Neural Network Based Architecture for Fatigue Detection Based on the Facial Action Coding System

Mihai Gavrilescu(✉) and Nicolae Vizireanu

Department of Telecommunications, University "Politehnica" of Bucharest,
1-3 Iuliu Maniu Blvd, 06107 Bucharest 6, Romania
mike.gavrilescu@gmail.com, warticol@gmail.com

Abstract. We present a novel non-invasive neural network based three layered system for detecting fatigue by analyzing facial expressions evaluated using the Facial Action Coding System. We analyze 16 Action Units pertaining to eye and mouth regions of the face. We define an Action Units map containing Action Unit intensity levels for each frame in the video sequence and we analyze this map in a pattern recognition task via a feed-forward neural network. We show that emotion-induced frontal face recordings offer more information in the training stage, while for testing stage the random dataset can be used with no major impact on accuracy, specificity and sensitivity. We obtain over 88% accuracy in intra-subject tests and over 83% for inter-subject tests and we show that our system surpasses the state-of-the-art in terms of accuracy, specificity, sensitivity and response time.

Keywords: Neural networks · e-health · Bioengineering
Facial expression recognition · Image processing

1 Introduction

Most research conducted in the area of facial expression recognition is focused on face recognition or emotion detection. The current paper takes another challenge, proposing a novel non-invasive neural-network based system for fatigue detection using the Facial Action Coding System (FACS). Although this task has been researched in other papers, the current research novelty is in how the architecture is designed as well as the use of feed forward neural networks for this task. Such a system would prove useful in a variety of tasks as monitoring the physical fatigue of a subject reveals the health condition of the person, but also can be used for real-time driver fatigue detection, fatigue being one of the main reasons for car accidents around the globe.

In the following chapter we will present the state-of-the-art in the area of fatigue detection based on facial features.

O. Fratu et al. (Eds.): FABULOUS 2017, LNICST 241, pp. 113–123, 2018.
https://doi.org/10.1007/978-3-319-92213-3_18

2 Related Work

Although the vast majority of research papers in the area of facial expression recognition are focused on face recognition or emotion detection, there are several research papers where facial features were used for fatigue detection.

Researchers in [1] use Facial Action Coding System (FACS) evaluated using Gabor filters with different frequencies and orientations and classified by means of a cascaded AdaBoost with the purpose of determining driver's fatigue. The proposed method is tested on a database based on ground truth information and it shows promising results in the context of a driver monitoring system. Similarly [2] presents a novel method for detecting daily fatigue using color consistent area correction in the preprocessing phase to reduce the environment illumination. Two kinds of color spaces are determined in the grey level co-occurrence matrix and a backward propagation algorithm is used for detection. The accuracy of the system reaches up to 92.3% with a self-built image database.

Kawamura et al. [3] study the changes of luminance in facial images to determine fatigue. Because these changes are usually influenced by vital signs such as heart rate and blood pressure, the level of fatigue is considered predictable with high accuracy by combining these features with the changes of luminance in the facial area. 13 facial parts are used to estimate subject's fatigue using feature values based on luminance changes for each facial part. The results show accuracy of up to 92%. In [4], facial videos acquired in a realistic environment, with natural lighting, where subjects are allowed to voluntarily move their head were used in order to determine physical fatigue. Facial feature point tracking method was used by combining a "good feature to track" and "supervised descent method". The experimental results show the proposed system outperforms video-based existing systems for physical fatigue detection. Similarly, in [5] a fatigue monitoring system is presented which analyzes eye blinking, head nod and yawning. The method employed to extract facial characteristics in time and frequency domain is mean-variance, for eye blink a Haar-like cascade classifier is used, while for yawning the Canny Active Contour method is employed. The testing of this system shows promising results.

Optical imaging through digital cameras installed on car dashboard is another method used for detecting driver fatigue [6]. The camera detects and tracks the driver face and a non-contact photoplethysmography (PPG) method is applied to get multiple physiological signals (brainwave, cardiac and respiratory pulses) which are used for measuring fatigue levels. These are assessed by studying the alteration of facial feature such as eye, mouth, and head. In order to extract information from the facial features, supervised descent method (SDM) with scale-invariant feature transform (SIFT) is used, while, for classifying the fatigue levels, support vector machine (SVM) methods are employed. [7] presents another research that aims detecting driver fatigue levels by evaluating facial features based on statistical local features, Local Binary Patterns (LBP) being used for person-independent fatigue facial expression recognition. The research shows that LBP features perform stably and robustly over a broad range of fatigue-affected face images. AdaBoost is employed to learn the most discriminative

fatigue facial LBP features from a pool of LBP features. These Boost-LBP features show better performance than state-of-the-art.

Given the above described state-of-the-art, we are proposing a novel neural-network based system for determining fatigue levels based on facial features analyzed by means of the Facial Action Coding System. Details on the proposed architecture are presented in the next chapter.

3 Proposed Architecture

As previously mentioned, the current paper proposes a novel neural network-based system for detecting fatigue based on facial features collected via Facial Action Coding System (FACS).

The Facial Action Coding System (FACS) [8] is a framework developed by Eckman and Freisen which divides the face into a set of Action Units (AUs) that are correlated with the activity of different facial muscles. The AUs can be additive (meaning that if a specific AU is triggered it will determine the trigger of another AU) or non-additive (meaning that the triggering of a specific AU is independent from the triggering of other AUs). FACS has showed very good results in determining hidden emotions and we are using it in a fatigue detection task as it offers more reliability compared to other methods of analyzing the face.

In order to achieve the task of detecting fatigue based on FACS, we design an architecture on three layers and we will present each layer in the following paragraphs. The architecture is also depicted in Fig. 1.

The base layer has the main purpose of acquiring facial features from each region of the face and determine if a specific AU is present or not, and, if present, at which intensity. For this the video frame containing the frontal face is normalized and the face is detected by means of Viola-Jones face detection algorithm, then the same algorithm is used for detecting the face components. We analyze only 16 out of the 46 FACS AUs, choosing only the ones known to convey important information for detecting fatigue (mostly linked to eyes and mouth) in order to avoid overcomplicating the architecture as well as overfitting the neural network, therefore in the base layer three components are being detected: Eye, Brow, and Mouth. For each of these components we use specific classifiers to determine the presence/absence and intensity of the AUs pertaining to that specific region such as:

- *Eye component:* Gabor jets-based features have been successfully used for analyzing the eye features providing classification rates of over 90% as well as fast convergence, surpassing other state-of-the-art methods [9]. Because of these strong points, we use them in our work as well, alongside with Support Vector Machines (SVMs) for the AU classification task. The AUs classified in this component are: *AU5* (Upper Lid Raiser), *AU7* (Lid Tightener), *AU43* (Eyes Closed), *AU45* (Blink).
- *Brow component:* We use again the same Gabor Jets with Support Vector Machines (SVMs) method as the one used for the Eye component. The AUs classified in this component are: *AU1* (Inner Brow Raiser), *AU2* (Outer Brow Raiser), *AU4* (Brow Lowerer).

– *Mouth component:* We use active contour classifiers [10] to classify the AUs pertaining to this component. The AUs classified are: *AU8* (Lips toward each Other), *AU10* (Upper Lip Raiser), *AU12* (Lip Corner Puller), *AU15* (Lip Corner Depressor), *AU16* (Lower Lip Depressor), *AU20* (Lip Strecher), *AU23* (Lip Tightener), *AU25* (Lips Part), *AU28* (Lip Suck).

Each of the AU classifiers are previously trained so that they offer over 90% accuracy in cross database tests on Cohn-Kanade [11] and MMI [12] databases.

The three components presented above will fetch to an *intermediary layer* the presence/absence of a specific AU as well as their intensity levels, as follows: *A – Trace* (classification score between 15 and 30), *B – Slight* (classification score between 30 and 50), *C – Marked and Pronounced* (classification score between 50 and 75), *D – Severe or Extreme* (classification score between 75 and 85), *E – Maximum* (classification score over 85), *O – AU is not present* (under 15 classification score). All these scores will be used to compute an *AU activity map* which will have the following structure: (A1A, A2C, A4A, etc.) where A1A means that the Action Unit AU1 was classified with level A of intensity. This *AU activity map* computed in the intermediary layer will contain a row for each frame from the video sequence, each row describing the intensity scores for the analyzed action units. The map is fetched to the top layer which will take the final decision regarding whether the analyzed subjects shows signs of fatigue or not.

In *the top layer* we use a neural network which analyzes the map built in the intermediary layer, in a pattern recognition task, and based on that it determines if the subject is affected by fatigue or not. Because it's a pattern recognition task in a bottom-up layered architecture without feedback loops, the neural network used is a feed-forward neural network as it is efficient for pattern recognition tasks. The neural network has one input layer, one hidden layer, and an output layer. The input layer contains 30 consecutive rows from the AU activity map, hence it has 450 input nodes which are normalized in the [0, 1] interval, such that level A = 0.2, level B = 0.4, level C = 0.6, level D = 0.8, level E = 0.9, 0 – if AU not present. We choose 30 consecutive rows because we are considering a framerate of 30 frames/second, hence 30 consecutive rows pertain to 1 s of the video sequence which is high enough to catch microexpressions and low enough to avoid overfitting the neural network. The output layer has only one node with a binary result, 0 meaning that the subject doesn't show signs of fatigue while 1 means that signs of fatigue are detected. As backpropagation shows the best performance and fast convergence in pattern recognition tasks [13] we employ it as a method for training the neural network. The neural network activation function is determined to be the log sigmoid after trial-and- error. Also through trial and error, trying to minimize the Average Absolute Relative.

Error (AARE), the optimal number of hidden nodes is determined to be 780. Gradient descent algorithm is used for learning the weights and biases of the neural networks until AARE is as low as 0.005. The optimal learning rate is 0.5 and the optimal momentum is 0.02. 50000 training epochs are needed to train the system and it took an average of 3 h to complete on an Intel i7 testbed. Nguyen-Widrow weights initialization is used to evenly distribute the initial weights for each neuron in the input layer.

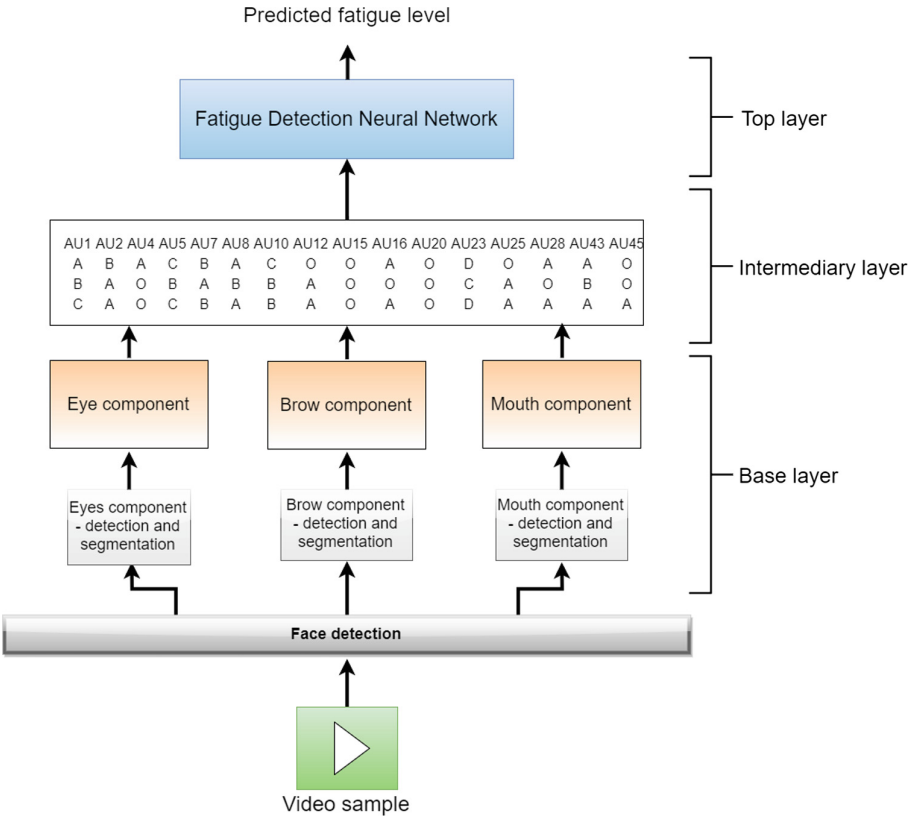

Fig. 1. Overall architecture.

The platform used for implementing the architecture is Scala (as programming language) using Spark MLib library. We used a Java Virtual Machine (JVM) and Eclipse as Integrated Development Environment (IDE). The application has a complexity of around 40.000 code lines and it took an average of 1.5 h to train the feed-forward neural network. The JVM is running on a system with Intel i7 processor, 8 GB of RAM memory, and using Linux Solaris 11.3 as an operating system.

4 Experimental Results

In order to thoroughly test the proposed architecture, we built our own database containing recordings of frontal facial expressions from 64 subjects when watching videos inducing the 6 basic emotions (Sadness, Fear, Happiness, Anger, Surprise, Disgust) from the LIRIS-ACCEDE database [14] as well as in random scenarios. Because the videos existing in the LIRIS-ACCEDE database typically have between 8

and 12 s, we combined more videos for the same emotion in a one minute video-sequence as we needed longer videos for both training and testing. The 64 subjects were recorded six times in three months, every time collecting one frontal face video recording of their reactions when watching each of the six emotion inducing videos (referred to as controlled dataset), as well as five frontal face video recordings in complete random scenarios (when no emotion was induced; referred to as random dataset). They were also asked to self-report their fatigue levels in each session when their face was recorded. Subjects were 32 males and 32 females with ages between 18 and 35, participating in accordance with the Helsinki Ethical Declaration. We have tested the system in both intra-subject and inter-subject methodologies and the results are presented in the following subchapters.

In order to assess the precision of the architecture, we compute sensitivity, specificity and accuracy for all tests conducted. Sensitivity is defined as the proportion of positives that are correctly identified and is calculated as the number of true positives divided by the sum of true positives and false negatives. Specificity is defined as the proportion of negatives that are correctly identified and is calculated as the number of true negatives divided by the sum of true negatives and false positives. Accuracy is calculated as the proportion of true positive and true negative results from the entire number of results.

4.1 Intra-subject Methodology

Intra-subject methodology refers to training and testing the system with samples pertaining to the same subject, but alternating the type of dataset used in training as well as testing stages (controlled, random, or controlled and random). The tests are repeated for all subjects and until all combinations of samples are exhausted. Averaged results are detailed in Table 1. As it can be observed, the higher accuracy is obtained when the controlled dataset is used for both training and testing, specifically when 24 samples are used for training and 12 for testing. In this case the accuracy reached 88.5%, while specificity is 94.5% and sensitivity is 97.2%. We notice that if we change the testing dataset from controlled to random the accuracy of the system shows a decrease of only 2% compared to when random dataset is used for training purposes when the accuracy dropped with 7%. This shows that the controlled dataset offers a lot more important information in the training stage, while for testing any random recordings of the subject can be used, offering similar accuracy, specificity and sensitivity. This is important as it offers the possibility for real-time monitoring of a subject, who will only need to watch emotion inducing videos when he/she first uses the application and further it can be monitored in random situations with over 88% accuracy. In terms of processing time, the time needed to detect fatigue when controlled dataset is used in training and random dataset for testing is no more than six seconds, making the system fast enough to be used for real-time monitoring.

We also conducted a further test for determining which of the six emotions ads the most value to the testing stage in order to detect fatigue with high accuracy. Results are detailed in Table 2.

Table 1. Fatigue detection accuracy, specificity and sensitivity in intra-subject tests.

Type of training samples	Type of test samples	Number of training samples/number of test samples	Accuracy (%)	Specificity (%)	Sensitivity (%)	Average time to converge to a result (s)
Controlled	Controlled	12/24	86.6	93.4	96	6
Controlled	Controlled	18/18	87.4	94	97.2	6
Controlled	Controlled	24/12	88.5	94.5	97.5	5
Random	Random	12/18	78.3	78	76.5	11
Random	Random	18/12	80.5	83	82	11
Random	Random	24/6	81.2	85	84.5	10
Controlled	Random	36/30	88.2	93	96.2	6
Random	Controlled	30/36	81.4	85	84.4	13
Controlled + Random	Controlled + Random	44/22	87.7	92.2	95.8	7

Table 2. Fatigue detection accuracy, sensitivity and specificity for intra-subject tests and for different emotions induced in the training dataset.

Emotion	Accuracy (%)	Sensitivity (%)	Specificity (%)
Happiness	**95.4**	**97.8**	**97.5**
Anger	79.4	80.3	84.2
Fear	**92.3**	**90.2**	94.3
Disgust	**95.4**	**95.2**	**94.1**
Surprise	70	70.1	75.4
Sadness	**95.2**	**98.5**	**98.2**

It can be observed that videos inducing emotions such as Happiness, Disgust and Sadness can be used to detect fatigue with over 95% accuracy, sensitivity, and specificity, while for other emotions results are lower.

4.2 Inter-subject Methodology

For inter-subject methodology we trained the system on multiple subjects and we tested it on a brand new subject, alternating both the number of subjects involved in training and testing as well as the type of datasets used (controlled or random). The tests were repeated using a leave-one-out approach until all combinations of subjects were exhausted. Averaged results are detailed in Table 3.

As it can be observed, results are similar with the ones obtained in intra-subject tests, in the sense that the highest accuracy is obtained when controlled dataset is used in both training and testing stages, specifically when 63 subjects were used in training and the system was tested on the remaining one, case when we obtained over 84% accuracy, and over 88% sensitivity and specificity. We make the same observation as in intra-subject tests that changing the test dataset to the random one only reduces the accuracy with 1%, as opposed to changing the training dataset to random when the accuracy is reduced with 8%. This shows again that the controlled dataset ads more value in the training stage, which is important if we consider building this application

Table 3. Fatigue detection accuracy, sensitivity and specificity in inter-subject tests.

Type of training samples	Type of test samples	Number of subjects involved in training/number of test subjects	Accuracy (%)	Sensitivity (%)	Specificity (%)	Average time to converge to a result (s)
Controlled	Controlled	32/32	81.2	86	86.2	9
Controlled	Controlled	48/16	83.3	87	87.2	8
Controlled	Controlled	63/1	84.3	88.5	88.4	8
Random	Random	32/32	73.3	76.1	75.4	18
Random	Random	48/16	75.2	78	77.4	16
Random	Random	63/1	76.2	79.2	78.8	15
Controlled	Random	63/1	83.2	87.4	86.8	9
Random	Controlled	63/1	76.4	80	80.2	15
Controlled + Random	Controlled + Random	63/1	82.5	85.5	84.5	10

for real-time monitoring, as the emotions need to be induced only when the application is first used, while it can be further assessed in totally random scenarios, completely ad-hoc.

The highest time needed to converge to a result is nine seconds when controlled dataset is used for training and random dataset for testing, which makes the approach fast and attractive for fatigue detection in real-time monitoring systems.

We have conducted a similar test as in intra-subject methodology to determine which emotion ca better be used to detect fatigue, and we reach similar results as in intra-subject tests, reaching over 89% accuracy, and over 90% specificity and sensitivity for Happiness, Disgust and Sadness, while for other emotions the accuracy is lower. These results are detailed in Table 4.

Table 4. Fatigue detection accuracy, sensitivity and specificity in inter-subject tests and for different emotions induced in the training dataset.

Emotion	Accuracy (%)	Sensitivity (%)	Specificity (%)
Happiness	**89.3**	**97.8**	**97.5**
Anger	77.2	75	78
Fear	**88.5**	**86.7**	89.2
Disgust	**91.2**	**90.2**	**90.5**
Surprise	68.4	63	64.2
Sadness	**89.4**	**93.4**	**93**

4.3 Comparison with State-of-the-Art

We have tested the other methods used in [2, 3, 7] on our dataset and our approach offered higher accuracy, sensitivity and specificity than the state-of-the-art as well as faster convergence. Results are detailed in Table 5.

Table 5. Comparison with state-of-the-art.

Work	Year	Method used	Accuracy (%)	Time to compute results (seconds)
[2]	2017	Color consistent area correction	85.3	15
[3]	2017	Luminance changes	86.2	16
[7]	2015	SVM with Boost- LBP features	83	25
Current work	2017	Feed-Forward Neural Network	*Intra-subject:* 88% *Intra-subject (emotion controlled testing):* 95% (Happiness, Sadness, Disgust) *Inter-subject:* 83% *Inter-subject (emotion controlled testing):* 89% (Happiness, Sadness, Disgust)	6–15

As it can be observed, our approach using a feed-forward neural network for determining the fatigue levels surpasses in terms of accuracy with 3% the results obtained when using color consistent area correction [2] and with 2% the results obtained when luminance changes [3] are used. Our approach also offers up to 5% more accuracy compared to the methods where SVM with Boost-LBP features [7] were used. In terms of execution time, our approach is faster than all other methods in state of the art, the time to compute results being lower than 15 s.

5 Conclusions

We presented a non-invasive neural network based system for fatigue detection by analyzing facial expressions acquired by means of the Facial Action Coding System. We only analyze 16 Action Units which are considered to be linked to fatigue and we propose a three layered architecture such that the base layer determines the facial action unit presence and intensity level, the intermediary layer builds a map containing AU details for each frame in the video sequence, and the top layer contains a feed-forward neural network trained to detect fatigue by analyzing the map built in the intermediary layer in a pattern recognition task.

We describe the database constructed by recording 64 subjects in both controlled (emotion is induced) and random (no emotion is induced) scenarios as well as self-reports of their fatigue level for each recording session. We have tested the system in both intra-subject and inter-subject methodologies and we have shown that emotion-induced frontal face recordings offer more information in the training stage, while for testing stage the random dataset can be used without impacting the accuracy, specificity

and sensitivity of the system too much. This is an important observation, as users will only have to watch emotion inducing videos when they first use this application, while further real-time monitoring can be done ad-hoc, in random scenarios. We obtain over 88% accuracy, over 93% specificity and over 96% sensitivity in intra-subject tests and over 83% accuracy, over 87% sensitivity and over 86% specificity for inter-subject tests when controlled dataset is used for training and random dataset for the testing stage. Results are computed in no more than 9 s, making such system fast and attractive for real-time monitoring applications. We have tested other methods from the state-of-the-art on our own database and have shown that our method surpasses them in terms of accuracy, sensitivity, specificity as well as response time. We have also analyzed which emotion used in the testing stage can offer the highest accuracies for fatigue detection and we concluded that these are Happiness, Disgust and Sadness, offering over 95% accuracy for intra-subject tests and over 89% accuracy for inter-subject tests. This information can be used to further tune the system by focusing on these three emotions for achieving higher accuracy which will be the direction of our future work.

References

1. Koon, L.Y., Suandi, S.A.: AU measurements from cascaded adaboost for driver drowsiness detection. In: 2013 8th IEEE Conference on Industrial Electronics and Applications (ICIEA), June 2013
2. Chen, J., et al.: Fatigue detection based on facial images processed difference algorithm. In: 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), February 2017
3. Kawamura, R., Takemura, N., Sato, K.: Mental fatigue estimation based on luminance changes in facial images. In: IEEE International Symposium on Systems Integration (SI), February 2017
4. Haque, M.A., Irani, R., Nasrollahi, K., Moeslund, T.B.: Facial video-based detection of physical fatigue for maximal muscle activity. IET Comput. Vis. **10**(4), 323–329 (2016)
5. Giao, Y., Zeng, K., Xu, L., Yin, X.: A smartphone-based driver fatigue detection using fusion of multiple real-time facial features. In: 2016 13th IEEE Annual Consumer Communications and Networking Conference (CCNC), March 2016
6. Tayibnapis, I.R., Koo, D.Y., Choi, M.K., Kwon, S.: A novel driver fatigue monitoring using optical imaging of face on safe driving system. In: 2016 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), January 2017
7. Zhang, Y., Hua, C.: Driver fatigue recognition based on facial expression analysis using local binary patterns. Optik – Int. J. Light Electron Opt. **126**(23), 4501–4505 (2015)
8. Ekman, P., Friesen, W.V.: Facial Action Coding System: Investigator's Guide. Consulting Psychologists Press, Palo Alto (1978)
9. Mikhail, M., Kaliouby, R.E.: Detection of asymmetric eye action units in spontaneous videos. In: 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, pp. 3557–3560 (2009)
10. Liu, X., Cheung, Y.M., Li, M., Liu, H.: A lip contour extraction method using localized active contour model with automatic parameter selection. In: 20th International Conference on Pattern Recognition (ICPR), pp. 4332–4335, August 2010

11. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 97–115 (2001)
12. Pantic, M., Valstar, M.F., Rademaker, R.: Web-based database for facial expression analysis. In: Maat, L. (ed.) Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2005), pp. 317–321, July 2005
13. Xiaoyuang, L., Bin, Q., Lu, W.: A new improved BP neural network algorithm. In: 2nd International Conference on Intelligent Computation Technology and Automation, pp. 19–22, October 2009
14. Baveye, Y., Dellandrea, E., Chamaret, C., Chen, L.: LIRIS-ACCEDE: a video database for affective content analysis. IEEE Trans. Affect. Comput. **6**(1), 43–55 (2015)