Driver inattention detection system using multi-task cascaded convolutional networks

Abdelfettah Soultana, Faouzia Benabbou, Nawal Sael, Soukaina Bouhsissin

Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'SIK, University Hassan II of Casablanca,
Casablanca, Morocco

Article Info

ABSTRACT

Driver inattention has emerged as a critical concern impacting road safety, resulting in an alarming surge in accidents and fatalities. This research introduces a novel system for detecting inattention, structured across six levels: perception, facial feature extraction, tracking driver face, and driver secondary task using pre-trained deep learning models, inattention detection, risk estimation, and alert. The system is based on image processing captured from two strategically positioned cameras that simultaneously capture the driver's activities while driving and their facial expressions. The second contribution concerns the driver facial features extraction using multi-task cascaded convolutional networks (MTCNN), and it is comparison with the histogram of gradient (HOG)-based frontal face detector, and haar feature-based cascade classifier. The algorithms were compared based on their runtime efficiency, robustness in handling varying lighting conditions, and various head movements. The MTCNN achieves high performance, reaching accuracy levels ranging from 96.4% to 99.5% on two datasets including realistic driving scenarios: the DrivFace dataset and, the driver drowsiness dataset. The comparative analysis sheds light on the strengths and weaknesses of each algorithm, providing valuable insights for selecting the most suitable face detection algorithm to use in our system.

Corresponding Author:

Abdelfettah Soultana
Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'SIK
University Hassan II of Casablanca
Casablanca, Morocco
Email: abdelfettah.soultana-etu@etu.univh2c.ma

1. INTRODUCTION

Road safety remains one of the most pressing concerns in our modern society. Road accidents caused by driver inattention continue to be costly regarding human lives and material resources. According to the World Health Organization (WHO) global status report, road traffic accidents cause 1.35 million deaths yearly. This is nearly 3,700 people dying on the world's roads daily [1]. In this context [2], driver inattention is defined as 'insufficient, or no attention, to activities critical for safe driving' and can be brought about through a number of different mechanisms such as 'driver-restricted attention' (e.g. due to biological states, such as drowsiness or fatigue) or 'driver misprioritized attention' (e.g. due to focusing attention on one aspect of driving to the exclusion of another which is more critical for safe driving). This can manifest in various forms, such as using a mobile phone, adjusting the radio, daydreaming, or engaging in other distracting activities. The consequences of such inattention are far-reaching and can have devastating impacts on road safety. It significantly impairs a driver's ability to react promptly to sudden changes in traffic conditions, increasing the likelihood of collisions and accidents. Based on 2017 police and hospital reports, the National

Highway Traffic Safety Administration (NHTSA) identified 91,000 car accidents caused by drowsy drivers [3]. A study by the American Automobile Association's Foundation for Traffic Safety estimated that more than 320,000 drowsy driving accidents happen yearly, including 6,400 fatal crashes [4]. Similarly, the report of the NHTSA in the USA concluded that around 64.4% of people lose life due to the diversion of attention from driving [5]. Moreover, their report also declared that somewhere between 94% and 96% of all motor vehicle accidents are caused by some human error, while many road accidents are due to the usage of electronic devices such as Bluetooth devices and mobile phones. The gravity of these statistics underscores the urgent need for comprehensive measures to combat driver inattention and forge safer roads for all.

While various factors contribute to driver inattention, distractions, fatigue, drowsiness, and emotional states are among the most prevalent. Distraction by a secondary task is one of the main factors impairing driving; common examples include using a mobile phone, eating or drinking, talking to passengers, grooming like applying makeup, adjusting controls, and reading. Additionally, fatigue and drowsiness are also common sources of inattention. Ceccacci et al. [6] indicate that the emotional state of a driver significantly influences their level of concentration. For example, Sterkenburg and Jeon [7] demonstrated that anger degrades driving performance as much as or more than other traditional distraction tasks. Driver's emotion has become a factor that cannot be ignored in traffic safety research [8], [9].

In response to these challenges, numerous research efforts have been undertaken to develop effective solutions for detecting and mitigating driver inattention. These efforts often focus on individual aspects of inattention, utilizing various technologies ranging from video image analysis to physiological signal monitoring. According to Deng and Wu [10], called the "DriCare" system, primarily targets detecting drivers' fatigue status through video images captured by a camera installed in the vehicle. By employing convolutional neural networks (CNN), this approach achieved an accuracy of 92% in identifying drowsiness and fatigue. Similarly, Chirra et al. [11] developed a novel deep-learning framework that detects driver drowsiness based on eye state. This approach achieved an accuracy of 96.42%. A different work was proposed in [12] to detect drowsiness based on electroencephalogram (EEG) signals. They used the Q-factor wavelet transform (TQWT) coupled with the extreme learning machine (ELM) for classification, resulting in an accuracy of 91.84%. Meanwhile, a multi-tasking CNN model is presented in [13], combining drowsiness and fatigue detection. Remarkably, their approach achieved a high accuracy rate of 98.81% while explicitly evaluating fatigue as 'very tired, less tired, and not tired.' Bakheet and Hamadi [14] proposed a framework for instantaneous driver drowsiness detection. They employed an adaptive variant of the histogram of oriented gradients (HOG) features to represent the eye region and utilized a naive Bayes (NB) model for classification. Their work was rigorously evaluated using the publicly available NTHU-DDD dataset, demonstrating the potential of their framework as a strong contender against several state-of-the-art baselines. Notably, their framework achieved a competitive detection accuracy of 85.62% while maintaining efficiency and stability.

Zhao et al. [15] also investigate driver distraction detection, focusing on the head pose. They employed the HPE_Resnet50 algorithm to achieve an accuracy of 95% in identifying instances of distraction. On the other hand, Jamsheed et al. [16] used a CNN-based method for developing driver action classifiers. Their research successfully classified distracted drivers into ten categories with an accuracy of 97%, utilizing the State Farm dataset. Furthermore, Panwar et al. [17] proposed a deep learning convolutional model to tackle distraction and drowsiness during driving. Their comprehensive approach, with an accuracy of 99.95%, categorized various inattention instances, such as un-attentive driving, mobile phone usage, frequent yawning, and sleeping. Li et al. [18] propose a novel algorithm for detecting manual distractions among drivers. This algorithm comprises two modules: the first predicts bounding boxes for the driver's right hand and right ear from RGB images, while the second module classifies the type of distraction based on these bounding boxes. They trained and tested the algorithm on a dataset consisting of 106,677 frames extracted from videos captured during simulated driving sessions with twenty participants. Notably, the framework yielded an F1-score of 0.84, 0.69, and 0.82, for classifying normal driving, touchscreen interaction, and phone conversation respectively. Moreover, in the case of [19], the random forest-based approach was implemented across physiological functional variables to take drivers' stress levels and accordingly categorize them. The analysis was performed on experimental data extracted from the drivedb open database. The physiological measurements of interest are electrodermal activity captured on the driver's left hand and foot, electromyogram (EMG), respiration, and heart rate; they achieved an accuracy of 81%. According to Leone et al. [20] a system was designed based on a low-cost camera to detect driver road rage through a meticulous analysis of the driver's facial expressions. What sets this approach apart is its sophisticated decision-making strategy, which relies on the temporal coherence of facial expressions categorized as "anger" and "disgust." This methodology yielded an accuracy of 84.56% when employing the support vector machine (SVM) algorithm.

Various sensors have been employed to detect different forms of driver inattention, including distraction, drowsiness, fatigue, and emotional state. Physiological approaches involve discretely placing

sensors on the driver's body to extract signals such as EEG, electrocardiogram (ECG), or EMG. Another approach relies on driving performance indicators, such as monitoring lane departure, pedal activity, and accelerator usage. However, the most widely adopted approach by drivers is vision-based, which utilizes cameras. This non-intrusive nature makes it a preferred choice, as it does not require physical contact with the driver and thus ensures a higher level of comfort and acceptance.

The literature contains a wealth of proposals for detecting driver inattention, as extensively described in our previous work [21]. These approaches have certainly provided valuable insights into understanding and addressing particular forms of inattention. However, these studies focus individually on specific aspects of driver inattention, such as distraction, fatigue, drowsiness, or emotional state. Recognizing these factors individually may lead to incomplete or less accurate assessments of driver inattention. An approach that takes into account the detection of different inattention factors could provide a more complete and nuanced understanding of the driver's condition.

This research aims to propose a novel system for detecting driver inattention, targeting four key aspects: distraction, fatigue, drowsiness, and emotions. This system is structured across six levels: perception, facial features detection, tracking driver using pre-trained deep learning models, inattention detection, risk estimation, and alert. Each type of distraction will be addressed by a distinct model designed to capture characteristics signals associated with these states of inattention. By combining these four models, the proposed system will accurately assess the level of driver inattention in real-time using a risk calculation. The system adopts an entirely non-intrusive approach, relying on image processing from two cameras installed in the car to monitor the driver's state. One of our key contributions lies in the selection of an optimal algorithm for level 2 of the system, which involves facial feature detection and tracking. To this end, we conducted a rigorous comparative analysis of several state-of-the-art algorithms, including multi-task cascaded convolutional networks (MTCNN), HOG with dlib library, and Haar feature-based cascade classifiers. This step is crucial for extracting the mouth region to detect signs of fatigue, while signs of drowsiness are identified by analyzing the eye region. Additionally, the system extracts the entire face region to detect and interpret facial expressions. The proposed system aims to fill the gap left by previous approaches by detecting any type of driver inattention and to improve road safety by providing relevant alerts corresponding to the assessed level of risk.

The paper is organized as follows: section 2 presents a novel system for driver inattention monitoring. Section 3 is devoted to the optimal algorithm proposed for facial features detection and feature extraction, which is MTCNN. Moving on to section 4, a comprehensive discussion is presented, offering insights into the rationale behind selecting the MTCNN algorithm. Additionally, this section outlines potential avenues for future refinement and deployment of the system. Finally, section 5 encapsulates the paper with a concluding summary.

2. NOVEL SYSTEM FOR DRIVER INATTENTION MONITORING

In this research, our primary focus revolves around the detection of driver inattentiveness, the generation of alert messages, and providing proactive assistance to the driver. To achieve these objectives, we have innovatively devised a comprehensive six-layered architecture specifically tailored for driver inattention detection, complemented by the implementation of a personalized assistance system, as visually depicted in Figure 1. Our six-layered architecture is designed to efficiently detect driver inattention. Each layer serves a distinct purpose: sensors collect diverse data, facial feature extraction identifies key facial elements, the tracking stage ensures continuous tracking of features, inattention detection via a temporal window, risk estimation assesses the level of risk, and the alert layer issues warnings in case of critical risk. This modular approach allows for independent enhancements and advancements, streamlining development and maintenance efforts.

2.1. Perception layer

In the initial layer, we strategically position two cameras: one on the driver's side to comprehensively capture secondary tasks and another directed frontally to focus on facial features. It is worth noting that the adopted approach is visually based, relying on camera usage. However, this architecture remains open to include other sensors when it is needed. For example, enabling physiological measures such as EEG, ECG, or EMG and integrating other predictive models to alert the driver when a problem arises.

2.2. Facial features extraction layer

Moving on to the second layer, the step of facial feature detection and extraction is applied to the stream of images from camera two, simultaneously extracting vital facial characteristics like the face, eyes, and mouth. It is crucial to emphasize that images from camera one bypass this step, as they are used in their entirety for identifying secondary tasks, eliminating the need for image segmentation. The market offers a plethora of commercial and non-commercial face detection and alignment algorithms; however, due to time constraints, evaluating them is impractical. Therefore, our evaluation will focus on the most employed algorithms in the commercial sector, those with widespread popularity, and open-source implementations freely available.

Considering the dual tasks of face detection and face alignment, our assessment will center on Haar Cascade-OpenCV [22], HoG-Dlib [23], and MTCNN [24]. These algorithms have garnered extensive adoption and have proven to be effective in accomplishing both objectives. They have emerged as highly viable options for real-world applications across various domains by striking a balance between accuracy and efficiency.
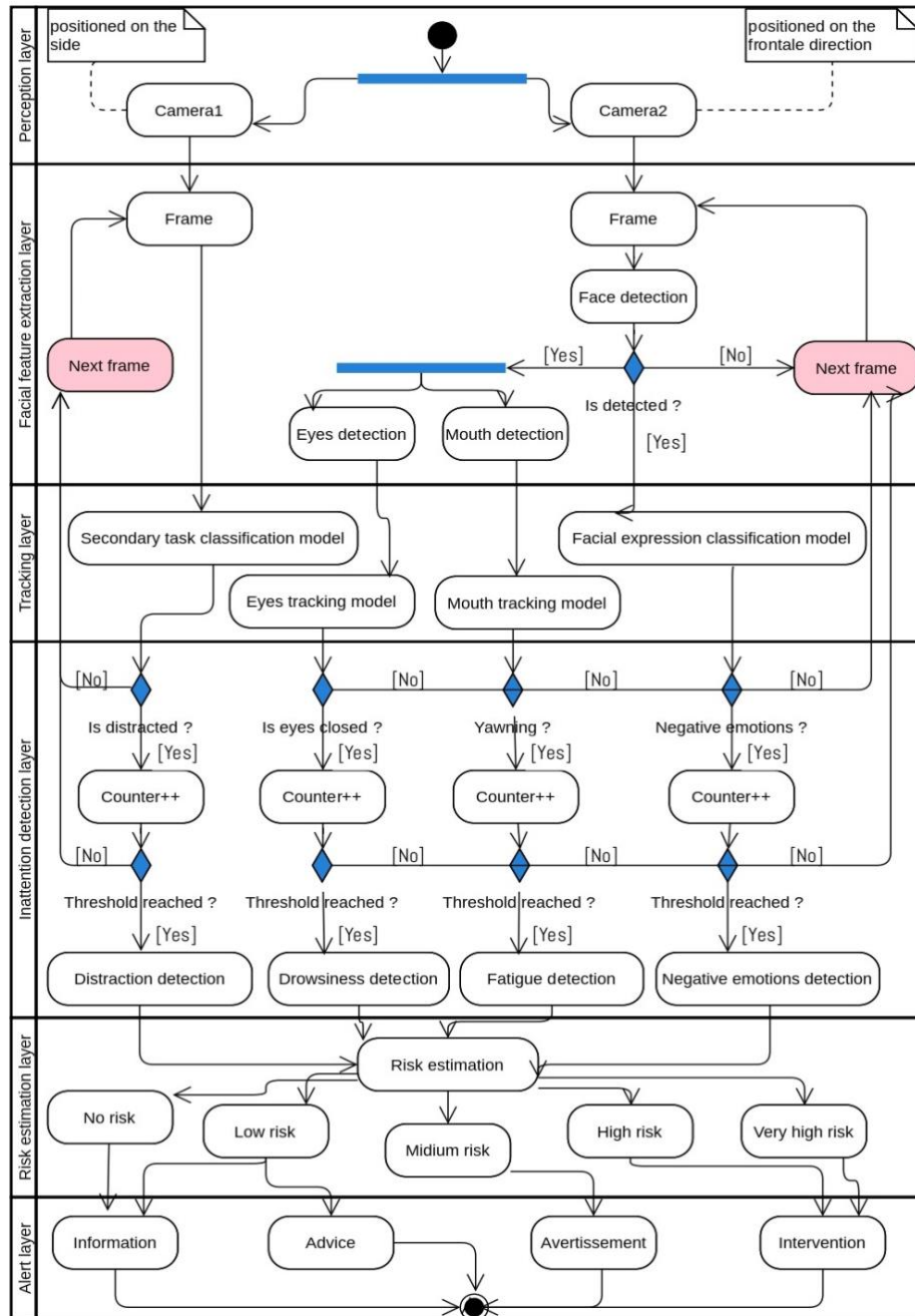

Figure 1. The flowchart of the proposed driver inattention monitoring system

2.3. Tracking using pre-trained models

In the third layer, the tracking stage utilizes deep learning models trained on extensive datasets offline. The initial model predicts secondary tasks using imagery from camera 1. These secondary tasks encompass activities like texting, calling, reaching_behind, hair_makeup, drinking, adjusting_radio, and talking_to_passenger. Another deep-learning model classifies facial expressions into seven distinct

categories: happiness, fear, anger, sadness, neutrality, surprise, and disgust. Additionally, a separate deep learning model monitors eye state (open or closed), while another model tracks mouth movement to ascertain whether the driver is yawning or not. Deep learning models for detecting driver distraction, drowsiness, fatigue, and emotions greatly benefit from the use of CNN. CNN are powerful tools for these tasks as they are specially designed to automatically extract meaningful features from visual data, such as images or videos. They are capable of identifying complex patterns, resisting variations in object position and size within the image, and learning features at different levels of abstraction. This makes them natural choices for detecting distracting behaviors, signs of drowsiness or fatigue, as well as various emotional expressions on the driver's face. These models undergo periodic updates based on driver imagery, allowing for continuous improvement in performance and personalized adaptation to individual driver profiles.

2.4. Inattention detection layer

In the fourth layer, driver inattention detection takes place, employing a temporal window to evaluate various forms of inattention. Relying solely on a single image (frame) does not reliably allow for the detection of driver inattention. Instead, it is more effective to base the analysis on the number of frames per second. This means that by analyzing multiple images per second, a more comprehensive and dynamic view of the driver's behavior can be obtained. However, this requires the establishment of a threshold, which is a predetermined value at which inattention is present. This threshold is crucial because it defines the point at which a specific state of inattention can be confirmed. For instance, if the number of frames displaying signs of inattention surpasses this established threshold, one can then conclude that the targeted form of inattention is present. The adaptive thresholding process incorporates a counter to monitor a specific number of successive frames that satisfy the criteria before issuing a warning. For example, Shakeel *et al.* [25] integrated a threshold mechanism: if the classifier consistently identifies ten consecutive instances of closed eyes, this observation suggests that the individual is exhibiting signs of drowsiness. Rafid *et al.* [26] established a specific criterion for determining drowsiness: if the classifier consistently identifies 30 consecutive instances of closed eyes, this observation indicates drowsiness. Fasanmade *et al.* [27] proposes the use of a sequence of frames to gauge the duration of driver distraction. The coding was configured such that once a threshold of 125 consecutive frames is reached (equivalent to 5 seconds), a classification decision is triggered. In the literature, diverse thresholds have been suggested for detecting driver inattention. Therefore, determining the appropriate number of consecutive frames necessitates rigorous testing and validation through realistic driving scenarios. This ensures that the chosen threshold effectively captures instances of inattention in practical driving situations.

2.5. Risk estimation layer

The fifth layer involves risk calculation based on the predicted classes. The scoring dictionaries play a pivotal role in quantifying the potential risk of driver inattention. Each dictionary corresponds to a specific category: distraction, drowsiness, yawning, and emotions. The system effectively captures the degree of inattention risk associated with the driver's actions by assigning predefined scores to various behaviors and expressions within each category. For instance, high-risk activities like texting receive higher scores, while neutral or positive emotions yield no additional risk. The individual scores from each category are then combined to calculate the global inattention risk score, providing a comprehensive assessment of the driver's attentiveness. This method enables a nuanced understanding of the driver's state. It facilitates the classification of risk levels, thereby contributing to the development of robust driver inattention monitoring systems and enhancing road safety. The scoring system utilized in these dictionaries draws its foundation from authoritative sources, particularly public reports such as those provided by the NHTSA. These reports furnish comprehensive statistics and insights into various forms of dangerous inattention that drivers may exhibit while on the road. By aligning our scoring criteria with the findings and assessments outlined in these reports, we aim to ensure that the risk assessments are rooted in well-documented and widely acknowledged data, thereby enhancing the accuracy and reliability of the safety evaluations.

2.5.1. Distraction scores

Many distractions appear to increase the relative risk of crashes and near-crashes, and distractions that require drivers to take their eyes off the road are potentially more of a safety problem than distractions that do not require drivers to take their eyes off the road [28]. Using a cell phone while driving creates enormous potential for deaths and injuries on roads [29]. Table 1 presents the distraction scores, providing insights into their relative risks.

2.5.2. Drowsiness and fatigue scores

A state of fatigue is often marked by the occurrence of frequent yawning. Yawning, along with the sensation of weariness, serves as clear indicators of both physical and mental exhaustion, posing a substantial threat to one's capacity to drive safely. Additionally, drowsiness is characterized by the involuntary closure

of the eyes, further heightening the risk. Recognizing the severity of this danger, fatigue and drowsiness have been assigned a high score of 3, underscoring the significant peril associated with these conditions. Table 2 provides a detailed breakdown of the drowsiness and fatigue scores, offering insights into their potential impact on driving safety.

Table 1. Distraction scores

Inattention activity (severity score)	Description
Texting (3 points)	Texting while driving is considered an extremely dangerous form of distraction. It requires significant visual, manual, and cognitive attention away from the road, making it a high-risk activity.
Calling (2 points)	While less distracting than texting, making a call still diverts attention from driving. It involves cognitive and manual distraction as the driver must hold the phone and engage in conversation.
Reaching behind (3 points)	This action involves significant manual and visual distraction, as the driver's attention is focused away from the road while reaching for an object.
Hair and makeup (1 point)	Though not as severe as texting, grooming activities still require manual and visual attention away from driving, making it a moderate-risk behavior.
Drinking or eating (1 point)	Taking a drink or eating while driving can lead to momentary distraction, particularly if the driver has to reach for a container.
Adjusting radio (1 point)	Adjusting the radio can lead to brief manual and visual distraction, but it is generally considered a lower-risk behavior.
Talking to passenger (1 point)	Conversations with passengers can cause some cognitive distraction, but it's generally a common and relatively lower-risk behavior.-

Table 2. Drowsiness and fatigue scores

Inattention activity (severity score)	Description
Drowsy (3 points)	Drowsiness significantly impairs a driver's ability to react quickly and make sound judgments. It's a high-risk condition as it increases the likelihood of accidents.
Yawning (3 points)	Yawning is often indicative of drowsiness or fatigue, which can severely impair a driver's ability to focus and react on time.
No drowsiness (0 points)	When a driver is alert and not drowsy, there is no additional risk associated with this factor.
No yawning (0 points)	When a driver is not yawning, there is no additional risk associated with this factor.

2.5.3. Emotions scores

Numerous studies have highlighted a noteworthy association between driving-related anger and specific high-risk driving practices, including instances of speeding, aggressive driving, and disregarding traffic signals [30]. Furthermore, scholars have underscored that feelings of anxiety and fear can also serve as predictors for engaging in risky driving behaviors [31]. Table 3 presents the emotion scores, outlining their potential contribution to risky driving practices.

Table 3. Emotions scores

Inattention activity (severity score)	Description
Anger (3 points)	Anger can lead to cognitive distraction, aggressive driving behaviors, and road rage, posing significant risks to road safety.
Fear and sadness (2 points each)	These emotions can lead to cognitive distraction and, in some cases, physical reactions that may affect driving performance.
Surprise and disgust (1 point each)	While these emotions may momentarily distract a driver, their impact is generally considered lower compared to fear and sadness.
Neutral and happiness (0 points)	When a driver is emotionally neutral or experiencing happiness, there is no additional risk associated with these factors. Positive emotions may even contribute to a more alert and focused state.

2.5.4. Total risk

The risk categorization algorithm is based on a simple "if-then" approach. We can determine the risk level associated with each combination by using the sum of points obtained from different combinations of scores for distracting activities, drowsiness levels, yawning occurrences, and the driver's emotions. Once we calculate the total points for each variety, we pass it to the categorization algorithm. If the total points are less than or equal to 2, the risk category is considered "low." If the total points fall between 3 and 7 (inclusive), the risk category is defined as "medium." Lastly, if the total points fall between 8 and 10

(inclusive), the risk category is labeled "high." If the total points exceed 10, the risk category is labeled "very high." Figure 2 provides a visual representation of the categorization algorithm.

```python
# Function to categorize risk level
def categorize_risk(total_points):
    if total_points <= 2:
        return "Low"
    elif 3 <= total_points <= 7:
        return "Medium"
    elif 8 <= total_points <= 10:
        return "High"
    else:
        return "Very High"
```

Figure 2. Risk function compute

2.6. Alert layer

In the culmination of the system's functionality, the final layer assumes the vital role of driver alerting. It employs a nuanced risk assessment approach to determine the appropriate level of alertness required. The alert stage represents a proactive approach to enhancing driver safety by providing timely and relevant feedback and, when necessary, intervening to prevent potential hazards associated with inattention. It serves as a valuable tool in promoting responsible and attentive driving behavior.

3. DRIVER FACIAL FEATURES DETECTION PROPOSITION

A primary and significant challenge in our system lies in extracting relevant features from images (driver facial features extraction layer). In the context of this study, we will place special emphasis on this crucial component. Figure 3 illustrates our approach using the MTCNN algorithm for extracting facial features such as eyes and mouth using the camera in front of the driver.

This choice is justified by two experiments on three datasets FE, DrivFace, and driver drowsiness dataset. The first focuses on low lighting and different head movements (across a 180° view), while the second experiment aims to assess the performance of the MTCNN algorithm on a real-world dataset of driving scenarios. To ensure precise analysis of facial features, MTCNN is compared to a histogram of gradient (HOG)-based frontal face detector, and a Haar feature-based cascade classifier. These algorithms were extensively compared for their adaptability to varying lighting conditions, and capacity to accommodate different head movements in experiment 1. Furthermore, in experiment 2, their adeptness in accurately detecting essential facial components such as the face, eyes, and mouth was meticulously assessed using expansive datasets obtained from real-world driving scenarios and performance metrics including true positives (TP), true negatives (TN), false positives (FP), false positives (FN), accuracy, precision, F1-score, and recall.
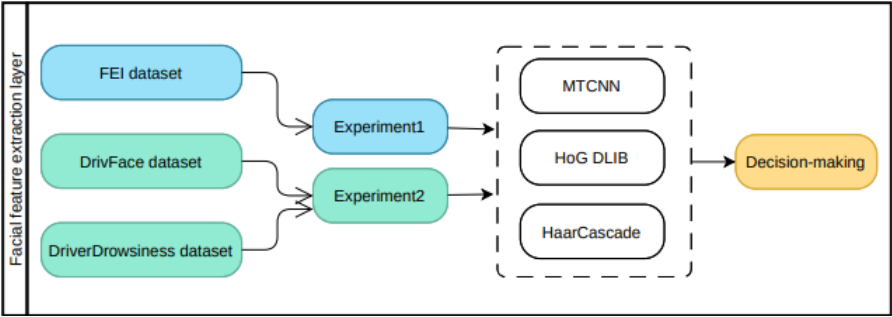

Figure 3. Proposal of incorporating MTCNN in facial feature extraction layer

3.1. Datasets

In this study, we utilized three datasets. The FEI dataset was employed to assess the performance of techniques, taking into account head rotation and lighting conditions. The FEI face dataset is a collection of

Brazilian face images captured between June 2005 and March 2006 at the Artificial Intelligence Laboratory of FEI in São Bernardo do Campo, São Paulo, Brazil [33]. This dataset consists of 2,800 colorful images, with 14 images for each of the 200 individuals. The photographs were taken against a white homogeneous background, showing the subjects in an upright frontal position with profile rotation of up to approximately 180 degrees. The images' original size is 640×480 pixels, with a possible variation in scale of about 10%. The faces featured in the dataset belong to individuals aged between 19 and 40 years, mainly comprising students and staff at FEI, displaying diverse appearances, hairstyles, and adornments. Notably, the dataset includes an equal number of male and female subjects, each totaling 100.

The other two datasets were utilized to evaluate the robustness of detections in real driving scenarios. The public DrivFace dataset [32] contains image sequences of subjects while driving in real scenarios. It comprises 606 samples of 640×480 pixels each, acquired over different days from 4 drivers (2 women and 2 men) with various facial features obstructions, such as glasses and facial hair. The driver drowsiness dataset [33] contains 1448 photos, divided into two groups: 723 images labeled "yawning" and 725 images labeled "No yawning". It is important to note that each of the 'yawning' and 'No yawning' subsets will be used separately for evaluation purposes. Figure 4 showcases examples of images from the three datasets.

Figure 4. Examples of images from the three datasets

3.2. Facial feature extraction techniques
3.2.1. Haar cascade-opencv

In 2001, Viola and Jones [22] proposed an effective object detection technique known as Haar feature-based cascade classifiers. It is a machine learning-based method that trains the classifier with many positive photos (with faces) and negative images (without faces). In their study, various extremely basic or weak facial traits are learned using the AdaBoost model to create a robust classifier for each face. The Viola-Jones detector is one of the earliest methods. It functions on grayscale images by interpreting the image as a collection of Haar features (lighter and darker rectangles). There are numerous Haar feature types with various placements of the rectangle's soft and dark areas. They can be computed very quickly using an integral image method. The integral image is a computing technique that enables the rapid and efficient calculation of the sum of pixel values, achieving constant time complexity and little computational overhead. The process involves generating an image of equal dimensions as the original image, referred to as a supplemental area table. The summation of the pixels located to the left and above each given pixel (x, y) in the original image is calculated using in (1).

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{1}$$

The function ii(x, y) represents the pixel values of the integral image, while I (x, y) represents the pixel values of the original image at point (x, y). The computation of the total pixel values within a rectangle region can be simplified by utilizing only four values from the integral image, rather than summing the values of all individual pixels. If A, B, C, and D represent the values at the corners of the table being totaled, the total sum within this rectangular region can be calculated using in (2).

$$sum = D + A - B - C \tag{2}$$

Cascading classifiers are employed to efficiently eliminate non-face instances and minimize superfluous computational operations. Cascading involves the construction of a cascade classifier, which is comprised of multiple stages, each housing a robust trained classifier. In the event of a sub-window experiencing failure at any point, it will promptly be eliminated. The Adaboost algorithm is employed in the creation of a cascade by meticulously constructing each individual stage. The process involves the integration of multiple classifiers to form the stages.

3.2.2. HoG-Dlib

The HoG-Dlib provides an approach for face detection based on HOG and Linear SVM [23]. The idea of the HOG descriptor is to create a vector of features so that the vector can be fed into a classification algorithm like SVM to predict the result. To calculate the HOG descriptor, we need to calculate the gradients of the x-axis- and y-axis gradients. The calculation of the gradient vector is as (3) and (4):

$$G_x(x,y) = H(x+1,y) - H(x-1,y) \tag{3}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1) \tag{4}$$

In (3) and (4) represent the horizontal gradient of the image pixel, Gx(x, y), and the vertical gradient, Gy(x, y), respectively. In (5) and (6) respectively represent the magnitude and direction of the gradient at pixel (x,y).

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{5}$$

$$a(x,y) = tan^{-1}\left(\frac{G_x(x,y)}{G_y(x,y)}\right) \tag{6}$$

These gradients capture the image's direction and magnitude of pixel intensity changes. By analyzing these gradients, Dlib can construct a feature vector that describes the characteristic patterns of faces. In the context of Dlib, five HOG filters are used for face detection: front-looking, left-looking, right-looking, front-looking but rotated left, and front-looking but rotated right. These filters help capture variations in facial orientation and ensure robust face detection even when faces are rotated or captured from different angles.

3.2.3. Multi-task cascaded convolutional neural network

The MTCNN stands as a prominent milestone in the realm of computer vision and face alignment. Unveiled by Kaipeng Zhang and fellow researchers in 2016 [24]. The MTCNN is well-known for its state-of-the-art performance on a variety of benchmark datasets as well as its landmark detection ability, which enables it to identify additional facial features like the eyes and mouth. The network employs a cascade structure with three networks. The image is first resized to various sizes (referred to as an image pyramid). The first model, the proposal network (P-Net), proposes candidate facial regions; the second model, the refine network (R-Net), filters the bounding boxes, and the third model, the output network (O-Net), proposes facial landmarks.

Upon the detection of a face, the P-Net algorithm provides the coordinates of a bounding box. The operation will be repeated in a section-wise manner, with the 12×12 kernel being shifted 2 pixels to the right or down at each iteration. The displacement of 2 pixels is commonly referred to as the stride. The facial features seen in most of the images exhibit a size exceeding 2 pixels. The likelihood of the kernel failing to detect a face is relatively low. The R-Net incorporates the precise coordinates of the updated bounding boxes, which are more accurate than previous versions. Table 4 provides technical information about the three algorithms, providing insights into their CPU and GPU usage capabilities, support for color information, and recommended image sizes.

The table presents technical information on three widely used face detection and alignment methods: Haar Cascade-OpenCV, HoG-Dlib, and MTCNN. All three methods support CPU usage, making them accessible for standard computing devices. While Haar Cascade and HoG-Dlib also offer GPU support, MTCNN stands out by utilizing color information for enhanced accuracy, while the other two methods operate on grayscale images. Recommended image sizes vary, with Haar Cascade suggesting 24×24, HoG-Dlib > 80×80, and MTCNN > 20×20. Each method has unique strengths, making it essential to

consider specific application requirements and available computational resources when selecting the most suitable algorithm for face detection and alignment tasks.

Table 4. Technical information about the three algorithms

	Haar Cascade	HoG-dlib	MTCNN
CPU	Yes, available	Yes, available	Yes, available
GPU	Certain implementations (Not OpenCV)	Yes, available	Yes, available
Using Colors	No	No	Yes
Image size	Recommends 24x24	>80x80	>20x20

3.3. Experiments and results

We undertook two distinct experiments to assess the performance of three algorithms across varied contexts. In the first experiment, we gauged performance by quantifying the successful face detections achieved. For the second experiment, we employed widely recognized metrics to thoroughly evaluate the performance of these three algorithms. These experiments utilize Python scripts with various libraries for face detection in images. The first script employs OpenCV, iterating through a directory of images and utilizing the Haar Cascade classifier to detect faces. The total count of detected faces is then printed. Similarly, the second script uses the Dlib library alongside OpenCV to detect faces in images. It iterates through the image directory, applies face detection using Dlib, and prints the total count of detected faces. Finally, the third script employs the MTCNN model for face detection, again iterating through the image directory, applying face detection using MTCNN, and printing the total count of detected faces. Overall, these experiments demonstrate different approaches to face detection using various libraries in Python.

To comprehensively evaluate the performance of the algorithms using various metrics, we augmented the existing datasets by introducing additional images that did not contain the driver. By including these images without drivers present, we could calculate TN and, FP, and other evaluation metrics. This approach allowed us to gain insights into the algorithms' effectiveness in detecting faces and aligning features precisely in driving-related scenarios, thus providing a more robust assessment of their performance in a real-world context. Here is an explanation of each metric: Precision is a metric that quantifies the accuracy of positive predictions made by a model: precision=TP/(TP+FP). Recall, also known as sensitivity or TP rate, measures the model's ability to correctly identify positive instances from all the actual positive instances in the dataset. Recall=TP/(TP+FN). The F1-Score is the harmonic mean of precision and recall. F1-Score=2×(Precision×Recall)/(Precision+Recall). Accuracy is a metric that measures the overall correctness of the model's predictions. Accuracy=(TP+TN)/(TP+TN+FP+FN).

3.3.1. Face detection on fei dataset

The objective of the experiment is to demonstrate how well each method can detect faces on different head movements and in low lighting conditions. We assessed face detection in the 10 head positions and different lighting conditions separately to understand how each of the algorithms performs in these scenarios. In this experimental setup, the aim is to detect faces using the FEI dataset, where each distinct head movement state consists of an equal number of 200 images. Additionally, there is a set of 200 images specifically representing low-light conditions. The experiment involves applying three different face detection algorithms to each state within the dataset. For every state and every face detection classifier, the code counts the total number of faces detected. The goal is to analyze and compare the effectiveness of these algorithms in various scenarios within the FEI dataset, including different head movements and lighting conditions. Table 5 summarizes the comparison results of the three algorithms from the initial evaluation experiment. It displays the outcomes achieved for each section of the dataset. States 1 to 10 represent the driver's head in various 180° head movement scenarios, while the final state, 11, depicts the driver in low-lighting conditions.

Table 5. Detection rate for on FEI

Algorithms	1	2	3	4	5	6	7	8	9	10	11
Haar Cascade	40	182	197	**200**	**200**	**200**	198	197	164	5	154
HoG-dlib	159	**200**	**200**	**2000**	**200**	**200**	**200**	**200**	196	95	177
MTCNN	154	188	197	197	198	197	197	196	187	86	155

The results presented in Table 5 demonstrate that for states 3 to 8, all three algorithms exhibit similar performance. However, for states 1, 2, 9, 10, and 11, noticeable differences emerge. While the Haar Cascade algorithm experiences a performance decrease, the other two algorithms also exhibit a reduction in their performance, albeit to a greater extent, much like Haar Cascade. This indicates that the Haar Cascade algorithm no longer performs effectively in cases where the face is extremely non-frontal, and also under conditions of low lighting.

3.3.2. Face detection in the context of driving

Table 6 presents the outcomes of the second experiment, which concentrated on two datasets that encompass real-world scenarios. These datasets are abundant in images featuring complex lighting conditions, diverse head orientations, a wide array of facial expressions, and various forms of occlusion, including glasses and sunglasses. The table provides a detailed analysis of the performance of three face detection algorithms on three different datasets. The datasets used are the "DrivFace dataset," "Yawn dataset," and "No Yawn dataset."

Table 6. Evaluation of different metrics on public datasets

Dataset	Algorithms	Tp	Tn	Fp	Fn	Precision	Recall	Accuracy	F1-score
DrivFace dataset	MTCNN	590	606	0	16	1.000	0.974	0.980	0.974
	DLIB	172	606	0	434	1.000	0.283	0.641	0.441
	HAAR CASCADE	148	606	0	458	1.000	0.244	0.622	0.392
Yawn subset	MTCNN	659	723	0	64	1.000	0.911	0.964	0.953
	DLIB	252	723	0	471	1.000	0.348	0.674	0.516
	HAAR CASCADE	248	723	0	475	1.000	0.343	0.712	0.511
No yawn subset	MTCNN	719	725	0	6	1.000	0.992	0.995	0.996
	DLIB	256	725	0	469	1.000	0.353	0.676	0.521
	HAAR CASCADE	160	725	0	565	1.000	0.221	0.560	0.363

MTCNN demonstrated remarkable performance across all three datasets. It achieved high precision rates of 100% for all datasets, indicating successful avoidance of FP. Additionally, it is recall was generally high, surpassing 90% in each case, highlighting it is ability to detect most TP efficiently, even in challenging conditions such as non-frontal images. DLIB ranks second in terms of overall performance. Although its precision is also 100%, it is recall was lower than that of MTCNN for all datasets. This suggests that DLIB encountered difficulties detecting specific positive faces, particularly in more complex scenarios where looks are not frontally aligned. Finally, Haar Cascade obtained the lowest performance among the three algorithms. While its precision was 100%, it is recall was considerably lower than the other two algorithms. This indicates that Haar Cascade struggled to detect many TP, which may be attributed to its limitations in face detection across more diverse scenarios. The precision of 1.00 for all datasets and the three algorithms is indeed related to the fact that the images added to the dataset that do not contain faces show no FP. This indicates that the algorithms successfully identified images without a face. The fact that TN is at 100% indicates that all cases where there are no faces were correctly detected as such, contributing to perfect precision.

In terms of runtime, the OpenCV Haar Cascade method outperforms others, achieving an impressive 30 fps (frames per second) [34]. However, it does suffer from the significant drawback of generating numerous false predictions and may require more efficient handling of different head orientations. The HoG face detector in Dlib is also quite fast, achieving a frame rate of 19 fps [34]. It excels in detecting faces, even in low-light conditions. However, its performance dips when faced with extremely non-frontal angles. In comparison, MTCNN emerges as the most accurate and robust approach, achieving a frame rate of 7 fps [34]. It exhibits exceptional capability in handling various lighting conditions and head orientations, a fact supported by two separate experiments. Another study conducted by [35] posits that an optimization of the MTCNN algorithm could potentially achieve an impressive 33 fps (frames per second).

These observations determined that the system requires a face detection and alignment algorithm with the following characteristics: Low false detection rate, precisely targeted and extracted facial regions, fast execution time (runtime), and challenges with extremely frontal faces. Consequently, the decision was made to utilize the MTCNN algorithm in the system for detecting faces and extracting eyes and mouth, which enables precise targeting of facial components. It is essential to note that the algorithm selection depends on the application's specific requirements. Each algorithm has its strengths and weaknesses, and the final choice will be based on the constraints and priorities of the project.

4.   DISCUSSION

In this paper, we present a context-aware inattention detection system based on MTCNN algorithm for face detection and alignment, aimed at enhancing road safety. The goal is to address most inattention forms that lead to numerous accidents. Utilizing two cameras, strategically positioned in front and on the right side of the vehicle, our system is structured with six layers to ensure both flexibility and scalability. Our study investigates the effects of different forms of driver inattention and proposes a unified system to detect and monitor these states in real-time. While earlier studies have focused on specific aspects of inattention, such as distraction or drowsiness, our approach comprehensively tackles these factors within a single framework. We find that integrating MTCNN for facial feature detection provides significant advantages over traditional methods, enabling robust detection of multiple forms of inattention. The adaptability of MTCNN to diverse driving contexts enhances the system's effectiveness in real-world scenarios, without compromising accuracy or reliability. This capability allows for the extraction of crucial features such as mouth movements, eye behavior, and overall facial expressions, enabling the system to identify signs of fatigue, drowsiness, or emotional states exhibited by the driver. Our study highlights the importance of ongoing research efforts to refine and optimize inattention detection systems, the incorporation of a physiological approach, using non-intrusive devices like smartwatches, offers a promising avenue for enhancing the system's capabilities. These devices can monitor the driver's internal state, including factors like heartbeat and heart rate variability.

The utilization of temporal windows for detection, based on real-world experiments is crucial for establishing precise thresholds and enhancing overall system effectiveness. For example, a window of 6-10 consecutive frames, may be employed for drowsiness detection [11], [36], while a window of 6 seconds could be used for distraction detection [37]. Nonetheless, the timing of the tracking step is critically important to minimize the occurrence of FP. Further investigation is warranted to strike a balance between meeting real-time requirements and ensuring the system's reliability.

The development of an efficient alert mechanism is of paramount importance. This mechanism should harness audio cues or visual messages on the dashboard to promptly notify the driver of potential inattention. However, it is crucial to design these alert messages with the utmost care, ensuring that they do not inadvertently become a new source of distraction for the driver. Striking the right balance between timely alerts and minimizing distractions is a nuanced challenge that warrants meticulous consideration in the system's design and development.

In conclusion, our research contributes to the development of an advanced context-aware inattention detection system leveraging MTCNN. By addressing various forms of driver inattention within a unified framework, our system offers a promising approach to enhancing road safety. Future endeavors will focus on refining the system's capabilities, incorporating additional sensors, and optimizing alert mechanisms to effectively mitigate risks associated with driver inattention.

5.   CONCLUSION

Our research proposes a novel context-aware inattention detection system that comprehensively tackles various aspects of driver inattention, including fatigue, drowsiness, distraction, and negative emotions, within a unified framework. This non-intrusive approach harnesses the power of image processing in conjunction with advanced face detection algorithms, ensuring precise analysis of the driver's state. By integrating these elements, our system offers a holistic view of inattention, allowing for accurate assessments and timely interventions. This research marks a significant stride in the pursuit of enhanced road safety, to reduce accidents stemming from driver inattention. Our conclusions are supported by the outcomes of our experiments, which affirm the efficacy of employing MTCNN for facial region extraction. For instance, we aim to enhance our system's performance by investigating various deep-learning models for more robust tracking of the driver's state. This endeavor will involve fine-tuning existing models and potentially developing novel architectures tailored to the specific nuances of driver inattention. While our system shows promise in addressing driver inattention, several limitations need to be considered: head movement tracking: the proposed system does not incorporate head motion as an indicator of drowsiness or distraction. Integrating deep learning models to track head movements could enhance the system's effectiveness in detecting these critical states. Model robustness: while our study demonstrates the efficacy of deep learning models for detecting fatigue, drowsiness, distraction, and emotional states, it is essential to recognize the need for continuous refinement. Training these models on diverse datasets and updating them regularly will improve their robustness and reliability. Threshold definition: defining precise thresholds for each form of inattention is crucial for accurate detection. Future research should conduct extensive real-world experiments to establish these thresholds based on empirical evidence. Message delivery: the delivery of alerts to the driver requires careful consideration, especially in relation to their emotional state. Future work should focus

on developing concise and effective messages that are sensitive to the driver's emotional condition. Despite these limitations, our proposed system represents a significant step towards enhancing road safety by addressing driver inattention.

REFERENCES

[1] World Health Organization, "Global status report on road safety 2018," *Global Report*, 2018. Accessed: Sep. 08, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789241565684

[2] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: Definition, relationship and taxonomy," *Accident Analysis and Prevention*, vol. 43, no. 5, pp. 1771–1781, Sep. 2011, doi: 10.1016/j.aap.2011.04.008.

[3] National Highway Traffic Safety Administration, "Drowsy driving," *Report a Safety Problem*, 2023. Accessed: Sep. 08, 2023. [Online]. Available: https://www.nhtsa.gov/risky-driving/drowsy-driving

[4] S. Martin, "Drowsy driving 2024 facts and statistics," *Bankrate*, 2024. Accessed: Jul. 19, 2024. [Online]. Available: https://www.bankrate.com/insurance/car/drowsy-driving-statistics/

[5] US National Highway Traffic Safety Administration (NHTSA), "Risky Driving," *Report a Safety Problem*, 2023. Accessed: Sep. 08, 2023. [Online]. Available: https://www.nhtsa.gov/risky-driving

[6] S. Ceccacci *et al.*, "Designing in-car emotion-aware automation," *European Transport-Trasporti Europei*, no. 84, pp. 1–15, 2021, doi: 10.48295/ET.2021.84.5.

[7] J. Sterkenburg and M. Jeon, "Impacts of anger on driving performance: A comparison to texting and conversation while driving," *International Journal of Industrial Ergonomics*, vol. 80, Nov. 2020, doi: 10.1016/j.ergon.2020.102999.

[8] D. H. -Fernández and S. F. a-Baeza, "Angry thoughts in Spanish drivers and their relationship with crash-related events. The mediation effect of aggressive and risky driving," *Accident Analysis and Prevention*, vol. 106, pp. 99–108, Sep. 2017, doi: 10.1016/j.aap.2017.05.015.

[9] S. R. Bogdan, C. Măirean, and C. E. Havârneanu, "A meta-analysis of the association between anger and aggressive driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 42, pp. 350–364, 2016, doi: 10.1016/j.trf.2016.05.009.

[10] W. Deng and R. Wu, "Real-time driver-drowsiness detection system using facial features," *IEEE Access*, vol. 7, pp. 118727–118738, 2019, doi: 10.1109/ACCESS.2019.2936663.

[11] 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.

[12] V. Bajaj, S. Taran, S. K. Khare, and A. Sengur, "Feature extraction method for classification of alertness and drowsiness states EEG signals," *Applied Acoustics*, vol. 163, Jun. 2020, doi: 10.1016/j.apacoust.2020.107224.

[13] B. K. Savaş and Y. Becerikli, "Real time driver fatigue detection system based on multi-task ConNN," *IEEE Access*, vol. 8, pp. 12491–12498, 2020, doi: 10.1109/ACCESS.2020.2963960.

[14] S. Bakheet and A. A. -Hamadi, "A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification," *Brain Sciences*, vol. 11, no. 2, pp. 1–15, 2021, doi: 10.3390/brainsci11020240.

[15] Z. Zhao *et al.*, "Driver distraction detection method based on continuous head pose estimation," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–10, Nov. 2020, doi: 10.1155/2020/9606908.

[16] V. A. Jamsheed, B. Janet, and U. S. Reddy, "Real time detection of driver distraction using CNN," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, IEEE, 2020, pp. 185–191, doi: 10.1109/ICSSIT48917.2020.9214233.

[17] P. Panwar, P. Roshan, R. Singh, M. Rai, A. R. Mishra, and S. S. Chauhan, "DDNet- A deep learning approach to detect driver distraction and drowsiness," *Evergreen*, vol. 9, no. 3, pp. 881–892, Sep. 2022, doi: 10.5109/4843120.

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

[19] N. El Haouij, J. M. Poggi, R. Ghozi, S. S. -Ghalila, and M. Jaïdane, "Random forest-based approach for physiological functional variable selection for driver's stress level classification," *Statistical Methods and Applications*, vol. 28, no. 1, pp. 157–185, 2019, doi: 10.1007/s10260-018-0423-5.

[20] A. Leone, A. Caroppo, A. Manni, and P. Siciliano, "Vision-based road rage detection framework in automotive safety applications," *Sensors*, vol. 21, no. 9, Apr. 2021, doi: 10.3390/s21092942.

[21] A. Soultana, F. Benabbou, N. Sael, and S. Ouahabi, "A systematic literature review of driver inattention monitoring systems for smart car," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 16, pp. 160–189, 2022, doi: 10.3991/ijim.v16i16.33075.

[22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1-9, 2001, doi: 10.1109/cvpr.2001.990517.

[23] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[24] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[25] M. F. Shakeel, N. A. Bajwa, A. M. Anwaar, A. Sohail, A. Khan, and Haroon-ur-Rashid, "Detecting driver drowsiness in real time through deep learning based object detection," in *Advances in Computational Intelligence*, Springer International Publishing, 2019, pp. 283–296, doi: 10.1007/978-3-030-20521-8_24.

[26] A. U. I. Rafid, A. I. Chowdhury, A. R. Niloy, and N. Sharmin, "A deep learning based approach for real-time driver drowsiness detection," in *2021 5th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2021*, IEEE, Nov. 2021, doi: 10.1109/ICEEICT53905.2021.9667944.

[27] A. Fasanmade *et al.*, "A Fuzzy-logic approach to dynamic bayesian severity level classification of driver distraction using image recognition," *IEEE Access*, vol. 8, pp. 95197–95207, 2020, doi: 10.1109/ACCESS.2020.2994811.

[28] M. Vegega, B. Jones, and C. Monk, "Understanding the effects of distracted driving and developing strategies to reduce resulting deaths and injuries: A report to congress," *U.S. Department of Transportation: National Highway Traffic Safety Administration*, 2013. Accessed: Oct. 15, 2023. [Online]. Available: https://www.nhtsa.gov/document/understanding-effects-distracted-driving

[29] National Highway Traffic Safety Administration, "Distracted driving," *Report a Safety Problem*, 2022. Accessed: Oct. 15, 2023. [Online]. Available: https://www.nhtsa.gov/risky-driving/distracted-driving

[30] C. A. M. -Juárez, C. G. -Hernández, E. O. -Ávila, and G. R. D. -Grijalva, "Approaching to a structural model of impulsivity and driving anger as predictors of risk behaviors in young drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 72, pp. 71–80, 2020, doi: 10.1016/j.trf.2020.05.006.

[31] P. M. Brown, A. M. George, and D. J. Rickwood, "Rash impulsivity, reward seeking and fear of missing out as predictors of

texting while driving: Indirect effects via mobile phone involvement," *Personality and Individual Differences*, vol. 171, 2021, doi: 10.1016/j.paid.2020.110492.

[32] A. H. -Sabat, A. Lpez, K. D. -Chito, "DrivFace," *UC Irvine: Machine Learning Repository,* 2016, doi: 10.24432/C5XC7Q.

[33] S. Raju, "Yawn eye dataset new," *Kaggle*, Accessed: Jul. 29, 2023. [Online]. Available: https://www.kaggle.com/datasets/serenaraju/yawn-eye-dataset-new

[34] M. Arora, S. Naithani, and A. S. Areeckal, "A web-based application for face detection in real-time images and videos," *Journal of Physics: Conference Series*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012071.

[35] S. Huang and H. Luo, "Attendance system based on dynamic face recognition," in *Proceedings - 2020 International Conference on Communications, Information System and Computer Engineering,* 2020, pp. 368–371, doi: 10.1109/CISCE50729.2020.00081.

[36] M. K. Sri and P. N. Divya, "Detection of drowsy eyes using viola jones face," *International Journal of Current Engineering and Scientific Research*, vol. 5, no. 4, pp. 375–380, 2018.

[37] R. J. Hanowski, R. L. Olson, J. S. Hickman, and J. Bocanegra, "Driver distraction in commercial motor vehicle operations," in *Driver Distraction and Inattention: Advances in Research and Countermeasures*, CRC Press, 2013, pp. 141–156, doi: 10.1201/9781315578156-10.

BIOGRAPHIES OF AUTHORS

Abdelfettah Soultana recently obtained his doctorate in computer science. He completed his master's degree in software quality at Hassan II University, Casablanca, Morocco, in 2015. His thesis, titled "Towards a contextual system for smart car management," marked a significant milestone in his academic journey. Currently, he is deeply engaged in his doctoral studies at the Laboratory of Information Processing and Modeling (LTIM) within the Ben M'sik Faculty of Science. His research focuses on machine learning, deep learning, and the application of internet of things technologies for driver context monitoring. He can be contacted at email: soultana.abdelfettah@gmail.com.

Faouzia Benabbou teacher-researcher since 1994, Authorized Professor since 2008 and Professor of Higher Education in the Department of Mathematics and Computer Science at the Ben M'Sick Faculty of Sciences in Casablanca since 2015. She is a member of the Information Technology and Modeling Laboratory and leader of the Cloud Computing, Network and Systems Engineering (ICCNSE) team. His research areas include cloud computing, data mining, machine learning, and natural language processing. She can be contacted at email: faouzia.benabbou@univh2c.ma.

Nawal Sael teacher-researcher since 2012, Authorized Professor since 2014 and Professor of Higher Education in the Department of Mathematics and Computer Science at the Ben M'Sick Faculty of Sciences in Casablanca since 2020 and her engineer degree in software engineering from ENSIAS, Morocco, in 2002. Here research interests include data mining, educational data mining, machine learning, deep learning, and internet of things. She can be contacted at email: saelnawal@hotmail.com.

Soukaina Bouhsissin received a B.Sc. degree in mathematical sciences and computer science from the Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, Morocco, in 2018, and an M.Sc. degree in data science andbig data from Hassan II University of Casablanca, in 2020, where she is currently pursuing the Ph.D. degree in computer science. Her research interests include driver behavior classification, intelligent transport systems, machine learning, deep learning, image classification, and temporal series. She can be contacted at email: bouhsissin.soukaina@gmail.com.