describe reported works for the detection of a driver's level of drowsiness and/or distraction detection. Some of them only present computational results without any implementation, while others are implemented in some mobile devices, such as smartphone. Phan et al.'s proposed driver's drowsiness detection system is based on ResNet-50V2 [13]. They used transfer learning to retrain only fully connected layers of ResNet. The detection accuracy obtained was 97%. All experiments were carried out using the GPU-based workstation. In [14], the authors proposed a driver's drowsiness detector using Haar-based face detector [15] for face region delimitation and a 6-layered CNN composed of 2 Conv layers, 2 Pooling layers, and 2 fully connected layers. Under the illumination variation environment, they reported 88% accuracy in the test set.

Anber et al. proposed a driver's fatigue detector, including the detection of a driver's drowsiness and distraction. They used fine-tuned AlexNet to obtain the flatten vector before the fully connected layer as a output vector [17]. Then, the non-negative matrix factorization (NMF) is used to reduce the dimensionality of the vector, which is introduced to the support vector machine (SVM) to classify the input frame into the driver's fatigue state or alert state, with an accuracy of 99.65%. This system cannot be installed in any devices with limited computational capability. Considering the many operations and large memory space required by AlexNet, the system is not easy to implement in some portable devices.

In [18], the authors proposed a driver's drowsiness detector, using EfficientnetB0 for feature extraction from the driver's face; then, a fuzzy-based system is used to classification. The authors reported 93% accuracy for their system, under a laboratory environment. Additionally, EfficientnetB0 requires high computational power with a large space of memory, which can be an impediment to installing it into portable devices. Jobbar et al. proposed a driver's drowsiness detection system and its implementation in mobile devices [19]. First, face landmark coordinates are extracted from the face image using the Dlib library of the OpenCV and introduced to the artificial neural networks (ANN) with dropout techniques. The developed algorithm is installed in the Android to work in the mobile devices. The global accuracy reported in [19] is 81%.

Uma and Eswari [20] proposed a system with several sensors, including a camera, using Raspberry-Pi to collect the sensing data, such as the density of alcohol and smoke, transmitting them to the cloud environment, in which the sensing data are processed to distinguish between safe driving and dangerous driving, including driver's drowsiness and distraction levels [20]. The accuracy of the binary classification is 96.5%. The authors of [21] proposed a driver's drowsiness and distraction detector, in which the deformable part model (DPM), local binary patterns (LBP), and its variation are used to extract several features from face images. The extracted features are introduced to the SVM for the driver's state classification. The proposed system is installed in mobile devices, specifically smartphones, with 93.11% as the global accuracy [21].

Pattarapongsin et al. proposed a driver's drowsiness and distraction detector based on two deep learning models [22]. The first model is the single-shot detector (SSD) with a backbone of ResNet-10, which is used to detect the face region of the video frame. The second model is a CNN, which receives the face region as an input image and obtains 98 facial landmark points. Using these landmark points, the eye aspect ratio (EAR) and mouth aspect ratio (MAR) are calculated to determine the driver's levels of drowsiness and distraction. They used a Jetson Nano and GPU-based server to carry out their work.

They reported only accuracies of EAR and MAR, while the global accuracies for a driver's drowsiness detection level and a driver's distraction level are not reported. Hashemi et al. proposed a driver's drowsiness detection algorithm, in which the driver's face region is detected by Haar-based face detector, and then using 68 face landmarks, one of the eyes regions is cropped [23]. The cropped eye region is introduced to the trained fully designed neural network (FD-NN). The FD-NN is trained using the ZJU Eyeblink dataset [24], obtaining an accuracy of 98.15%.

3. Proposed System

The software part of the proposed device SOMN_IA is composed of five stages, as shown in Figure 1. The summary of these five stages is as follows:

Stage-1: "Frame separation". In this stage, independently of the video encoding algorithm used in the optical sensor, each video frame is separated to generate a sequence of images, which is analyzed in the following stage (Stage 2).

Stage 2: "Analysis of consecutive results". This stage appears two times (first part and second part) in the proposed system. In the first part of this stage, the consecutively face detection results are analyzed to determine the driver's distraction level. This section includes the face-detection algorithm, "MediaPipe face detector", which is selected due to its high accuracy and high operation speed. In the second part of this stage, the results obtained by the S-CNN (Stage 3) from the consecutive video frames are analyzed to differentiate between "normal blinking" and "real drowsiness". In the "real drowsiness", the driver has their eyes closed during a certain period exceeding a predetermined threshold (1.5 s). This threshold value is determined considering the difference between a normal blinking and a micro-sleep or drowsiness [25]. In the distraction analysis, a similar technique used for real drowsiness detection is employed.

Stage 3: "Implementation of a Shallow CNN (S-CNN)". Having the detected face region, a shallow convolutional neural network (S-CNN) is implemented, which is specifically designed to minimize the computational cost for portable device SOMN_IA while keeping the accuracy as high as possible. The S-CNN classifies the face region into three states: "open eye", "closed eye", and "distraction". In this stage, SOMN_IA does not make any decision about the driver's dangerous state, which will be carried out in the next stage, the second part of the Stage 2.

Stage 4: "Isolated prediction error correction". This stage corrects some inconsistent S-CNN classification errors due to the nonuniformity illumination and variation of the sensor position in a real driving situation. The inconsistent classification errors usually happen separately from the classification results of the previous and following frames. Then, using the results of the previous and following frames, several inconsistent classification errors are detected and corrected. This stage avoids the inadequate interruption of the alarm systems, which can cause a lamentable accident.

Stage 5: "Activation of alarms". According to the results obtained from the previous stage, the system can generate four different types of alarms. The first one is a sound alarm, which warns the driver that he or she is driving recklessly or dangerously. The second one is a visual alarm, which activates certain lighting systems to wake up the driver, and the third one refers to an informative alarm that sends alerts using the graphic user interface (GUI) of the system. If the first three alarms are not enough to alert the driver, the fourth and last alarm is activated, where SOMN_IA sends an SMS message in real time to a third party, notifying the driver's extremely dangerous state together with the driver's current location and time.

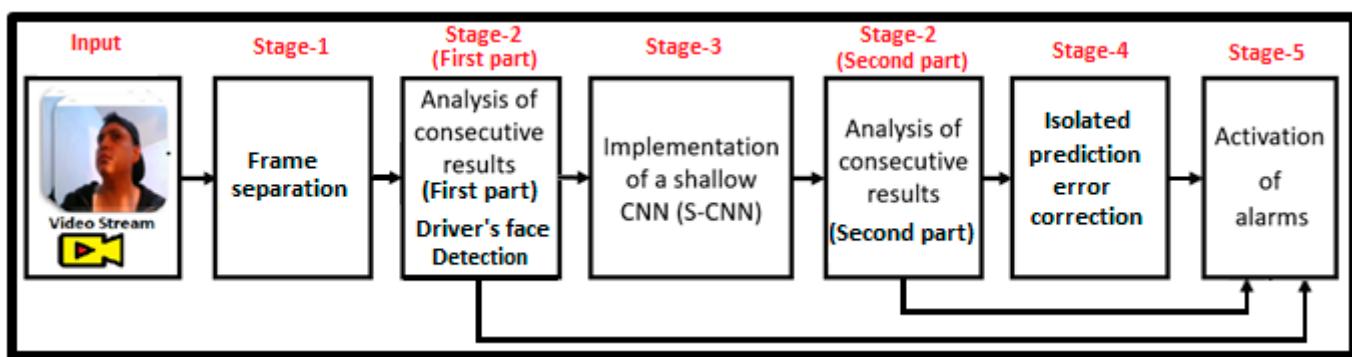


Figure 1. Proposed driver's drowsiness and distraction detection system.

3.1. Frame Separation (Stage-1)

To obtain the video data, we use an optical sensor with two LDR photoresistors, which capture the illuminance level in the environment, and according to this level one of two spectrums—the visible spectrum or infrared spectrum—is selected. Figure 2 shows the video acquisition scheme used in SOMN_IA. Once the input video sequence is obtained, independently of the video encoding algorithm, each frame is separated to generate a time sequence of images for further operation.

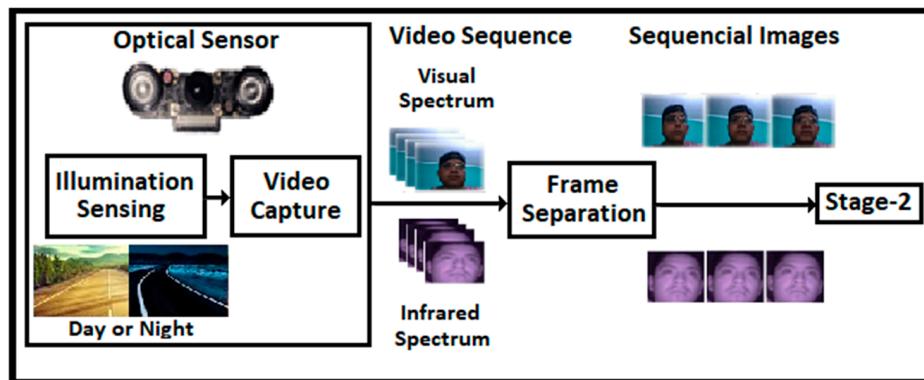


Figure 2. Process of obtaining images in different lighting conditions.

3.2. Analysis of Consecutive Results (Stage 2)

The analysis of consecutive results is carried out in two different parts, as mentioned above. The first part is carried out after the video frames are separated, and the second part of this stage is performed after the S-CNN (Stage 3), as shown in Figure 1. In this section, we describe both parts of the stage.

3.2.1. Analysis of Consecutive Results (First Part)

Before the face region is delimited to be introduced into the trained S-CNN, we must analyze if the face detector can detect the driver's face or not in each video frame because if the driver's distraction level is high, for example, the driver is looking back or the driver falls asleep, then the face detector will not be able to detect the driver's face. To analyze the face detector response for consecutive video frames, each frame is labelled as "Yes" if the face is detected; otherwise, the frame is labelled as "No". Using the label of each frame, an analysis of consecutive results is carried out for the driver's distraction level detection. The main idea of this stage is discriminate real distraction from short-time distraction, which normally occurs. This process is described in Figure 3.

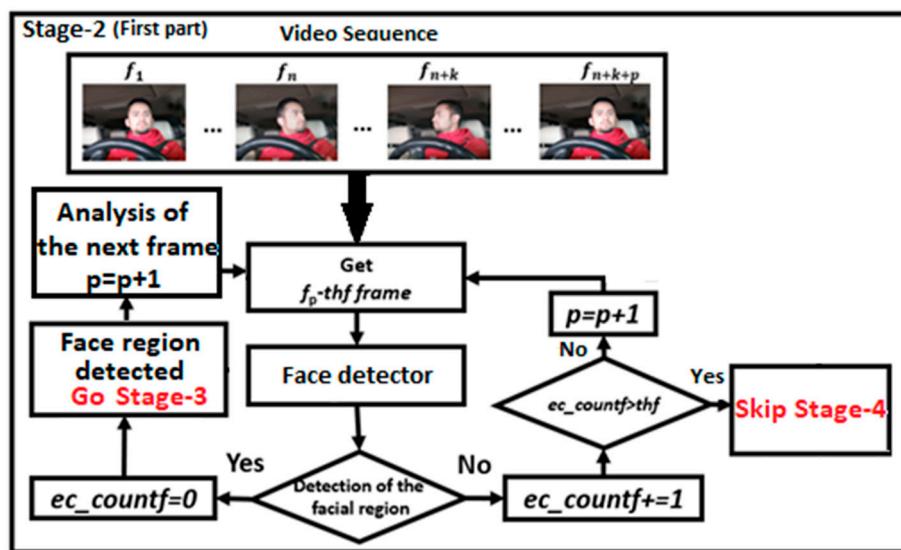


Figure 3. Analysis of consecutive results (First part of Stage 3).

In this process, we introduce a counter “*ec_countf*”, which increases if the frame is consecutively labelled as “No”. If the counter “*ec_countf*” exceeds a predetermined threshold *thf*, then the process skips to the Stage 4: “Isolated prediction error correction”. We set the value of *thf* equal to 1.8 s because the distance that the driver travels during 1.8 s with the speed of 80 Km/h is 40 m, and this distance can be considered as the upper limit to avoid a possible accident. If the system detects the face at any moment, automatically “*ec_countf*” is initialized (*ec_countf* = 0) and continues the process until the system finishes monitoring the driver.

To localize the face region, we used the MediaPipe face detector [16] because it is designed for mobile implementation, providing an accurate localization with low computational complexity [12]. We compared the MediaPipe face detector with a Haar-based face detector [15] from computer complexity and accuracy points of view. The comparison results show that the MediaPipe face detector provides a higher detection accuracy with fewer false positive errors and slightly faster detection speed [12]. In the MediaPipe face detector, six reference points (Figure 4) are used for a guideline to generate a bounding box. Using these points, four coordinates—upper-left (x_1, y_1), upper-right (x_2, y_1), lower-left (x_1, y_2) and lower-right (x_2, y_2)—are obtained to delimitate the face region. If the MediaPipe face detector cannot obtain more than four reference points of the six, the bounding box cannot be created, and as consequence of this, the face is not detected. Once the face region is delimitated by the bounding box, we expand to the three channel images copying the gray-scale face region, if the face region is gray-scale. Finally, the delimited face region is resized to 64×64 pixels in each channel. This process is depicted in Figure 5. The final image size is $64 \times 64 \times 3$, as shown by Figure 5d, which is the input image of the pretrained S-CNN.

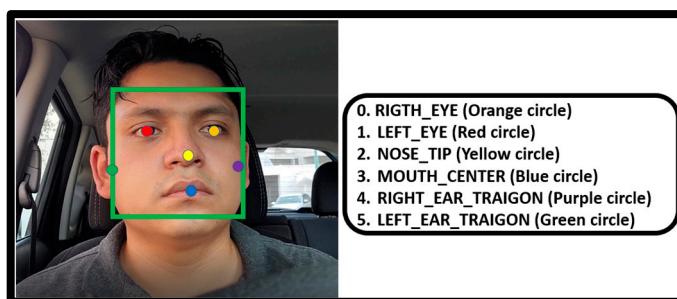


Figure 4. Reference points used to detect a face region using the MediaPipe face detection.

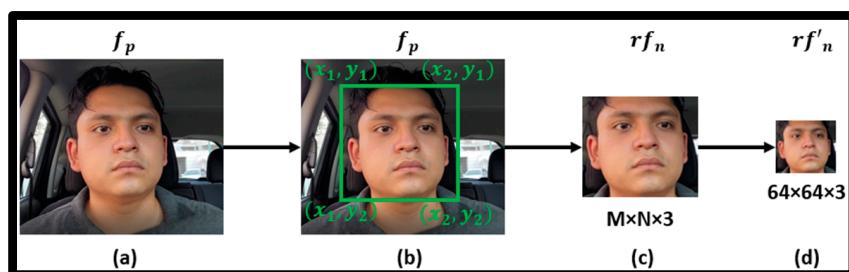


Figure 5. Detection process, fragmentation, and resizing of the driver’s face. (a) denotes the input frame, (b) shows the detected face, (c) shows the segmented face and (d) the decimated face image.

3.2.2. Analysis of Consecutive Results (Second Part)

The resized image rf'_n with three channels (Figure 5d) is introduced into the pretrained S-CNN (Stage 3), whose prediction results consist of three values: the probability of the driver paying attention (open eye), the probability of the drowsiness in the driver (closed eye), and the probability that the driver is distracted. The prediction results of consecutive video frames are analyzed to determine the real state of the driver.

The distraction detection algorithm using S-CNN is shown by Figure 6, which is similar to the “analysis of consecutive results (first part)” shown in Figure 3 and described as follows: if the highest probability predicted by the S-CNN is “driver’s distraction level”, the counter “ec_countf” is increased by one, and if this value exceeds the predetermined threshold thf, which is the same value used in Section 3.2.1, then the process passes to Stage 5; otherwise, the counter “ec_countf” is initialized to continue the analysis.

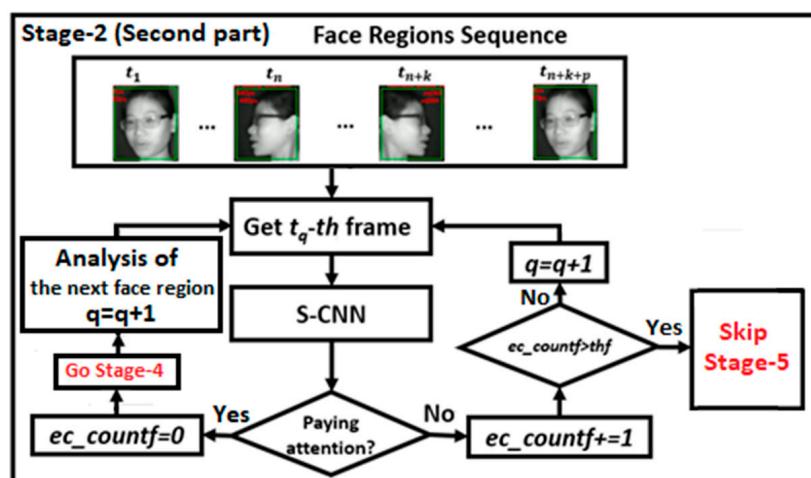


Figure 6. Distraction detection by classification.

3.2.3. Drowsiness Detection Using S-CNN’s Prediction Results

The last process of the analysis of consecutive results is the determination of real drowsiness in the driver using the consecutive prediction results of the S-CNN. The proposed S-CNN determines the probability of “open eye” or “closed eye”, and based on these predictions we have implemented the driver’s drowsiness detection algorithm, as shown in Figure 7. The principal objective of this algorithm is to distinguish between “normal blinking” and “real drowsiness”. In the algorithm, we introduce a new counter “ec_count”, whose initial value is zero. If the prediction result is “closed eye”, then the counter “ec_count” increases by one; otherwise, the counter is initialized. If the S-CNN predicts “th + 1” consecutive frames as “closed eye”, then the process is passed to Stage 4.

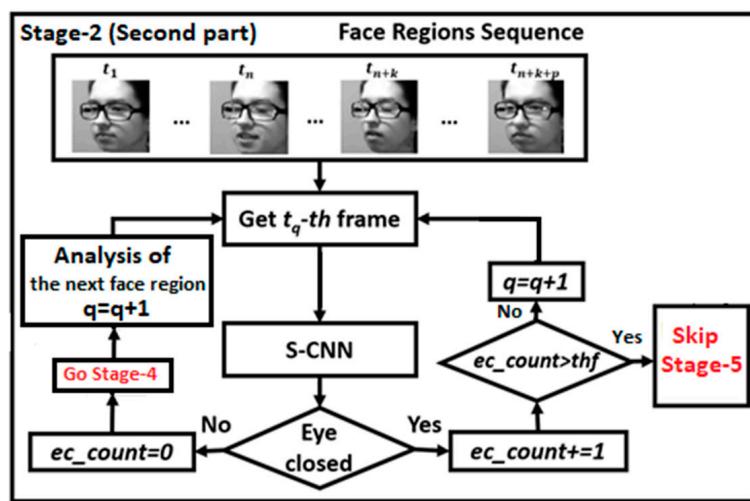


Figure 7. Proposed driver's drowsiness detection algorithm.

Here, “th” is the threshold value, which is determined considering the normal adult blinking duration. According to [25], the duration of the normal adult blinking is around 290 to 750 milliseconds. Taking account of the average speed of the proposed system, which is 21 fps, we decided to set the threshold value “th” equal to 36, which is equivalent to approximately 1.7 s. It is worth noting that the processes shown by Figures 6 and 7 are basically same; the only difference is the threshold value. Figure 8 shows an example when a driver performs normal blinking.

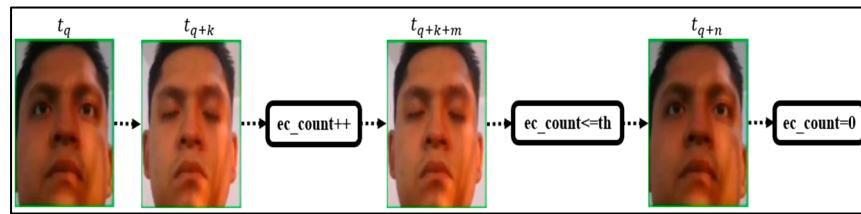


Figure 8. An example of driver's normal blinking. If the closed eye continues m frames and m is smaller than the threshold “th”, then the driver performs his/her normal blinking.

3.3. Implementation of a S-CNN (Stage 3)

This section provides a description of the construction of the database used for the training and evaluation of the proposed system, together with the implementation of the proposed system.

3.3.1. Database Construction

In this third stage, using a CNN-based classifier, nominated as S-CNN, we classify three patterns from the face region extracted by the previous stage. The three patterns are “open eye”, “closed eye”, and “distraction”. An appropriate construction of the database to train the S-CNN, which must have enough capability of generalization in a real driving situation, is very important. Then, we analyze the relevant patterns that indicate drowsiness and distraction in drivers. After analyzing several video sequences, three driver's state patterns can be distinguished: drivers have fallen asleep, they have clear signs of drowsiness, or they present a pattern of distractions. From this analysis, we define three features of the face region: a face without drowsy features, a face with drowsy features, and a face with distraction features. Figure 9 shows these three features, and Table 2 shows the features used in the driver's three states.

Table 2. Specific features considered in each of three states of the driver.

Face Without Drowsy Features	Face with Drowsy Features	Face with Distraction Features
<ol style="list-style-type: none"> 1. Position of eyes (Figure 9a-1). 2. Position of the eyebrows (Figure 9a-2). 3. Expression lines (Figure 9a-3). 4. Position of the face (Figure 9a-4, a-6). 5. Mouth position (Figure 9a-5). 	<ol style="list-style-type: none"> 1. Position of eyes (Figure 9b-1). 2. Position of the eyebrows (Figure 9b-2). 3. Position of the face (Figure 9b-3). 4. Mouth position (Figure 9b-4). 	<ol style="list-style-type: none"> 1. Position of the face (Figure 9c) 2. Position of Eyes (Figure 9c)

When the driver is drowsy, the position of the face is determined by driver's nose point (NOSE_TIP in Figure 4). If the camera is collocated in the front of the driver, then this position is supported by two points: RIGHT_EAR_TRAIGON and LEFT_EAR_TRAIGON, as shown by Figure 9a; otherwise, only the driver's nose point is used to determine the face position, as shown by Figure 9b. In the features that indicate the driver's distraction level, the following conditions are taken into account.

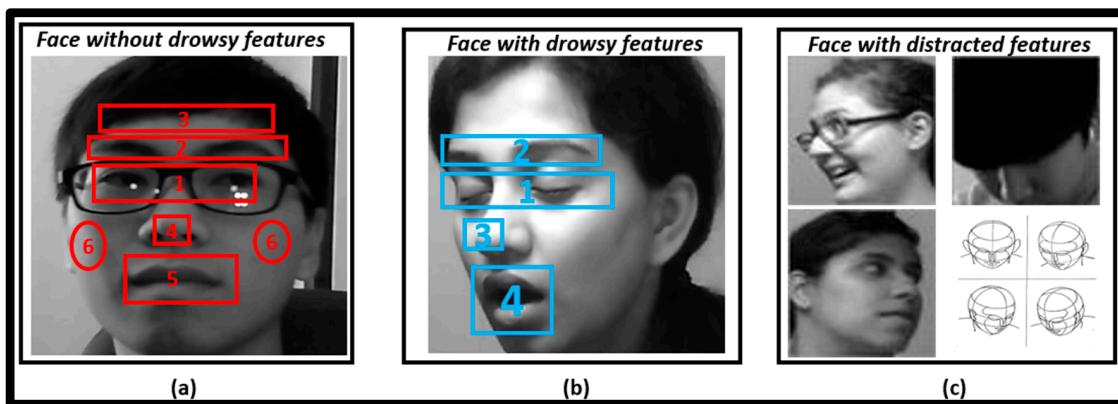


Figure 9. Three of the driver's state features for database construction. (a) Face with 7 points detected, (b) face image with 4 points detected and (c) face with distracted features.

Position of the face (Figure 9c): to determine that the face has patterns of distraction, the analyzed face region must contain at most three landmark points with respect to all existing points (six points) in a totally visible face (see Figure 4).

Position of the eyes (Figure 9c): the position of the eyes directly indicates a deviation of a driver's face from their front, which can be considered as distraction, although this feature must be analyzed further in the next stage.

Based on the main features of drowsiness and distraction, our database for training and testing is constructed from the NTHU-DDD dataset [26], labeled "open eye" for "faces without drowsy features", "closed eye" for "faces with drowsy features", and "distraction" for "faces with distraction features". In total, we construct the database with 4800 labeled face images, i.e., 1600 face images per class. All the face regions are normalized to be the same size 64×64 , with three color channels in order of blue, green, and red (BGR). Although the NTHU-DDD dataset provides gray-scale images, we used three channels for future improvement. Some examples of the constructed database are shown in Figure 10.



Figure 10. Some examples of face regions in database. (**Top**): “eyes open” class; (**middle**): “eyes close”; and (**bottom**): “distraction”.

Our database is divided randomly, keeping the same proportions by class, into the training set and the test set. The training set contains 75% of all images in the database, i.e., 3600 images, and the other 1200 images are included in the test set. The training set is further divided into a real training set and validation set. We used a five-fold cross-validation technique to attain the best hyperparameters. Table 3 shows the best three configurations from accuracy and the number of trainable parameters points of view. We selected a CNN in the first row of the table as the proposed S-CNN, considering the best accuracy and small number of trainable parameters.

Table 3. Best three candidate configurations for the proposed S-CNN.

Models	Architecture	Additional Settings	Model Size	Trainable Params	Accuracy (%)
Selected Model	Input Image (64,64,3) 3 × (2D Conv, BN, 2 × 2 MaxPooling) Flatten, 2 × FC (128), Dp (0.2), FC (3)	2D Conv	32		
			32		
			64		
		Max Pooling	6 × 6 × 64	3.98 MB	340,835
			31 × 31 × 32		
			14 × 14 × 32		
Model 2	Image input (64,64,3) 3 × (2D Conv, BN, 2 × 2 MaxPooling) Flatten, 2 × FC (128), FC (3)	2D Conv	32		
			32		
			128	5.42 MB	700,547
		Max Pooling	31 × 31 × 32		
			14 × 14 × 64		
			6 × 6 × 128		
Model 3	Image input (64,64,1) 3 × (2D Conv, BN, 2 × 2 MaxPooling) Flatten, 2 × FC (128), Dp (0.2), FC (3)	2D Conv	32		
			32		
			64	3.98 MB	340,259
		Max Pooling	31 × 31 × 32		
			14 × 14 × 32		
			6 × 6 × 64		

3.3.2. Architecture of the Proposed S-CNN

The selected architecture of the proposed S-CNN is given in Figure 11. As shown in the figure and mentioned in Section 3.2.1, the input image is $64 \times 64 \times 3$. To facilitate the training process and alleviate overfitting to the relatively small size of training set, we introduce the preprocessing with data augmentation techniques, which consists of rescaling of 1/255, shearing with 20%, zooming with 20%, horizontal shifting with 20%, vertical shifting with 20%, a rotation range with 20%, and horizontal flipping. In all convolutional layers, the ReLU activation functions are introduced.

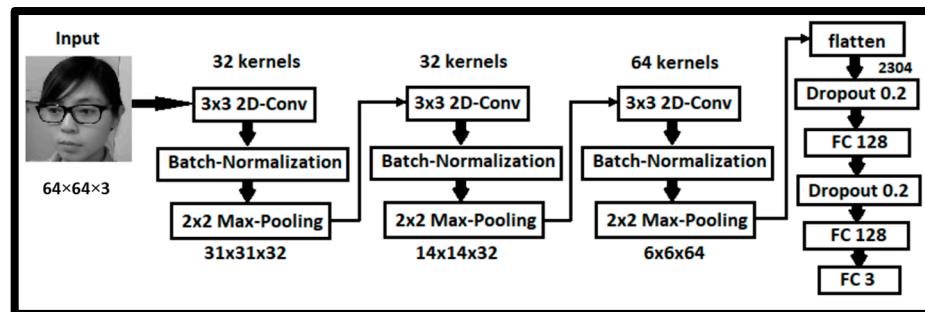


Figure 11. Proposed S-CNN architecture.

In the training process, we introduce two dropouts with a dropout rate of 20% in the fully connected layers to alleviate the overfitting. As an optimizer, we used the Adam optimizer with a hyper-parameter of $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of $\alpha = 0.001$. The batch size used was 64 samples. Figure 12 shows the training behavior of the proposed S-CNN, in which we can observe that the overfitting is minimal. We consider that the shallow CNN configuration and techniques such as data augmentation and dropout introduced in the training process alleviate the overfitting.

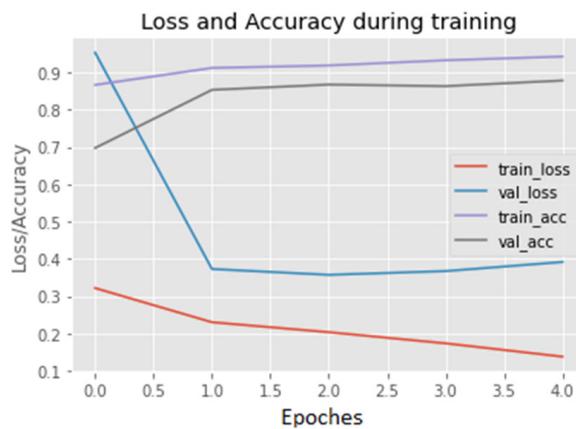


Figure 12. Loss and accuracy during training.

3.4. Isolated Prediction Error Correction (Stage 4)

In the implementation in the real situation, any machine learning classifier, including deep learning classifiers, presents few errors due to nonuniformity in environmental conditions such as the illumination variation and the optical sensor position. In this stage, some isolated errors caused by the classifier are corrected using the consecutive classification results. Figure 13 shows a clear example of this type of error if this stage is omitted in the system. The classifier determines the driver's drowsiness level by considering the classification results of their previous frames (Figure 13a); however, the next frame is erroneously classified as the driver's "open eye" (Figure 13b) due to some environmental variations. This error causes the initialization of the counter "ec_count", and until the next n consecutive frames are classified correctly as the driver's "closed eye" (see Figure 8), the

activation of the alarm is interrupted because the system considers this isolated error as the driver's normal blinking (Figure 13c) if this it is not corrected.

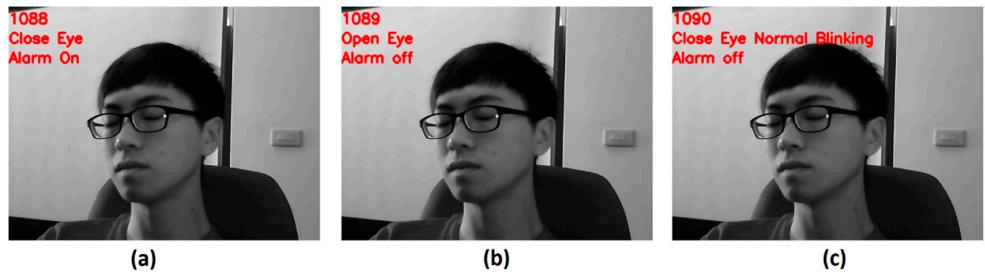


Figure 13. An example of isolated prediction error on three consecutive frames. (a) Classified as driver drowsiness; (b) isolated error caused by an environmental nonuniformity; and (c) classified as normal blinking due to isolated error in the previous frame (b).

Based on the above-mentioned problem, a new variable called prediction error correction (PEC), given by (1), is introduced. The PEC is used as a threshold value to correct isolated errors that occur during frame-by-frame real-time monitoring.

$$PEC = \frac{fps}{4} \quad (1)$$

where fps is the operation speed of the device. The PEC is used to discriminate between normal blinking and an isolated inconsistent error caused by S-CNN misclassification. According to [25], the duration of normal blinking is between 290 and 750 milliseconds, so that less than a quarter of a second (250 ms) indicated by [11] can be considered as an isolated error, while the same classification results during PEC consecutive frames are considered as a real state change. Figure 14 shows the block diagram of this process, in which if the current prediction result is different from the previous consecutive results, then the counter called “countN” increases by one; otherwise, this value is initialized. If the counter “countN” exceeds the PEC given by (1), then the new prediction result is considered as a real state change; otherwise, this prediction result is considered as isolated error of the classifier S-CNN, and then this error is corrected by assigning the previous prediction to the current prediction.

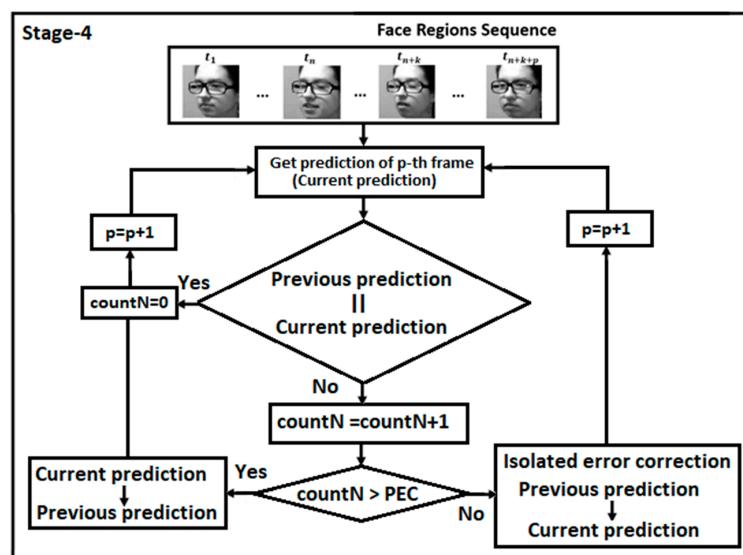


Figure 14. Process for correcting isolated prediction error, caused by the S-CNN.

3.5. Activation of Relevant Alarms

The alarms in SOMN_IA are both internal and external configurations. The external alarms are the light alert subsystem (LAS), the sound alert subsystem (SAS), and the tracking Subsystem (TS), as shown in Figure 15. As mentioned before, SOMN_IA generates four different types of alarms: a sound alarm, a visual alarm, an informative alarm based on the graphic user interface (GUI), and TS. The first three alarms are activated to directly alert the driver, and the fourth alarm based on TS transmits the SMS message to notify the third party, if first three alarms are not effective to modify the driver's state. In the following subsections, we describe the activation of the alarms depending on the driver's state, such as distraction or drowsiness.

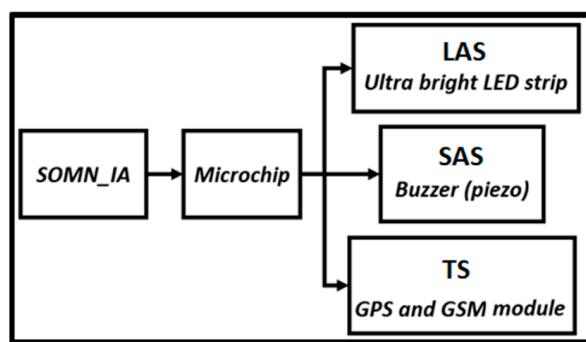


Figure 15. LAS, SAS, and TS connection and activation process.

3.5.1. Alarms for Driver's Distraction Level

When the previous stage of the SOMN_IA detects that the driver is distracted, two LED strips in the LAS, as well as four piezoelectric in the SAS, are activated. The intensity of both alarms increases until the driver is aware of his/her distraction and pays attention to his/her front or reaches the alarms' maximum intensities. If both alarms reach their maximum intensities without driver's any reaction, the TS is activated to notify the driver's dangerous situation to the third party. The procedure implemented to manage these three alarms is given in Figure 16, in which the counter “ec_countf” introduced in Sections 3.2.1–3.2.3 is used. To manage the LAS and SAS alarms, a new counter called the “MDAS_counter”, which determines the intensity of both alarms, and a threshold named “MDAS_th”, which is the maximum intensity of both alarms, are introduced. The intensity of both alarms indicated by “MDAS_counter” increases each 2 s while the driver continues to be distracted. If the driver corrects his/her dangerous state, all alarms are deactivated and “MDAS_counter” is initialized. The GUI-based informative alarm displays each video frame, together with the driver's distraction level in the touchscreen monitor (See Section 4—Hardware Implementation).

3.5.2. Alarms for Driver's Drowsiness Level

The management of three alarms for a driver's drowsiness level is very similar to the case for a driver's distraction level, whose procedure is given by Figure 17. Instead of two LED strips used for a driver's distraction level, here all four LED stripes are used to wake up the driver, and four piezoelectric in the SAS are used. We introduce “SMAS_counter” and “SMAS_th” to control intensity of LAS and SAS alarms.

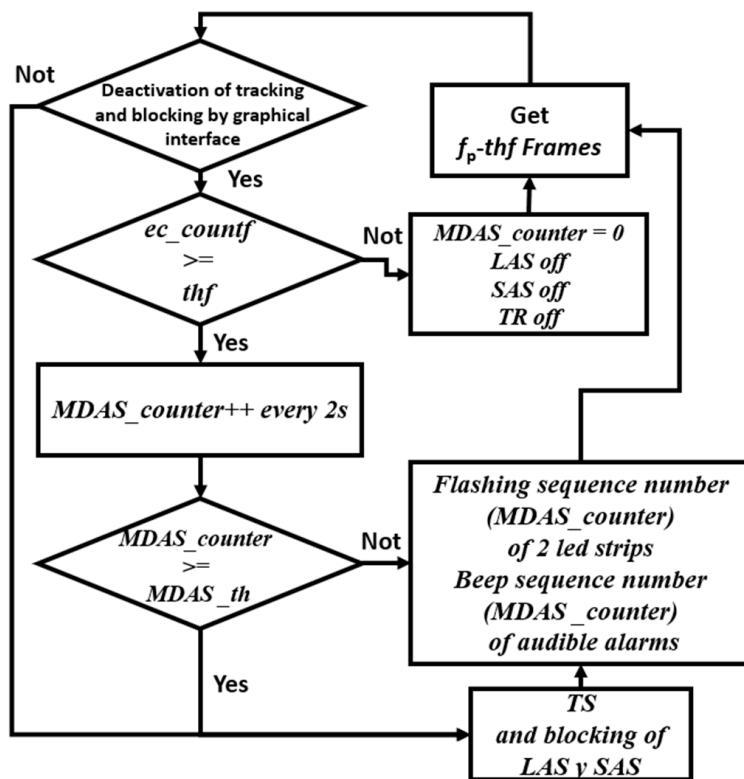


Figure 16. Activation and deactivation process of LAS, SAS, and TS in the driver's distraction detection.

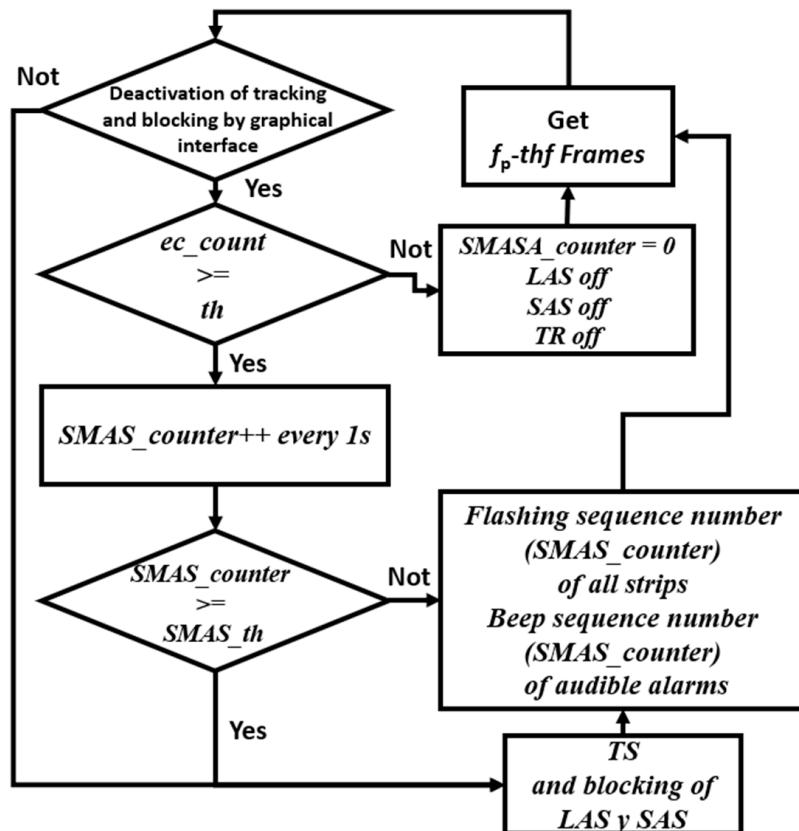


Figure 17. Activation and deactivation process of LAS, SAS, and TS for drowsiness detection.

In the activation and deactivation of three alarms for the driver's drowsiness level, we use the counter “ec_count” and threshold “th” used in the Section 3.2.3. If “ec_count” exceeds “th”, the driver's drowsiness level is considered and activates LAS and SAS alarms, the intensity of both alarms “SMAS_counter” increases each 1 s while the driver continues to sleep until these intensities reach their maximum intensities, determined as “SMAS_th”. If the driver is awoken by these two alarms, “SMAS_counter” is initialized and all alarms are deactivated. If driver does not awake when LAS and SAS alarms have their maximum intensities, the TS alarm is activated to inform the driver's dangerous situation to the third party. Again, the GUI-based informative alarm displays each video frame together with the driver's drowsiness level in the touchscreen monitor (See Section 4—Hardware Implementation).

3.5.3. Information Transmitted by TS Alarm

The TS alarm is activated when the intensity of the LAS and the SAS alarms exceeds the thresholds determined for the driver's distraction level and driver's drowsiness level, respectively, which are “MDAS_th” and “SMAS_th”, respectively. The TS alarm transmits several information about the driver and the vehicle, such as the driver's name, the vehicle type, the vehicle plate, and the driver's state (distraction or drowsiness) and coordinates the vehicle to the telephone number preregistered by the driver. The coordinate of the vehicle is updated in real time by the GPS.

4. Hardware Implementation

The SOMN_IA is a wireless portable device that has been specifically developed to detect drowsiness and distraction in drivers. Based on all the algorithms previously described in Section 3, in this section we describe the main components, their implementations to construct the SOMN_IA, and its implementation inside of the vehicle.

4.1. Components That Make Up the Front Part of the SOMN_IA

The components in the front part of the SOMN_IA are given in Figure 18, in which each component used is described. It consists of four lights alarms of the LAS, which are denoted as 1–4. A case, denote by 5, which is the container of the system created by 3D printer. It is made of recycled PLA (polylactic acid) plastic. The proposed system has also a touchscreen of 7 inches, denoted by (6) in Figure 18. Finally, SOMN_IA has a built-in DC-DC converter for input 12 v–24 v DC and output 5 v at 4 A minimum. The type of connector used is a 5.5 × 2.1 mm male Jack.

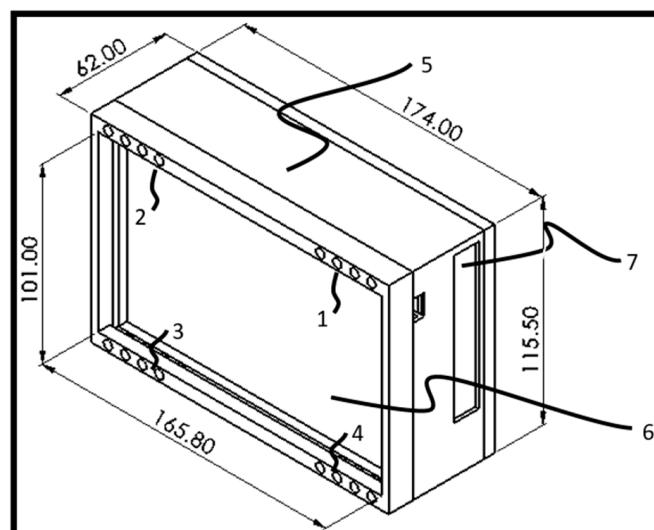


Figure 18. Design of front part of the SOMN_IA.

4.2. Components That Make Up the Back Part of the SOMN_IA

The components used for the back part of the SOMN_IA are given by Figure 19. It contains an antenna, denoted by 1, a heatsink, 2, and an IRIS camera connector by ISP, denoted by 3; as well as GSM and control modules denoted by 4 and 5 respectively. It also has a motherboard, 6, as well as a GPS module and a GPS antenna denoted by 7 and 8, respectively. Finally, it has four screws and the system cover, denoted by 9 and 10 respectively.

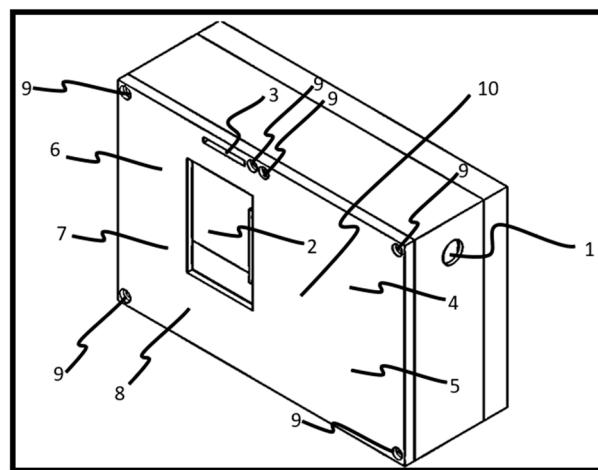


Figure 19. Design of back part of the SOMN_IA, with its main 10 components.

4.3. Internal Composition of SOMN_IA

The internal composition of the SOMN_IA is given in Figure 20. Each module composed by SOMN_IA is described in detail.

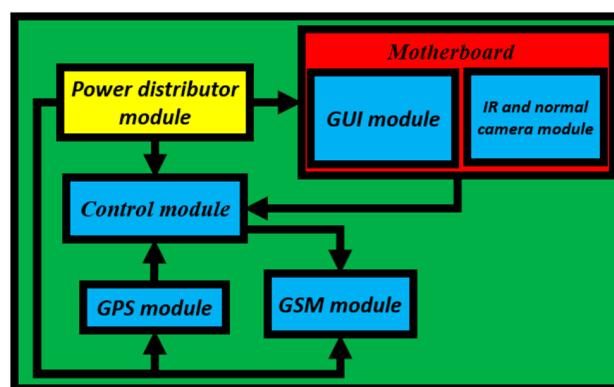


Figure 20. Block diagram of the intercommunication of the internal modules of SOMN_IA.

Power distributor module: This module provides the power to the different modules of the SOMN_IA; the main characteristic of this module is that it generates the 12 v dc that comes out of the car battery and reduces it to 5 v–5 A dc continuously, which ensures the proper functioning of the system since for SOMN_IA to work without limitations, a constant supply of 20 w must be maintained.

Motherboard: This is the most important module because in this section the general software of SOMN_IA is implemented; here, the entire monitoring process of the driver is carried out, along with the application of the intelligence algorithms and the commands of the alarms. The information is sent to the graphical interface, where it receives the data from the optical sensor.

Control module: This module is the second most important; it monitors the results given by the main card and thus is able to activate or deactivate the sound and light

alarms, in addition to the fact that it is in constant communication between the GSM and GPS modules. In this module, the predesigned messages have been encoded, as well as the overseeing of the decoding the GPS address, so that, later, the message is sent to the pre-loaded number with the help of the GSM module.

GPS module: It is a small module in charge of receiving the data in real time of the location of the system, with the help of its antenna; later, it communicates with the control module to be able to provide the necessary information, which will be decoded and sent with the support of the GSM module to the preloaded number chosen by the user.

GSM module: This internal module in the system is responsible for connecting to the cellular network of the designated telephone company, and with this it is able to send previously encrypted messages.

4.4. Connection Diagram in the Vehicle

As we can observe in Figure 21, the installation of SOMN_IA is extremely simple since the only thing required for its proper functioning is the direct connection of the eliminator to the vehicle battery. Although the driver can select other types of connection, the direct connection is recommended for its better function. Additionally, the connection of the optical sensor (camera) is as direct as possible. The position of the optical sensor must be adequate to capture the whole driver's face because if the SOMN_IA does not detect at least four reference points of the six established points of the driver's face, due to wrong position of optical sensor, the system determines the driver's distraction level erroneously (See Section 3.2.1).

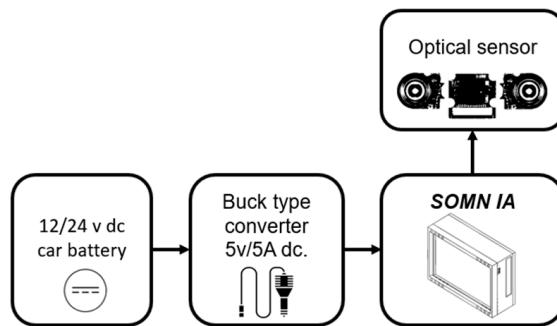


Figure 21. Connection diagram in the car.

5. Experimental Results

The principal contribution of the paper is the hardware construction of a portable system nominated as SOMN_IA, to detect the driver's distraction and drowsiness levels using a CNN, and if the system detects these dangerous situations in the driver, several alarms are activated to warn the driver to avoid a lamentable accident. The alarms are activated only if the SOMN_IA considers the real drowsiness and/or real distraction levels. Almost all previous systems that detect the driver's drowsiness or distraction levels are only computational systems, in which any hardware limitations, such as computational power and memory space, are not considered. As mentioned before, the SOMN_IA is a portable system that can be installed in any type of vehicles. To achieve the portability and universality of the SOMN_IA, the first four stages shown in Figure 1 and described in Section 3, which are equivalent to the software part of the SOMN_IA, must be adequate.

In this section firstly, we provide the performance of the software part of the SOMN_IA, which is equivalent to the first four stages of the SOMN_IA described in Section 3. Next, the global performance of the SOMN_IA is shown, providing the installation state of the SOMN_IA in a vehicle and several links of video scripts to show its real-time performance.

5.1. Software Performance of the SOMN_IA

In this subsection, we provide the classification performance of the S-CNN whose configuration is given by Figure 11. Table 4 and Figure 22 show a performance comparison among off-the-shelf CNNs, such as MobileNetV2, VGG16, ResNet, GoogleNet, and Xception [27–31], and the proposed S-CNN, from several points of view. The operation speed is the most important issue for a real-time operation, in which two face-detection algorithms, Mediapipe face detection algorithm [16] and Haar-based face detection algorithm [15], are considered for comparison purposes. The off-the-shelf CNNs are pretrained for the ImageNet challenge and are fine-tuned by retraining some latest layers using the same training set employed for the proposed S-CNN. The accuracies of the off-the-shelf CNNs are of optimum value after the fine-tuning operation. From Table 4 and Figure 22, the classification accuracy of the proposed S-CNN is slightly lower than the other CNNs; however, the model size and total trainable parameters of the S-CNN are much smaller than those of other CNNs, which makes possible the portable and real-time implementation of the S-CNN. The proposed S-CNN provides the operation speed with 21 fps, when the Mediapipe algorithm is used for face detection, which allows for adequate operation without any delay. Additionally, the vehicle's power supply is enough for the SOMN_IA with the S-CNN.

Table 4. Comparison of the proposed S-CNN with fine-tuned off-the-shelf CNNs.

Model	Accuracy	Input Image Size	Model Size (MB)	Total Params	Operation Speed (fps)	
					Mediapipe	Haar
Proposed	0.9577	64 × 64	3.98	341,091	21	21
MobileNetV2 [27]	0.9675	64 × 64	10.30	2,590,083	17	16
VGG16 [28]	0.9777	64 × 64	56.80	14,850,179	15	16
ResNet50 V2 [29]	0.9760	64 × 64	92.50	24,116,419	14	13
InceptionV3 [30]	0.9600	75 × 75	84.50	21,938,275	13	12
Xception [31]	0.9713	64 × 64	81.90	21,390,187	16	12

Table 5 shows the performance of the S-CNN under different conditions in both normal light and infrared light (IR). Table 5 shows the accuracy provided by the proposed system when the driver uses glasses and does not use glasses, under both normal visible light and infrared light (IR).

Table 5. Performance of the proposed S-CNN under four different conditions.

Condition	Detection Accuracy (%)
Without glasses under normal illumination	96.67
With glasses under normal illumination	95.83
Without glasses under IR illumination	95.91
With glasses under IR illumination	94.67
Global performance	95.77

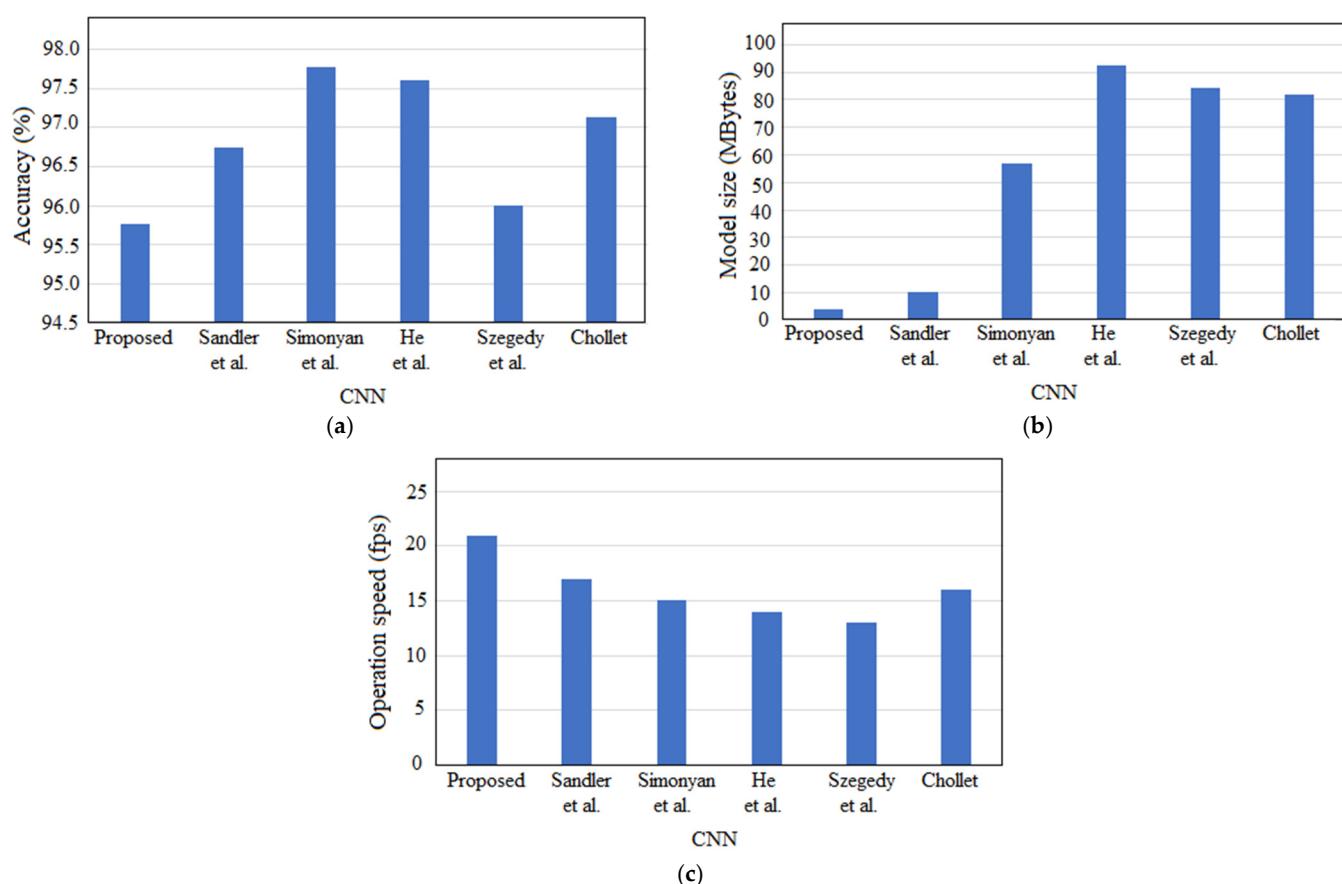


Figure 22. Performance comparison of proposed S-CNN with other off-the-shelf CNN structures proposed by Sandler et al. [27], Simonyan et. Al [28], He et al. [29], Szegedy etr al. [30] and Chollet [31] (a) Accuracy comparison, (b) model size, and (c) processing speed in frames per second.

From Table 5, it follows that the better performance is obtained when the driver does not use glasses and the video frames are captured under normal visible light, while the worst performance when driver uses glasses under IR illumination is only 2% less. It is totally acceptable because the proposed system SOMN_IA uses several consecutive video frames to correct isolated classification errors provided by the S-CNN, as described in Section 3.4.

The confusion matrices of the S-CNN under four different conditions are provided by Figure 23a–d; they are (a) without glasses under normal illumination, (b) with glasses under normal illumination, (c) without glasses under IR illumination, and (d) with glasses under IR illumination.

5.2. Global Performance of the SOMN_IA

In this section, we show the global performance of the proposed system SOMN_IA using a link of the video generated by us. Firstly, we show the outward appearance of the SOMN_IA, together with its installation inside of the vehicle in Figure 24. Figure 25 shows that the SOMN_IA is working during driving, and Figure 26 shows the functionality of the alarm LAS (Light Alert Subsystem) of the SOMN_IA, when driver is distracted by his smartphone (Figure 26a) and the driver is asleep during driving (Figure 26b), under an environment with visible light. The functionality of the SOMN_IA in the different environment under visible light and IR light is shown in Figure 27. It is worth noting that for both the driver's drowsiness level and driver's distraction level, the alarm SAS (Sound Alert Subsystem) also works. Finally, Figure 28 shows an example of the information transmitted by the tracking subsystem (TS) after the alarms LAS and SAS reach their maximum intensities.

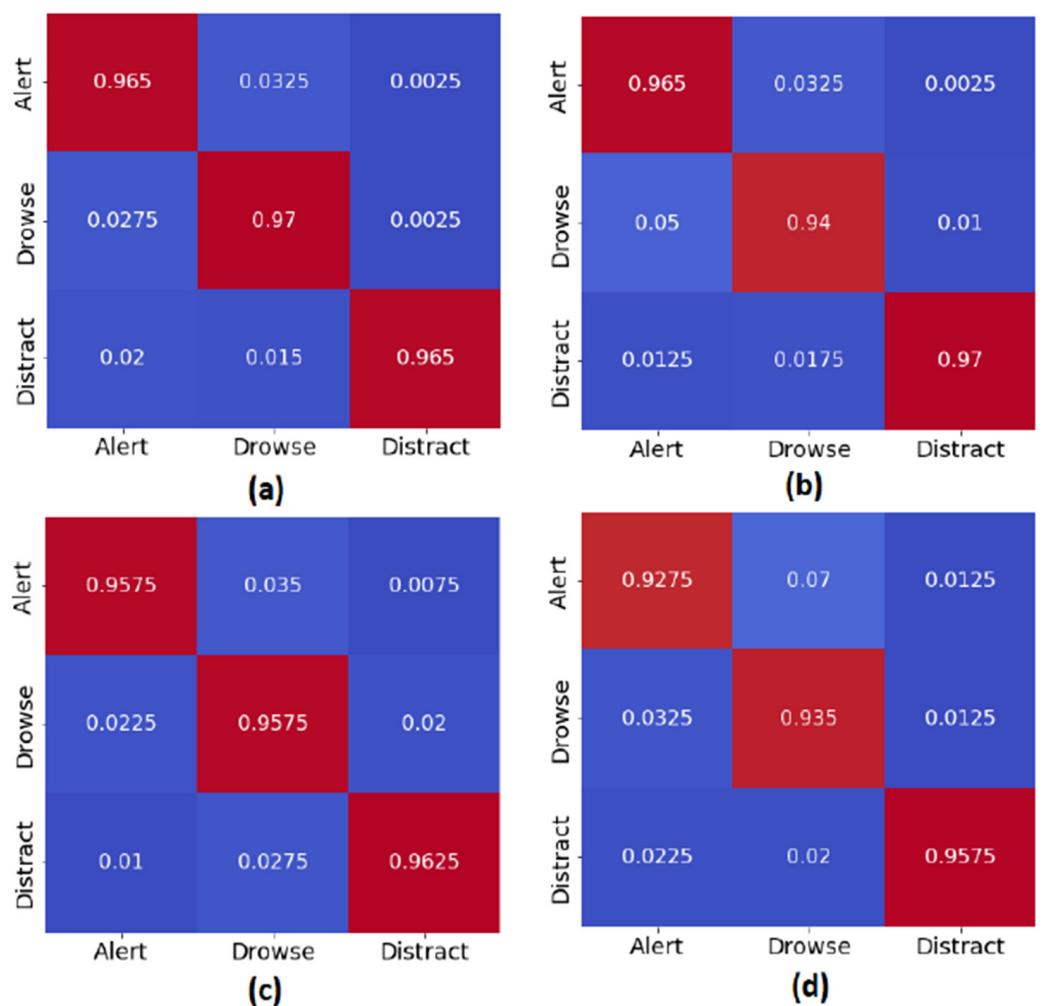


Figure 23. Confusion matrices under four different conditions. (a) Without glasses under normal illumination, (b) with glasses under normal illumination, (c) without glasses under IR illumination, and (d) with glasses under IR illumination.



(a)



(b)

Figure 24. SOMN_IA. (a) The outward appearance of the SOMN_IA (black version); (b) the installation of the SOMN_IA (green version) inside of a vehicle.



Figure 25. The SOMN_IA is working during driving. The graphical user interface (GUI) shows driver's face and his state.

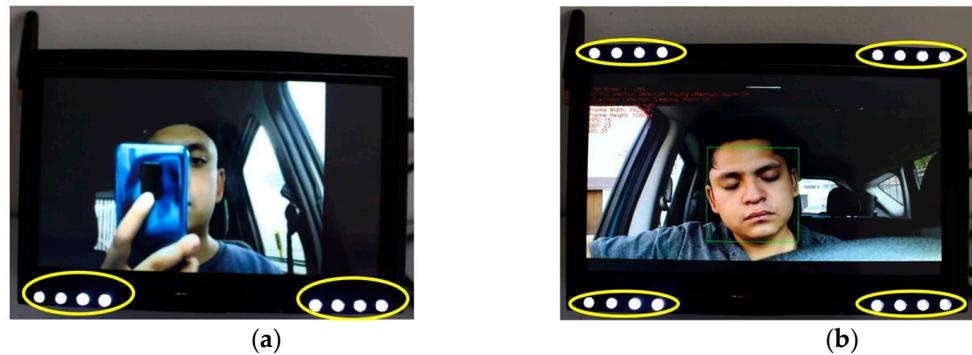


Figure 26. The functionality of the light alert subsystem (LAS), (a) activation of the LAS when the driver is distracted, and (b) activation of the LAS when the driver is asleep.



Figure 27. The functionality of SOMN_IA, shown in its GIU. The upper rows show three different driver states under the visible light. The lower rows show three different driver states in the night using the IR camera. (a) The driver is drowsiness, (b) the driver is alert, (c) the driver is distracted.

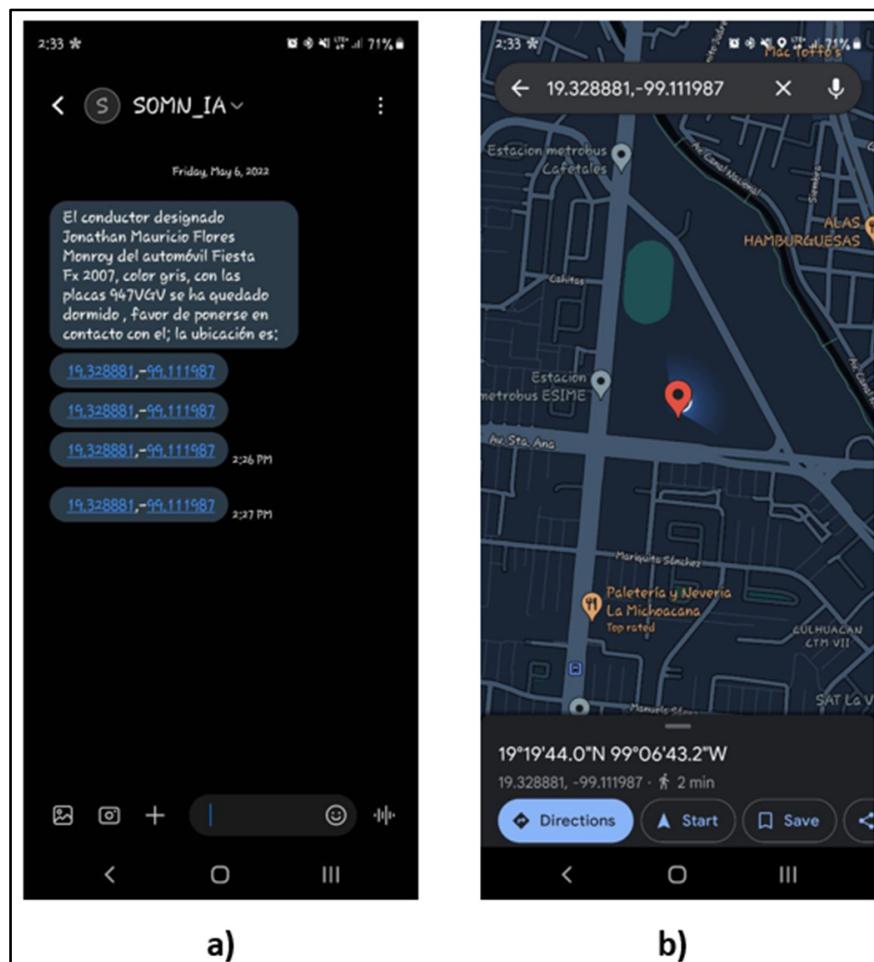


Figure 28. The information transmitted by the TS alarm, in which GSP information is updated in real-time, as shown in (a) and in (b) the map, indicates the localization of vehicle.

The global functionality of the SOMN_IA (supplementary material) can be observed in the following link to a video created by us. <https://youtu.be/035Qq5egiS8> (produced on May 2022)

5.3. Performance Comparison

As mentioned before, almost all systems for a driver's drowsiness and distraction system are not implemented in the specific portable hardware, so we cannot perform a fair comparison with the proposed SONM_IA, whose principal contribution is a construction of the portable and universal device that can be installed in any type of vehicle. However, to analyze the performance and functionality of the SONM_IA, we provide a comparison in Table 6. In this table, we compare several issues, which are detection accuracy, operation speed measured by frame per second (FPS), whether the system is implemented in hardware or not, the database used for training and evaluation, and whether the system can be adapted to the different lighting conditions in real time or not. Specially, the operation speed is important for real-time operation and real-time alert to the driver if he/she is driving in a dangerous state.

Some systems provide slightly better detection accuracy [13,17,20,23]; however, these systems used high computational power with a large memory space to perform off-the-shelf CNNs, such as AlexNet and EfficientNet. As shown in Table 4, these off-the shelf CNNs require a large memory space and high computational power, which make them difficult to implement in a compact portable system with limited resources. The proposed device provides a higher speed operation, 21 FPS, which can operate in real time. Additionally, the

SOMN_IA equips two types of sensors; one of them activates automatically depending on the lighting condition in the environment. It is worth noting that the SOMN_IA provides several alarms to alert the driver, while in other systems in the table, the alarm system is not mentioned.

Figure 29 shows a comparison of the performance of the proposed drowsiness detection system with other previously proposed systems in terms of detection accuracy and processing rate, measured in terms of frames per second, when the systems are implemented in different hardware devices, as shown in Table 6. The evaluation results show that even the accuracy provided by the proposed system is slightly lower than other previously proposed schemes; its processing rate is much higher, enabling faster detection because its processing rate, 21 fps, is closed to the frame rate provided by many digital cameras (about 25 fps). It is worth mentioning that the proposed system also detects the driver's distraction level; however, because almost all reported systems only detect the driver's drowsiness level, only drowsiness detection is reported in Figure 29, although it is provided in Table 6.

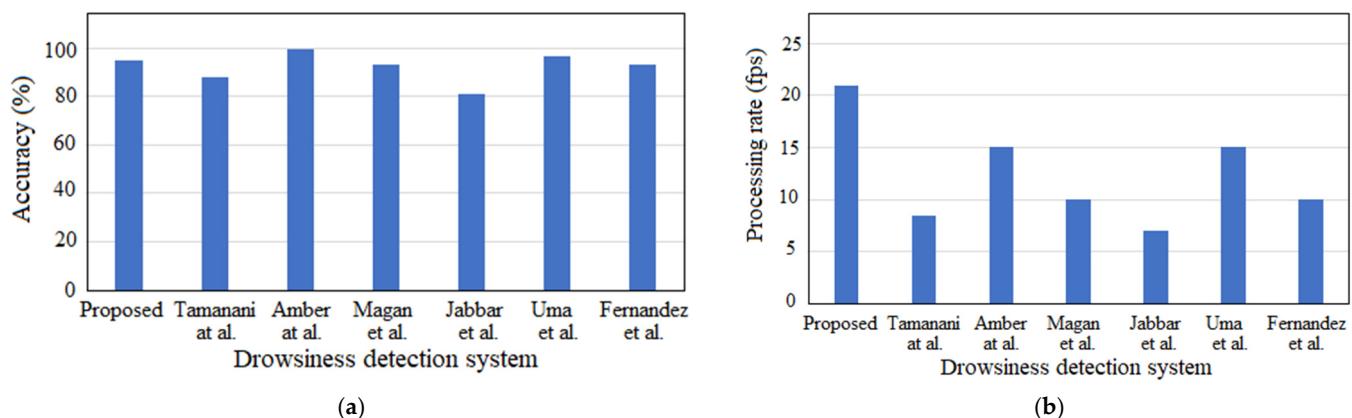


Figure 29. Performance of proposed drowsiness detection system compared with other schemes previously by Tamanin et al. [14], Amber [17], Magan et al. [18], Jabbar et al. [19], Uma et al. [20], Fernandez et al. [21] in terms of (a) its detection accuracy and (b) the processing rate measured in terms of the number of frames per seconds when it operates in real time.

Table 6. Performance comparison for the detection of several driver's drowsiness and/or distraction levels. × means that the function is not available, ✓ the function is available, - not reported.

System/ Research	Accuracy				Monitoring in Any Lighting Condition		
	Drowsiness	Distraction	FPS	Hardware Implementation	Database	Real Time	Video
Proposed	0.9506	0.9638	21	SOMN_IA (Portable specific device)	NTHU-DDD	✓	✓
[13]	0.97	×	-	No	Self-built	×	✓
[14]	0.88	×	8.4	No	UTA-RLDD + Self-built	×	✓
[17]	0.9965		15	No	NTHU-DDD	×	✓
[18]	0.93	×	10	No	UTA-RLDD	×	×
[19]	0.81	×	7	Mobile device (Android Smartphone)	NTHU-DDD	×	✓

Table 6. Cont.

System/ Research	Accuracy					Monitoring in Any Lighting Condition	
	Drowsiness	Distraction	FPS	Hardware Implementation	Database	Real Time	Video
[20]	0.965	×	15	Raspberry Pi + cloud computing	Self-built	✓	✓
[21]	0.9311	×	10	Mobile device (Smartphone)	CEW	✗	✓
[22]	0.95	×	-	JetsonNano + GPU server	FDDDB WFLW	✗	✓
[23]	0.9815	×	-	No	ZJU Eyeblink dataset	✗	✓

5.4. Evaluation of Proposed System in Real-World Conditions

To evaluate the proposed system in a more realistic situation, we used the real-life drowsiness dataset (UTA-RLDD) [32], which has 30 h of video and 60 subjects. A principal characteristic of this database is the recoding of real driving, and drivers themselves reported their drowsiness level, which is the ground truth of this database. The videos are captured at 30 fps by different optical sensors with different formats and resolutions. The optical sensors used can capture only the visible spectrum, meaning that there are no IR videos. The average duration of each video is approximately 10 min. Table 7 shows the drowsiness detection accuracy of the proposed system using the database UTA-RLDD, together with the performance of [14,18], which used the same database. From the table, we can observe that the proposed system provides higher accuracy compared with other two systems ([14,18]). It is worth noting that the UTA-RLDD is used only as a testing process in the proposed system, which is trained by NTHU-DDD [26].

Table 7. Drowsiness detection accuracy of the proposed system using dataset UTA-RLDD [32].

System	Accuracy	FPS
Proposed	0.96	21
[14]	0.88	8.4
[18]	0.93	10

6. Conclusions

In this paper, we proposed a portable and universal system to automatically detect a driver's distraction and drowsiness levels and alert to the driver of his/her dangerous state using equipment alarms. This system is denominated as SOMN_IA, which can be installed in any vehicle without additional hardware. The SOMN_IA is composed of a software part and a hardware part, and in the software part, there are five stages, which are (1) the frame separation and detection of the region to be analyzed, in which each video frame is extracted (the face region of each frame is detected using the Mediapipe face detection algorithm); (2) the implementation of a shallow convolutional neural network (S-CNN), in which firstly we designed several configurations of CNN that satisfy some requirements regarding the hardware limitations for the portability and universality of the SOMN_IA; (3) the analysis of consecutive results, in which the drowsiness is differentiated from the normal blinking and the distraction is differentiated from the normal actions of the driver to pay attention his/her right and left; (4) the prediction error correction due to a real driving situation, in which isolated errors of the S-CNN classifier due to a nonuniform environment are corrected to avoid alarm interruption; and (5) the activation of relevant alarms, in which three types of alarms are managed to alert the driver's dangerous state,

and if the driver does not correct his/her state after the predetermined time, information such as a driver's name, vehicle type, GPS, and so on are transmitted to the third party.

The hardware component and the implementation of the SOMN_IA were described in detail, which can be installed in any type of vehicles without any additional component. The video scripts that show its correct functionality under different environments are provided, which can be observed by accessing the links to a video script provided in Section 4.3. From the video script, we can observe the correct functionality of SOMN_IA under different environments.

The principal contribution of this work is the design and construction of a portable and universal system for a driver's distraction and drowsiness detection, operating in real time. To attain all the requirements, all algorithms used in the system, such as face detection, the CNN-based classifier, and adjustment using consecutive classification results, are optimized. As far as we know, there is not any similar system with which to carry out a fair comparison with our system. Almost all proposals for the detection of a driver's drowsiness and/or distraction are algorithmic proposal without any consideration of their implementation, and the performances reported is obtained in a laboratory environment, where hardware limitation does not exist.

A wider evaluation of the SOMN_IA using the different types of vehicles and a different environment is considered as future work. In this evaluation, a driving simulator with some sensors to capture the driver's physiological signal such as EEG, which can measure the driver's drowsiness or distraction level as a ground truth, is required. Additionally, we consider developing some methods to measure the driver's drowsiness grade using the number of normal blinks or number of yawns during a predetermined period. In this case, the sequential information must be used; then, the use of the recurrent neural network, such as LSTM, can be considered.

Supplementary Materials: The following supporting information, the functionality of the SOMN_IA, can be downloaded at <https://www.mdpi.com/article/10.3390/electronics11162558/s1>, Video S1: SOMN_IA.mp4.

Author Contributions: All authors participated in the research whose contributions are: conceptualization, J.F.-M. and M.N.-M.; methodology and software, J.F.-M. and H.P.-M.; validation, E.E.-H., H.P.-M. and M.N.-M.; formal analysis, G.S.-P.; investigation, M.N.-M.; data analysis, G.S.-P. and M.N.-M.; hardware development, J.F.-M. and E.E.-H., writing and original draft preparation, J.F.-M. and M.N.-M.; and review, H.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: The persons involved in the real time evaluation as well as in the video <https://youtu.be/035Qq5egiS8> reported in this research are the authors and then it is not required any informed consent statements. Other evaluation were performed using public data bases.

Data Availability Statement: The public database used in this paper is NTHU CVlab Driver Drowsiness Detection Dataset, which can be found in <http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD> (accessed on 1 January 2022).

Acknowledgments: The authors would like to thanks to the National Council of Science and Technology and to the National Polytechnic Institute of Mexico for the support provided during the realization of this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mexico Ranked Seventh Worst Place in the Traffic Accident in the World (Spanish), Instituto Nacional de Salud Pública. Available online: <https://www.insp.mx/avisos/4761-seguridad-vial-accidentes-transito.html> (accessed on 22 May 2022).
2. Causes of Traffic Accidents in Mexico (Spanish), by the National Institute of Statistics, Geographics and Informatics in Mexico (INEGI). Available online: https://www.inegi.org.mx/app/tabulados/interactivos/?px=ATUS_2&bd=ATUS&idrt=168&opc=t (accessed on 22 May 2022).

3. Accidents and Their Causes (Spanish), Mexican National Committee for Security (CNS). Available online: http://www.cns.gob.mx/portalWebApp/appmanager/portal/desk?_nfpb=true&_pageLabel=portals_portal_page_m2p1p2&content_id=830068&folderNode=830052&folderNode1=810277 (accessed on 22 May 2022).
4. 2019 Traffic Safety Culture Index, Foundation for Traffic Safety. 2020. Available online: <https://aaafoundation.org/2019-traffic-safety-culture-index/> (accessed on 10 April 2022).
5. Chacon-Murguia, M.; Prieto-Resendiz, C. Detecting driver drowsiness: A survey of system designs and technology. *IEEE Consum. Electron. Mag.* **2015**, *4*, 107–119. [CrossRef]
6. Wang, J.; Zhu, S.; Gong, Y. Driving safety monitoring using semisupervised learning on time series data. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 728–737. [CrossRef]
7. Wu, B.-F.; Chen, Y.-H.; Yeh, C.-H.; Li, Y.-F. Reasoning-based framework for driving safety monitoring using driving event recognition. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1231–1241. [CrossRef]
8. Kokonozzi, A.K.; Michail, E.M.; Chouvarda, I.C.; Maglaveras, N.M. A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects. In Proceedings of the Computing in Cardiology, Bolonya, Italia, 14 September 2008.
9. Vidente, V.; Laguna, P.; Bartra, A.; Bailón, R. Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* **2016**, *54*, 827–937.
10. Zhang, C.; Wang, H.; Fu, R. Automated detection of driver fatigue based on entropy and complexity measures. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 168–177. [CrossRef]
11. Flores-Monroy, J.; Nakano-Miyatake, M.; Perez-Meana, H.; Sanchez-Perez, G. Visual-based real time driver drowsiness detection system using CNN. In Proceedings of the International Conference on Electrical Engineering, Computing Science and Automatic Control, IEEE, Mexico City, Mexico, 10 November 2021.
12. Flores-Monroy, J.; Nakano-Miyatake, M.; Perez-Meana, H.; Escamilla-Hernandez, E.; Sanchez-Perez, G. A CNN-based driver's drowsiness and distraction detection system. In Proceedings of the 14th Mexican Conference on Pattern Recognition, Chihuahua, Mexico, 22–25 June 2022.
13. Phan, A.-C.; Nguyen, N.-H.-Q.; Trieu, T.-N.; Phan, T.-C. An Efficient Approach for Detecting Driver Drowsiness Based on Deep Learning. *Appl. Sci.* **2021**, *11*, 8441. [CrossRef]
14. Tamanani, R.; Muresan, R.; Al-Dweik, A. Estimation of driver vigilance status using real-time facial expression and deep learning. *IEEE Sens. Lett.* **2021**, *5*, 6000904. [CrossRef]
15. Viola, P.; Jones, M. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [CrossRef]
16. Bazarevsky, V.; Kartynnik, Y.; Vakunov, A.; Raveendran, K.; Grundmann, M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. In Proceedings of the Computer Vision & Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
17. Anber, S.; Alsaggaf, W.; Shalash, W. A Hybrid Driver Fatigue and Distraction Detection Model Using AlexNet Based on Facial Features. *Electronics* **2022**, *11*, 285. [CrossRef]
18. Magán, E.; Paz-Sesmero, M.; Alonso-Weber, J.M.; Sanchis, A. Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images. *Appl. Sci.* **2022**, *12*, 1145. [CrossRef]
19. Jabbar, R.; Al-Khalifa, K.; Kharbeche, M.; Alhajyaseen, W.; Jafari, M.; Jiang, S. Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques. *Procedia Comput. Sci.* **2018**, *130*, 400–407. [CrossRef]
20. Uma, S.; Eswari, R. Accident prevention and safety assistance using IOT and machine learning. *J. Reliab. Intell. Environ.* **2021**, *8*, 79–103. [CrossRef]
21. Fernández-Villán, A.; Usamentiaga-Fernández, R.; Casado-Tejedor, R. Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas. *Rev. Iberoam. Autom. Inform. Ind.* **2017**, *14*, 307–328. [CrossRef]
22. Pattarapongsin, P.; Neupane, B.; Vorawan, J.; Sutthikulsombat, H.; Horanont, T. Real-time drowsiness and distraction detection using computer vision and deep learning. In Proceedings of the ACM International Conference Proceeding Series, 1, Toronto, ON, Canada, 22–24 July 2020.
23. Hashemi, M.; Mirrashid, A.; Shirazi, A.B. Driver safety development: Real-time driver drowsiness detection system based on convolutional neural network. *SN Comput. Sci.* **2020**, *8*, 1–10. [CrossRef]
24. Pan, G.; Sun, L.; Wu, Z.; Lao, S. Eyeblink-based anti-spoofing in face recognition from a generic webcam. In Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
25. Kwon, K.A.; Shipley, R.; Edirisasinghe, M.; Ezra, D.G.; Rose, G.; Best, S.M.; Cameron, R.E. High-speed camera characterization of voluntary eye blinking kinematics. *J. R. Soc. Interface* **2013**, *10*, 1–6. [CrossRef] [PubMed]
26. Weng, C.H.; Lai, Y.H.; Lai, S.H. Driver drowsiness detection via a hierarchical temporal deep belief network. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 117–133.
27. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted residual-als and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–29 June 2016; pp. 770–778.

30. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhouckel, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
31. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
32. Ghoddoosian, R.; Galib, M.; Athitsos, V. A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 178–187.