This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3173004, IEEE Transactions on Industrial Informatics

Drowsiness detection in the era of Industry 4.0: are we ready?

Carmen Bisogni, *Member IEEE,* Fei Hao, Vincenzo Loia, *Senior Member IEEE*, and Fabio Narducci, *Member IEEE*

***Abstract*—Interconnectivity and smart automation of Internet of Things in recent times have led to the concept of Industry 4.0. Together with the improvement in productivity and new business models, employment conditions should take advantage of these new technologies. Safety in workplace is one of the most sensitive topics on matters, that needs targeted and accurate solutions. The safety can be guaranteed by investigating the attention states of the workers and, in particular, their drowsiness levels. Several technologies have faced this problem by using biometrics, but how many of them are applicable in a real-case-use scenario of Industry 4.0? This work aims to answer this question by discussing available data and methods that can be used in specific workplaces. We highlight their limitations and accuracy to sketch out the recent literature which may contribute to worker safety in Industry 4.0. Finally, we point out a gap that needs to be filled in order to implement these strategies on a large scale.**

***Index Terms*—Industry 4.0, Biometrics, Drowsiness, Workplace safety, commercial driving.**

I. INTRODUCTION

ALERT states in humans are mainly related to drowsiness. Drowsiness is a state of low vigilance associated with several physical and behavioral biometrics. When a user is less alert, sleepy, or falling asleep, his physical state can be detected by analyzing Electrocardiogram (ECG), Electromyogram (EMG), Skin conductance (GSR), and respiratory signals

On the other hand, the behavior of the subject changes when in a drowsy state. Eye blink has shown its effectiveness in the detection of drowsiness, in particular by studying frequency, duration, eyelid speed, and percentage of eye closure. Other behavioral signs of drowsiness are the variation in the yawning number and the anomalies in pupil dilatation. It is clear that drowsiness detection finds its applicability in safety. From recent literature, it can be observed that the higher accuracy in detecting drowsiness is provided by physiological parameter-based techniques [1]. However, those techniques are more invasive since the majority of sensors are wearable. Since one of the more thriving application fields of drowsiness detection is the driver scenario, By combining data provided by those sensors with other information such as vehicular data (speed, acceleration, etc.), it is possible to provide the alert necessary to avoid car accidents. If in the case of the driver scenario, this variety of sensors is available and contact sensors can still be tolerated by the users, the application to workplace safety is quite different. Drowsiness and workplace safety are strongly related, but the scenario settings are different from driving. It has been demonstrated that over-sleeping sleepy employees are 70% more likely to be involved in workplace accidents with respect to colleagues that have no sleep deprivation [2]. In some employments, as medical residents and shift workers, the drowsiness of the user becomes a problem for the others, other than the user itself. For example, a study of 100 nurses revealed a number of mathematical errors in annotations for drug administrations that were 32% higher during the night shift [3].

Inspired by those observations, we want to present, in this work how modern drowsiness detection technologies can be applied for workplace safety in Industry 4.0. The increase of automated machines and robots in the industry leads to new frontiers of investigation for safety. As an example, it has been demonstrated that robots able to detect human fatigue and adapt to the latter can improve the quality of the cooperative work [4]. The aim is to correctly manage safety to avoid new devices causing more problems than they solve [5]. From a biometrics point of view, we will investigate:

- available datasets and their limitations to be used in drowsiness detection applications for Industry 4.0;
- accuracies and computational costs of recent drowsiness detection methods for the discussed scenarios;
- the link between accuracies and subject-dependent architectures.

By analyzing the above-mentioned aspects, we will draw considerations about the real applicability of recent drowsiness techniques to workplace safety.

The remaining of this paper is organized as follows: In Section II we present all the available data that can be applied to solve the drowsiness classification problem in this scenario. In Section III all the recent literature is presented with details about the data involved, their accuracy, and computational costs. In IV we investigated the dependency of the architectures by the subjects. In Section V we discuss some challenges on this topic. Finally, in VI we have drawn our conclusions about existing solutions and open challenges.

Carmen Bisogni and Fabio Narducci are with the Department of Computer Science, University of Salerno, Salerno, Italy, 84084 e-mail:cbisogni, fnarducci@unisa.it

Fei Hao is with the School of Computer Science, Shaanxi Normal University, Xi'an 710119, China. Email:fhao@snnu.edu.cn

Vincenzo Loia is with the Dipartimento di Scienze Aziendali - Management Innovation Systems, University of Salerno, 84084 Fisciano (SA), Italy.

Manuscript received ; revised

II. DROWSINESS DETECTION DATA

Since drowsiness can be detected from a wide variety of data, we introduce here the most popular ones, discussing their usability in the workplace scenario.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3173004, IEEE Transactions on Industrial Informatics

A. Hidden biometrics data

The hidden biometrics data more used in drowsiness is undoubtedly electroencephalographic (EEG). The motivations behind this choice are mainly due to the high accuracy reached by biometrics. In this area, we can find both homemade datasets that reflect the innovativeness of acquisition devices and consolidated ones. Some of them, like the MIT-BIH Polysomnographic Database [6] used in the work of [7], are well-known datasets in the study of sleep disorders with a wide set of annotations. However, even if in [7], its applicability can be demonstrated to detect drowsiness, the real use of those data, collected two decades ago, can not be compared to recent devices that are less invasive on one hand, but that can lose accuracy on the other. The same happens for other works that use homemade data recorded with devices that are not applicable in real environment.

Devices as proposed in [8] are more utilizable since they are wearable, like a head band, and they are connected to a smartwatch that demonstrates the user's drowsiness state. However, here, the ground truth of the collected EEG in terms of drawsiness does not come from other EEG signals, but from PERCLOS, [8] that subdivided the alert state from eyelid closure.

For those reasons, even if accurate, that kind of data is not usable in a workplace where the user should be free to move and, eventually, to walk.

Due to new devices, a physical parameter more usable is Heart Rate Variability (HRV). HRV is not only an indicator of the cardiorespiratory system but also of the activity in the sympathetic and parasympathetic nervous systems. The more reliable way to detect this data is by EEG, as proposed in several papers we will introduce later. However, this kind of acquisition has the same limitations in applicability as the ones previously introduced. Even when data recorded comes from a realistic scenario, as used in the work of [9], sensors are quite invasive. As an example, the dataset proposed in [10]is particularly reliable because the subjects collected were professional drivers, 13 in total, with an age between 29 and 56, and more than 100 hours of driving have been detected. They recorded various paths. However, the monitoring system is quite invasive. It consists of biomedical monitors using an inductive band located at the middle of the trunk above the diaphragm to study the HRV and four EEG single electrodes located in the vertex zone of the cranium. It is remotely possible that all of the professional/commercial drivers can install those kinds of invasive sensors on their vehicles. In addition, since sensors are invasive, and in the proposed study dataset the subjects were asked to drive for 8 hours, it is possible that the sensors are a cause of stress for the drivers. This demonstrates that even in conditions in which data is collected in a realistic scenario, the sensors are far from being applicable on a huge scale in a driving scenario.

B. Behavioural biometrics data

One of the most commonly used behavioural data sets is represented by the eye blink. The eye blink can be spontaneously, voluntarily, or in response to an external stimulus [11]. During the majority of blinks, the phase in which the eye is closed lasts a few milliseconds. However, the closed phases last longer when the user is drowsy. A recent dataset, named ”mEBAL” (Multimodal database for Eye Blink detection and Attention Level estimation) [12] has been developed with the aim of using multiple sources to evaluate drowsiness. EEG is here registered in parallel with Near Infrared (NIR) and RGB cameras. There are in total 38 participants and 6,000 samples, each sample of 21 frames for a total number of images of 756,000. Here, as the opposite of PERCLOS, the EEG has been used to correctly classify blinks and provide ground truth. In the same work, the author provided a method to perform blink detection. For the application of mEBAL in the workplace scenario, we can consider two main limitations. The subjects are all students, for this reason we expected that the age range is limited. Since the blink number and duration also vary with age, this represents a limitation. The subjects were registered when seated, in a static position when performing actions that only involved writing. This could make this dataset ideal if the application field involves a workplace in which it is proper to monitor attention during filling out documents, or solving other tasks.

The limited position during registration is common to other datasets as well, as the University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD) [13] where subjects sit in front of a camera and see a video. However, here, the subjects are passive instead of active, and the different levels of drowsiness were labeled based on the moment of the day in which the data was collected. In total, here we have around 30 hours of RGB videos of 60 healthy participants for each of three different classes: alertness, low vigilance, and drowsiness. Each video lasts about 10 minutes and is registered at 30fps, for a total of more than 3 million frames. Even if the number of participants is higher, the action performed is the same as in each recording, i.e., seeing a video reproduced on a display. Also in this case, the limitation is the same as in the previous case, but UTA-RLDD is best suited for workplace surveillance monitoring since the subjects are passive.

To solve the problem of a more dynamic scenario, we can find the HUST-LEBW dataset [14], which is a collection of 20 commercial movies. Even if it is in the wild, the drowsiness annotation is missing, and only eyeblink information is available. It was mainly developed to detect blinks instead of detecting drowsiness. However, it could be useful for clustering methods or more generally for methods that do not need labeled data.

A more dynamic kind of data is represented by the video collected under a driving scenario. Those kinds of videos and the correspondent drowsiness detection have already been studied in the case of commercial drivers as introduced in [15]. The data here is from the DMS vendor database, [16] which monitored 70 professional drivers from different companies. However, this database only provides the number of alerts generated, and not the raw information about the blink or other behavioural signals that led to the alert.

The NTHU Computer Vision Lab dataset [17] is one of the only available public datasets on the driving scenario that contains a large number of frames. Here there are 36 subjects recoded with and without glasses, during normal driving,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3173004, IEEE Transactions on Industrial Informatics

yawning, slow blink rate, falling asleep, bursts of laughter, etc. In total, there are more than 700k frames to perform the training, more than 100k frames to perform the validation and more than 700k frames to perform the test. The label is binary, e.g. drowsy or not drowsy.

For the dataset we introduced in the previous section, the HRV Dataset, we have added the availability of the IR Driving video. For this reason, some of its characteristics fall into this section. The authors classified the subjects in the video as: attentive if they are fast reactive to road events, with good lateral and longitudinal control; fatigued when reactions are slower, yawns are present, as well as large body movements; and drowsy if there is less attention to the reading, driving errors, and loss of facial expressivity.

Another recent dataset based on driving videos has been proposed in [18]. The Invedrifac dataset is composed of 30 subjects, all males and professional drivers, between 20 and 40 years old. The videos have been recorded during both day and night, in RGB and NIR. Each video lasts between 15 and 20 minutes and is registered at 30fps.

To the best of our knowledge, and from the work of the authors, there is no other reliable database on this field that is publicly available. For this reason, all the homemade data used in other works will be discussed directly in the architecture description to enhance their specific contribution to the literature.

In table I we summarized the available datasets to perform drowsiness detection together with the application field of Industry 4.0 in which they well fit.

TABLE I
DATASETS AVAILABLE TO PERFORM DROWSINESS DETECTION AND THEIR APPLICATION TO INDUSTRY 4.0. N/A IS FOR NOT AVAILABLE IN CASE OF INFORMATION AND NOT APPLICABLE IN CASE OF APPLICATIONS.

Database	Year	Data	Modality	App. Ind. 4.0
MIT-BIH [6] Polysomno-graphic	2000	EEG	Sleep monitoring	N/A
HRV Dataset [10]	2011	EGG+EOG +IR video	Driving	Commercial Driving
mEBAL [12]	2020	EEG+RBG +IR Videos	Seated writing	Active static workplace
UTA-RLDD [13]	2019	RGB Videos	Seated watching	Passive static workplace
HUST-LEBW [14]	2019	RGB Movies	Various/ Commercial Movies	Various (no drowsiness annotations)
DMS Driving Database [16]	N/A	RGB Videos	Driving	Commercial Driving (to buy)
NTHU Computer Vision Lab [17]	2016	RGB Videos	Driving	Commercial Driving
Invedrifac [18]	2019	RGB+IR Videos	Driving	Commercial Driving

In Figure 1 there are some examples of available datasets. For MIT-BIH and HRV we can not shown images, in the first case because there are no video data, in the second one for privacy.

III. DROWSINESS DETECTION ARCHITECTURES

In contrast to the large number of publicly available datasets, there is a large number of architectures well described in the literature. To enhance the contribution of works that are applicable in Industry 4.0, we will exclude from this analysis work that uses invasive wearable devices to collect and analyse data. As we discussed in the previous section, those works are not applicable to the selected scenario. The majority of the remaining works use homemade datasets, for this reason their results should be evaluated by taking into account the amount of data and their heterogeneity, other than performances and computational time, as usual.

To highlight the contribution of those works to Industry 4.0 scenarios, we will classify algorithms in terms of application fields. As in the case of datasets, we can classify them as: commercial driving scenarios; passive static workplaces; and active static workplaces.

A. Commercial driving scenarios

In terms of datasets and literature, the commercial driving scenario is the most thriving of the three we will present. This can be explained on one hand by the recent developments in smart devices for cars and on the other hand by the number of car accidents that are caused by a driver's drowsiness. These can also be very serious and involve several people.

The focus of those methods is their real-time applicability. In [19], to reach this aim, an offline and an online modules have been developed. A deep cascaded convolutional neural network (DCCNN) detects the face, and then the Dlib library detects the landamrks. The Eyes Aspect Ratio (EAR) is computed, and then a Support Vector Machine (SVM) classifies the drowsiness state. The homemade dataset to perform the test is composed of 50 videos, each of two minutes. The subjects are only 4 captured at 30fps.

Recent applications involve smartphones alerting the driver and/or passengers about the driver's drowsiness, as proposed in [18]. Here, three stages are involved: PERCLOS is computed from the front camera of the smartphone; acquires speech data and computes the voice to unvoiced ratio (VUR) if the PERCLOS threshold is reached; proposes a reaction time test to obtain the final result. The three stages make the system more reliable. The tests have been performed on the Invedrifac dataset.

In[20] multimodal information as the posture of the driver, the blinks, the vehicular information, and HRV are studied to detect both slight and severe drowsiness. The model used to classify those indices is the Radial Basis Function-kernel based-support vector regression. The results are presented over a homemade dataset of 49 participants in a virtual road environment, at 60 fps videos.

In [17] a hierarchical temporal Deep Belief Network (HT-DBN) is used to detect drowsiness states in drivers by using both high-level facial features and head features. Hidden Markow Models, in particular two, are here used in the network to highlight relations between eyes, mouth and head movements. The dataset used here is from the NTHU ComputerVision Lab.

A Neural Network architecture is also used in [21] where motions are analyzed by spatio-temporal representation learning. Here, the interesting additional step the authors performed is an automatic understanding of the scene. An optimization algorithm obtains the balance between drowsiness detection

Fig. 1. Drowsiness detection available datasets. (a) HUST-LEBW, (b) DMS Driving, (c) NTHU Computer Vision Lab, (d) mEBAL, (e) UTA-RLDD, (f) Invedrifac

and scene understanding. Also here, the dataset used is the NTHU ComputerVision Lab. For this reason, the results are fully comparable with those of the previous method.

Spatio-temporal representations are also used in [22], where the framework proposed is the resultant of a hybridization between 3D conditional generative adversarial network and two-level attention bidirectional long short-term memory network (3DcGAN-TLABiLSTM). The first part is useful for capturing short-term information and the second part is for learning long-term dependencies. Finally, they put in place temporal smoothing to refine the prediction. Also, in this case, the dataset used is the NTHU ComputerVision Lab.

The high performances of hybridization were also involved in [23], a very recent publication. Here, the system tracks the vehicle statistics to provide more accurate results. The fusion has been put in place by emotion detection based on SVM, which combines information obtained by yawning, eye movements, and lip gestures. Even if the overall idea seems promising, there is not a well-defined dataset on which it was tested. There are no information about the amount of video or subjects, for this reason, the accuracy obtained from the authors is not comparable with other techniques on a real case-use.

Distinguish drowsiness by events is involved in [24]. The driver's eyes and mouth motions are analyzed from the event streams to obtain event-based drowsiness-related features. The novelty of this method is the use of neuromorphic vision sensors that capture only the motion in the scene as an asynchronous sequence of events. The authors also proposed a dataset called EDDD using this sensor. However, even if it should be public, it appears to be no longer available online.

In Table II we presented all the methods applicable in a commercial driving scenario. Considering the number of subjects involved, the work in [20] is more complete. However, the dataset is not publicly available, the F1 score is provided but not the accuracy, and speed is not provided. The works using NTHU are the second ones to use a large set of subjects. Between them the best performances in terms of both accuracy and speed have been reached by the work proposed in [22], which is also the most recent one using NTHU.

If we focus on the most realistic application in performances of Table II, we can claim Invedrifac to have the more realistic scenario. For this reason, in this case, the work in [18] is the more reliable from this point of view.

TABLE II
PERFORMANCES AND DATA DESCRIPTION OF METHODS IN THE COMMERCIAL DRIVING SCENARIOS. HM IS FOR HOMEMADE AND HM-M IS FOR HOMEMADE-MULTIMODAL.

Method	Data	#Subj	#Vid	Acc(%)	F-1(%)	fps
[17] (2016)	NTHU	36	360	84.82	84.79	20
[21] (2019)	NTHU	36	360	76.2	76.25	38.1
[19] (2019)	HM	4	50	94.8	N/A	20.1
[18] (2019)	Invedrifac	30	N/A	93.33	94.11	14.67
[20] (2020)	HM-M	49	49	N/A	56.5	N/A
[22] (2020)	NTHU	36	360	91.2	N/A	39.21
[24] (2020)	HM	26	260	96.2	N/A	0.1

B. *Passive static workplaces*

In Passive Static Workplace, the subjects are mainly involved for surveillance purposes. They have to maintain a level of attention adequate to the details they have to observe. In this section, we also refer to remote learning algorithms since they well fit the problem in which a great effort is required not only to maintain attention but also to understand the lesson. Drowsiness assumes here a more broad concept of attention, which can be lost not only by drowsy users but also by fatigued ones.

Recently, in [25], this topic has been investigated. Here, a machine learning technique for detecting expressions and a counter are added to the method to detect the number of yawns. The participants involved are 20 students and 30 computer scientists. With a questionnaire, the indices detected are compared with the answers of the subjects. This study has been presented more focused on differences between the perception of drowsiness by students and professors and the collected indices. Accuracy is not available.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3173004, IEEE Transactions on Industrial Informatics

For the same kind of data, there is the work in [26]. Here the authors focused on body poses extracted by OpenPose. A convolutional neural network uses the keypoints detected to classify the attentivity state. Five classes have been contemplated: attentive; head rested on hand; leaning back; writing; not looking at the screen. The tests have been performed on 10 subjects using static images collected from students.

Eye blink has been used to detect vigilance when e-learning was not already a necessity. In [27] an eye tracker has been used to study how the blink frequency and duration change over time. Here the blood flow is used like a ground truth to relate the latter to eye blink. To perform tests, 16 subjects have been used. Also, in this case, accuracy is not provided but only the relations between indices.

Eye-blink rate is a matter of study also in [28], where an adaptive threshold based on the blink can detect fatigue and low attention. Here the eye candidate is found by using the Viola-Jones algorithm, and then a threshold for open eyes is obtained. Finally, if the duration of blinks is anomalous, it is classified as drowsiness. Results are performed on a homemade dataset for which specifics are not provided. In the same manner, an average accuracy of eye and eye-blink detection has been provided, but not drowsiness detection accuracy.

Finally, the work that better represents the surveillance tasks is the one proposed in [13]. Here, the RLDD dataset is presented and to validate the latter, we alsopresent a drowsiness detection method. The architecture is based on a Hierarchical Multiscale Long Short-Term Memory (HM-LSTM) Network, that uses blink features to classify the sequence into three states: alert; low vigilant; and drowsy.

As can be observed by Table III, there are very few works using passive static data that report accuracy. Even if the first one in the table obtained very high accuracy, the use of images instead of videos, the poor number of subjects, and the reported speed lead us to classify the second one as more reliable to work in this area. In addition, the data used in [13] are quite realistic because the subjects are not at a fixed distance from the camera and the recordings were made by using different devices.

TABLE III
PERFORMANCES AND DATA DESCRIPTION OF METHODS IN THE PASSIVE STATIC WORKPLACES. HM IS FOR HOMEMADE.

Method	Data	#Subj	#Record	Acc(%)	F-1(%)	fps
[26] (2020)	HM	10	17k img	99.82	N/A	0.03
[13] (2019)	RLDD	60	180 vid	65.2	N/A	57.5

C. Active static workplaces

Inactive workplaces are required to perform actions during a monitoring task. Due to the data that the workers in this section use to perform experiments, we can assume that the active workplace is static. As an example, the user is seated or is stationary in a position.

The authors of the mEBAL dataset we previously introduced, also provided a benchmark in their work. By using a convolutional neural network trained on RGB images, they detect eye blink. Then, the level of attention of the student during e-learning. However, in this case, the students have to demonstrate their cognitive abilities in different situations. However, here, the visual information is used only to detect blinks and the EEG band to estimate attention. mEBAL is also used in [29], where, however, the authors used only blinks to estimate the attention. Also, in these cases, they used a convolutional neural network.

In some cases, a dataset is registered under multi-modal conditions. For this reason, it can be used in both passive and active workplaces. This is observable in works using Invedrifac. The authors, already mentioned in the dataset description, also presented their attention estimation results when the users have to perform actions. In their ”Experiment II”, they used 60 participants to perform actions on 12 different tasks.

The work in [30] has been specifically intended for workplaces. Here the subjects used computers, and for this reason, the data acquired involved mouse velocity and acceleration, click duration, time between clicks, and so on, for a total of 13 mouse and keyboard features. Different thresholds obtained from samples are used to classify the attention level of the worker. In total, 15 students were involved in the experiments, each one in a task of varying duration between 70 and 95 minutes. Unfortunately, even if the experiments are interesting, no computational time or accuracy is provided.

Analysis of students' attention was also investigated in [31]. Here, a scanpath, a collection of eye movements, is used. The method to analyze the features is MultiMatch, which compares all possible pairs of scanpaths There are five features in total for the scanpath: shape; direction; length; position; and duration. Unfortunately, relations between features are reported but accuracy in detecting attention level and computational details are not available.

Finally, machine learning has been involved in [32]. Here, a student's behavior pattern is collected and analyzed. Several classification models were used to classify attention from keystroke dynamics, mouse dynamics, and attention performance metrics. Participants were 48, and the best model to perform classification results was Random Forest.

In Table IV we presented the available performances of methods that can be involved in active static workplaces. In this case, information is quite fragmentary and there is not a method that reports all of the required indices to perform an overall analysis of costs and benefits. If we consider the use of a wide dataset and the accuracy obtained, the work in [18] can be considered the most reliable in this application. In addition, the dataset used by this work, Invedrifac, was also collected in a real scenario, making the latter even more reliable.

TABLE IV
PERFORMANCES AND DATA DESCRIPTION OF METHODS IN THE ACTIVE STATIC WORKPLACES. HM IS FOR HOMEMADE, N/A IS FOR NOT AVAILABLE.

Method	Data	#Subj	#Record	Acc(%)	F-1(%)	fps
[29] (2021)	mEBAL	38	756k frames	74.47	N/A	N/A
[18] (2019)	Invedrifac	60	N/A	97.58	97.56	14.67
[32] (2019)	HM	48	N/A	87.5	N/A	N/A

In Figure 2 we summarized the performances of all the methods presented in this work, regardless of the application field and the number of subjects involved in the experiments.

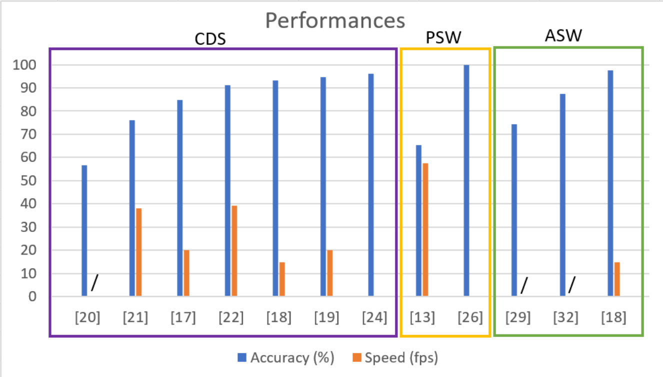


Fig. 2. Performances in terms of accuracies and speed of the discussed methods. In purple Commercial Driving Scenario papers (CDS), in yellow Passive Static Workplace papers (PSW) and in green Active Static Workplace papers (ASW). Time data not available are indicated with '/'. Where the time column is not visible, it is a value lower than 1.

In general, from all the work analysed in this section, we can observe again a strong unbalance in methods focused on driving scenarios and methods focused on workplaces. The latter uses students in their experiments and not professional workers. Also, the experiments conducted are less rigorous and often essential information is missing. It is undoubtedly that those applications are gaining interest only recently and do not gain the same interest in the scientific community, but especially in the commercial community, compared to commercial driving. On the other hand, commercial drivers represent only a small percentage of workers involved in scenarios when drowsiness affects performance, or worse, the safety of the user.

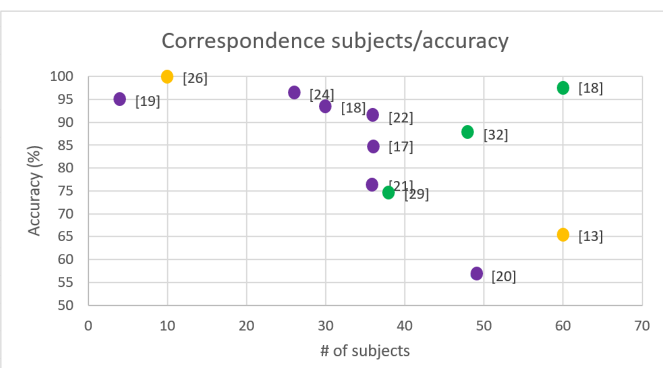


Fig. 3. Comparisons between number of subjects evaluated and accuracy obtained. In purple Commercial Driving Scenario papers, in yellow Passive Static Workplace papers and in green Active Static Workplace papers.

In Figure 3 we put in relation the amount of subjects used in the experiments and the accuracy obtained by the methods. This data is particularly interesting since the method in [26] that gain a particularly high accuracy has also a very limited number of subjects, and the dataset is not available online. On the other hand, here we can observe how reliable is the work in [18]. Even if from fig. 2 its accuracy is similar to others, here we can appreciate how wide is the dataset they used. In addition, the dataset is composed of professional workers and it make the results of [18] even more reliable. The maximum number of subject considered is, in each cases at most 60. This means that there is a lack in literature about the performances in wider scenario (as an industry), especially if subject-dependent methods are considered.

Where possible, we considered the devices the authors used to performed the experiments and put them in relation with the speed and the accuracy obtained. The result of this comparisons are in Table V. Here we can notice that authors in [13] developed a very speed method, even if the system they used is not the most performing.

TABLE V
METHODS COMPARED WITH ACCURACY, TIME AND COMPUTATIONAL CHARACTERISTICS. CSD IS FOR COMMERCIAL DRIVING SCENARIO, PSW IS FOR PASSIVE STATIC WORKPLACES AND ASW IS FOR ACTIVE STATIC WORKPLACES.

Method	Use	Acc (%)	fps	Core	GHz	RAM
[17]	CDS	84.82	20	i5	3.1	16GB
[21]	CDS	76.2	38.1	i7	3.4	16GB
[19]	CDS	94.8	20.1	i5	3.4	16GB
[18]	CDS	93.33	14.67	i5	2.2	8GB
[22]	CDS	91.2	39.21	i7	N/A	N/A
[13]	PSW	65.2	57.5	i4	3.5	16GB
[18]	ASW	97.58	14.67	i5	2.2	8GB

Finally, by considering the kind of data involved in the qualitative analysis, we observed that:

- The use of glasses reduce the performances of the methods. In particular in the work of [17], the use of glasses decrease the performances of about 5%.
- Night acquisition are less accurate than day acquisitions. We notice this trend in particular for the work in [22] (around 10% of difference) and [21] (around 3% of difference). An exception can be observed for [24] where day performance are lower than night. The first two works used the NTHU dataset in which the subjects are frontal and the last one an Homemade dataset in which the subjects are in profile.
- The use of sunglasses decrease the performance more than the use of glasses. In [17] the environment with sunglasses has around 5% of accuracy lower than the environment with glasses. The same happen in the work in [21].
- When blink is used, the method could be influenced by users' eyes. Is the case of [19], in which the smaller the eyes are, the lower the EAR is.
- A median class between low vigilant and alert, if present, is the most difficult to detect. This happen in [13] where the class of low vigilant is 50% lower in accuracy compared to the detection of alert and drowsiness. It is interesting that humans, even if with a general lower accuracy, can detect low vigilant with a decreasing of only 20% respect to alert and drowsiness.

IV. DISCUSSION ON SUBJECT'S DEPENDENCY

In general, biometrics trait that not involve subject recognition, are preferable not subject-dependent. However, in the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3173004, IEEE Transactions on Industrial Informatics

case of safety, as in drowsiness detection, it may not be a problem to use a subject-dependent architecture in which accuracy is higher than the corresponding subject-independent one. This is, again, a consideration that should be made depending on the application. To perform a more useful comparison, we will report only methods that declare their accuracy in the drowsiness/attention detection task.

In some workplaces, the user must be identified in order to perform actions. As an example, to use a specific machine, or to access a digitalised archive. In those cases, the identity of the user is known and the use of a personalized classifier is possible without leading to the generalization requirement of the system. Subject-dependent methods are also applied in circumstances in which the identity of the subject is unknown but an enrollment phase can be put in place without affecting the fluency of the performances. An example can be an enrollment step before a long ride in the case of a commercial driving scenario.

In this section, we will analyze the methods that use subject-dependency by outlining if they need an enrollment step or subject recognition.

In [19] the individual differences of the driver are enhanced in the EAR part, for which a specific SVM is trained for each subject. For this reason, here the dependency of the subjects is off-line and the subject needs to be recognized in order to use in his/her videos the corresponding SVM classifier.

In [13], where the UTA-RLDD dataset is presented, the first third of the blinks of the alert state has been used to compute the mean and standard deviation of each individual for each feature. The values obtained are used to normalize the blinks in all the states. In this case, this operation can be done online during an enrollment phase.

Those two methods are the only ones that take advantage of the subject dependencies, and are summarized in Table VI.

TABLE VI
METHODS THAT USE SUBJECT-DEPENDENT CLASSIFIERS.

Model	**Acc(%)**	**Speed(fps)**	**Enrollment**	**Application**
[19] (2019)	94.8	20.1	Offline	Commercial driving scenarios
[13] (2019)	65.2	57.5	Online	Passive static workplaces

The remaining methods are subject-independent. We can observe that the majority of methods belong to this category, independently of the application scenario.

In terms of accuracy, there is not a significant difference between subject-dependent and subject-independent methods. On the other hand, in terms of speed, we can observe that the methods in [13] are the faster ones and also subject-dependent. However, since the dependency is made online, the speed of this method cannot be imputed to the subject dependency but to the model itself.

Given that the accuracies are generally relatively close, one could wonder why a subject-dependent method would be favoured over a subject-independent one. We assumed that the methods for detecting drowsiness, poor alertness, or stress worked the same way for all subjects. However, in a work environment, some subjects could be at more risk than others due to physical or psychological pathologies, medicaments, age, etc. For this reason, it is reasonable that the level of sensitivity of the method and the consequent action put in place when a potential dangerous situation is detected may vary from one subject to another. This motivates the study of subject-dependent methods, especially in an environment in which subject-dependent characteristics are enhanced.

V. EMERGING CHALLENGES

The topic we discussed is relatively new for the applications in Industry 4.0. However, in drowsiness and attention detection itself, there are some challenges that are even newer and could be of future interest to Industry 4.0.

We discussed the use of RGB images and sequences to detect drowsiness, but also infrared and thermal cameras could be used for this purpose. In [33], Tipprasert et al. use infrared images to deal with the problem, detecting drowsiness in drivers in low light conditions. Their method consists of face, eye, and mouth detectors, combined with eye closure and yaw detection. Over the 3760 images they studied, they obtained a 98% and 92.5% of accuracy on eye closure and yawing detection, respectively. The work of Tashakori et al [34] focused on thermal cameras to evaluate the forehead temperatures of 30 subjects in a driving simulated scenario. In this case, SVM, KNN, and the regression tree classifiers were used, obtaining 82% of accuracy, 85% of sensitivity, 90% of specificity, and 84% of precision to detect drowsiness. Using both infrared and thermal information, we find the work of Cardone et al [35]. They used a low-cost and high-resolution thermal infrared technology to evaluate 10 sleep-deprived subjects in a one-hour driving simulation task. Here, the thermal camera evaluates the facial skin temperature and the infrared camera, the PERCLOS. They used a multivariate machine learning approach based on a three-level SVM and obtained a classification accuracy of 65% $\pm$ 9%.

Another challenge derived from recent times, is the necessity to perform driver fatigue detection while wearing protective masks. Wearing protecting mask could be itself a cause of stress for the workers. In fact, in a recent study focused on COVID-19 effects, it has been demonstrated that wearing a mask for several hours (e.g. ¿8 hours) can increase psychological stress [36]. This aspect should be taken into account, especially if we consider that detecting attention, drowsiness, and stress when only a portion of the face is visible is even more challenging. In this case, it is not sufficient to consider algorithms that take only the periocular zone into account to detect drowsiness, since most of them use all the face to detect this part. Unfortunately, there is not a specific literature on drowsiness and stress detection while wearing masks, but some references to the topic can be found in more recent works. In [37] the experiments that led to around 90% of accuracy in driver drowsiness detection were also conducted by wearing facial masks. This demonstrates that it is possible to solve this task even if not specifically studied.

VI. Conclusion

Drowsiness detection is undoubtedly a necessary task to create a safe workplace. In Industry 4.0, several technologies are involved that can also detect drowsiness. In this paper, we investigate how drowsiness detection techniques can be applied to this new scenario. We presented the available datasets that can be used to solve this task, with the number of subjects involved, sensors used, and modalities. We highlighted the scenario that each of these datasets simulates, obtaining three possible applications: commercial driving scenarios; Active static workplaces; Passive static workplace. From this first analysis, we can already see how a dynamic scenario is not contemplated in the currently available data. A user that moves in a room or a building, using multiple workplaces, cannot be evaluated for drowsiness by using existing datasets.

Then, we presented several methods that evaluate drowsiness. Methods that simulated driving results to be more thriving and in a more advanced state. Benchmark datasets (as NTHU) are often used, and the studies differentiate between the number of sensors involved, illumination variations, hours of acquisition, and classification modalities. On the other hand, work on active static workplaces and passive static workplaces results in a more preliminary stage. We notice from several works that the main part of the face involved to detect drowsiness is the eye area and the mouth area. For works using hand crafted features or machine learning, the reason is that they are focused on blink detection and yawn detection. Even if in a more indirect manner, also deep learning techniques are strongly influenced by those parts of the face. In fact, also in their case the use of sunglasses of glasses decrease the accuracies of the methods. This is observable also in combined techniques as GANs or Bi-LSTM.

Several works, even if interesting, do not yet discuss accuracy in drowsiness detection and other aspects such as computational costs or the number of samples used. This literature is more recent since the pandemic state due to COVID-19 increased the use of e-learning and the necessity for professors to know the effectiveness of the lessons. However, in Industry 4.0, paying more attention to those kinds of applications could expand drowsiness detection to new scenarios, improving the safety of workplaces. Finally, we investigated the contribution of subject-dependent architecture to the overall accuracy, and we concluded that a dependency is not necessarily related to best performances.

Acknowledgment

Fei Hao's work was funded in part by the Natural Science Basic Research Plan in Shaanxi Province of China (Grant No. 2022JM-371), and the Fundamental Research Funds for the Central Universities (Grant No. GK202103080).

References

[1] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61 904–61 919, 2019.

[2] H. R. Colten and B. M. Altevogt, "Sleep disorders and sleep deprivation: An unmet public health problem." [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK19958/

[3] D. Kaliyaperumal, Y. Elango, A. Murali, and I. Santhanakrishanan, "Effects of sleep deprivation on the cognitive performance of nurses working in shift," *Journal of Clinical and Diagnostic Research*, vol. 11, pp. CC01–CC03, 08 2017.

[4] L. Peternel, N. Tsagarakis, D. Caldwell, and A. Ajoudani, "Robot adaptation to human physical fatigue in human–robot co-manipulation," *Autonomous Robots*, vol. 42, pp. 1–11, 06 2018.

[5] Z. Liu, K. Xie, L. Li, and Y. Chen, "A paradigm of safety management in industry 4.0," *Systems Research and Behavioral Science*, vol. 37, no. 4, pp. 632–645, 2020.

[6] A. Goldberger, L. Amaral, L. Glass, S. Havlin, J. Hausdorg, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, H. Stanley, and P. Physiobank, "Components of a new research resource for complex physiologic signals," *PhysioNet*, vol. 101, 01 2000.

[7] U. Budak, V. Bajaj, Y. Akbulut, O. Atila, and A. Sengur, "An effective hybrid model for eeg-based drowsiness detection," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7624–7631, 2019.

[8] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable eeg system for driver drowsiness detection," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7169–7180, 2015.

[9] J. Vicente, P. Laguna, A. Bartra, and R. Bailon, "Drowsiness detection using heart rate variability," *Medical Biological Engineering Computing*, vol. 54, 01 2016.

[10] N. Rodríguez-Ibáñez, M. A. García-González, M. Fernández-Chimeno, and J. Ramos-Castro, "Drowsiness detection by thoracic effort signal analysis in real driving environments," in *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2011, Boston, MA, USA, August 30 - Sept. 3, 2011*. IEEE, 2011, pp. 6055–6058.

[11] M. Johns, "The amplitude-velocity ratio of blinks: A new method for monitoring drowsiness," 05 2003, pp. A51–A52.

[12] R. Daza, A. Morales, J. Fierrez, and R. Tolosana, "Mebal: A multimodal database for eye blink detection and attention level estimation," in *Companion Publication of the 2020 International Conference on Multimodal Interaction*, ser. ICMI '20 Companion. New York, NY, USA: Association for Computing Machinery, 2020, p. 32–36.

[13] R. Ghoddoosian, M. Galib, and V. Athitsos, "A realistic dataset and baseline temporal model for early drowsiness detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[14] G. Hu, Y. Xiao, Z. Cao, L. Meng, Z. Fang, J. T. Zhou, and J. Yuan, "Towards real-time eyeblink detection in the wild: Dataset, theory and practices," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2194–2208, 2020.

[15] S. Ferreira, Z. Kokkinogenis, and A. Couto, "Using real-life alert-based data to analyse drowsiness and distraction of commercial drivers," *Transportation Research Part F Traffic Psychology and Behaviour*, vol. 60, pp. 25–36, 10 2018.

[16] [Online]. Available: https://dmdatabases.com/

[17] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *ACCV Workshops*, 2016.

[18] A. Dasgupta, D. Rahman, and A. Routray, "A smartphone-based drowsiness detection and warning system for automotive drivers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4045–4054, 2019.

[19] F. You, X. Li, Y. Gong, H. Wang, and H. Li, "A real-time driving drowsiness detection algorithm with individual differences consideration," *IEEE Access*, vol. 7, pp. 179 396–179 408, 2019.

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

[21] J. Yu, S. Park, S. Lee, and M. Jeon, "Driver drowsiness detection using condition-adaptive representation learning framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4206–4218, 2019.

[22] Y. Hu, M. Lu, C. Xie, and X. Lu, "Driver drowsiness recognition via 3d conditional gan and two-level attention bi-lstm," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4755–4768, 2020.

[23] A. Altameem, A. Kumar, R. C. Poonia, S. Kumar, and A. K. J. Saudagar, "Early identification and detection of driver drowsiness by hybrid machine learning," *IEEE Access*, vol. 9, pp. 162 805–162 819, 2021.

[24] G. Chen, L. Hong, J. Dong, P. Liu, J. Conradt, and A. Knoll, "Eddd: Event-based drowsiness driving detection through facial motion analysis

with neuromorphic vision sensor," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6170–6181, 2020.

[25] A. F. Abate, L. Cascone, M. Nappi, F. Narducci, and I. Passero, "Attention monitoring for synchronous distance learning," *Future Generation Computer Systems*, vol. 125, pp. 774–784, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X21002880

[26] A. Revadekar, S. Oak, A. Gadekar, and P. Bide, "Gauging attention of students in an e-learning environment," in *2020 IEEE 4th Conference on Information Communication Technology (CICT)*, 2020, pp. 1–6.

[27] L. K. McIntire, R. A. McKinley, C. Goodyear, and J. P. McIntire, "Detection of vigilance performance using eye blinks," *Applied Ergonomics*, vol. 45, no. 2, Part B, pp. 354–362, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687013000938

[28] Z. A. Haq and Z. Hasan, "Eye-blink rate detection for fatigue determination," in *2016 1st India International Conference on Information Processing (IICIP)*, 2016, pp. 1–5.

[29] R. Daza, D. DeAlcala, A. Morales, R. Tolosana, R. Cobos, and J. Fierrez, "Alebk: Feasibility study of attention level estimation via blink detection applied to e-learning," 2021.

[30] D. Duraes, D. Carneiro, J. Bajo, and P. Novais, "Modelling a smart environment for nonintrusive analysis of attention in the workplace," *Expert Systems*, vol. 35, no. 5, p. e12275, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12275

[31] S. Stranc and K. Muldner, *Scanpath Analysis of Student Attention During Problem Solving with Worked Examples*, 06 2020, pp. 306–311.

[32] D. Durães, "Student attention evaluation system using machine learning for decision making," in *EPIA*, 2019.

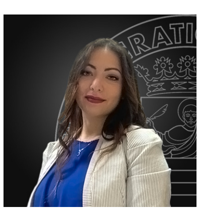
[33] W. Tipprasert, T. Charoenpong, C. Chianrabutra, and C. Sukjamsri, "A method of driver's eyes closure and yawning detection for drowsiness analysis by infrared camera," in *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 2019, pp. 61–64.

[34] M. Tashakori, A. Nahvi, and S. E. H. Kiashari, "Driver drowsiness detection using facial thermal imaging in a driving simulator," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 236, no. 1, pp. 43–55, 2022, pMID: 34477030.

[35] D. Cardone, C. Filippini, L. Mancini, A. Pomante, M. Tritto, S. Nocco, D. Perpetuini, and A. Merla, "Driver drowsiness evaluation by means of thermal infrared imaging: preliminary results," 08 2021, p. 25.

[36] Z. Tian, B.-Y. Kim, and M.-J. Bae, "A study on the effect of wearing masks on stress response," *International Journal of Engineering Research and Technology*, vol. 13, p. 807, 04 2020.

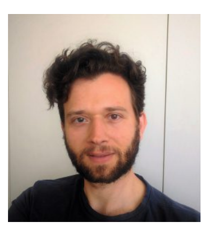
[37] A.-U.-I. Rafid, A. I. Chowdhury, A. R. Niloy, and N. Sharmin, "A deep learning based approach for real-time driver drowsiness detection," in *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2021, pp. 1–5.



Carmen Bisogni received the B.S. degree and M.S. degree (cum Laude) in Mathematics from University of Salerno in 2015 and 2017, respectively. She received the Ph.D. in Computer Science in 2021, from University of Salerno. She is currently a research fellow at the Biometric and Image Processing Laboratory (BIPLAB) at University of Salerno, Italy. Her research interests include applied mathematics for Machine Learning, Biometrics, Image Processing and Statistical Analysis. Dr. Bisogni is member of IEEE and GIRPR/IAPR, member of the Editorial Board of Electronics (MDPI) and Associate Editor of the IEEE Biometrics Council Newsletter.

Fei Hao received the B.Sc. degree in Information and Computing Science and the M.Sc. degree in Computer Software and Theory from Xihua University, China, in 2005 and 2008, respectively, and the Ph.D. degree in Computer Science and Engineering from Soonchunhyang University, South Korea, in 2016. He is currently an Associate Professor with the School of Computer Science, Shaanxi Normal University, China. From 2020 to 2022, he was a Marie Curie Fellow at the University of Exeter, Exeter, United Kingdom. His research interests include social computing, ubiquitous computing, big data analysis and processing, knowledge graph and mobile edge computing.

Vincenzo Loia is Full professor in Computer Science and Rector of University of Salerno. Graduated in Computer Science at the University of Salerno in 1985, and received his PhD in Computer Science in 1989 at the Universite Pierre Marie Curie Paris VI, France. Over the years, the pillars of his scientific research have been applied in various application domains, from medicine to law, and more recently in the cyber-intelligence field. It is exactly in this latter context, that he shows his great aptitude to multidisciplinary, by coordinating the scientific committee of the "Multidisciplinary Observatory for the fight against organised crime and terrorism". Moreover, his competences are used to support research related to the 4th industrial revolution, as demonstrated by his scientific responsibility and participation in national projects on smart manufacturing, and international projects on cyber physical systems. He is Editor-in-Chief and Founder of the international journal "Ambient Intelligence and Humanized Computing", Springer- Verlag; Editor-in-Chief of the international journal "Evolutionary Intelligence", Springer-Verlag; Editor Responsible for Special Issues of the international journal "Soft Computing", SpringerVerlag; Co-Editor in Chief of the international journal "Information Processing Systems". He wrote about 400 scientific papers, which have been published in proceedings of international conferences, book chapters, international journals.



Fabio Narducci is an assistant professor at the University of Salerno. He received a Ph.D. in Computer Science at the Virtual Reality Lab of the University of Salerno in 2015. He served as TPC of international conferences and as reviewer of several international journals. He was Guest Editor of Special Issues on IEEE Cloud Computing, Elsevier Pattern Recognition Letters and Multimedia Tools and Applications and he is currently Section Editor of Multimedia Tools and Application, associate Editor of IET Electronic Letters and Editorial Board Member of Inderscience International Journal of Embedded Systems and MDPI Electronics. His research interests mainly include Machine Learning and Deep Learning, Image processing and Pattern Recognition, Biometrics, Mobile and Wearable computing.