ISSN 2007-9737

Non-Intrusive Drowsiness Detection for Accident Prevention Using a Novel CNN

David Hiram Vázquez-Santana, Amadeo José Argüelles-Cruz*

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

{dvazquezs2019, jamadeo}@cic.ipn.mx

Abstract. Drowsiness detection problem is not only complex, but also very important for accident prevention. In this paper, we propose a non-intrusive drowsiness detection method using the right eye and mouth. Face detection is performed using HOG + SVM method and facial features are segmented using 8 landmarks obtained by an ensemble of regression trees and classified using a novel convolutional neural network that we call Dozy-Net. Then, drowsiness detection is carried out using three behavioral parameters: PERCLOS, blink frequency, and yawning duration. Two state-of-the-art and one self-constructed dataset were used to train, test, and compare Dozy-Net's performance against other six state-of-the-art convolutional neural networks, being Dozy-Net significantly faster. Drowsiness detection model was tested on a real-life dataset performing 75.8% accuracy and an average speed of 21.51 FPS. Compared to other existing models, our proposal has the advantage of having been tested in conditions similar to those to be expected in a real environment.

Keywords. Drowsiness detection; convolutional neural network; behavioral measurement; deep learning; computer vision.

1 Introduction

In recent years, our sleep habits have changed negatively due to the amount of time we dedicate to work, hobbies, and transportation. Moreover, the incorporation of new technologies into our lifestyle has also contributed to reducing the time we dedicate to sleep [3].

There are several studies [24, 27, 35, 43] that confirm that between 10% and 13% of the global population frequently suffer from fatigue. Lack of sleep in a minimum of recommended hours [22] drives to several conditions such as drowsiness, irritability, depression, low alertness, anxiety, tension, headache, among others [4, 19, 23, 32]. Drowsiness results in a decreased awareness and both physical and mental performance. It increases the risk of suffering traffic and work accidents [54, 20].

Car accidents tend to be more fatal when caused by drowsiness compared with other causes [13] and employees with sleep problems are 1.62 times more likely to suffer an injury compared to their peers without sleep problems [46].

Social and economic demands have soared in recent years. Nowadays, we spend little time at home and plenty of time at work or in our vehicles, driving from home to work (and vice versa), shopping, going on vacation or visiting other people. In addition, there are several dangerous practices that have become commonly accepted, such as working night shifts or staying up late for leisure after a long workday [15].

Due to our lifestyle, it is important to detect and alert drowsy people to avoid accidents while driving or performing different tasks at work. Nowadays, there are several techniques to detect drowsy people using different techniques [38] based on behavioral parameters, environmental parameters,

ISSN 2007-9737

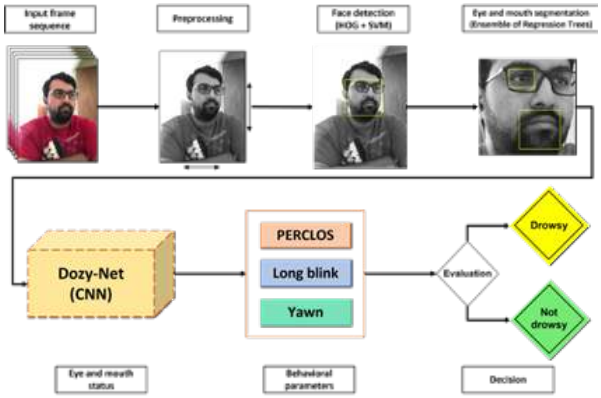


Fig. 1. General diagram of the proposed method for drowsiness detection

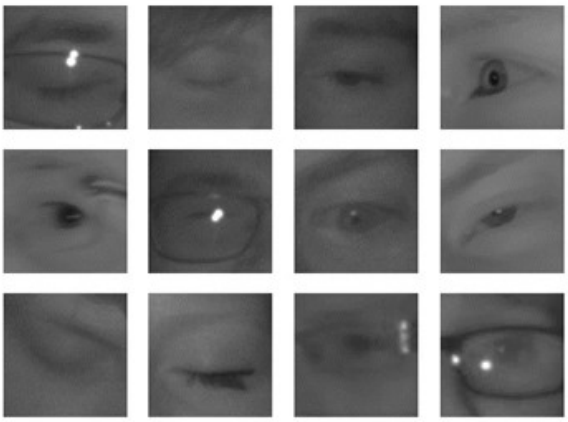


Fig. 2. Sample images from the MRL eye dataset

or physiological parameters. This research will focus on drowsiness detection using behavioral parameters obtained from the right eye and mouth of people in different video sequences.

The aim is to classify everyone into one of two classes: drowsy and non-drowsy. This document is organized as follows: in Section 2, related research and state-of-the-art works will be presented.

Section 3 will introduce the methodology used in this research, along with the datasets used, the novel Convolutional Neural Network (CNN) model used for image classification, and the detailed drowsiness detection algorithm.

Section 4 will present the results obtained when detecting drowsy people. Section 5 will discuss the results from this research, and a comparison between state-of-the-art CNN models and our proposal will be presented. Finally, in Section 6, conclusions and future work will be established.

2 Related Works

Related work will be divided into three parts: those related to object recognition and tracking, those related to face and facial feature detection, and those describing drowsiness detection methods.

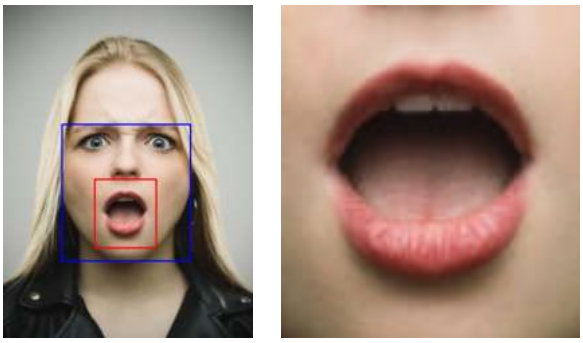
2.1 Object Recognition and Tracking

Object recognition is a fundamental task of computer vision and its goal is to recognize particular instances of a certain class in a more complex image. As one of the fundamental tasks of computer vision, object recognition is involved in many other tasks, such as object description or tracking. A vast literature on object detection is available but, in this paper only a few of the most recent works will be mentioned.

The Histogram of Oriented Gradients (HOG) [12] is a feature descriptor which, when combined with a classifier, such as a support vector machine (SVM), can detect objects. The HOG is still widely used despite being a pre-deep learning approach. For example, a vehicle recognition method using a combination of HOG and SVM was recently presented [36] and also a proposal on diabetic retinopathy detection using HOG, and a k-nearest neighbor classifier (KNN) [21].

With the rise of deep learning, it became possible to create more accurate, precise, faster, and more capable models. In this context, the You Only Look Once model (YOLO) is suitable for object tracking due to its high speed on detecting objects [39]. The evolution of YOLO continues to these days and their different versions have been used in a wide variety of applications.

A model for detecting and counting olive fruit flies [29], a vehicle recognition and tracking system in real time [6] and a real time recognition and tracking people during nighttime based on YOLOv3 [30] are some examples.



(a) Face of a woman **(b)** Extracted mouth region

Fig. 3. Mouth segmentation example

Fig. 4. Sample images from the mouths dataset

2.2 Face Detection

Face detection is a particular case of object recognition. Its aim is to identify faces in an image regardless of the identity of the person to whom it belongs. Face detection is important to develop different applications, such as a model that detects faces through a hybrid model that combines the Haar cascade classifier and a skin detector for developing an automatic video surveillance [17].

New techniques have been developed to improve facial expression recognition, such as feature descriptors based on geometrical moments [44] or reducing the HOG vector corresponding to the eyes and mouth area using the graph signal processing (GSP) [31].

Thanks to Deep Learning techniques, it has become possible to detect faces even under very difficult conditions, as demonstrated in works where the importance lies in the design of a CNN capable of detecting faces with various alterations that make them difficult to detect, such as blurring, noise or low illumination [7].

2.3 Drowsiness Detection

Drowsiness detection methods can be classified into three main categories: techniques based on behavioral parameters [14, 33, 53], environmental parameters [5, 9], or physiological parameters [10, 40, 52].

Environmental parameters-based methods work using data collected by one or more sensors. An example of fatigue detection using this method is the design of a steering wheel that can detect fatigue through eleven features calculated from two driving parameters: steering wheel angle and steering wheel angular velocity.

Both parameters are obtained at a rate of 25 Hz. Steering wheel was tested on ten people (3 women and 7 men) using a driving simulator and the prediction was performed using three models: an SVM, a multilevel ordered logit (MOL) and a neural network. The best accuracy obtained was 74.95% using MOL [9].

The main limitations for using this type of methods are related to the sensors. Installation and cost of the sensors could be high, and external factors can affect the measurements. For example, in driver drowsiness detection, measurements can be affected by pavement or weather conditions.

Physiological fatigue detection methods have been widely used due to the good results obtained with this type of measurements. Electroencephalograms (EEGs) are one example of the physiological measurements and parameters. For example, a work was carried out in the Chinese province of Liaoning, where 16 men took part and EEG signals were obtained through the Emotiv EPOC headset.

The EEGs obtained were processed through a twelve-layer convolutional neural network (CNN) getting an average accuracy of 97.02% [10]. Fatigue detection through physiological parameters is intrusive due to the need of wearable devices, which can cause discomfort in users on everyday activities, making them difficult to implement in vehicles or work scenarios.

Methods based on behavioral parameters have been gaining popularity in recent years due to the capability of non-intrusive drowsiness detection using parameters such as the PERcentage of

Fig. 5. Sample images (extracted frames) from the YawDD dataset

Fig. 6. Sample images (extracted frames) from the UTA-RLDD dataset

eyelid CLOSure (PERCLOS) over the pupil, facial expressions, blinks, head position, yawns, analysis of physiological patterns over time series [28]. One example of drowsiness detection through PERCLOS and yawning is a model which detects drowsiness if within two minutes a person has yawned at least 3 times or if the PERCLOS value is greater than 0.4.

This method was tested using 10 videos from the YawDD dataset [1] plus one video belonging to one of the authors, achieving an accuracy of 95.91% [49]. Despite the high accuracy percentage, few videos were used to test the model.

Another example is a method that uses the mouth and eye region, PERCLOS and Frequency Of Mouth (FOM) parameters and a Multi-tasking Convolutional Neural Network for drowsiness detection.

This method was tested on the YawDDD [1] and NTHU-DDD [50] datasets achieving an accuracy of 98.72% and 98.91% on YawDD and NTHU-DDD datasets, respectively. Although the performance is good, drowsiness detection is not performed in real time, and the datasets do not take into account truly drowsy people [41].

An example of a work [37] that uses modern recurrent neural networks (RNNs) to classify segments of videos from the eyes and then detects drowsiness with an overall accuracy of 82% with RNNs and 95% with convolutional RNNs. A recent research work in which three descriptors are used to extract information from facial images: the histogram of oriented gradients (HOG); the covariance descriptor (COV); and the local binary pattern were used.

The results provided by each of the descriptors are processed by an SVM and the final decision is reached by merging the individual decisions of each one. The model was tested on the NTHU-DDD [50] dataset which is divided into three sets: training set, evaluation set and test set, scoring an accuracy of 79.84% [33].

In the same way, our proposal focuses on a non-intrusive method that uses several behavioral parameters. Our major contribution in this work is a novel and minimalist CNN, called Dozy-Net, that classifies eye blinking to compute PERCLOS index. that same CNN classifies mouth opening to detect if a person is yawning. As a result, a combination of features allows us to detect drowsy people using only video footage of their faces, trying to avoid further accidents in work or everyday life scenarios.

3 Methodology / Proposal

Drowsiness usually exhibits characteristics such as yawning, closed eyes for large periods of time and increased frequency of blinking. This drowsiness detection method is based on three behavioral parameters obtained through video frame analysis: PERCLOS, long blinks and yawns. Considering that the average blink duration is between 174 and 191 milliseconds [34] it is possible to detect them without processing every single frame of the video, thus in this model half of the frames

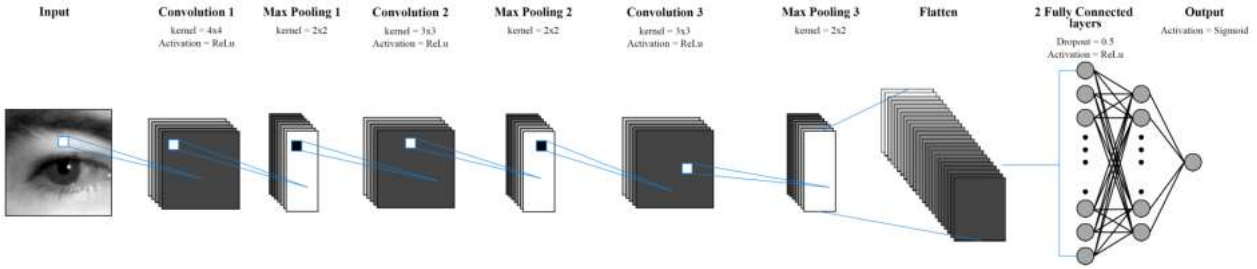


Fig. 7. Dozy-Net model

are processed. For example, in a 10-frame sequence $(f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10})$, frames $(f_1, f_3, f_5, f_7, f_9)$ will be omitted and frames $(f_2, f_4, f_6, f_8, f_{10})$ will be processed. As shown in 1, the face is detected using HOG + SVM then, the right eye and mouth are segmented using 8 landmarks obtained through an ensemble of regression trees [25].

Both extracted images of facial features are resized and normalized before being classified through our custom designed Dozy-Net, to determine their status. Then, drowsiness is detected through three parameters: Yawning, PERCLOS and long blinks. These behavioral parameters are obtained from the number of continuous frames in which the mouth or eye has been open.

The right eye was tracked since the camera was placed at the right in most of the UTA-RLDD videos [18], making easier to detect and track it. PERCLOS is calculated every 200 frames and since not every frame is processed, it is needed to modify the way that the number of frames is counted and how the PERCLOS value is calculated:

$$\text{PERCLOS} = \frac{f_c \cdot 2}{f_t} \cdot 100. \tag{1}$$

PERCLOS is computed according to equation 1, where f_c and f_t are the amount of closed eye frames and total amount of elapsed frames respectively. In this paper, a person is drowsy if the eye remains closed for 7 or more consecutive processed frames, if PERCLOS is higher than 0.07 or if the mouth remains open for 60 or more consecutive frames. If any of these conditions are present, then the person is considered to be in a vigilant state. As seen in 1, the drowsiness detection system is fully described. Each of the components, including the datasets used for this research, will be described below.

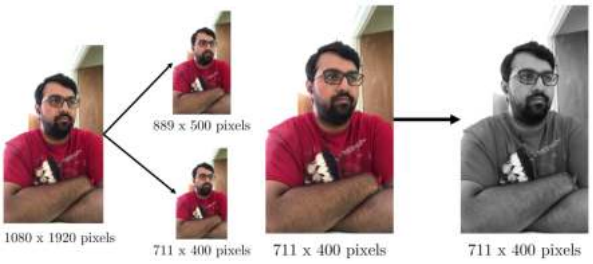
3.1 Datasets

3.1.1 Eyes Dataset

To train our model to classify and detect eye blinking, we have used the MRL Eye Dataset [16]. This dataset contains 84,898 infrared images from 37 persons from which there are 41,945 images of closed eyes and 42,953 of open eyes. 24,001 images are from people that wear glasses. Also, from the total of images, 66,060 do not contain any reflections; 6,129 images contain low reflection levels; and 12,709 contain high reflection levels. From the total of images, 53,630 have poor illumination conditions and the remaining 31,268 have good illumination conditions. MRL Eye Dataset is publicly available at http://mrl.cs.vsb.cz/eyedataset. 2 shows a sample of eye images from the dataset.

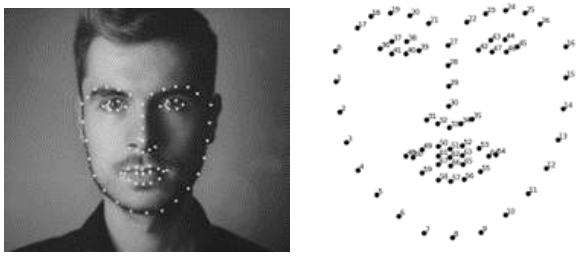
3.1.2 Mouths Dataset

There are no publicly available datasets of yawning people that segment the mouth region. Therefore, we build our own dataset from internet-based images from Google, Yandex, and Getty images. We obtained 5,657 images from people with open and closed mouths.



(a) Frame resizing **(b)** RGB to grayscale

Fig. 8. Image preprocessing



(a) Example image **(b)** Enumerated POI

Fig. 9. POI extraction

Then, we segmented the region of the mount 3 to build the Mouths dataset which contains 2,805 images from open mouths, and 2,852 images from closed mouths. This dataset is publicly available at https://www.kaggle.com/davidvazquezcic/yawn-dataset. 4 shows an example of the images from the dataset.

3.1.3 Yawn Dataset

We used the YawDD dataset [1], to train our model to detect if a person is yawning or is simply opening its mouth for another casual activity (talking, singing). This dataset contains videos from 57 men and 50 women from different ages and different ethnicities.

Length of videos are between 15 and 40 seconds. The class “yawn” contains 58 videos, and the class “no-yawn” contains 82 videos. In addition, it is important to mention that all videos of YawDD dataset are from people “acting”. Therefore, real-life behavior could vary and that is the reason we only used this dataset to detect yawning and not for detecting drowsy people from the “yawn” class. YawDD is publicly available at https://ieee-dataport.org/open-access/yawdd-yawning-detection-dataset. 5 shows an example of the content of the YawDD dataset.

3.1.4 Drowsiness Dataset

In order to evaluate the effectiveness of our proposal, we used the University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD) [18]. This dataset provides videos from tired people that guarantee realness in the expressions of the participants. UTA-RLDD contains RGB videos of approximately 10 minutes and were obtained from web cameras and smartphone cameras.

There are videos from 51 men and 9 women of ages from 20 to 59 years old and different ethnicities. There are a total of 180 videos (60 per class) divided in three classes: alert, low vigilant, and drowsy. UTA-RLDD is publicly available at https://sites.google.com/view/utarldd/home. 6 shows an example of the content of the UTA-RLDD dataset.

3.2 Network Architecture

There are a variety of algorithms capable of determining the state of selected facial features such as associative memories [2] or associative classifiers [47], in this article, a convolutional neural network is used based on the LeNet classical network, Dozy-Net identifies eye and mouth closure by classifying images of these facial features.

Unlike some famous neural network models belonging to families such as EfficientNet or ResNet which can also achieve good results in facial feature image classification, Dozy-Net is considerably more compact and performs facial feature image classification in a significantly shorter time. The complete Dozy-Net architecture is shown in 7. In CNNs, convolution is used to extract features [26]. The input image is considered as a matrix of size MxN and represented as $W_(m,n)$.

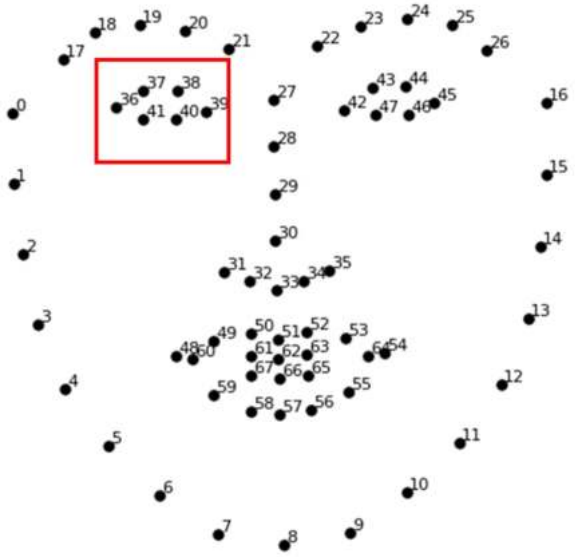


Fig. 10. POI used to extract the right eye

This input is convolved with a kernel $k(p,q)$ of size PxQ. Dozy-Net is formed by three convolutional layers, the number of layers was chosen empirically aiming to reduce the network size as much as possible but preserving feature-extraction capabilities. The kernel size of the first layer is 4x4, while for the second and third layer the kernel size is 3x3.

The number of kernels for the first, second and third layers are 32, 16 and 8 respectively. A one-pixel stride is applied for all convolutional layers. No padding was used. After each convolutional layer, a max-pooling layer with a kernel size of 2x2 was added. After the max-pooling 3rd layer, data is converted to a 1-dimensional array through flattening.

Feature integration is achieved using two fully connected (FC) layers. The first FC layer has 40 neurons, and the second FC layer has 12 neurons. Finally, classification is performed through a sigmoid activation function neuron. To prevent overfitting, we used the Dropout technique [45] with a dropout probability of 50% in the two FC layers.

3.3 Preprocessing and Segmentation

Each frame from the input video is processed according to the following steps:

1. Resize the frame to two different sizes (Display size and learning size, shown in 8a), according to 2.
2. Convert learning size from RGB image format to grayscale image (8b):

$$w' = \left[w \cdot \frac{h'}{h} \right], h' = \left[h \cdot \frac{w'}{w} \right]. \tag{2}$$

Once the preprocessing is complete, we segmented the face from the grayscale image using HOG+SVM. Moreover, once we obtained the face, we used a regression tree-based algorithm [25] contained in the dlib library for Python programming language to get 68 points of interest (POI) from the extracted face 9.

All images are segmented as shown in 9. Therefore, once the POI is obtained, we extract the face features used in this research to detect drowsiness.

3.4 Eye and Blinking Detection

We segmented the right eye from each extracted face through a box that contains the eye, using points 18, 21, 38, and 40 (10). Then, we adjust the height of the box 3 to guarantee that the eye is completely extracted:

$$\begin{aligned} y_1' &= y_1 - 8h \cdot 0.11), \\ y_2' &= y_2 - (h \cdot 0.16), \end{aligned} \tag{3}$$

where y_1 and y_2 correspond to points 38 and 40 respectively and h is the height of the box containing the eye. Once the box is computed, we extract from the input frame the region of the eye, normalize the image and pass it to the Dozy-Net to classify each frame as “open” or “closed”.

We used a threshold of 7 frames (equivalent to approximately 233.8 ms) to determine if a blink was longer than normal [34]. We also use PERCLOS, which is one of the most used behavioral parameters for drowsiness detection [51]. If the PERCLOS value is greater than 0.07, the person is considered to be drowsy.

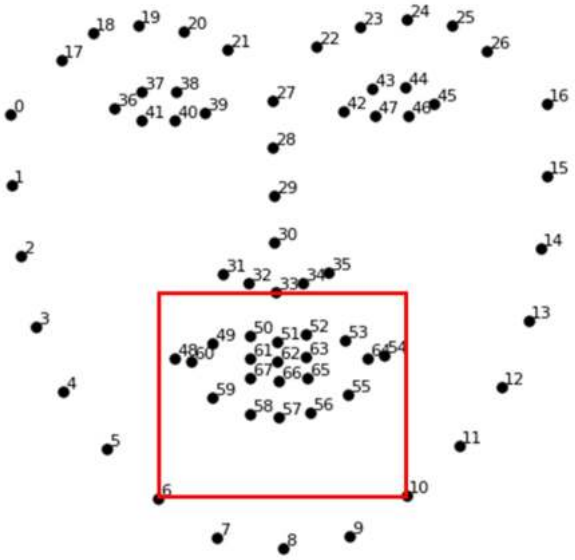


Fig. 11. POI used to extract the mouth

3.5 Mouth and Yawning Detection

Like the eye detection, we segmented the mouth from the face, using points 6, 10 and 33 11. Once again, we pass the normalized extracted image to the Dozy-Net to classify each frame as “open” or “closed”. We used a threshold of 60 continuous frames (equivalent to approximately 2 seconds) to determine if a mouth opening is associated with the sensory yawning peak, which is the most open phase of a yawn [48].

4 Results

4.1 Experimental Framework

All experiments presented in this paper were performed on a PC with Intel i7-9750H processor; 8 GB of RAM; 512 SSD storage and an Nvidia GPU GTX 1650 with 4 GB GDDR5 memory. The drowsiness detection model was built in Python 3.7.0 and the main libraries used were Tensorflow-GPU 2.3.0 and Keras 2.4.3 for CNNs implementation, training, and testing; OpenCV 4.5.1 for video and image processing; and dlib 19.21.1 for face landmarks estimation.

Motivated by the good results that VGG-16, ResNet-50, MobileNet-V1, DenseNet-121, NASNet-Mobile and EfficientNet-B2 are able to achieve in image classification [8, 11] (; ; Fulton et al., 2019; Hong et al., 2021; Kundu et al., 2021; Umair et al., 2021; Wang et al., 2020; Yang et al., 2021), we compared the results obtained by the above networks against Dozy-Net.

To ensure a fair comparison, neither transfer learning nor fine-tuning were used, and synaptic weights were randomly initialized. In addition, we used the same learning hyperparameters for all neural networks except for the input size.

All neural networks were trained using a learning rate of 0.001 and binary cross-entropy as the cost function. In addition, we used Adam as an optimization algorithm with parameters $b_1 = 0.9$, $b_2 = 0.999$, and $\epsilon = 1x10^{-7}$, and we set a batch size of 180 to train the models.

4.2 Validation Method

MRL Eye Dataset and mouth dataset were used to train and test the facial feature classification models. Therefore, they were divided into three sets following the hold-out validation method: training, validation, and testing.

YawDD was used to determine whether our drowsiness detection model can distinguish yawning from other activities such as singing or talking, and two classes of the UTA-RLDD were used to test our drowsiness detection model. 1 shows the number of patterns per set for each dataset, and the type of data used.

4.3 Metrics

Each dataset used in this work is balanced, meaning that there are about the same instances per class. We computed the correctly recognized examples (true positives); the correctly recognized examples that do not belong to each class (true negatives); the examples that were incorrectly assigned to the positive class (false positives); and the examples that were incorrectly assigned to the negative class (false negatives).

Table 1. Sets distribution after hold-out validation method

Dataset	Set	Class 1	Class 2	Examples type
MRL Eye Dataset	Train	24,873	24,539	Images
	Validation	16,582	16,358	
	Test	1,497	1,497	
Mouths dataset	Train	1,611	1,629	Images
	Validation	1,074	1,086	
	Test	120	137	
YawDD	Test	82	58	Videos
UTA-RLDD	Test	60	60	Videos

Table 2. Classification results on the eye dataset for all CNN models

Model	Accuracy			Time per epoch (s)
	Training	Validation	Test	
VGG-16	50.32	50.33	50.00	91.82
ResNet50	98.81	88.37	**95.22**	100.53
DenseNet 121	**98.87**	85.42	94.09	75.25
MobileNet	98.59	89.07	94.26	53.77
NASNet-Mobile	98.81	58.79	87.11	74.10
EfficientNet B2	98.72	88.96	94.49	97.31
Dozy-Net	96.16	**92.26**	95.02	**49.07**

Therefore, we used accuracy (4) to evaluate the overall effectiveness of different classifiers in a binary classification task [42]:

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn}. \tag{4}$$

In addition, we also have evaluated the time per epoch that each model takes when training.

4.4 Eye Detection Results

The MRL Eye dataset is a balanced set. Therefore, accuracy was used to measure the performance of classification. 2 shows the results from the baselines compared with our Dozy-Net. From 2, we can observe that on validation and testing, our model takes the 1st and 2d place respectively. Moreover, there is a substantial difference in training time compared to our proposal that is twice as fast compared to the best result.

4.5 Mouth Detection Results

The own created Mouth dataset is balanced. Therefore, accuracy was also used to measure the performance of classification. 3 shows the results from the baselines compared with our Dozy-Net.

From 3, we can see that our proposal gives us competitive results and it is the fastest among all models. Dozy-Net is up to three times faster than the closer result and up to 2.5 times faster than the best result.

4.6 Drowsiness Results

In this section, we present the results obtained by our drowsiness detection model. We use Dozy-Net for facial feature classification in our drowsiness detection model, since it is the fastest.

The first 11,400 frames of each video were analyzed and PERCLOS value was estimated every 200 frames. Through several tests, we found the following judgment conditions for detecting drowsiness:

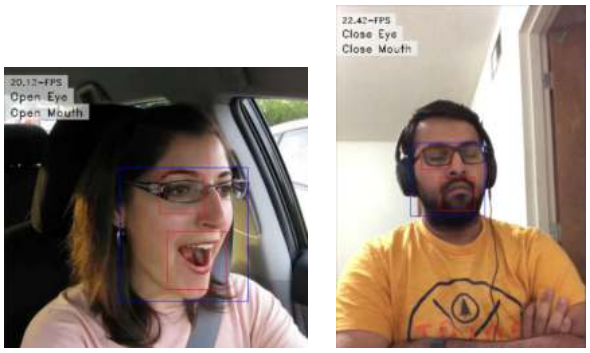
$$\begin{cases} \text{Drowsy}, & (\text{PERCLOS} \geq 0.7) \,||\\ & (F_1 \geq 7\,\text{frames}) \,||\\ & (M_1 \geq 60\,\text{frames})\,, \\ \text{Alert}, & \text{others}, \end{cases} \tag{5}$$

where F_1 and M_1 are the number of consecutive frames processed from closed eye and open mouth, respectively. 12 shows the result of the classification by Dozy-Net of the facial features segmented through the use of an ensemble of regression trees. We achieved an accuracy of 75.8% in the UTA-RLDD with an average speed of 21.51 FPS.

From the confusion matrix, shown in 13, we can observe that of the 60 cases, 42 people were correctly classified. 18 were misclassified. From the 60 cases where people were fully aware (non-drowsy class), 11 were misclassified and 49 correctly classified using our Dozy-Net. Therefore, the most difficult class for our model was the drowsy class.

Table 3. Classification results on the mouths dataset for all CNN models

Model	Accuracy			Time per epoch (s)
	Training	Validation	Test	
VGG-16	50.19	50.28	53.31	14.47
ResNet50	96.32	87.13	96.50	14.56
DenseNet 121	**97.31**	89.62	**98.83**	12.91
MobileNet	96.34	89.83	98.44	5.15
NASNet-Mobile	97.10	53.43	90.66	19.48
EfficientNet B2	96.59	90.59	97.67	13.06
Dozy-Net	95.02	**93.62**	95.33	**4.60**



(a) YawDDD dataset **(b)** UTA-RLDD

Fig. 12. Detection and classification results

4.7 Size Comparison

Compared to baseline models, Dozy-Net is tiny and has very few parameters. 4 shows the characteristics of baseline and Dozy-Net models.

5 Discussion

Our proposal focuses on the detection of drowsiness from fatigued people. We trained our novel Dozy-Net to classify eyes from being open or closed. From 2, we observed our model obtained the best result of accuracy on the validation set with a 92.26% of accuracy. In addition, our proposal achieved the best result on the test set with 95.05% only after the ResNet50 with 95.22%. However, our model achieved that result from a training two times faster than ResNet50 with only 49.07 seconds per epoch, being the fastest among all CNNs.

Apart from eye detection, our methodology required to detect and classify the mouths of the people from being closed or open. From 3, again we can see that our model achieved the best classification on the validation set with a score of 93.62%.

On the test set we achieved fifth place (95.33%) after the DenseNet 121 (98.83%), MobileNet (98.44%), EfficientNet B2 (97.67%), and the ResNet50 (96.50%). We again obtained the faster training with only 4.60 seconds compared with the models with higher scores DenseNet 121 (12.91s), MobileNet (5.15s), EfficientNet B2 (13.06), and the ResNet50 (14.56).

Our proposal was 2.8 times faster than the best model and 3.1 times faster than the ResNet50. Therefore, in order to obtain the fastest model for fast classification, we performed the evaluation of the UTA-RLDD dataset using the Dozy-Net for eye and mouth classification. We evaluated different conditions for the behavioral parameters mentioned, such as PERCLOS, blink duration, and mouth opening.

We found that best results were obtained using the conditions from equation 5, from which we use a PERCLOS rate greater or equal to 0.07 (7%); for blinking, a number greater than 7 continuous frames; and a mouth opening equal or greater than 60 continuous frames. Only one of the above conditions is needed to classify drowsy people. As a result, we achieved an accuracy of 75.8% on the UTA-RLDD dataset.

On the other hand, from 4 there is no doubt that our Dozy-Net, in addition to being the fastest, has the fewest parameters with only 32.7 thousand compared to the 4.2 million (MobileNet-V1) which is the smallest of the reference models. In other words, our proposal has 128 times fewer parameters than the next in size.

On top of that, the storage size of Dozy-Net is 32 times smaller than the MobileNet-V1 with only 0.5MB compared to 16MB. The smaller storage size and fewer parameters makes our proposal more suitable to obtain faster processing rates and faster FPS in a final real-time application. Therefore, our proposal could be easily implemented on mobile and embedded

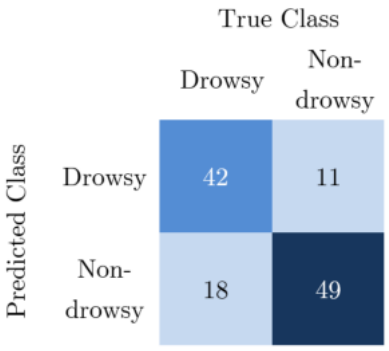


Fig. 13. Confusion matrix for our proposed drowsiness detection model

Table 4. Number of parameters and size of baseline models and Dozy-Net

Model	Model size	Parameters
VGG-16	528 MB	138.3 million
ResNet-50	98 MB	25.6 million
MobileNet-V1	16 MB	4.2 million
DenseNet-121	33 MB	8.1 million
NasNet-Mobile	23 MB	5.3 million
EfficientNet-B2	36 MB	9.1 million
Dozy-Net	**0.5 MB**	**32.7 thousand**

devices with a fraction of processing power compared to a full-size GPU.

6 Conclusions and Future Work

Drowsiness detection is very important for accident prevention. In this paper we introduce a novel drowsiness detection model based on two facial features and a novel compact convolutional neural network architecture suitable for eye and mouth image classification which we named Dozy-Net.

Dozy-Net proved to be competitive by achieving the shortest classification time among all tested models. It achieved 95.02% and 95.33% accuracy in the eye and mouth image classification test set, respectively, with only 32.7 thousand parameters compared to 25.6 million and 8.1 million parameters of ResNet-50 and DenseNet-121, which were the best performing models in eye and mouth classification, respectively. In addition, the size of Dozy-Net is extremely small being 32 times smaller than MobileNet-V1 and 1056 times smaller than VGG-16.

Finally, our drowsiness detection model combines three behavioral parameters obtained from two facial features. Dozy-Net can detect drowsiness at an average speed of 21.51 FPS using an entry-level GPU. We achieved an accuracy of 75.8% on the UTA-RLDD.

This dataset was selected because, compared to the most commonly used datasets in drowsiness detection, the UTA-RLDD contains information from real drowsy people and thus the exposed facial features are more accurate than synthetic datasets. As a result, our proposal is suitable for real-life scenarios and can positively help prevent and avoid different types of accidents in several real-life scenarios.

The contributions of this work are threefold: (1) the presented convolutional neural network model (Dozy-Net) showed its feasibility to identify critical fatigue features in selected facial features, having 128 times fewer parameters than MobileNet-V1 and 4,229 fewer than VGG-16.; (2) the exploration of thresholds and behavioral parameters suitable for non-intrusive drowsiness detection in real scenarios; and (3) a dataset suitable for yawning detection composed of 5,657 images; 2,805 of open mouths and 2,852 of closed mouths.

Based on the results of the study, we propose to increase the processing speed of Dozy-Net by (1) implementing a reflection removal algorithm on the obtained eye images, (2) binarizing the eye and mouth images through a threshold, and (3) implementing multi-thread processing and a queue data structure for frame storage.

Acknowledgments

The authors gratefully acknowledge the Instituto Politécnico Nacional (Secretaría Académica, Comisión de Operación y Fomento de Actividades Académicas, Secretaría de Investigación y Posgrado, Centro de Investigación en Computación, and Centro de Innovación y Desarrollo Tecnológico en Cómputo), the Consejo

Nacional de Ciencia y Tecnología (CONACyT), and Sistema Nacional de Investigadores for their economic support to develop this work.

References

1. **Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., Hariri, B. (2014).** YawDD: A yawning detection dataset. Proceedings of the 5th ACM Multimedia Systems Conference, pp. 24–28. DOI: 10.1145/2557642.2563678.
2. **Acevedo-Mosqueda, M., Yáñez-Márquez, C., López-Yáñez, I. (2007).** Alpha-beta bidirectional associative memories: Theory and applications. Neural Processing Letters, Vol. 26, No. 1, pp. 1–40. DOI: 10.1007/s11063-007-9040-2.
3. **Ahmad-Kamran, M., Naeem-Mannan, M. M., Yung-Jeong, M. (2019).** Drowsiness, fatigue and poor sleep's causes and detection: A comprehensive study. IEEE Access, Vol. 7, pp. 167172–167186. DOI: 10.1109/ACCESS.2019.2951028.
4. **Anaya, F., Abu-Alia, W., Hamoudeh, F., Nazzal, Z., Maraqa, B. (2022).** Epidemiological and clinical characteristics of headache among medical students in palestine: A cross sectional study. BMC Neurology, Vol. 22, No. 1, pp. 1–8. DOI: 10.1186/s12883-021-02526-9.
5. **Arefnezhad, S., Samiee, S., Eichberger, A., Frühwirth, M., Kaufmann, C., Klotz, E. (2020).** Applying deep neural networks for multi-level classification of driver drowsiness using vehicle-based measures. Expert Systems with Applications, Vol. 162, pp. 113778. DOI: 10.1016/j.eswa.2020.113778.
6. **Azimjonov, J., Özmen, A. (2021).** A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways. Advanced Engineering Informatics, Vol. 50, pp. 101393. DOI: 10.1016/j.aei.2021.101393.
7. **Ben-Fredj, H., Bouguezzi, S., Souani, C. (2021).** Face recognition in unconstrained environment with CNN. The Visual Computer, Vol. 37, No. 2, pp. 217–226. DOI: 10.1007/s00371-020-01794-9.
8. **Chaddad, A., Kucharczyk, M. J., Desrosiers, C., Okuwobi, I. P., Katib, Y., Zhang, M., Rathore, S., Sargos, P., Niazi, T. (2020).** Deep radiomic analysis to predict gleason score in prostate cancer. IEEE Access, Vol. 8, pp. 167767–167778. DOI: 10.1109/ACCESS.2020.3023902.
9. **Chai, M., Li, S. W., Sun, W. C., Guo, M. Z., Huang, M. Y. (2019).** Drowsiness monitoring based on steering wheel status. Transportation Research Part D: Transport and Environment, Vol. 66, pp. 95–103. DOI: 10.1016/j.trd.2018.07.007.
10. **Chen, J., Wang, S., He, E., Wang, H., Wang, L. (2021).** Recognizing drowsiness in young men during real driving based on electroencephalography using an end-to-end deep learning approach. Biomedical Signal Processing and Control, Vol. 69, pp. 102792. DOI: 10.1016/j.bspc.2021.102792.
11. **Chen, J., Zhang, D., Suzauddola, M., Zeb, A. (2021).** Identifying crop diseases using attention embedded MobileNet-v2 model. Applied Soft Computing, Vol. 113, pp. 107901. DOI: 10.1016/j.asoc.2021.107901.
12. **Dalal, N., Triggs, B. (2005).** Histograms of oriented gradients for human detection. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, pp. 886–893. DOI: 10.1109/CVPR.2005.177.
13. **Davidović, J., Pešić, D., Lipovac, K., Antić, B. (2020).** The significance of the development of road safety performance indicators related to driver fatigue. Transportation Research Procedia, Vol. 45, pp. 333–342. DOI: 10.1016/j.trpro.2020.03.024.
14. **Elamrani-Abou-Elassad, Z., Mousannif, H., Al-Moatassime, H., Karkouch, A. (2020).** The application of machine learning

techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence, Vol. 87, pp. 103312. DOI: 10.1016/j.engappai.2019.103312.

15. **Eurofound and International Labour Organization (2019).** Working conditions in a global perspective. op.europa.eu/en/publication-detail/-/publication/19767879-917b-11e9-9369-01aa75ed71a1/language-en. (Accessed on 03/19/2024).
16. **Fusek, R. (2018).** Pupil localization using geodesic distance. Proceedings of the Advances in Visual Computing: 13th International Symposium, ISVC 2018, Springer International Publishing, Cham, pp. 433–444. DOI: 10.1007/978-3-030-03801-4_38.
17. **Gautam, K. S., Thangavel, S. K. (2019).** Video analytics-based intelligent surveillance system for smart buildings. Soft Computing, Vol. 23, No. 8, pp. 2813–2837. DOI: 10.1007/s00500-019-03870-2.
18. **Ghoddoosian, R., Galib, M., Athitsos, V. (2019).** A realistic dataset and baseline temporal model for early drowsiness detection. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society, pp. 178–187. DOI: 10.1109/CVPRW.2019.00027.
19. **Griggs, S., Harper, A., Hickman, R. L. (2022).** A systematic review of sleep deprivation and neurobehavioral function in young adults. Applied Nursing Research, Vol. 63, pp. 151552. DOI: 10.1016/j.apnr.2021.151552.
20. **Hanifah, M. S. A., Ismail, N. (2020).** Fatigue and its associated risk factors: A survey of electronics manufacturing shift workers in malaysia. Fatigue: Biomedicine, Health & Behavior, Vol. 8, No. 1, pp. 49–59. DOI: 10.1080/21641846.2020.1739806.
21. **Hatua, A., Subudhi, B. N., Veerakumar, T., Ghosh, A. (2021).** Early detection of diabetic retinopathy from big data in hadoop framework. Displays, Vol. 70, pp. 102061. DOI: 10.1016/j.displa.2021.102061.
22. **Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Adams-Hillard, P. J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M. V., Ware, J. C. (2015).** National sleep foundation's updated sleep duration recommendations: Final report. Sleep Health, Vol. 1, No. 4, pp. 233–243. DOI: 10.1016/j.sleh.2015.10.004.
23. **Hudson, A. N., Van-Dongen, H. P. A., Honn, K. A. (2020).** Sleep deprivation, vigilant attention, and brain function: A review. Neuropsychopharmacology, Vol. 45, No. 1, pp. 21–30.
24. **Jason, L. A., Evans, M., Brown, M., Porter, N. (2010).** What is fatigue? pathological and nonpathological fatigue. PM&R, Vol. 2, No. 5, pp. 327–331. DOI: 10.1016/j.pmrj.2010.03.028.
25. **Kazemi, V., Sullivan, J. (2014).** One millisecond face alignment with an ensemble of regression trees. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874. DOI: 10.1109/CVPR.2014.241.
26. **Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998).** Gradient-based learning applied to document recognition. Proceedings of the IEEE, Vol. 86, No. 11, pp. 2278–2324. DOI: 10.1109/5.726791.
27. **Loge, J. H., Ekeberg, O., Kaasa, S. (1998).** Fatigue in the general norwegian population: Normative data and associations. Journal of Psychosomatic Research, Vol. 45, No. 1, pp. 53–65. DOI: 10.1016/S0022-3999(97)00291-2.
28. **López-Yáñez, I., Sheremetov, L., Yáñez Márquez, C. (2014).** A novel associative model for time series data mining. Pattern

Recognition Letters, Vol. 41, No. C, pp. 23–33. DOI: 10.1016/j.patrec.2013.11.008.

29. **Mamdouh, N., Khattab, A. (2021).** YOLO-based deep learning framework for olive fruit fly detection and counting. IEEE Access, Vol. 9, pp. 84252–84262. DOI: 10.1109/ACCESS.2021.3088075.
30. **Manssor, S. A. F., Sun, S., Elhassan, M. A. (2021).** Real-time human recognition at night via integrated face and gait recognition technologies. Sensors, Vol. 21, No. 13. DOI: 10.3390/s21134323.
31. **Meena, H. K., Joshi, S. D., Sharma, K. K. (2021).** Facial expression recognition using graph signal processing on hog. IETE Journal of Research, Vol. 67, No. 5, pp. 667–673. DOI: 10.1080/03772063.2019.1565952.
32. **Meerlo, P., Sgoifo, A., Suchecki, D. (2008).** Restricted and disrupted sleep: Effects on autonomic function, neuroendocrine stress systems and stress responsivity. Sleep Medicine Reviews, Vol. 12, No. 3, pp. 197–210. DOI: 10.1016/j.smrv.2007.07.007.
33. **Moujahid, A., Dornaika, F., Arganda-Carreras, I., Reta, J. (2021).** Efficient and compact face descriptor for driver drowsiness detection. Expert Systems with Applications, Vol. 168, pp. 114334. DOI: 10.1016/j.eswa.2020.114334.
34. **Navascues-Cornago, M., Morgan, P. B., Maldonado-Codina, C., Read, M. L. (2020).** Characterisation of blink dynamics using a high-speed infrared imaging system. Ophthalmic and Physiological Optics, Vol. 40, No. 4, pp. 519–528. DOI: 10.1111/opo.12694.
35. **Patel, V., Kirkwood, B., Weiss, H., Pednekar, S., Fernandes, J., Pereira, B., Upadhye, M., Mabey, D. (2005).** Chronic fatigue in developing countries: Population based survey of women in india. BMJ (Clinical research ed.), Vol. 330, No. 7501, pp. 1190. DOI: 10.1136/bmj.38442.636181.E0.
36. **Qu, Z., Chen, Z. (2021).** An intelligent vehicle image segmentation and quality assessment model. Future Generation Computer Systems, Vol. 117, pp. 426–432. DOI: 10.1016/j.future.2020.12.002.
37. **Quddus, A., Zandi, A. S., Prest, L., Comeau, F. J. (2021).** Using long short term memory and convolutional neural networks for driver drowsiness detection. Accident Analysis & Prevention, Vol. 156, pp. 106107. DOI: 10.1016/j.aap.2021.106107.
38. **Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., Mahmood, A. (2019).** A survey on state-of-the-art drowsiness detection techniques. IEEE Access, Vol. 7, pp. 61904–61919. DOI: 10.1109/ACCESS.2019.2914373.
39. **Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016).** You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, pp. 779–788. DOI: 10.1109/CVPR.2016.91.
40. **Rundo, F., Rinella, S., Massimino, S., Coco, M., Fallica, G., Parenti, R., Conoci, S., Perciavalle, V. (2019).** An innovative deep learning algorithm for drowsiness detection from EEG signal. Computation, Vol. 7, No. 1, pp. 13. DOI: 10.3390/computation7010013.
41. **Savaş, B. K., Becerikli, Y. (2020).** Real time driver fatigue detection system based on multi-task ConNN. IEEE Access, Vol. 8, pp. 12491–12498. DOI: 10.1109/ACCESS.2020.2963960.
42. **Sokolova, M., Lapalme, G. (2009).** A systematic analysis of performance measures for classification tasks. Information Processing & Management, Vol. 45, No. 4, pp. 427–437. DOI: 10.1016/j.ipm.2009.03.002.
43. **Son, C. G. (2012).** Review of the prevalence of chronic fatigue worldwide. The Journal of Korean Medicine, Vol. 33, No. 2, pp. 25–33.
44. **Sossa-Azuela, J., Yáñez-Márquez, C., de-León-S, J. (2001).** Computing geometric moments using morphological erosions.

Pattern Recognition, Vol. 34, No. 2, pp. 271–276. DOI: 10.1016/S0031-3203(99)00213-7.

45. **Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014).** Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, Vol. 15, No. 1, pp. 1929–1958.
46. **Uehli, K., Mehta, A. J., Miedinger, D., Hug, K., Schindler, C., Holsboer-Trachsler, E., Leuppi, J. D., Künzli, N. (2014).** Sleep problems and work injuries: A systematic review and meta-analysis. Sleep Medicine Reviews, Vol. 18, No. 1, pp. 61–73. DOI: 10.1016/j.smrv.2013.01.004.
47. **Villuendas-Rey, Y., Rey-Benguría, C., Ferreira-Santiago, Á., Camacho-Nieto, O., Yáñez-Márquez, C. (2017).** The Naïve associative classifier (NAC): A novel, simple, transparent, and accurate classification model evaluated on financial data. Neurocomputing, Vol. 265, pp. 105–115. DOI: 10.1016/j.neucom.2017.03.085.
48. **Walusinski, O., Deputte, B. L. (2004).** Le bâillement: phylogenèse, éthologie, nosogénie. Revue Neurologique, Vol. 160, No. 11, pp. 1011–1021. DOI: 10.1016/S0035-3787(04)71138-8.
49. **Wang, Y., Qu, R. (2021).** Research on driver fatigue state detection method based on deep learning. Proceedings of the 2020 International Conference on Mechanical Automation and Computer Engineering (MACE 2020), Vol. 1744, No. 4, pp. 042242. DOI: 10.1088/1742-6596/1744/4/042242.
50. **Weng, C. H., Lai, Y. H., Lai, S. H. (2017).** Driver drowsiness detection via a hierarchical temporal deep belief network. Computer Vision – ACCV 2016 Workshops, Springer International Publishing, Cham, pp. 117–133. DOI: 10.1007/978-3-319-54526-4_9.
51. **Wierwille, W. W., Wreggit, S. S., Kirn, C. L., Ellsworth, L. A., Fairbanks, R. J. (1994).** Research on vehicle-based driver status/performance monitoring: Development, validation, and refinement of algorithms for detection of driver drowsiness. final report. Research on vehicle-based driver status/performance monitoring : development, validation, and refinement of algorithms for detection of driver drowsiness. Final report, pp. 247.
52. **Wu, E. Q., Deng, P., Qiu, X., Tang, Z., Zhang, W., Zhu, L., Ren, H., Zhou, G., Sheng, R. S. F. (2021).** Detecting fatigue status of pilots based on deep learning network using EEG signals. IEEE Transactions on Cognitive and Developmental Systems, Vol. 13, No. 3, pp. 575–585. DOI: 10.1109/TCDS.2019.2963476.
53. **Yu, J., Park, S., Lee, S., Jeon, M. (2019).** Driver drowsiness detection using condition-adaptive representation learning framework. IEEE Transactions on Intelligent Transportation Systems, Vol. 20, No. 11, pp. 4206–4218. DOI: 10.1109/TITS.2018.2883823.
54. **Zhipeng Peng, H. Z., Wang, Y. (2021).** Work-related factors, fatigue, risky behaviours and traffic accidents among taxi drivers: A comparative analysis among age groups. International Journal of Injury Control and Safety Promotion, Vol. 28, No. 1, pp. 58–67. DOI: 10.1080/17457300.2020.1837885. PMID: 33108968.

Article received on 16/05/2024; accepted on 01/07/2024.
**Corresponding author is Amadeo José Argüelles-Cruz.*