







Emotion Analysis in Hospital Bedside Infotainment Platforms Using Speeded up Robust Features

A. Kallipolitis^{1,2} , M. Galliakis^{1,2} , A. Menychtas^{2,3} ,
and I. Maglogiannis¹ 

¹ Department of Computer Science and Biomedical Informatics,
University of Thessaly, Volos, Greece

{nasskall, michaelgalliakis, imaglo}@unipi.gr

² Department of Digital Systems, University of Piraeus, Piraeus, Greece
amenychtas@bioassist.gr

³ BioAssist S.A, Kastritsiou 4, 26504 Rion, Greece

Abstract. Far from the heartless aspect of bytes and bites, the field of affective computing investigates the emotional condition of human beings interacting with computers by means of sophisticated algorithms. Systems that integrate this technology in healthcare platforms allow doctors and medical staff to monitor the sentiments of their patients, while they are being treated in their private spaces. It is common knowledge that the emotional condition of patients is strongly connected to the healing process and their health. Therefore, being aware of the psychological peaks and troughs of a patient, provides the advantage of timely intervention by specialists or closely related kinsfolk. In this context, the developed approach describes an emotion analysis scheme which exploits the fast and consistent properties of the Speeded-Up Robust Features (SURF) algorithm in order to identify the existence of seven different sentiments in human faces. The whole functionality is provided as a web service for the healthcare platform during regular Web RTC video teleconference sessions between authorized medical personnel and patients. The paper discusses the technical details of the implementation and the incorporation of the proposed scheme and provides initial results of its accuracy and operation in practice.

Keywords: Healthcare platforms · Affective Computing ·
Hospital bedside infotainment systems · WebRTC ·
Speeded Up Robust Features (SURF) · Emotion analysis

1 Introduction

While the relation between the psychological status of human beings and their health was acknowledged in numerous studies [4–6] in the past years, conventional medicine failed to exploit this notion. In practice, it is only recently that medical experts, in parallel with the routine treatment, are investing in the improvement of the emotional status of their patients to reinforce the effects of provided therapy. Towards the same direction, bioinformatics researchers are investigating methods to better interpret,

distinguish, process and quantify sentiments from various human expressions (body posture [3], speech [1], facial expression [2]), all summarized in what is called Affective Computing (AC) or Artificial Emotional Intelligence. Depending on the source of the human expression, affective computing is divided in three main categories: a. Facial Emotion Recognition (FER), b. Speech Emotion Recognition (SER), c. Posture Emotion Recognition (PER).

The importance of affective computing systems is highlighted by the engagement of many IT colossi (Google [7], IBM [9], Microsoft [8]) to implement systems of real-time affective analysis of multimedia data depicting human faces and silhouettes. As far as healthcare platforms are concerned, integrating equivalent schemes in systems which bear the responsibility of monitoring and managing patients' biosignals is of great significance to the healing procedure, especially in the case of chronic diseases. In brief, the generation of positive emotions assists in keeping the patient in a stable psychological condition, which is the basis for fast and efficient treatment [32], whereas negative ones have the opposite effect. Apart from the integration of affective systems in healthcare platforms, rapid development of emotional AI techniques has been reported in a wide range of areas, namely Virtual Reality, Augmented Reality, Advanced Driver Assistance and Smart Infotainment as part of a general trend leading towards the alignment with human-centered computing.

In this paper, we describe the design and deployment of a FER system, incorporated in a healthcare management system as a web service to provide functionalities through the entire lifecycle of the Medical Staff – Kinsfolk – Patient interaction. Motivated by the improved results that a treatment can have when combined with the psychological management of the patient, this system will provide the ability of real time measurement and quantification of the patient's emotions for the medical staff to assess and act upon. Moreover, correlating the emotion measurements with health-related markers collected by the system may lead to important newly discovered knowledge.

The remainder of this paper is structured in 6 sections, as follows: Sect. 2 presents the related research works, while Sect. 3 describes the proposed emotion analysis system architecture. Section 4 describes the system in practice and Sect. 5 reports the experiments conducted and the corresponding results. Finally, Sect. 6 concludes the paper.

2 Related Work

As stated earlier, the analysis, recognition and evaluation of human sentiment via pattern recognition techniques does not rely solely in the processing of facial expressions, but in the quantification of body posture and speech as well. Focusing on SER, several approaches have been proposed in the literature for the extraction of vocal features and their exploitation in forming appropriate classifying models. Methods based on the extraction of low-level features like raw pitch and energy contour [11, 12] are outperformed by high level features utilizing Deep Neural Network [13] to an extend of 20% better accuracy. PER is the least examined territory related to the field of AC. The interpretation of human emotions from body posture in an attempt to assist

individuals that suffer from autism spectrum disorder is described in [14], while an approach based on theoretical frameworks investigates the correlation between patterns of body movement and emotions [15]. On the other hand, FER methodologies vary from the exploitation of Deep Belief Networks combined with Machine Learning Data Pipeline features [16], the utilization of a Hierarchical Bayesian Theme Model based on the extraction of Scale Invariant Feature Transform features [17], the capitalization of Online Sequential Extreme Learning Machine method [18] to Stepwise Linear Discriminative Analysis with Hidden Conditional Random Fields [19]. In addition, hybrid implementations of all the above-mentioned approaches that combine FER and SER are available in the literature to complete the puzzle of Affective computing methodologies [20].

In general, affective computing has been widely deployed in the blooming field of electronic healthcare. As examples of such applications, patients' breathe is managed via emotion recognition carried out by Microsoft Kinect sensor in [21], while in [22] sentiments are analyzed via a facial landmark detecting algorithm from patients suffering from Alzheimer. Another application in electronic healthcare systems is the detection of potential Parkinson patients by recognizing facial impairment when certain expressions are formed with the generation of specific emotions [23].

An innovative notion concerning healthcare solutions is the hospital bedside infotainment systems (HBIS). These systems are designed to enhance medical staff - patient communication and promote patients' clinical experience. Comprising internet, video, movies, radio, music, video or telephone chatting with authorized personnel or kinsfolk and biosignals monitoring in one device and connected to the Electronic Health Record (EHR), it can be proven a productive tool for healthcare ecosystems [10]. Furthermore, constant monitoring of patients can assist in the improvement of their health status and lead to early detection of potential setbacks like detection of outliers [33], poor medication adherence, changes in sleep habits.

Despite the fact that HBIS and emotion analysis services exist as stand-alone cloud-based applications, the combination of the aforementioned advances in a platform is a newly breed idea with positive effects concerning the timely intervention of specialists and kinsfolk when negative emotions or depression is detected.

3 System Architecture

3.1 Overview

The FER Restful web service is built to provide functionality as an additional feature of an existing hospital bedside infotainment system and assisted living solution [24]. The target group of this system are patients who suffer from chronic diseases or are obliged to stay in rehabilitation centers for long periods due to the reduced mobility. Another group of people affected are the elders who live independently or in far regions and conduct routine medical teleconsultations with doctors and caregivers [25]. Although the existing system provides numerous features like the monitoring of patients'

biosignals through a mobile application while conducting measurements via wearables and Bluetooth enabled devices as illustrated in Fig. 1, the contribution of this paper is focused on the real-time video communication functionality through which the ability of communication with their medical experts and kinsfolk in a 24/7 basis is rendered. The FER service operates in parallel with the video communication functionality and is called upon request of the doctor. As mentioned earlier, the importance of automated analysis of facial emotion expression is high, especially to patients and elderly people whose health status is strongly connected to their psychological condition and emotion management. In reference to the FER service, it is divided in two modules: (a) the face extraction module, (b) the emotion recognition module. The face extraction module takes place in the web browser of the client side, while the emotion recognition module occurs on the cloud platform (server side).

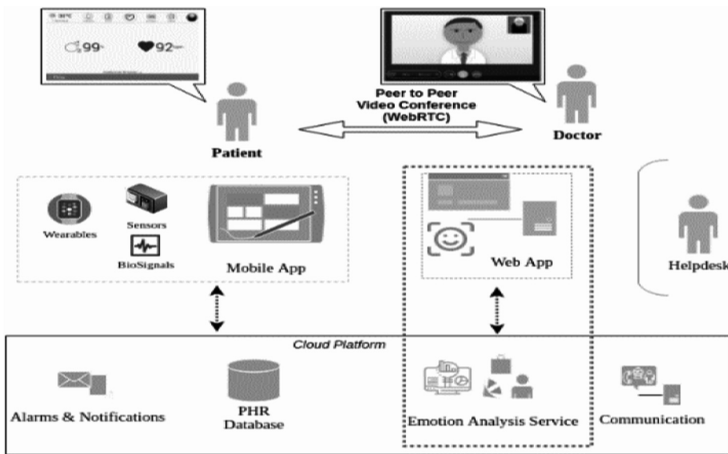


Fig. 1. Overview of the homecare platform after the integration of the FER service

3.2 The Emotion Analysis Process

In general, the basic skeleton of methodologies related to FER consists of five steps: (a) Preprocessing of images, (b) Face's acquisition, (c) Landmarks acquisition (if necessary) (d) Facial Feature extraction, (e) Facial Expression classification. The proposed method, specifically, comprises six steps as described in the pseudocode in Fig. 2 and as follows: (a) frame extraction from the real-time streaming video, (b) face detection, (c) cropping of picture to the dimensions of the detected face (Fig. 4), (d) resizing the face picture to 256×256 pixels (if needed), (e) analysis of the face picture for emotions, (f) presentation of the emotions to the medical expert during the video conference, (g) storage of generated results to the patient's personal health record. The analysis of facial images and their classification in seven different

sentiments (anger, disgust, happiness, neutral, sadness, surprise, fear) is accomplished by the extraction of Speed Up Robust Features (SURF) which form a k dimensional vector as a result of the Bag of Words technique to the extracted features. Given a collection of r images, an algorithm that extracts local features is utilized to create the visual vocabulary (Visual Vocabulary). In our case the Speeded Up Robust Features (SURF) algorithm [28] extracts n 64-dimensional vectors where n is the interest points which are automatically detected by the algorithm and, in turn, described by using a Fast Hessian Matrix (SURF Descriptor) in each one of the r images (Fig. 4). Upon completion of the feature extraction process from the r images, a collection of $r \times n$ 64-dimensional vectors is formed, which represent corresponding points in a 64-dimensional space. This collection is grouped using a clustering algorithm (Kmeans ++ is utilized) in k groups. The centroid of each group represents the visual word, resulting in the formation of a visual vocabulary of k visual words. The process of extracting SURF features is implemented utilizing ImageJ [26], face detection is based on the OpenCV library [27], while the processes of clustering and classification are using the WEKA tool [29].

```
while (videoconference) {
  frame ← extractFrame ()
  If (detectFace(frame)) {
    face ← cropFace(frame)}
  resizedFace ← resize (face, 256, 256)
  emotions ← analyzeFace(resizedFace)
  f (emotions) {
    showEmotions(emotions)
    storeEmotions(emotions)}}}
```

Fig. 2. Emotion Analysis Process as pseudocode. Green color is indicative of the functions performed client-side, red color shows the operations performed server-side (Color figure online)

The emotion recognition service is called during a video call (Fig. 5). A sequence of image frames (1 frame per second) is captured during the WebRTC video conference. In order to avoid additional overload on the network, the Face detection module is executed locally on the web browser. Cropping the image to a face bounding box reduces the amount of data being sent from client to server which in turn results to overall improved performance of the system. This is accomplished by the utilization of the recent implementation of OpenCV library in JavaScript, which provides the functionality of OpenCV models in the JavaScript runtime environment of web browsers.



Fig. 3. Jaffe Image Dataset depicting seven different emotion expressions

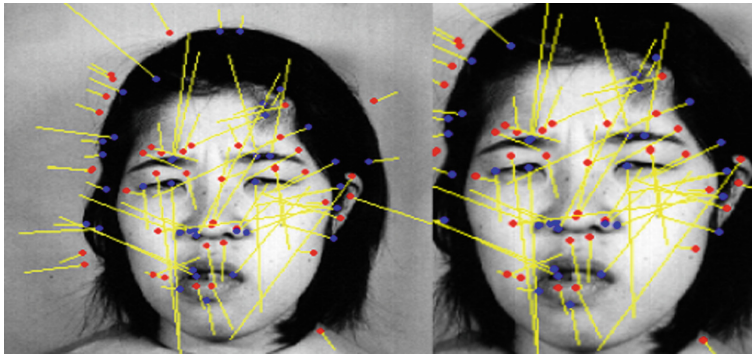


Fig. 4. (left) Interest Points detected in dataset image utilizing Speeded Up Robust Features-imageJ, (right) Cropped image utilizing haar cascade classifier-opencv

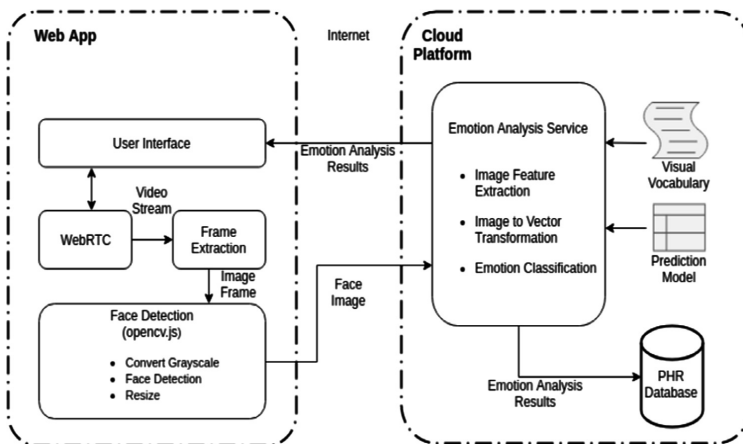


Fig. 5. Emotion analysis operational scenario

4 The System in Practice

The functionality of the proposed solution takes place transparently as far as the users are concerned and upon selection of the medical experts. This provides the discreet capability of monitoring and registering emotional status of the patients while performing a regular video conference ‘visit’ (Fig. 6).

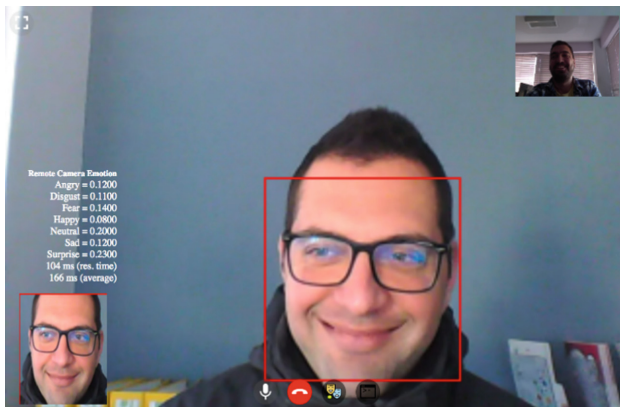


Fig. 6. Medical expert interface showing the FER results of the patient during a video conference

The results of FER are returned from the cloud service in JSON format (Fig. 7) and consequently, visualized in the user interface.

```
{prediction": {  
  angry": 0.12  
  disgust": 0.11,  
  fear": 0.14,  
  happy": 0.08,  
  neutral": 0.20  
  sad": 0.12,  
  surprise": 0.23},  
status": 0}
```

Fig. 7. Sample response of the FER Service

Testing the system in practice was performed by conducting 50 video sessions of 1-min duration. In these sessions, the client-side burden was handled by a PC with Quad core Intel Core i5-7400, while the server-side (cloud services) was deployed to an IaaS Cloud environment with two cores of Intel Xeon CPU E5-2650. Necessary internet connection for the communication along the two sides was provided by a

typical 24 Mbps ADSL connection. The images that are captured by the camera had a resolution was set to 640×480 pixels. Average time in milliseconds for basic operations conducted in client side and server side are measured and depicted in Tables 1 and 2 respectively. Average time allotment for the uploading of the cropped image file from client to server is 70 ms. Observation of the measurements in both sides demonstrates that the most time-consuming operation is the feature extraction from the cropped image (server side), followed by the uploading of the image file in the client side. In addition, operations performed in the server side are far more expensive in time than those in client side, which was expected and strategically planned for the discharge of all the computationally demanding tasks from the web browser.

Table 1. Average time allotment for performing operations in the client side

FER operations (Client Side)					
Frame extraction	Face detection	Face bounding box generation	Face cropping	Image file creation	Time (sum)
10 ms	105 ms	3 ms	5 ms	6 ms	129 ms

Table 2. Time allotment for performing operations in the server side

Web service invocation	FER operations (Server Side)					
	Feature extraction	Visual vocabulary loading (only once)	Image to vector transformation	Prediction model loading (only once)	Image classification	Time (sum)
Beginning of session	200 ms	178 ms	142 ms	157 ms	1 ms	678 ms
All other calls	200 ms	–	142 ms		1 ms	243 ms

Further experimentation on the requirement of running face detection on the front end is presented in Table 3. The Table illustrates the produced overhead for network, browser's memory and CPU for the two scenarios, one for the image size set to 320×240 indicated as s (s for small) and the other set to 640×480 accordingly indicated with (l).

Table 3. Resources allotment for face detection on the web browser.

	Face detection specs (Front End)				
	Idle (l)	Frame (l)	Face (l)	Frame (s)	Face (s)
Memory	3.9 GB	4.1 GB	4.1 GB	4.0 GB	4.0 GB
CPU	~30%	~33%	~45%	~33%	~45%
Data	–	~400 KB	~65 KB	~122 KB	~18 KB

When idle, image is processed in 640×480 ; therefore, idle for (s) scenario does not exist. The experiment was conducted using the Mozilla Firefox browser (version 0.66), but the module also operates in Opera 58.0.3135.117 (64 bit) and Google Chrome 73.0.3683.86 (64-bit) without any issues. The experiments demonstrated that memory consumption is insignificantly influenced in all scenarios, whereas large variations are evidenced in data length as expected.

5 Experimental Results

While the main objective of this paper is the presentation of integration of a FER web service into a homecare platform, initial results for two scenarios of the classification of the JAFFE [30] dataset are provided. The first scenario splits the dataset into two emotional categories (positive and negative emotions, an assumption is made that anger, fear, disgust, sadness are negative emotions) and the second scenario seven emotional categories (anger, fear, disgust, neutral, happiness, surprise, sadness-Fig. 3) are provided with the utilization of various classifiers. The procedure is conducted following the 10-fold cross-validation of the whole JAFFE dataset (214 images, 256×256 pixels, grayscale). In order to discover the more efficient space representation of the training dataset, extensive testing of the Bag of Visual Word scheme was conducted. Kmeans++ method (350 clusters, 70 seeds) was selected among Kmeans, Canopy and Farthest First WEKA's implementations for its ability to better distinguish inter-class and intra-class relationships. Kmeans++ improves the initialization phase of the Kmeans clustering algorithm by selecting strategically the initial seeds [31].

Table 4. Classification accuracy results for JAFFE dataset

Classifier	Experiments	
	Scenario 1 (Two emotional categories)	Scenario 2 (Seven emotional categories)
Random forest	92,01%	77,46%
Naïve Bayes	87,79%	54,92%
MLP	93,48%	73,23%
Sequential minimum optimization	93,42%	74,64%
Logistic	92,01%	71%
K Star	93,42%	84,03%
LMT	89,67%	67,43%

The accuracy of the emotion detection module is provided in Table 4. A Multilayer Perceptron (learning rate: 0.3, momentum rate: 0.2, epoch number: 500, threshold for number of consecutive errors: 20) reaches a 93,48% classification accuracy for the first scenario while the selection of K Star classifier (manual blend: 20%, value missing replaced with average) from the Weka library, achieves the best accuracy (84,03%) for the second scenario.

6 Conclusion

Whereas other affective computing systems operate as stand-alone applications, this paper presents an innovative facial emotion recognition web service, integrated in a healthcare information system for monitoring and timely management of emotional fluctuations of elders and patients with chronic diseases as part of a human-centric treatment. The value of the provided functionality to classify faces into corresponding sentiments real-time during video communication sessions is of great significance especially in cases of patients with diseases related to their psychosomatic condition. Future work will be focused in the realization of a service that can execute emotion recognition in the web browser. This feature will liberate the application from its cloud based imposed restrictions. Concerning the classification performance towards the improvement of accuracy of the current prediction model, other schemes of Bag of Words techniques will be tested in an effort to provide weighted and localized information of the Visual words. Although results are promising, further testing with the utilization of larger and Caucasian oriented labeled dataset should be performed towards more thorough evaluation of the system. Correlating emotion recognition results along with information related to the biosignals and everyday routine activities of individual patients can lead to the discovery of specific patterns and valuable knowledge to the medical community.

Acknowledgment. This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (SISEI: Smart Infotainment System with Emotional Intelligence, project code: T1EDK-01046).

References

1. Gunawan, T., Alghifari, M.F., Morshidi, M.A., Kartiwi, M.: A review on emotion recognition algorithms using speech analysis. *Indonesian J. Electr. Eng. Inf.* **6**, 12–20 (2018)
2. Ko, B.C.: A brief review of facial emotion recognition based on visual information. *Sensors* **18**(2), 401 (2018)
3. Dael, N., Mortillaro, M., Scherer, K.: Emotion expression in body action and posture. *Emotion* **12**, 1085 (2011). <https://doi.org/10.1037/a0025737>
4. DuBois, C.M., Lopez, O.V., Beale, E.E., Healy, B.C., Boehm, J.K., Huffman, J.C.: Relationships between positive psychological constructs and health outcomes in patients with cardiovascular disease: a systematic review. *Int. J. Cardiol.* **195**, 265–280 (2015). <https://doi.org/10.1016/j.ijcard.2015.05.121>. ISSN 0167-5273
5. Burger, A.J., et al.: The effects of a novel psychological attribution and emotional awareness and expression therapy for chronic musculoskeletal pain: a preliminary, uncontrolled trial. *J. Psychosom. Res.* **81**, 1–8 (2016)
6. Huffman, J.C., Millstein, R.A., Mastromauro, C.A., et al.: *J. Happiness Stud.* **17**, 1985 (2016)
7. Google Cloud Vision API Homepage: <https://cloud.google.com/vision/>
8. Microsoft Cognitive Services Homepage: <https://azure.microsoft.com/en-us/services/cognitive-services/>

9. IBM Watson Visual Recognition Homepage: <https://www.ibm.com/watson/services/visual-recognition/>
10. Dale, Ø., Boysen, E.S., Svagård, I.: One size does not fit all: design and implementation considerations when introducing touch-based infotainment systems to nursing home residents, computers helping people with special needs. In: Miesenberger, K., Bühler, C., Penaz, P. (eds.) ICCHP 2016. LNCS, vol. 9758, pp. 302–309. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41264-1_41
11. Schuller, B., Rigoll, G., Lang, M.: Hidden markov model-based speech emotion recognition. In: Proceedings of IEEE ICASSP 2003, vol. 2, pp. I–II. IEEE (2003)
12. Nwe, T.L., Hieu, N.T., Limbu, D.K.: Bhattacharyya distance based emotional dissimilarity measure for emotion classification. In: Proceedings of IEEE ICASSP 2013, pp. 7512–7516. IEEE (2013)
13. Han, K., Yu, D., Tashev, I.: Speech emotion recognition using deep neural network and extreme learning machine. *Interspeech* **2014**, 223–227 (2014)
14. Libero, L.E., Stevens, C.E., Kana, R.K.: Attribution of emotions to body postures: an independent component analysis study of functional connectivity in autism. *Hum. Brain Mapp.* **35**, 5204–5218 (2014)
15. Dael, N., Mortillaro, M., Scherer, K.R.: Emotion expression in body action and posture. *Emotion* **12**, 1085–1101 (2012)
16. Uddin, M.Z., Hassan, M.M., Almogren, A., Zuair, M., Fortino, G., Torresen, J.: A facial expression recognition system using robust face features from depth videos and deep learning. *Comput. Electr. Eng.* **63**, 114–125 (2017)
17. Mao, Q., Rao, Q., Yu, Y., Dong, M.: Hierarchical Bayesian theme models for multipose facial expression recognition. *IEEE Trans. Multimed.* **19**(4), 861–873 (2017)
18. Cossetin, M.J., Nievola, J.C., Koerich, A.L.: Facial expression recognition using a pairwise feature selection and classification approach. In: 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016, pp. 5149–5155. IEEE (2016)
19. Siddiqi, M.H., Ali, R., Khan, A.M., Park, Y., Lee, S.: Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Trans. Image Process.* **24**(4), 1386–1398 (2015)
20. Ekman, P.: Facial expression and emotion. *Am. Psychol.* **48**(4), 384 (1993)
21. Dantcheva, A., Bilinski, P., Broutart, J.C., Robert, P., Bremond, F.: Emotion facial recognition by the means of automatic video analysis. *Gerontechnol. J. Int. Soc. Gerontechnol.* **15**, 12 (2016)
22. Tivatansakul, S., Chalumporn, G., Puangpontip, S., Kankanokkul, Y., Achalaku, T., Ohkura, M.: Healthcare system focusing on emotional aspect using augmented reality: emotion detection by facial expression. In: *Advances in Human Aspects of Healthcare*, vol. 3, p. 375 (2014)
23. Almutiry, R., Couth, S., Poliakoff, E., Kotz, S., Silverdale, M., Cootes, T.: Facial behaviour analysis in parkinson’s disease. In: Zheng, G., Liao, H., Jannin, P., Cattin, P., Lee, S.-L. (eds.) MIAR 2016. LNCS, vol. 9805, pp. 329–339. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-43775-0_30
24. Menychtas, A., Tsanakas, P., Maglogiannis, I.: Automated integration of wireless biosignal collection devices for patient-centred decision-making in point-of-care systems. *Healthc. Technol. Lett.* **3**(1), 34–40 (2016)
25. Panagopoulos, C., et al.: Utilizing a homecare platform for remote monitoring of patients with idiopathic pulmonary fibrosis. In: Vlamos, P. (ed.) GeNeDis 2016. AEMB, vol. 989, pp. 177–187. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57348-9_15
26. ImageJ Homepage: <https://imagej.net>

27. Bradski, G., Kaehler, A.: *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media Inc, Sebastopol (2008)
28. Bay, H., Tuytelaars, T., Gool, V.G.: Speeded up robust features. *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008)
29. Weka 3, Data Mining Software in Java Homepage: <https://cs.waikato.ac.nz/ml/weka>
30. Lyons, M.J., Akemastu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
31. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*, Philadelphia. Society for Industrial and Applied Mathematics, pp. 1027–1035 (2007)
32. Chakhssi, F., Kraiss, J.T., Sommers-Spijkerman, M., Bohlmeijer, E.T.: The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatry*. **18**(1), 211 (2018)
33. Fouad, H.: Continuous health-monitoring for early detection of patient by web telemedicine system. In: *International Conference on Circuits, Systems and Signal Processing*, 23–25 September 2014. Saint Petersburg State Politechnical University, Russia (2014)