Facial expression detection using Viola-Jones algorithm in the learning environment

Hadhami Aouani (Hadhamiaouani22@gmail.com)
MIRACL, University of Sfax
Yassine Ben Ayed
MIRACL, University of Sfax

Research Article

Keywords: Face detection, Viola-Jones, emotion, HOG, SVM

Posted Date: April 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2822747/v1

Additional Declarations: No competing interests reported.

Facial expression detection using Viola-Jones algorithm in the learning environment

Hadhami Aouani[1] and Yassine Ben Ayed[1]

[1]Multimedia InfoRmation systems and Advanced Computing Laboratory, MIRACL , University of Sfax, Sfax, Tunisia.

Contributing authors: Hadhamiaouani22@gmail.com; yassine.benayed@isims.usf.tn;

Abstract

Emotion recognition using facial expression is an active research topic in the field of computer vision. In this paper, our system is based on a three-step approach, namely face detection, feature extraction and classification. Face detection takes photos/videos for information and finds face areas in these images. Facial extraction finds important highlights positions (eyes, mouth, nose and ocular temples) within a distinguished face using the Viola-Jones algorithm. In order to extract the faces, we have built a database of face images. We propose two systems: our first facial emotion recognition system supports the classification of the raw face inputs, the second extracts the Histogram of Oriented Gradient (HOG) from the face image. We use Support Vector Machines (SVM) for the classification phase. The experiments are conducted at the Ryerson Multimedia Laboratory (RML) dataset.The results of our experiments showed good accuracy compared to previous studies.

Keywords: Face detection, Viola-Jones, emotion, HOG, SVM

1 Introduction

The face is the most visible part of the human body, with which the person's identity can also be detected, as well as their age or even their gender. As many studies on facial recognition show, facial emotion recognition is increasing and becoming familiar. Scientific research has shown that facial emotion recognition can be used in a wide range of applications. Many companies offering a wide range of consumer products are leveraging facial emotion recognition technologies to recognize customer feedback and identify potential customers. These facial expressions become more valuable to the field of machine learning. In the field of education, facial emotions are used to determine the level of understanding of students. Teachers have the freedom to teach their students in the way that suits them best. They teach the courses using different platforms, digitally or traditionally. The main goal is that at the end of each lesson, the students understand the lesson. By recognizing facial emotions, teachers can assess whether or not students have followed the class. Based on the feedback determined using facial emotion patterns, the teacher can modify her strategy in presenting the lesson. Expression recognition systems, especially emotional ones, consist of three main steps [1]:

1. Face Acquisition: This step can be done by face detection or by head pause estimation.
2. Extraction of facial information related to expression: this information is related either to the appearance of the expression or to the geometry of the deformations. In the context of an image sequence, information related to expression dynamics is also useful.
3. Recognition of the expression: this step is carried out by means of a classification.

As part of our work, we focus on face detection from a video database, followed by descriptor extraction and finally emotion recognition.

In this article, we first proposed the use of Viola and Jones for face detection from video frames, and then considered these faces as inputs to our classification system. In order to make a comparison we extract from the detected faces the Histogram of Oriented Gradients (HOG) characteristic. We used Support Vector Machines (SVM). Our system is evaluated on the RML database.

The remainder of the article is organized as follows: Section 2 presents the most recent studies on facial emotion recognition. Section 3 describes the methods of our proposed system. In section 4, the experimental results are presented. And finally section 5, we conclude our work.

2 Related works

Facial Emotion Recognition (FER) is an important part of human-computer interaction that allows computers to understand facial expressions based on human thought. According to the processing of facial expression recognition process, the facial expression recognition process can be divided into three important modules, face detection, feature extraction and classification. The face detection as the main technology face recognition [2, 3] with its rapid development it has a mature basis, which can be effectively extracted from the original facial image of the excellent features and correct classification features become a major factor affecting the recognition result. For example, Gao.al [4] obtained facial expression traits from facial images to predict emotional states based on changes in facial expressions.

Facial expression feature approaches fall into two categories: geometry and appearance. Initial attempts at facial expression recognition are based on the geometry of the face, as well as position, distance, and angles [5, 6, 7]. However, geometric feature-based focusing methods rely on accurate detection of face components and are difficult to apply under various real-time conditions. Appearance-based methods have overcome the problems of geometric feature-based methods. To extract edits that describe changes in facial texture, intensity, histograms, and pixel values. These methods apply an image filter to all or part of the face. Principal Component Analysis (PCA) [8], Independent Component Analysis (ICA) [9], Gabor Wavelet [10], and Local Binary Pattern (LBP) [11] are the algorithms considered to extract the descriptors of precise facial appearance.

Noroozi et al. [12] proposed a multimodal emotion recognition system where visual geometric features are merged with acoustic features using these random forest and Convolutional Neural Network(CNN)classifiers, SVM. The recognition rate for RML and SAVEE database is 31.67% and 36.10% using the SVM classifier.

On the other hand in [13], the authors present a framework in which the Hidden Markov Model (HMM) is applied to the active appearance characteristics to carry out the FER. They achieved an accuracy of 89.04% on the RML dataset.

In another study, H. Noushin et al [14] use the SVM classifier with Geometric Features: Whole Face Region, Eyes and Eyebrows Region, Nose and Mouth Region extracted from RML database with respective recognition rate 36.06%, 39.76% and 37.51%.

In this work, we first propose a system based on the face detected by Viola-Jones as a geometric feature and make the first emotion recognition system secondly we extract the face HOG features detected for the second system. SVM classification is used.

3 Proposed method

The general architecture diagram of the proposed FER system is shown in Fig.1.The main steps are face detection, feature extraction, and SVM classification.

3.1 Face detection

Face detection is the computer's ability to recognize a face. The Haar feature-based cascade classifier is simple and robust, which makes it a very famous face detection model [15]. The De Haar Cascade was used to detect the mouth and eyes only [16].

The Viola Jones algorithm is also a famous face detection model, proposed by Paul Viola and Michael Jones in 2001 [17]. It uses rectangular features to identify the human face in the image [18].

The input image is converted to a grayscale image. Face detection performed on a grayscale image using Viola-Jones algorithm. The steps of the Viola-Jones algorithm are shown in Fig 2 Viola Jones object detector is based on a binary classifier that produces a positive output when the search window consists of the desired object, otherwise it returns a negative output. The classifier can be used multiple times as the window slides over the tested image.

The binary classifier used in the algorithm is realized using several hierarchy layers that form an ensemble classifier [19]. Such classifier works by classifying images based on the value of simple features. It is observed that this works much faster than a system that bases the classification on a pixel-based system [4]. The Viola Jones algorithm exercises

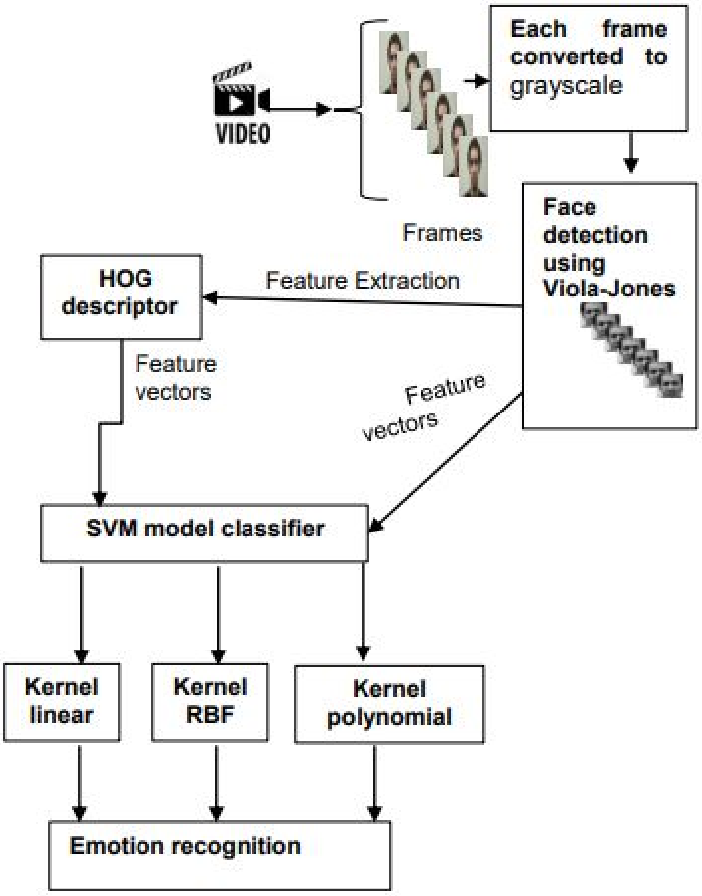
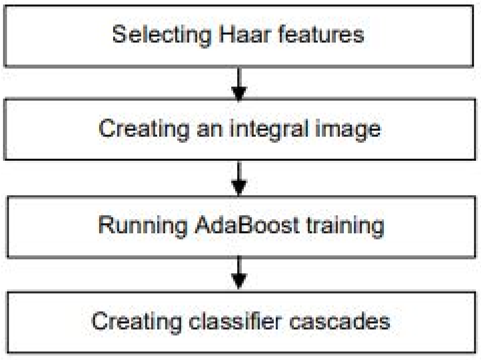


Fig. 1 Architecture of our system of emotion recognition.

control over three features dictated by Viola et al. in [17] namely, the functionality of two rectangles, the functionality of three rectangles and the functionality of four rectangles. The framework proposed by the group has the following steps:

- The Haar feature selection: is calculated using Haar basis functions which are based on the three features listed above and usually include the summation of the pixels of the adjacent rectangular areas involved, and then calculates the difference between these sums. A representation of the Haar characteristics with respect to the corresponding detection window is shown in Fig 3
- The integral image is then created and is used to evaluate rectangular features in a constant time. Since the number of features can vary greatly,
- The Adaboost or Adaptive Boosting algorithm is used to select the best features and to train the classifiers using them. This is responsible for creating a "strong"



Selecting Haar features
Creating an integral image
Running AdaBoost training
Creating classifier cascades

Fig. 2 The steps of the Viola-Jones algorithm.

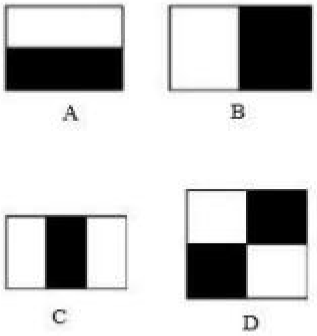


Fig. 3 Representation of the rectangular features displayed in relation to the detection window.

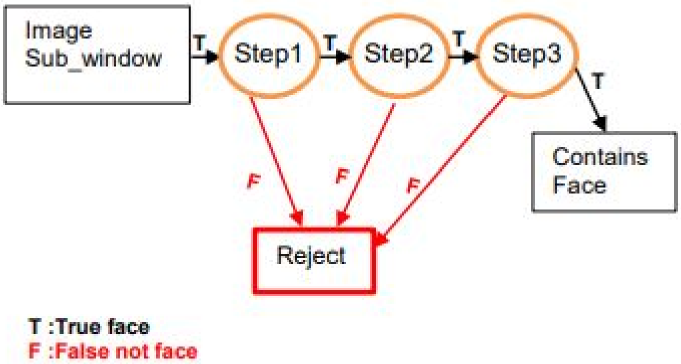


Fig. 4 Representation of classifier work flow in the Viola-Jones algorithm.

classifier which is considered as a linear weighted combination of simple ”weak” classifiers.

- Finally, in cascade, each step consisting of “strong” classifiers is grouped into several steps. Each step is responsible for determining whether a sub-window consists of a face or not, as shown in Fig.4.

3.2 Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) find applications in the field of object and pattern recognition as they are able to extract crucial information even from images obtained in scrambled environments [20]. Initially it was introduced by Dalal and Triggs [20] to find out things. Counts the number of times the gradient trend occurs in the local image correction. In this article, each 64x64 size image is divided into overlapping 8x8 blocks. This generates 196 blocks of size 8×8. Each 8x8 block is then represented by a 9-dimensional uniform pattern histogram to describe each image block. Then, these extracted features are concatenated into a single block to form a 1764-dimensional feature vector for the final facial appearance.

3.3 Support Vector Machines

Support Vector Machines (SVM) was introduced in 1964 and has developed rapidly since the 1990s with a series of improved and extended algorithms. The SVM is one of the most widely used algorithms for automatic classification, especially for emotion classification for global optimization. It is mainly used for binary classification, but can also be applied to multi-class detection. Indeed, multi-class SVMs are used in several domains and have proven to be effective in identifying the different classes of data presented to it [21] .
It is a supervised machine learning technique that is used for classification as well as regression. It attempts to classify the data by finding the appropriate hyperplane that can separate the data by the highest margin, that is, the best separation of the training data projected into the feature space by a kernel function K, the most used kernel functions, such as linear, polynomial, RBF, based on the training sets, the new values are separated and analyzed. Thus, the use of this classification method is essentially to select good kernel functions and adjust the parameters to achieve maximum identification rate. We will use the SVM with these three kernel functions, so that:

- Linear:
$$k(x_i, x_j) = x_i^T x_j \tag{1}$$
- Polynomial:
$$k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \tag{2}$$
- RBF:
$$k(x_i, x_j) = exp(-\gamma(\| x_i - x_j \|)^2), \gamma > 0 \tag{3}$$
With: d: degree of the polynomial, r: weighting parameter (used to control the weights), γ: kernel flexibility control parameter.

Then, the adjustment of the different parameters of the SVM classifier is done empirically, each time we change the type of SVM kernel to determine the values γ, r, d and c which are values chosen by the user, in order to find the most suitable kernel parameters for our research.

4 Experimental results and discussion

In our work, we used the RML[22] emotion database which contains 720 sample audiovisual emotional expressions that were collected at the Ryerson Multimedia Lab. Six basic human emotions are expressed: Anger, Disgust, Fear, Happy, Sad, and Surprise. The database was collected from subjects speaking six different languages such as English, Mandarin, Urdu, Punjabi, Persian and Italian). We only use the English language the number of audio become 241 audiovisual extracts (70% dedicated to learning and 30% to tests). For each video, we extract the images and do the face detection by the Viola- Jones. We obtain our new base which contains a total of 6797 faces. The experiments are performed in the Python environment. We perform the first facial emotion recognition system with these 64*64 dimensional images as feature vectors using SVM Fig.5.The second FER is the same but we propose to extract the HOG characteristic of the face detected image Fig. 6

The Table 1 presents a summary of the best recognition rate found for the three SVM kernels for our first and second system FER. The results obtained show that the RBF kernel gives the best performance compared to the linear for the two systems but the same results with polynomial and RBF for the second system with the HOG characteristic. The best results for different emotions using the kernel RBF of SVM are summarized in the following Table 2

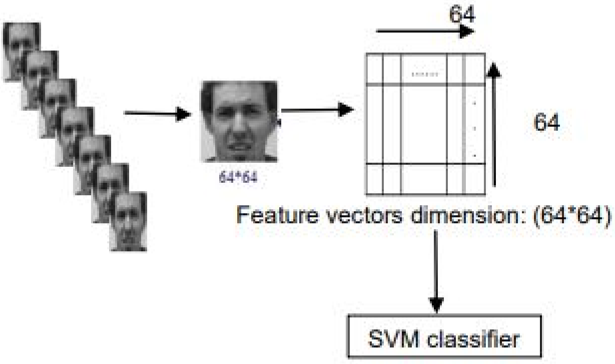


Fig. 5 Our first system of facial emotion recognition.

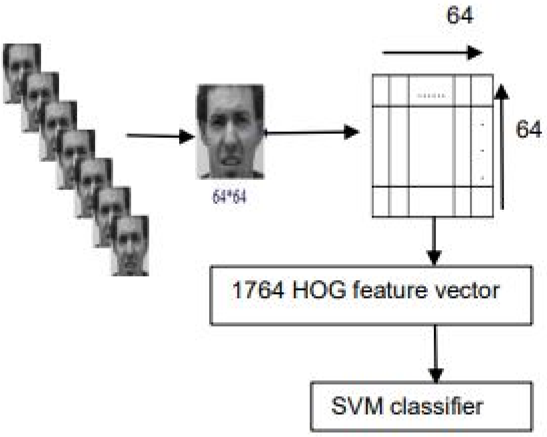


Fig. 6 Our second system of facial emotion recognition.

Table 1 The recognition rates on the test corpus obtained with the SVM for our two systems.

	Kernel Linear	Kernel Polynomial	Kernel RBF
The first System	97,30%	98,23%	98,82%
The second system	96,96%	99,46%	99,46%

Table 2 The recognition rates on the test corpus obtained with the SVM for our two systems.

Systems	First system :4096 features	Second system: HOG
Emotion		
Angry	97.85%	98.57%
Disgust	100%	99.75%
Fear	97,85%	98.77%
Happy	99.41%	100%
Sad	99.41%	99.70%
Surprise	97.92%	99.70%

Table 3 shows the effectiveness of our two proposed systems in identifying facial emotions, as it outperforms other advanced techniques.

Table 3 The recognition rates on the test corpus obtained with the SVM for our two systems.

State of the art	Methods and features	Accuracy(%)
Avots et al.[23],(2019)	frame-based emotion recognition obtained using CNN using RML database	60.20
Xianzhang Pan [24],(2020)	SVM using RML database	
	CNNs	
	for extracting useful information from each original image frame	
	(displacement, scale and deformation invariance features)+HOGfeatures	65.12
H.Noushin et al [14],(2021)	Geometric characteristics:	
	Whole Face Region	36.06
	Eyes and eyebrows region	39.76
	Nose and mouth region	37.51
	SVM using RML database	
Sharafi et al.[25],(2022)	Spatial and temporal:	
	convolutional neural networks(Deep Temporal Network (DTN),	
	Deep Spatial Network (DSN)),	97.33
	Bidirectional Long Short-Term Memory network(BiLSTM),	
	RML dataset	
Our proposed system	SVM	
	A facial Expression Vector(EV) of dimension 4096	98.82
	1764 dimension HOG feature of face image	99.46

5 Conclusion

In this article, we have presented the performance of our proposed systems based on the face image we propose based on one as entered in the SVM model and the other we will extract HOG features. The results of the two systems show that the proposed algorithm is robust. Our comparative study of the four emotion classification systems shows the effectiveness of our propositions.
In the future, we can think of using other types of features and apply our system on other broader bases and use methods for the reduction of the dimension of the features like the autoencoder, finally we can also consider to perform emotion recognition using an audiovisual base and in this case to benefit from descriptors of speech and others of the image. This allows us to improve the recognition rate of each emotion.

Declarations

- Funding : This research received no external funding.
- Conflict of interest/Competing interests: The authors declare no conflict of interest.
- Ethics approval: not applicable.
- Availability of data and materials:not applicable.
- Authors' contributions: All authors have read and agreed to the published version of the manuscript.

References

[1] Yingli Tian, Takeo Kanade, and Jeffrey. Cohn. "Facial Expression Recognition". In: *Handbook of Face Recognition*. Ed. by E. Zaimis. Vol. 42. Verlag: Springer, 1/2011, pp. 487–519. DOI: 10.1007/978-0-85729-932-1_19.

[2] Adjabi Insaf et al. "Past, Present, and Future of Face Recognition: A Review". In: *Electronics* 9 (July 2020), p. 1188. DOI: 10.3390/electronics9081188.

[3] Liping Zhang; Linjun Sun; Lina Yu; Xiaoli Dong; Jinchao Chen; Weiwei Cai; Chen Wang; Xin Ning. "ARFace: Attention-Aware and Regularization for Face Recognition With Reinforcement Learning". In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* (Jan. 2022).

[4] Hao Gao and Bo Ma. "A Robust Improved Network for Facial Expression Recognition". In: *Frontiers in Signal Processing* 4 (Oct. 2020). DOI: 10.22606/fsp.2020.44001.

[5] Annabelle Redfern. "Fernández-Dols, J.-M., Russell, J. A. (Eds.). The Science of Facial Expression Fernández-DolsJ.-M., RussellJ. A. (Eds.). The Science of Facial Expression. Oxford, England: Oxford University Press, 2017; 540 pp.: ISBN: 978-0-19-061350-1, £84.00 Hardback." In: *Perception* 47 (Jan. 2018), p. 030100661775196. DOI: 10.1177/0301006617751965.

[6] Yingli Tian, Takeo Kanade, and Jeffrey Cohn. "Facial Expression Recognition". In: Jan. 2011, pp. 487–519. DOI: 10.1007/978-0-85729-932-1_19.

[7] Michel Valstar, Stefanos Zafeiriou, and Maja Pantic. "Facial Actions as Social Signals". In: May 2017, pp. 123–154. ISBN: 9781107161269. DOI: 10.1017/9781316676202.011.

[8] Leonardo Franco and Alessandro Treves. “A Neural Network Facial Expression Recognition System using Unsupervised Local Processing”. In: (Apr. 2001).

[9] Md. Zia Uddin, J.J. Lee, and Tae-Hun Kim. “An Enhanced Independent Component-Based Human Facial Expression Recognition from Video”. In: *Consumer Electronics, IEEE Transactions on* 55 (Dec. 2009), pp. 2216–2224. DOI: 10.1109/TCE.2009.5373791.

[10] G. Hegde. “Subspace based Expression Recognition Using Combinational Gabor based Feature Fusion”. In: *International Journal of Image, Graphics and Signal Processing* 9 (Jan. 2017), pp. 50–60. DOI: 10.5815/ijigsp.2017.01.07.

[11] Sajid Khan, Ayyaz Hussain, and Muhammad Usman. “Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features”. In: *Multimedia Tools and Applications* 77 (Jan. 2018). DOI: 10.1007/s11042-016-4324-z.

[12] Fatemeh Noroozi et al. “Audio-Visual Emotion Recognition in Video Clips”. In: *IEEE Transactions on Affective Computing* PP (June 2017), pp. 60–70. DOI: 10.1109/TAFFC.2017.2713783.

[13] Hernán García, Mauricio Álvarez, and Alvaro Orozco. “Dynamic Facial Landmarking Selection for Emotion Recognition using Gaussian Processes”. In: *Journal on Multimodal User Interfaces* 11 (Nov. 2017). DOI: 10.1007/s12193-017-0256-9.

[14] Noushin Hajarolasvadi, Enver Bashirov, and Hasan Demirel. “Video-based person-dependent and person-independent facial emotion recognition”. In: *Signal, Image and Video Processing* (2021), pp. 1–8.

[15] A. Alanazi . A.S. Aljaloud H. Ullah. “Facial emotion recognition using neighborhood”. In: *Int. J. Adv. Computer Sci. Appl.* 11 (2020), pp. 299–306.

[16] D. Yang et al. “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment”. In: *Procedia Computer Science* 125 (Jan. 2018), pp. 2–10. DOI: 10.1016/j.procs.2017.12.003.

[17] Paul Viola and Michael Jones. “Robust Real-Time Face Detection”. In: *International Journal of Computer Vision* 57 (May 2004), pp. 137–154. DOI: 10.1023/B:VISI.0000013087.49260.fb.

[18] Emre Dandıl and Ridvan Ozdemir. “Real time Facial Emotion Classification Using Deep Learning”. In: 2 (July 2019), pp. 13–17.

[19] Charles Lo and Paul Chow. “A high-performance architecture for training Viola-Jones object detectors”. In: Dec. 2012, pp. 174–181. ISBN: 978-1-4673-2846-3. DOI: 10.1109/FPT.2012.6412131.

[20] Navneet Dalal and Bill Triggs. “Histograms of Oriented Gradients for Human Detection”. In: vol. 1. July 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177.

[21] Frank Dellaert, Thomas Polzin, and Alex Waibel. “Recognizing Emotion In Speech”. In: *International Conference on Spoken Language Processing, ICSLP, Proceedings* 3 (Dec. 1996).

[22] Yongjin Wang and Ling Guan. “Recognizing Human Emotional State From Audiovisual Signals*”. In: *Multimedia, IEEE Transactions on* 10 (Sept. 2008), pp. 936–946. DOI: 10.1109/TMM.2008.927665.

[23] Egils Avots et al. “Audio-Visual Emotion Recognition in Wild”. In: *Machine Vision and Applications* 30 (July 2019). DOI: 10.1007/s00138-018-0960-9.

[24] Xianzhang Pan. “Fusing HOG and Convolutional Neural Network spatial-temporal features for video-based facial expression recognition”. In: *IET Image Processing* 14 (Jan. 2020). DOI: 10.1049/iet-ipr.2019.0293.

[25] Masoumeh Sharafi et al. “A novel spatio-temporal convolutional neural framework for multimodal emotion recognition”. In: *Biomedical Signal Processing and Control* 78 (2022). DOI: 10.1016/j.bspc.2022.103970.