EMOTION RECOGNITION USING AUTOENCODERS AND CONVOLUTIONAL NEURAL NETWORKS

Luis Antonio Beltrán Prieto, Zuzana Komínková Oplatková

Tomas Bata University in Zlín
Faculty of Applied Informatics
Nám. T. G. Masaryka 5555, 76001 Zlín
Czech Republic
beltran_prieto@fai.utb.cz

Abstract: *Emotions demonstrate people's reactions to certain stimuli. Facial expression analysis is often used to identify the emotion expressed. Machine learning algorithms combined with artificial intelligence techniques have been developed in order to detect expressions found in multimedia elements, including videos and pictures. Advanced methods to achieve this include the usage of Deep Learning algorithms. The aim of this paper is to analyze the performance of a Convolutional Neural Network which uses AutoEncoder Units for emotion-recognition in human faces. The combination of two Deep Learning techniques boosts the performance of the classification system. 8000 facial expressions from the Radboud Faces Database were used during this research for both training and testing. The outcome showed that five of the eight analyzed emotions presented higher accuracy rates, higher than 90%*.

Keywords: *Emotion Recognition, Convolutional Neural Networks, Deep Learning, AutoEncoders*.

1 Introduction

Facial emotion detection is the process of identifying the feeling that a person is expressing at a particular moment. Potential applications of emotion recognition include the improvement of student engagement [1], the design of smart health environments [2], the analysis of customers' feedback [3], and the evaluation of quality in children's games [4], among others. Face recognition within multimedia elements, including images and videos, has been identified as one of the current challenges pursued by artificial intelligence. Several powerful techniques from the Deep Learning field have been recently implemented seeking an improvement in emotion detection, such as Convolutional Neural Networks (CNN) [5], Deep Belief Networks (DBN) [6], and Auto Encoders [7]. Deep Learning is a novel research area in the machine learning territory which focuses on the learning of high-level data representations and abstractions, namely images, sounds, and text by using hierarchical architectures, being neural networks, convolution networks, belief networks, and recurrent neural networks the most well-known in several artificial intelligence areas, for instance image classification [8], speech recognition [9], handwriting recognition [10], computer vision [11], and natural language processing [12].

Identifying the sentiment expressed by a person is one of the side objectives achieved by face detection. Recent research [13] has proven that emotion recognition can be accomplished by implementing machine learning and artificial intelligence algorithms. To assist in this task, several open-source libraries and packages, being OpenCV [14], TensorFlow [15], Theano [16], Caffe [17] and the Microsoft Cognitive Toolkit (CNTK) [18] the most notorious examples, cut down the process of building deep-learning-based algorithms and applications. Emotions including anger, disgust, happiness, surprise, and neutrality can be distinguished.

The aim of this paper is to analyze the performance of a Convolutional Neural Network which uses AutoEncoder Units for emotion-recognition in human faces. The combination of two Deep Learning techniques boosts the performance of the recognition system despite of the complexity introduced by both algorithms. 8000 facial expressions from the Radboud Faces Database were examined in different phases of the experiments for training and evaluation purposes.

This paper is organized as follows. Background information introducing Emotion Recognition, Convolutional Neural Networks, AutoEncoders, and the Radboud Faces Database is presented as part of the theoretical background section. Afterwards, the problem solution is described by explaining the methods and methodology that were used for this comparison. Evaluation results are shown subsequently. Finally, conclusions are presented in the final section.

2 Theoretical Background

2.1 Emotion Recognition

Emotions are a strong representation of feelings about people's situations and relationships with other people. The most basic process a human has to reflect how they feel is by using facial expressions. Speech, gestures, and behavior are also used to describe a person's current state. Emotion recognition can be defined as the process of detecting the feeling expressed by a human being. Basic emotions include anger, happiness, sadness, fear, surprise, and disgust. It has been demonstrated that humans are able to identify facial emotions even since early ages [19]. Following machine learning's objective of imitating human thinking, algorithms have been developed for this purpose. Emotions play a key role in decision-making and human-behavior, as many actions are determined by how a person feels at some point.

Typically, these algorithms consider either a picture or a video (which can be considered as a set of images) as input, then they proceed to detect and focus their attention on a face and finally, specific points and regions of the face are analyzed in order to detect the affective state. Machine Learning algorithms, methods and techniques can be applied to detect emotions from a picture or video. For instance, a deep learning neural network can perform effective human activity recognition with the aid of smartphone sensors [20].

2.2 Convolutional Neural Networks

A convolutional neural network (CNN) is a class of deep, feed-forward artificial neural networks that has been applied in visual images analysis. CNNs were inspired by biological processes in which the connectivity pattern between neurons is influenced by the organization of the animal visual cortex. Individual neurons respond to stimuli only in a restricted region of the receptive field. Receptive fields of different neurons partially overlap to top the entire visual field. CNNs use relatively little pre-processing compared to other image classification algorithms, meaning that the network learns the filters which in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. They have been applied for image and video recognition, recommender systems [21] and natural language processing [22].

A CNN contains three main layers: convolutional layers, pooling layers and fully-connected layers, each one with their own purpose. In the convolutional layers, several kernels convolve, i.e. entwine, an image and intermediate feature maps, creating new feature maps, as shown in Fig. 1. Benefits of convolution include a minimal number of parameters, correlation learning between neighboring pixels due to local connectivity, an unchanged object location, and a faster learning process than for fully-connected layers. Secondly, a pooling layer usually follows a convolutional layer in order to reduce feature maps dimensions. This layer is translation invariant as well because the computation considers neighboring pixels by using strategies such as average pooling or max pooling. As an example, an 8x8 feature map is reduced to a 4x4-dimensional output with a max pooling operator of 2x2 size and 2 for the stride value. Finally, a fully connected layer is similar to a traditional neural network and uses about 90% of the parameters in a CNN. It makes possible to feed forward the neural network into a vector with predefined length and can be used for image classification or follow-up processing.

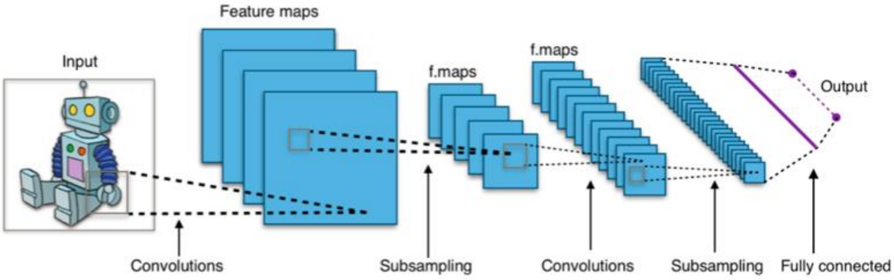

Fig. 1: A Convolutional Neural Network architecture [23]

2.3 AutoEncoders

AutoEncoders are feedforward neural networks where both the input and output are the same; they can be used to reconstruct their own input in a lower dimensional space. By assuming that one image is sent to this type of network, random representations in its center-most hidden layer are generated at first. By feeding the network with more and more similar images, the network will be benefited, developing a unique construct which contains the elements of the subject's face encoded in it. Leveraging that intuition, the concept is that an autoencoder network is capable of learning a specific emotion shape for different classes in the training set. In other words, by feeding the autoencoder network with different images of people smiling, the network would be able to learn that the feature to encode is the emotional distinctiveness of happiness.

2.4 Radboud Faces Database

The Radboud Faces Database (RaFD) is a high-quality set of pictures of 67 models (between male, female, Caucasian, Moroccan, children and adults) displaying 8 emotional expressions (anger, disgust, fear, happiness, sadness, surprise, contempt, and neutrality), as displayed in Fig. 2. The database contains 28,709 faces. This initiative by the Behavioural Science Institute of the Radboud University Nijmegen is available for research purposes upon request. Each emotion was shown looking to three directions (left, front, and right), with five camera angles. However, frontal images were selected only for this research.

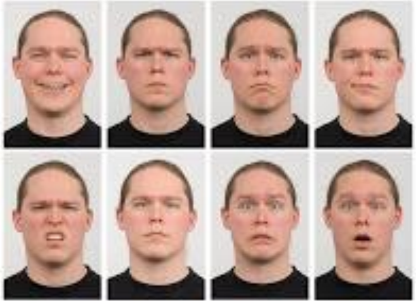

Fig. 2: Sample images from the RaFD

3 Methods And Methodology

This research focuses on the development of a combination of two deep learning techniques, an AutoEncoder used as an input for the Convolutional Neural Network for emotion detection in human faces.

Convolutional autoencoders are a special variant of CNNs which encode their input using convolution into a compact internal representation which is then decoded using de-convolution into the original space, trying to minimize the reconstruction error. Such encoder/decoder pairs can be optionally stacked, as described in [24], to increase the complexity of the internal representations. Past halfway into the first decade of the 21st century [25], research demonstrated that autoencoders could be used in neural networks pre-training and overcome the deep neural networks training obstacles, including: the differences between magnitudes of gradients in the lower and higher layers., the landscape of the objective function is difficult for stochastic gradient descent to find a good local optimum, and the fact that deep networks have various parameters, which can easily recall training data, thus complicating the generalization process.

Pre-training process of a deep network is divided into two steps: Pretraining and Fine-tuning. Pretraining consists of training a sequence of autoencoders, greedily one layer at a time, using unsupervised data. For Fine-tuning, first the last layer is trained by using supervised data. After that, backpropagation is used to fine-tune the entire network using supervised data.

Researchers have demonstrated that this pretraining idea improves deep neural networks as pretraining is done one layer at a time, meaning it does not suffer from the difficulty of full supervised learning. This approach was widely accepted as it resembles the brain’s likeliness for unsupervised learning, which is also more appealing since it does not make use of expensive unlabeled data.

Most of the research in emotion detection focuses on the usage of a deep learning technique alone [26, 27, 28, 29]. The experiment consists of two parts. First, an autoencoder network with one hidden layer containing 300 and 500 nodes was developed, with Rectified Linear Unit (ReLU) designed as the activation function, which accelerates the convergence of gradient descents faster than a sigmoid function. To overcome the ReLU problem, referred as “dead neurons”, i.e., neurons which are never activated across the dataset, a - 0.01 threshold slope was considered. The AutoEncoder network was trained on the RaFD database, as previously mentioned, with 8000 images, all of which were normalized to 48x48 dimensions and grayscale colors, while categorizing each of them per emotion depicted on. The implementation of the AutoEncoder was done using the Tensorflow deep learning framework with Python code.

Afterwards, the generated images were used as input for an 8-layer designed CNN which included three convolutional layers, three pooling layers, and two fully connected layers. For each of the convolutional layers, a 5x5 filter size was used, except for the third convolutional network, which considered a 3x3 filter. The stride value was set to 1, while the filter count was stablished at 512. Between the first two convolutional layers, an average pooling was considered in order to reduce complexity and maximize the low-level features extraction, such as edges. Between the last two convolutional layers, a max pooling over a 3x3 window was considered in order to extract specific features inside the face, such as the mouth, the eyes, and the sentiment in general. Accordingly, the Tensorflow framework was used to implement the CNN in Python.

The Fold cross-validation technique was used during the experiment to get a training set and 10 sets for testing. For each emotion, 10 folds were generated, each one consisting of 240 images: 120 elements included the evaluated sentiment while the rest incorporated the other seven emotions (20 images per sentiment were included in order to balance the distribution between all classes).

For each emotion, we generated a confusion matrix for each fold from the analysis. Examples of results sets are shown in Tables 1 and 2 for Neutral and Contempt emotions, respectively. Furthermore, Tables 3-10 present the statistical measures obtained from the performance of our classification model when analyzing each emotion. The following metrics are exhibited in these tables: Accuracy (Acc), Sensitivity (Sens), Specificity (Spec), Precision (Prec), Negative Predictive Value (NPV), and F1 Score. We considered these calculations as the most significative ones during our analysis.

Table 1: Confusion Matrices for Neutral Emotion analysis

Fold #1 – Neutral		Actual class		**Fold #6 – Neutral**		Actual class	
		Neutral	Non-Neutral			Neutral	Non-Neutral
Prediction	Neutral	130	10	Prediction	Neutral	132	8
	Non-Neutral	5	135		Non-Neutral	7	133
Fold #2 – Neutral		Actual class		**Fold #7 – Neutral**		Actual class	
		Neutral	Non-Neutral			Neutral	Non-Neutral
Prediction	Neutral	127	13	Prediction	Neutral	128	12
	Non-Neutral	11	129		Non-Neutral	6	134
Fold #3 – Neutral		Actual class		**Fold #8 – Neutral**		Actual class	
		Neutral	Non-Neutral			Neutral	Non-Neutral
Prediction	Neutral	129	11	Prediction	Neutral	132	8
	Non-Neutral	15	125		Non-Neutral	4	136
Fold #4 – Neutral		Actual class		**Fold #9 – Neutral**		Actual class	
		Neutral	Non-Neutral			Neutral	Non-Neutral
Prediction	Neutral	136	4	Prediction	Neutral	133	7
	Non-Neutral	2	138		Non-Neutral	6	134
Fold #5 – Neutral		Actual class		**Fold #10 - Neutral**		Actual class	
		Neutral	Non-Neutral			Neutral	Non-Neutral
Prediction	Neutral	131	9	Prediction	Neutral	126	14
	Non-Neutral	13	127		Non-Neutral	10	130

Table 2: Confusion Matrices for Contempt Emotion analysis

Fold #1 – Contempt		Actual class		**Fold #6 – Contempt**		Actual class	
		Contempt	Non-Contempt			Contempt	Non-Contempt
Prediction	Contempt	96	44	Prediction	Contempt	99	41
	Non-Contempt	55	85		Non-Contempt	48	92
Fold #2 – Contempt		Actual class		**Fold #7 – Contempt**		Actual class	
		Contempt	Non-Contempt			Contempt	Non-Contempt
Prediction	Contempt	92	48	Prediction	Contempt	95	45
	Non-Contempt	51	89		Non-Contempt	46	94
Fold #3 – Contempt		Actual class		**Fold #8 – Contempt**		Actual class	
		Contempt	Non-Contempt			Contempt	Non-Contempt
Prediction	Contempt	88	52	Prediction	Contempt	100	40
	Non-Contempt	49	91		Non-Contempt	48	92
Fold #4 – Contempt		Actual class		**Fold #9 – Contempt**		Actual class	
		Contempt	Non-Contempt			Contempt	Non-Contempt
Prediction	Contempt	86	54	Prediction	Contempt	93	47
	Non-Contempt	42	98		Non-Contempt	44	96
Fold #5 – Contempt		Actual class		**Fold #10 - Contempt**		Actual class	
		Contempt	Non-Contempt			Contempt	Non-Contempt
Prediction	Contempt	85	55	Prediction	Contempt	97	43
	Non-Contempt	45	95		Non-Contempt	39	101

Table 3. Statistical measures of the performance of Neutral emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	94.64	96.3	93.1	92.86	96.43	94.55
2	91.43	92.03	90.85	90.71	92.14	91.37
3	90.71	89.58	91.91	92.14	89.29	90.85
4	97.86	98.55	97.18	97.14	98.57	97.84
5	92.14	90.97	93.38	93.57	90.71	92.25
6	94.64	94.96	94.33	94.29	95	94.62
7	93.57	95.52	91.78	91.43	95.71	93.43
8	95.71	97.06	94.44	94.29	97.14	95.65
9	95.36	95.68	95.04	95	95.71	95.34
10	91.43	92.65	90.28	90	92.86	91.3

Table 4. Statistical measures of the performance of Contempt emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	64.64	63.58	65.89	68.57	60.71	65.98
2	64.64	64.34	64.96	65.71	63.57	65.02
3	63.93	64.23	63.64	62.86	65	63.54
4	65.71	67.19	64.47	61.43	70	64.18
5	64.29	65.38	63.33	60.71	67.86	62.96
6	68.21	67.35	69.17	70.71	65.71	68.99
7	67.5	67.38	67.63	67.86	67.14	67.62
8	68.57	67.57	69.7	71.43	65.71	69.44
9	67.5	67.88	67.13	66.43	68.57	67.15
10	70.71	71.32	70.14	69.29	72.14	70.29

Table 5. Statistical measures of the performance of Happiness emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	98.21	97.2	99.27	99.29	97.14	98.23
2	97.5	98.54	96.5	96.43	98.57	97.47
3	97.5	97.16	97.84	97.86	97.14	97.51
4	97.14	97.14	97.14	97.14	97.14	97.14
5	97.86	97.18	98.55	98.57	97.14	97.87
6	97.5	97.84	97.16	97.14	97.86	97.49
7	97.14	96.48	97.83	97.86	96.43	97.16
8	98.21	98.56	97.87	97.86	98.57	98.21
9	96.43	95.77	97.1	97.14	95.71	96.45
10	97.5	96.5	98.54	98.57	96.43	97.53

Table 6. Statistical measures of the performance of Sadness emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	98.21	97.87	98.56	98.57	97.86	98.22
2	97.86	97.18	98.55	98.57	97.14	97.87
3	96.79	96.45	97.12	97.14	96.43	96.8
4	96.79	98.52	95.17	95	98.57	96.73
5	98.21	97.87	98.56	98.57	97.86	98.22
6	98.21	98.56	97.87	97.86	98.57	98.21
7	96.79	97.12	96.45	96.43	97.14	96.77
8	94.64	94.96	94.33	94.29	95	94.62
9	97.86	97.18	98.55	98.57	97.14	97.87
10	97.5	95.86	99.26	99.29	95.71	97.54

Table 7. Statistical measures of the performance of Disgust emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	86.43	85.92	86.96	87.14	85.71	86.52
2	86.43	84.46	88.64	89.29	83.57	86.81
3	83.93	84.67	83.22	82.86	85	83.75
4	85.36	85.61	85.11	85	85.71	85.3
5	84.29	85.82	82.88	82.14	86.43	83.94
6	88.21	87.94	88.49	88.57	87.86	88.26
7	82.5	81.82	83.21	83.57	81.43	82.69
8	85	84.51	85.51	85.71	84.29	85.11
9	84.64	83.45	85.93	86.43	82.86	84.91
10	84.29	83.8	84.78	85	83.57	84.4

Table 8 Statistical measures of the performance of Anger emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	93.93	94.24	93.62	93.57	94.29	93.91
2	93.21	91.72	94.81	95	91.43	93.33
3	91.79	91.49	92.09	92.14	91.43	91.81
4	94.64	93.71	95.62	95.71	93.57	94.7
5	91.79	90.91	92.7	92.86	90.71	91.87
6	94.29	93.06	95.59	95.71	92.86	94.37
7	92.86	93.48	92.25	92.14	93.57	92.81
8	94.64	94.33	94.96	95	94.29	94.66
9	90.71	90.14	91.3	91.43	90	90.78
10	94.29	92.47	96.27	96.43	92.14	94.41

Table 9. Statistical measures of the performance of Surprise emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	92.5	93.43	91.61	91.43	93.57	92.42
2	95	95.65	94.37	94.29	95.71	94.96
3	95.36	94.41	96.35	96.43	94.29	95.41
4	93.57	94.2	92.96	92.86	94.29	93.53
5	92.5	93.43	91.61	91.43	93.57	92.42
6	91.43	92.03	90.85	90.71	92.14	91.37
7	94.64	94.33	94.96	95	94.29	94.66
8	91.43	91.43	91.43	91.43	91.43	91.43
9	91.79	90.91	92.7	92.86	90.71	91.87
10	94.64	94.96	94.33	94.29	95	94.62

Table 10. Statistical measures of the performance of Fear emotion classification

Fold	Acc %	Sens %	Spec %	Prec %	NPV %	F1 %
1	73.93	74.81	73.1	72.14	75.71	73.45
2	74.29	74.64	73.94	73.57	75	74.1
3	73.93	72.79	75.19	76.43	71.43	74.56
4	72.5	71.72	73.33	74.29	70.71	72.98
5	72.86	72.86	72.86	72.86	72.86	72.86
6	76.43	75.69	77.21	77.86	75	76.76
7	71.79	71.03	72.59	73.57	70	72.28
8	74.29	75.37	73.29	72.14	76.43	73.72
9	71.07	71.53	70.63	70	72.14	70.76
10	73.57	73.24	73.91	74.29	72.86	73.76

4 Discussion

In order to get a detailed overview of each emotion, a confusion matrix was generated for each sentiment. Furthermore, Accuracy, Sensitivity, Specificity, Precision, the Negative Predictive Value rate, and the F1 score were the statistical measures of the performance of our model selected for this analysis and included in the tables. As each test is binary, i.e. the predicted sentiment belongs or not to a specific class, a more specific analysis based on the false positive, false negatives true positive, and true negative rates was allowed rather than a simple accuracy report. We detected during he analysis that five of the eight emotions were greatly benefited due to the usage of the proposed combination of deep learning techniques. The surprise, anger, neutral, happiness, and sadness sentiments present sensitivity and specificity values greater than 90%, meaning that the model is able to identify these feelings with a considerable level of trust and also report those emotions who do not belong to these classes. Since accuracy involves both specificity and sensitivity in its calculations, the accuracy for these five emotions are close to 90% as well. In a previous research [30], our findings were no higher than 80% for all emotions accuracy. Another improvement was detected for the disgust feeling, which presents an approximate 85% value in both accuracy and precision. However, contempt and fear feelings got 35% and 30% error rate values -the fraction of misclassified cases- respectively, which means the model is not perfect at all for these sentiments. As explained in [31], action units are taken into consideration in order to detect an emotion. An action unit is a specific characteristic of the face, for instance, the chin raise, the lip press, and the brow lower, among others. Contempt is an emotion with only one action unit: the dimple. However, it can be so small that even a person would be confused whether the expressed sentiment is neutrality or contempt. Likewise, fear and sadness share two action units: the brow lower and inner brow raise, making the distinction between any of the feelings not an easy task. An example of this complication is illustrated by Fig. 3. The following future work will be aimed to solve this. Finally, the F1-score was included as well in the results table as it is considered as the harmonic mean of both precision and accuracy and also is in agreement with our findings.

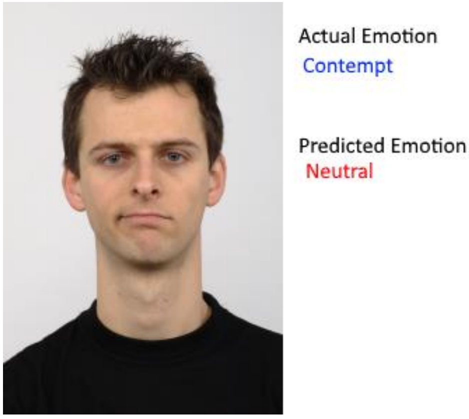

Fig. 3: A contempt emotion mistakenly classified as neutral

5 Conclusions

This paper deals with the results of a research about a proposed combination of two deep learning techniques for emotion detection in human faces. While work has been conducted by using a deep learning algorithm for both face detection and emotion recognition, few analyses implementing more than one technique have been performed. AutoEncoders and Convolutional Neural Networks were the selected models for this research. The achievements include an improvement from our previous work, for instance, 90% and higher accuracy rates for five sentiments were obtained. Several other statistical measures including sensitivity, specificity, and F1-score were also included in the analysis, with values in agreement to the accuracy obtained. Future research will include emotion detection in videos by using the proposed combined deep learning techniques, as well as a broad exploration of the contempt and fear emotions, both of which obtained a low accuracy rate in the current investigation.

Acknowledgement: This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the National Sustainability Programme Project no. LO1303 (MSMT-7778/2014), further by the European Regional Development Fund under the Project CEBIA-Tech no. CZ.1.05/2.1.00/03.0089 and by Internal Grant Agency of Tomas Bata University under the Projects no. IGA/CebiaTech/2018/003. This work is also based upon support by COST (European Cooperation in Science & Technology) under Action CA15140, Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO), and Action IC1406, High-Performance Modelling and Simulation for Big Data Applications (cHiPSet). The work was further supported by resources of A.I.Lab at the Faculty of Applied Informatics, Tomas Bata University in Zlin (ailab.fai.utb.cz). L.A.B.P author also thanks the doctoral scholarship provided by the National Council for Science and Technology (CONACYT) and the Council for Science and Technology of the State of Guanajuato (CONCYTEG) in Mexico.

References

[1] Garn, A. C., Simonton, K., Dasingert, T., Simonton, A.: Predicting changes in student engagement in university physical education: Application of control-value theory of achievement emotions. Psychology of Sport and Exercise (29), 93-102 (2017).

[2] Fernandez-Caballero, A., Martinez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., Lopez, M. T., Zangroniz, R., Latorre, J. M., Fernandez-Sotos, A.: Smart environment architecture for emotion detection and regulation. Journal of Biomedical Informatics (64), 57-73 (2016).

[3] Felbermayr, A., Nanopoulos, A.: The Role of Emotions for the Perceived Usefulness in Online Customer Reviews. Jounal of Interactive Marketing (36), 60-76 (2016).

[4] Gennari, R., Melonio, A., Raccanello, D., Brondino, M., Dodero, G., Pasini, M., Torello, S.: Children's emotions and quality of products in participatory game design. International Journal of Human-Computer Studies (101), 45-61 (2017).

[5] Campos, V., Jou, B., Giró-i-Nieto, X.: From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction. Image and Vision Computing (65), 15-22 (2017).

[6] Mannepalli, K., Sastry, P. N., Suman, M.: A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition. Alexandria Engineering Journal (56), 485-497 (2017).

[7] Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., Li, Y.: Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. Computers in Biology and Medicine (79), 205-214 (2016).

[8] Affonso, C., Rossi, A. L. D., Vieira, F. H. A., Ferreira de Carvalho, A. C. P. de L.: Deep learning for biological image classification. Expert Systems with Applications (85), 114-122 (2017).

[9] Fayek, H. M., Lech, M., Cavedon, L.: Evaluating deep learning architectures for Speech Emotion Recognition. Neural Networks (92), 60-68 (2017).

[10] Roy, S., Das, N., Kundu, M., Nasipuri, M.: Handwritten isolated Bangla compound character recognition: A new benchmark using a novel deep learning approach. Pattern Recognition Letters (90), 15-21 (2017).

[11] Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., Agrawal, A.: Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials (157), 322-330 (2017).

[12] Shin, H., Lu L., Summers, R. M.: Deep Learning for Medical Image Analysis. Academic Press, USA (2017).

[13] Yogesh, C. K., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., Polat, K.: Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech. Applied Soft Computing (56), 217-232 (2017).

[14] OpenCV Homepage, https://opencv.org, [Online; accessed 05-Feb-2018]

[15] Tensorflow Homepage, https://www.tensorflow.org/, [Online; accessed 05-Feb-2018]

[16] Theano Homepage, http://deeplearning.net/software/theano/, [Online; accessed 05-Feb-2018]

[17] Caffe Homepage, http://caffe.berkeleyvision.org/, [Online; accessed 05-Feb-2018]

[18] CNTK Homepage, https://www.microsoft.com/en-us/cognitive-toolkit/, [Online; accessed 05-Feb-2018]

[19] Lawrence, K., Campbell R., Skuse D.: Age, gender, and puberty influence the development of facial emotion recognition. Frontiers in Psychology (6), 1-14 (2015).

[20] Ronao, C. A., Cho, S.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications (59), 235-244 (2016).

[21] Zuo, Y., Zeng, J., Gong, M., Jiao, L.: Tag-aware recommender systems based on deep neural networks. Neurocomputing (204), 51-60 (2016)

[22] Liao, S., Wang, J., Yu, R., Sato, K., Cheng, Z.: CNN for situations understanding based on sentiment analysis of twitter data. Procedia Computer Science (111), 376-381 (2017).

[23] Typical cnn.png. file, https://commons.wikimedia.org/w/index.php?title=File:Typical_cnn.png [Online; accessed 09-Feb-2018]

[24] Masci, J., Meier, U., Cires¸an, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In 21th International Conference on Artificial Neural Networks on Proceedings, pp. 52–59. Springer Berlin Heidelberg, Berlin (2011).

[25] Bengio, Y., Lamblin, P., Popovici, D., Larochelle. H.: Greedy layer-wise training of deep networks. In Advances in Neural Information Processing Systems. MIT Press, USA (2007).

[26] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., Dobaie, A. M.: Facial expression recognition via learning deep sparse autoencoders. Neurocomputing (273), 643-649 (2018),

[27] Mayya, V., Pai, R. M., Pai, M. M. M.: Automatic Facial Expression Recognition Using DCNN. Procedia Computer Science (93), 453-461 (2016).

[28] Pitaloka, D. A., Wulandari, A., Basaruddin, T., Liliana D. Y.: Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition. Procedia Computer Science (116), 523-529 (2017).

[29] Kaya, H., Gürpınar, F., Salah, A. A.: Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image and Vision Computing (65), 66-75 (2017).

[30] Beltrán Prieto, L. A., Komínkova-Oplatková, Z.: A performance comparison of two emotion-recognition implementations using OpenCV and Cognitive Services API. MATEC Web of Conferences (125), 1-5 (2017)

[31] Langner, O., Dotsch, R., Bijlstra G., Wigboldus, D. H. J., Hawk, S. T., van Knippenberg, A: Presentation and validation of the Radboud Faces Database. Cognition & Emotion 24 (8), 1377-1388 (2010)