

# Detecting exercise-induced fatigue using thermal imaging and deep learning

Miguel Bordallo Lopez , Carlos R. del-Blanco , Narciso Garcia

**Abstract**— Fatigue has adverse effects in both physical and cognitive abilities. Hence, automatically detecting exercise-induced fatigue is of importance, especially in order to assist in the planning of effort and resting during exercise sessions. Thermal imaging and facial analysis provide a mean to detect changes in the human body unobtrusively and in variant conditions of pose and illumination. In this context, this paper proposes the automatic detection of exercise-induced fatigue using thermal cameras and facial images, analyzing them using deep convolutional neural networks. Our results indicate that classification of fatigued individuals is possible, obtaining an accuracy that reaches over 80% when utilizing single thermal images.

**Keywords**— Fatigue detection, facial expression, deep learning, thermal imaging

## I. INTRODUCTION

Fatigue is often defined as an impairment of performance that includes both an increase in the perceived effort necessary to exert a desired force and an eventual inability to produce this force [1] [2] [3]. Exercise-induced fatigue is a type of fatigue induced by physical activity usually planned structured and repetitive for the purpose of conditioning. Although exercise is commonly used to improve health, maintain fitness and as a means of physical rehabilitation, several studies performed in healthy adults have shown that exercise-induced local fatigue can temporally and adversely affect physical skills, such as balance, postural control, strength or hydration capability and cognitive skills (e.g. decision making or attention). Hence, to properly assess the practice and planning of further exercise or a period of resting, fatigue monitoring is of great importance, and has applicability in numerous scenarios.

Computer vision techniques and facial image analysis can provide an unobtrusive way of detecting and measuring fatigue from a distance, without the need of complementary wearable sensors. However, this is a very challenging problem that needs to deal with images and sequences taken in various conditions of pose and illumination. Moreover, many times, subjects only show subtle changes in expression during the acquisition process, making the inference of the fatigue condition even more complex.

Automatic fatigue detection from visual information, has mostly been related to the detection of drowsiness, blinking and yawning of drivers [4] [5]. The most common approaches rely on the extraction of facial expression cues that model a possible state of tiredness. However, these models are based on effects mostly due to psychological states and they are

unable to detect physical or muscular exercise-induced fatigue. Although the assessment of muscular fatigue is of increasing interest, it is usually performed using bio-signals or clinical tests [6] [7], while its assessment from visual data is still mostly unexplored.

This might be due to the fact that research on face analysis has mainly been focusing on analyzing color images obtained from regular cameras and visible light. Hence, ignoring information at other wavelengths that can be useful for discriminating characteristics that show in the face. While there exist facial analysis methods that have shown promising results on individual data sets for different health-related tasks [8], the generalization capabilities of these methods for other applications, such as fatigue detection, have been questionable.

In this context, if we take a look at facial images related to resting and fatigued individuals at typical wavelengths, such as visual or near-infrared (NIR), (Figure 1), it is basically impossible to explicitly name any textural differences between them. However, when facial images are obtained from a thermal camera at mid-infrared wavelength, this textural differences become very apparent.



Fig. 1: RGB, near infrared and thermal images of a resting (up) and fatigued (down) face. In the thermal images, darker pixels corresponds to colder and lighter to hotter.

Inspired by the aforementioned observations, we propose the use of thermal cameras to detect and recognize fatigue in variable conditions. Thermal imaging is a non-contact and non-invasive imaging method which provides information on human body temperature by assessing the infrared spectrum of the subject. In this imaging method, thermal cameras are able to detect radiation in the infrared range of the electromagnetic spectrum, usually from 5 to 20  $\mu\text{m}$  (mid-infrared bands), converting the amount of radiation into visible images. Since infrared radiation is emitted by all objects, thermal images make it possible to see objects even without visible illumination. Thermographies obtained from facial images have been used for several applications related to health monitoring and diagnosis. Since the amount of radiation emitted by an object is directly proportional to its temperature, the nature of the information provided by this method makes it highly relevant for applications related to clinical medicine [9]. Indeed, not only the thermal-print provides information on the shape of the face [10], but also multiple contributing thermal factors can be assessed from the body (e.g., blood flow, cell metabolism, sweat gland activation) as they would cause local changes in superficial skin temperature. More specifically, different reasons such as inflammatory processes [11] or fever [12] can be responsible for changes in the skin temperature. Due to the increase of thermal camera accuracy and resolution [13], increasing research has been performed lately to evaluate how much information the distribution of heat in the face can provide. For example, various applications related to mass-screening have been using thermal imaging as a fever or respiratory problem detection [14].

In this work, we propose the detection of exercise-induced fatigue using thermal facial images obtained from healthy individuals in both resting and fatigue conditions. The detection approach is based on the utilization of facial analysis and deep learning. Our contributions can be summarized as follows:

- We propose for the first time the use of thermal facial images to assess the exercise-induced fatigue of healthy people.
- We present a new and meaningful data set composed of 5700 thermal images from 19 subjects.
- We show that classification of fatigue from facial images can be performed using pre-trained deep convolutional neural networks and support vector machines, obtaining an accuracy over 80%.
- In addition, we study the facial regions that contribute more to the correct classification, and discuss the limitations of the data set and imaging modality.

The rest of this paper is organized as follows: Section II describes our proposed methodology based on the extraction of deep features and classification. Section III depicts the experimental setup. Section IV shows the obtained results and their analysis. Finally, Section V concludes the paper and offers some future directions.

## II. ADOPTED METHODOLOGY

Our fatigue detection system consists of three main steps. It starts with detecting, segmenting and aligning the thermal facial images based on eye and nose coordinates. The results of this step are a set of aligned facial thermal pictures and regions of interest (ROIs). The alignment tries to minimize possible differences across subjects and images. The second step relies on two different large convolutional neural networks that extract fixed-length deep feature vectors from facial image or region. These features are then utilized on the third and last step, a set of Support Vector Machines (SVMs) which determine the state of a subject as fatigued or resting. In this section, we describe in detail the design decisions for each one of the steps, detailing how their parameters are learned and utilized.

### A. Face pre-processing

To mitigate the influence of inconsistent poses and displacements of the faces across the images included in the database, the first step of our approach consists on segmenting the face regions from each video sequence.

For that purpose, we have employed a correlation map obtained from the first image of the set belonging to a subject and a pre-defined template of a centered face [15]. Based on this correlation map, the regions containing frontal faces are then cropped from every frame. Finally, the face-regions are aligned using the nose tip as the center, and scaled further to preserve the interpupillary distance.

In order to compute features for a region proposal, we must first normalize the size of the face related to the image regions in order to make it compatible with the input of the CNNs, 227 x 227 or 224 x 224 pixel sizes. Of the many possible transformations of our arbitrary-sized regions, we opt for the simplest. Regardless of the size or aspect ratio of the candidate region, we warp all pixels in a squared bounding box of the required size [16].

### B. Deep features

Deep neural networks have been recently outperforming the state of the art in various classification tasks. Particularly, convolutional neural networks (CNNs) demonstrated impressive performance in object classification in general and face recognition in particular. However, deep neural networks require a huge amount of training data to learn efficient features. This is not the case for currently available thermal imaging data sets [15].

To extract meaningful deep features, an alternative to CNNs created from scratch is to use a pre-trained network. A number of very deep pre-trained architectures has already been made available to the research community. In this work, we use two well known pre-trained CNNs, Alexnet and VGG (both its variants VGG-16 and VGG-19). This approach helps us overcoming the lack of enough training data.

Pre-trained convolutional neural networks can be used as a feature extractor or as part of a transfer learning strategy. For the feature extractor case, an intermediate output of a pre-trained network is directly the feature vector to be used as input in a new classifier. The reasoning after this strategy is that the first layers of a CNN contains useful and general visual features for image-based classification, while the own classification is performed by the last layers. On the other hand, for the transfer learning strategy [16], the pre-trained network is used as a starting point, which is then fine-tuned with new samples of the specific application (in our case, infrared face images for the fatigue detection) to adapt the network to the new domain application. This approach has the advantage of using a much smaller number of training images than that of training the network from scratch. In addition, it is trained much faster.

In this work, we have tested both variants. For the feature extractor case, we have used the following pre-trained CNNs: AlexNet, VGG-16, and VGG-19 (see below a description of them). The deep features are obtained for the intermediate output called 'fc7' (the output of a fully connected layer) in the previous network architectures, whose dimension is 4096. These 4096-dimensional features have been used independently and jointly along with a Support Vector Machine (SVM) classifier, as will be described later.

For the transfer learning strategy, the last layers of the original pre-trained networks, corresponding to fully connected layers, have been substituted for a new set of fully connected layers with the purpose to detect two classes (resting or fatigue states), instead of the original task performed by the AlexNet, VGG-16, and VGG-19 architectures: 1000-class image classification. The training of the new network has been performed by using a reduced set of samples of such new domain (facial thermal images). Stochastic gradient descent (SGD) algorithm with momentum has been used to train the new CNNs. In addition, the learning rate of the first layers of the network has been fixed to 0.001, which is a 1/10th of the learning rate of the rest of layers. This is done to preserve the learnt features of the original network (corresponding to the first layers), while focusing on training the last layers in charge of the classification task. In each SGD iteration, we uniformly sample 16 faces (over both classes) creating a mini-batch of size 32.

#### AlexNet features

AlexNet is a Convolutional Neural Network first proposed by Krizhevsky et al. [17]. The original pre-trained network utilized over 1.2 million images collected from the web and labeled by Amazon Mechanical Turk workers. The feature vectors are computed by forward propagating a mean-subtracted 227 x 227 RGB image through five convolutional layers and two fully connected layers. AlexNet, is based on layer-inputs provided the previous layer and does not rely on filter concatenation. The detailed parameters of the AlexNet CNN are provided by Table 1.

Num	Name	Type	Layer description
1	'data'	Image Input	227x227 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Norm.	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Norm.	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	softmax
25	'output'	Classification	crossentropyx with 'fatigue' and 'resting' classes

Table 1: AlexNet CNN architecture.

#### VGG features

VGG-16 and VGG-19 are two related deep convolutional network for object recognition developed and trained in Oxford's Visual Geometry Group. Originally trained on a subset of the ImageNet database [18], it was first introduced to compete in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [19]. VGG is pre-trained on more than a million images.

The detailed parameters of the VGG CNN are provided by Table 2. The input of the network is a RGB color image with a size of  $224 \times 224$  pixels. The network is composed of 13 (VGG-16) or 16 (VGG-19) linear convolution layers (*conv*), and two fully connected layers (*fc*), both outputting a vector of size 4096.

In this context, to extract deep face features for kinship verification, we input the facial images one by one to the CNNs and collect the feature vector issued by the fully connected layer fc7 (all the layers of the CNN except the class predictor fc8 layer and the softmax layer are used), using the resulting descriptor for further classification.

#### C. Classification and Fusion

After the extraction of features, the next step is the fusion of features and classification. In this work, we adopted early fusion, fusing features directly after the feature extraction process, and directly concatenating the different feature sets obtained by the different networks. The fused features are then fed to the classifier.

To estimate the fatigue class of each image, we use a binary linear Support Vector Machine (SVM). For fatigue detection, the setup was chosen to be a two-class problem, with the first class corresponding to a *fatigue* state and the second one as *resting* state.

The predictive performance of SVMs depends on parameter selection. In order to select the best error penalty parameter range (C), we split the training data leaving one subject out, and obtain the accuracy of 12 different sets of SVM, which



Num	Name	Type	Layer description
1	'input'	Image Input	224x224 images with 'zerocenter' normalization
2	'conv1-1'	Convolution	64 3x3x3 convolutions with stride [1 1] and padding [1 1]
3	'relu1-1'	ReLU	ReLU
4	'conv1-2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1]
5	'relu1-2'	ReLU	ReLU
6	'pool1'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
7	'conv2-1'	Convolution	128 3x3x64 convolutions with stride [1 1] and padding [1 1]
8	'relu2-1'	ReLU	ReLU
9	'conv2-2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1]
10	'relu2-2'	ReLU	ReLU
11	'pool2'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
12	'conv3-1'	Convolution	256 3x3x128 convolutions with stride [1 1] and padding [1 1]
13	'relu3-1'	ReLU	ReLU
14	'conv3-2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1]
15	'relu3-2'	ReLU	ReLU
16	'conv3-3'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1]
17	'relu3-3'	ReLU	ReLU
18	'pool3'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
19	'conv4-1'	Convolution	512 3x3x256 convolutions with stride [1 1] and padding [1 1]
20	'relu4-1'	ReLU	ReLU
21	'conv4-2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
22	'relu4-2'	ReLU	ReLU
23	'conv4-3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
24	'relu4-3'	ReLU	ReLU
25	'pool4'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
26	'conv5-1'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
27	'relu5-1'	ReLU	ReLU
28	'conv5-2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
29	'relu5-2'	ReLU	ReLU
30	'conv5-3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
31	'relu5-3'	ReLU	ReLU
32	'pool5'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
33	'fc6'	Fully Connected	4096 fully connected layer
34	'relu6'	ReLU	ReLU
35	'drop6'	Dropout	50% dropout
36	'fc7'	Fully Connected	4096 fully connected layer
37	'relu7'	ReLU	ReLU
38	'drop7'	Dropout	50% dropout
39	'fc8'	Fully Connected	1000 fully connected layer
40	'prob'	Softmax	softmax
41	'output'	Classification Output	'crossentropyex' with 'fatigue' and 'resting' classes

Table 2: VGG-16 CNN architecture.

corresponds to a search range from -6 to 16 with an interval of 2. The selected model is the one that presents the best accuracy on the training data. The final accuracy value is obtained after applying a mean pooling strategy over all different data splits.

### III. EXPERIMENTAL SETUP

#### A. Experimental data and capturing protocol

In order to test the classification of resting and fatigued individuals, we collected a new data set containing 5700 facial images acquired using a thermal camera. The data was obtained for 19 individuals (11 male, 8 female), with ages ranging from 23 to 37 years old. Nine of the individuals were from Chinese ethnicity. Six of them wore glasses, hence, from a the thermal wavelength perspective, their faces were partially occluded, since glass is opaque to far infrared thermal radiation.

The images were obtained in two different sessions using continuous shooting during approximately 20 seconds, collecting a total number of 150 images per session. The first session was recorded in resting state. The participants were asked to sit down and rest for several minutes, until their heart rate was below 80 beats per minute and their respiratory rate was below 12 breaths per minute.

To record the second session we induced fatigue in the subjects by asking them to do intense exercise. The subjects were asked to go up and down the floor stairs as fast as they could for at least three minutes, or until they felt fatigued. If their heart rate reached over 120 beats per minute and their respiratory rate reached over 15 breaths per minute, we started

recording. If not, we asked the subjects to exercise more.

The facial images were obtained using the Therm-App thermal camera mounted on a tripod. The Therm-app camera is a 17 $\mu$ m thermal detector, equipped with a 19mm lens and manual focus. Its resolution is 288x384 pixels in portrait mode, obtained at 8.7Hz. The camera can be plugged to a standard android device (in our case a Samsung Galaxy Edge 6 mobile phone) running a custom made application coded with the Therm-App Developer SDK. The application uses continuous shooting to save uncompressed images. The thermal data range was compressed to 256 levels (grayscale). The temperature dynamic range is clipped so that the average temperature of the image corresponds to the level 128, the 90% maximum temperature and above is set to 256, and the 10% minimum or below is set to 0. The facial images were taken at about 1 meter distance from the subject, a distance that maximizes the resolution of the face, with interpupillary distance of at least 100 pixels. The Therm-App camera is depicted in Figure 2.



Fig. 2: Therm-App mobile thermal camera: 19mm lens, 288x384 resolution, 8.7Hz, 17 $\mu$ m wavelength.

#### B. Evaluation protocol

To evaluate the performance of different methods, ensuring a proper generalization to unseen subjects, we adopted the *leave-one-subject-out* scheme. For each one of the subjects, a SVM classifier was trained using 18 users out of 19. Every trained SVM was then tested in the remaining "unseen" subject. Hence, in each experiment, 19 classification tasks were conducted and their results averaged, avoiding biased results. To ensure a fair comparison between methods, the results are reported with the mean accuracy and the standard deviation of single subjects.

#### C. Run-time analysis

To perform the training and testing, we have used a desktop computer featuring an Intel i7 processor, a Titan Xp GPU with 3840 CUDA cores running at 1.5GHz and Matlab 2015. In this environment, the extraction and training phase of our method lasts about 10 minutes per model on the GPU (about 1 hour in the CPU). The classification time is about 200ms per frame, with more than 90% of the time used in feature extraction. These times make our approach suitable for real-time applications, possibly even in embedded systems.

#### IV. RESULTS AND ANALYSIS

Table 3 shows the obtained results on the data set comparing the performance of three different pre-trained CNNs. To explore the complementarity of the networks, we also report the results obtained by the combination of features obtained from the networks using an early fusion scheme (feature level concatenation). From these results we can see that although the accuracy levels are comparable, AlexNet seems to show a better performance than VGG in both their variants. The combination of features does not seem to offer any performance advantage showing that all three networks compute similar features that do not complement each other properly.

Method	Accuracy (mean %)	Std. deviation (%)
AlexNet	81.51	0.83
VGG-16	70.29	2.56
VGG-19	78.71	18.75
AlexNet + VGG-16	77.43	5.72
AlexNet + VGG-16 + 19	77.86	0.84

Table 3: Classification accuracy (mean) for AlexNet, VGG-16, VGG-19 and fused features.

Figures 3 and 4 show examples of correctly and incorrectly classified subjects in all three networks. When observing correctly classified individuals, it can be seen that the texture of the facial temperature seem to follow marked patterns. For fatigued individuals the nose gets comparatively colder than the rest of the face, the mouth tends to be either open or colder, the eye region gets warmer and different textural patterns emerge on the skin, probably due to the blood irrigation and irregular sweat.

For the incorrectly classified subjects, the network seems to classify the most of the images belonging to a single individual as belonging to one of the classes (either *resting* or *fatigued*). This is in line with what is observed by visually inspecting the data. It is visually very difficult to guess to which class an single image belongs to when examples of both classes are not available for comparison. This misclassification is an argument for the training of "person-specific models" that could easily outperform the general model presented here.



Fig. 3: Examples of correctly classified subjects.

Table 4 shows the results derived from the AlexNet CNN: those results using AlexNet as feature extractor (using a



Fig. 4: Examples of incorrectly classified subjects.

SVM as classifier with different kernels) and those ones using AlexNet for transfer learning (changing the last layers and using fine-tuning). Linear-SVM seems to marginally improve the performance when compared with the original transfer learning strategy.

Method	Accuracy (mean %)
AlexNet transfer learning and fine tuning	80.48
AlexNet SVM-Gaussian	77.70
AlexNet Linear-SVM	81.51

Table 4: Classification accuracy (mean) for AlexNet using different classifiers

In order to know which parts of the face contribute more to the detection of fatigue, we have classified the data using three different facial regions. The eye region is selected to cover the eyes, part of the eyebrows and the nasal bridge. The nose region covers the central part of the face including the complete nose. The mouth region covers the lips and their surrounding area. Figure 5 shows examples of the extracted facial regions of two individuals, one with glasses and one without them, for both *resting* and *fatigued* states.

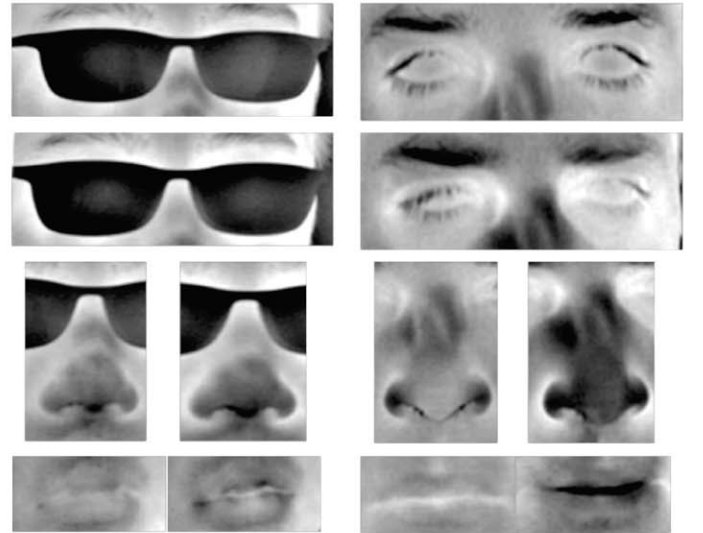


Fig. 5: Division of the face in regions: eyes (first and second row), nose (third row) and mouth (fourth row) for two different individuals (left and right columns)

When visually inspecting obtained images, it can be seen that the individuals wearing glasses show a very visible partial facial occlusion, due to the fact that glasses are "colder" and the IR radiation at the selected wavelength cannot go through them. This occlusion has a noticeable effect in the classification accuracy. The results of these experiments can be seen in Table 5.

Method	glasses	no glasses	whole set
AlexNet Linear-SVM whole face	76.33	83.89	81.51
AlexNet Linear-SVM eye region	55.83	71.33	69.69
AlexNet Linear-SVM nose region	54.30	67.34	63.44
AlexNet Linear-SVM mouth region	87.22	85.36	86.14

Table 5: Classification accuracy (mean) for AlexNet for different face regions

As expected, and due to the occlusion caused by the glasses, the performance obtained when using only the eye or mouth regions is small compared with the one obtained using whole face. Moreover, the split between individuals wearing glasses shows that no reliable classification can be done for occluded regions when looking only at the eye or nose regions. When classifying only individuals without glasses, the accuracy is noticeably better, but since the model was trained also with individuals using glasses, the accuracy it is still not comparable to the results using the whole facial region.

Supporting these observations, the mouth region, not influenced by occlusions, shows a classification performance comparable and even superior to the one obtained using only the whole face. Also, the splits between occluded and not occluded subjects yields comparable results. These experiments suggest that for data sets composed of individuals where the use of glasses cannot be always avoided, the best strategy is to focus in the lower part of the face. This fact could be possibly extended to other applications using facial thermal images.

## V. CONCLUSION

In this paper we have presented the first system able to recognize exercise-induced fatigue in healthy individuals from facial images. The use of a mobile thermal camera allows overcoming possible performance differences due to variant illumination. Deep convolutional neural networks show to be an adequate algorithm to perform the classification task, and the use of pre-trained and fine-tuned models allows for reasonably small training times and real-time classification.

To evaluate our approach, we have collected a meaningful data set composed of 5700 images, showing that classification achieves an accuracy over 80%. The shortcomings of the imaging method, namely the possible occlusions due to objects transparent to visible light, but opaque to thermal radiations are overcome by utilizing different facial regions.

This work has focused only in the detection of fatigue. Future work could aim for quantifying this fatigue in a continuous way, allowing for a better assessment of the exercise intensity. Also, a comparison between image modalities could address their performance differences and study their possible complementarity.

## ACKNOWLEDGMENT

This work has been partially supported by the Suomen Kulturi Rahasto central fund and the Ministerio de Economía, Industria y Competitividad (AEI/FEDER) of the Spanish Government under projects TEC2013-48453 (MR-UHDTV) and TEC2016-75981 (IVME).

## REFERENCES

- [1] Don B Chaffin. Localized muscle fatigue-definition and measurement. *Journal of Occupational and Environmental Medicine*, 15(4), 1973.
- [2] Roger M Enoka and Douglas G Stuart. Neurobiology of muscle fatigue. *Journal of applied physiology*, 72(5):1631–1648, 1992.
- [3] Gerome C Gauchard, Pierre Gangloff, Alexandre Vouriot, Jean-Pierre Mallie, and Philippe P Perrin. Effects of exercise-induced fatigue with and without hydration on static postural control in adult human subjects. *International Journal of Neuroscience*, 112(10):1191–1206, 2002.
- [4] Xiao Fan, Yanfeng Sun, Baocai Yin, and Xiuming Guo. Gabor-based dynamic representation for human fatigue monitoring in facial image sequences. *Pattern Recognition Letters*, 31(3):234–243, 2010.
- [5] Haisong Gu and Qiang Ji. An automated face reader for fatigue detection. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pages 111–116, 2004.
- [6] Robert S Foote, Justin D Pearlman, Alan H Siegel, and Kiang-Teck J Yeo. Detection of exercise-induced ischemia by changes in b-type natriuretic peptides. *Journal of the American College of Cardiology*, 44(10):1980–1987, 2004.
- [7] Romualdo Belardinelli, Francesca Lacalaprice, Flavia Carle, Adelaide Minnucci, Giovanni Cianci, GianPiero Perna, and Giuseppe D'Eusano. Exercise-induced myocardial ischaemia detected by cardiopulmonary exercise testing. *European heart journal*, 24(14):1304–1313, 2003.
- [8] Jerome Thevenot, Miguel Bordallo López, and Abdenour Hadid. A survey on computer vision for assistive medical diagnosis from faces. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1–14, 2017.
- [9] LJ Jiang, EYK Ng, ACB Yeo, S Wu, F Pan, WY Yau, JH Chen, and Y Yang. A perspective on medical infrared imaging. *Journal of medical engineering & technology*, 29(6):257–267, 2005.
- [10] Yufeng Zhao, Dong Zhang, and Yaxiang Wang. Automatic location of facial acupuncture-point based on content of infrared thermal image. In *Computer Science and Education (ICCSE), 2010 5th International Conference on*, pages 65–68. IEEE, 2010.
- [11] G Varju, CF Pieper, JB Renner, and VB Kraus. Assessment of hand osteoarthritis: correlation between thermographic and radiographic methods. *Rheumatology*, 43(7):915–919, 2004.
- [12] WT Chiu, PW Lin, HY Chiou, WS Lee, CN Lee, YY Yang, HM Lee, MS Hsieh, CJ Hu, YS Ho, et al. Infrared thermography to mass-screen suspected sars patients with fever. *Asia Pacific Journal of Public Health*, 17(1):26–28, 2005.
- [13] EFJ Ring and K Ammer. Infrared thermal imaging in medicine. *Physiological measurement*, 33(3):R33, 2012.
- [14] Eddie Y-k Ng and Rajendra U Acharya. Remote-sensing infrared thermography. *IEEE Engineering in Medicine and Biology Magazine*, 28(1):76–83, 2009.
- [15] Ramin Irani, Kamal Nasrollahi, Abhinav Dhall, Thomas B Moeslund, and Tom Gedeon. Thermal super-pixels for bimodal stress recognition. In *Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on*, pages 1–6. IEEE, 2016.
- [16] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [19] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.