Using neural and distance-based machine learning techniques in order to identify genuine and acted emotions from facial expressions

RANNOW BUDKE, Jaine and DA COSTA ABREU, Marjory
<http://orcid.org/0000-0001-7461-7570>

Published version

RANNOW BUDKE, Jaine and DA COSTA ABREU, Marjory (2021). Using neural and distance-based machine learning techniques in order to identify genuine and acted emotions from facial expressions. In: 11th International Conference of Pattern Recognition Systems (ICPRS 2021). IET, 121-126.

Using neural and distance-based machine learning techniques in order to identify genuine and acted emotions from facial expressions

Jaine Rannow Budke[1] **and Márjory Da Costa-Abreu**[2]

[1]Federal University of Rio Grande do Norte, Natal, BR
[2] Department of Computing, Sheffield Hallam University, Sheffield, U.K.

Keywords: Face analysis, emotion prediction, machine learning

Abstract

Facial expressions are part of human non-verbal communication. Automatically discriminating between genuine and acted emotion can help psychologists, judges, human-machine interface, and so on. The problems for researchers starts when there are few real emotion facial datasets available, and thus, most of experimentation for evaluation is done by using fake emotions from actors. Thus, this paper explores the problem of classifying emotions from facial expressions as genuine or acted. We propose to extract facial features from images and to classify using k-Means, k-Nearest Neighbor and Neural Network. The best results obtained presented a promising 98.6% of precision for happiness emotion and 92% for sadness emotion.

1 Introduction

Human communication can be done on verbal or non-verbal form. People from all cultures use non-verbal communication to express themselves, which includes posture types, voice modulation, facial expression and gestures [1] as part of expressing themselves. It is something that we internalise during our life. Facial expression recognition is, therefore, a fundamental characteristic of humans that helps on communication process.

There are several different ways of performing face analysis in an automated manner. One of the most popular is using what we call 'intelligent systems'. Intelligence systems are developed with aim on the automation of human tasks. In this context, there are many solutions on the literature for automatic identification of human emotions [2], [3], [4], [5], [6], [7]. On the other hand, there are situations that emotions can be necessarily not genuine, they can be acted, and have few classifiers to differentiate emotions from acted to genuine [8], [9], [10].

There are many important reasons to identify when emotions are genuine or acted, such as judging whether a person is lying or being honest [11], exploring the authenticity of the user experience [12], improving human-machine interface [13], [14], helping the psychologist with patients with manipulative symptom disorders [15], and so on.

However, the lack of datasets containing real emotions from facial images can be a big problem when developing systems that are going to attempt to classify real emotions. By looking at the literature, it is very hard to identify a facial dataset that has real emotions and most of the papers published in this area use fake emotions performed by actors, which might not represent the wide range of features of real emotions.

In this paper, we aim to investigate the features of facial expression that can be used to identify acted from genuine emotion with the goal to explore the possibility of developing a unified system to classify facial expressions of the same emotion (happy or sad) as genuine or acted.

This paper is organised as follows: in Section 2, we describe related works in emotions recognition; in Section 3, we explain our method and the datasets used; in Section 4, we present the algorithms used; in Section 5, we discuss the results and, finally, in Section 6, we conclude this paper.

2 The current literature on emotion prediction from face images

In general, the main topics on the literature regarding emotions are about facial expression recognition systems. Most of the approaches use deep learning models to train a dataset with acted emotions and to classify in one of the six universal emotions proposed by [16].

The work presented in [2], the authors proposed a facial expression recognition in real time on the six universal emotions based on detecting the face in an image using Viola Jones algorithm, applying the Supervised Descent Method (SDM) to detect facial points and measuring deformation considering the first and last frames, and finally using the backpropagation neural network technique to classify the data.

In [3], it was used Generative Adversarial Networks (GAN) and Convolutional Neural Networks (CNN) to analyse the same six emotions. They used three datasets with acted emotions and compare the results. Similarly, [4] extracted regions of interest with the face detection algorithm Multi-task CNN and trained the model with Deep Alignment Network (DAN). As for [5], they applied a Haar-Cascade filter to crop the input image faces and uses Convolutional Neural Network (CNN) to classify the six emotions in real time.

[6] suggested the use of the iPar-CLR method to find facial landmarks and apply euclidean distance at the points to detect the face muscle movements. The objective was to clas-

sify the emotions in real time on the six emotions using the database JAFFE and MUG. Based on the features obtained, they used Ensemble Neural Network (ENN), Multilayer Perceptron (MLP) and the supervised learning model Support Vector Machine (SVM).

[7] described a collected dataset wherein the participants were provoked by external stimuli (music and image), eliciting one of the emotions: fear, happiness and sadness. The image processing was conducted with face detection using Haar-like features with Adaboost algorithm, facial characteristics points were marked based on the geometric model, feature values were calculated based on facial geometric points and then action units are used to inference the muscle movements, classifying the emotions on the images.

Also, there were some studies on fake emotions identifying. A previous work conducted by [8] collected a database selecting Youtube videos where each emotion included 20 videos with 2 minutes. Genuine expressions were collected from reality shows and acted emotions from movies with similar scenes. The work was realised with three emotions: fear, anger and happiness. The classifying method used was a simple Feedforward Neural Network.

In [9], they collected a database by inducing emotional state on participants through videos exposed. In each video the participants started with a neutral emotion and then they expressed fake or genuine facial emotion. The images were processed by combining Haar-feature face detector and an MOSSE-based object tracker. After the face identifying stage, the facial landmarks were detected by DLib library. Then, a Recurrent Neural Network with Parametric Bias (RNN-PB) was applied to classify as fake or genuine emotion.

[10] suggested a classification technique with a new collected database that was called SASE-FE. This database was collected by eliciting the six universal emotions on participants, showing videos according with the target emotion. The study used Convolutional Neural Network (CNN) to learn a static representation from images and then extract some features along facial landmarks. Lastly, they used Support Vector Machine (SVM) for final classification.

In [17], it was proposed a system to classify genuine and posed pain. They used Weighted Spatio-temporal Pooling (WSP), a video summarisation method, to encode the video sequence into an image and then, they used a Residual Generative Adversarial Network (R-GAN) to identify genuine pain from posed pain.

A comparative summary of related works is presented in Table 1. It is possible to identify an interest by the researchers in using variations of Haar-features in order to perform the classification stage and the use of very high performance deep learning neural networks architectures, such as GAN or CNN. However, we believe that if we do not use the correct datasets, the results tend to be very costly.

Most of what can be found in the literature are techniques related to deep learning, thus the feature extraction approach they use considers the classification method, which is very different from what is being proposed on our work. The calculation of feature values and Action Units inference present by [7] is an approach that showed to be efficient for classification of emotions and was not tested on works of fake emotions detection.

Thus, in this paper, we will replicate the image processing method from [7] as well as the naturally generated emotions from face images, we will implement three very simple classification algorithms: k-Means, kNN and a Neural Network. Our aim is to show that we can get good results on fake emotion prediction by simply using a well collected dataset.

3 Natural facial emotions vs. fake emotions: a methodology

As already mentioned in the previous section, our aim with this work is to investigate the main differences between naturally generated emotions from fake emotions of public datasets. We have decided to compose our dataset using faces with fake emotions and faces with natural emotions, both from public datasets.

The fake emotions faces were acquired from the IMPA-FACE3D[1] and Cohn-Kanade [18] databases which are widely used in face emotion analysis experiments. The IMPA-FACE3D database was collected with 38 individuals and each sample contained images with neutral face with the six universal humans expressions proposed by [5]. This database is available online and was used for requested happiness and sadness emotions images which represents the focus of this paper.

The Cohn-Kanade database was collected with 123 individuals and each sample contained sequences of frames with the expressive emotions collected. Some samples contained the six universal emotions and others have only some of these emotions. We filtered just happiness and sadness emotions of the database and selected only one image with a neutral face and one image representing the expressiveness face of each emotion. After the processing, this database contains 111 participants, 73 samples of sadness emotion and 102 samples of happiness emotion.

For the comparison with genuine emotions, for the natural emotions faces, we had used the collected dataset by [7] where the participants were provoked by external stimuli (music and image), eliciting one of the emotions: fear, happiness and sadness. This dataset consisted of multiple frames collected with each of 101 participants which were divided on the 3 analysed emotions. For this work, we only considered the 62 participants that demonstrated happiness or sadness emotions and we did not consider all the frames, just the one with the neutral face and the one expressive emotion of each participant.

3.1 Pre-processing

Since we have used images from originally different datasets, we have an important pre-processing stage. Our detailed methodology can be described as follow:

1. detection of face on the images,
2. marking of facial landmark points,

[1] http://app.visgraf.impa.br/database/faces

Table 1. Summary of related works

References	Databases	Objective	Feature Extraction Technique	Classification	Results
[8]	Generated	Fake emotions identifying system. Emotions: fear, anger and happiness.	-	Simple Feed-forward Neural Network	Accuracy: 99%
[9]	Generated	Fake emotions identifying system. Emotions: happy, sad, fear, anger, disgust, and surprise.	Haar-feature face detector and an MOSSE-based object tracker within the OpenCV environment	RNN with Parametric Bias and Gradient Boosting Machine (GBM)	Accuracy: 66.7%
[10]	Generated	Fake emotions identifying system. Emotions: happy, sad, fear, anger, disgust, and surprise.	Convolutional Neural Network(CNN) to learn a static representation from still images and then use FV encoding	Support Vector Machine (SVM)	Anger: 80.1%; Disgust: 88.0%; Fear: 95.1%; Happiness: 89.7%; Sadness: 91.3%; Surprise: 92.7%
[17]	UNBC-McMaster; Shoulder Pain, BioVid; Head Pain, STOIC	Genuine and posed pain detect system.	Weighted Spatio Temporal Pooling	Residual Generative Adversarial Network (R-GAN)	UNBC: 91.34%; BioVid: 85.05%; STOIC: 96.52%
[2]	CK+, Oulu-CASIA VIS, JAFFE	Facial expression recognition systems. Emotions: happy, sad, fear, anger, disgust, and surprise.	Feature selection using CfsSubsetEval feature evaluator and Best First as search method	Backpropagation Neural Network	CK+: 99%; OULU-CASIA VIS: 84.7%; JAFFE: 93.8%
[3]	JAFFE, MMI, CK+	Facial expression recognition systems. Emotions: happiness, sadness, fear, anger, disgust, and surprise.	-	Generative Adversarial Networks (GAN) and Convolutional Neural Networks (CNN).	JAFFE: 59.62%; MMI: 61.86%; CK+: 76.58%
[4]	CK+, JAFFE, ISED	Facial expression recognition systems. Emotions: happiness, sadness, fear, anger, disgust, and surprise.	Extract regions of face with Multi-task CNN	Deep Alignment Network (DAN).	CK+: 73.6%; JAFFE: 46.5%; ISED: 62%
[5]	CK+, JAFFE	Facial expression recognition systems. Emotions: happiness, sadness, fear, anger, disgust, and surprise.	Haar-Cascade filter from OpenCV library	Convolutional Neural Network (CNN).	JAFFE: 62%; CK+: 90.7%
[6]	JAFFE, MUG	Facial expression recognition systems. Emotions: happiness, sadness, fear, anger, disgust, and surprise.	iPar-CLR method for extract the features and calculate the relative distances among the facial points.	Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Ensemble Neural Network (ENN)	JAFFE: NN (89.16%); SVM (77.03%); ENN (90.54%); e MUG: NN(92.70%); SVM (89.48%); ENN (95.50%)
[7]	Generated	Facial expression recognition systems. Emotions: happiness, sadness and fear.	Viola-Jones method for face detection, geometric model for extracting facial features.	Action Units (AUs) inference	Overall precision: 92%

3. calculation of feature values.

From the result of these steps, we have generated a new database with the result of the difference between neutral and expressive feature values, the expressive emotion (sadness or happiness) and a boolean field representing emotion type (acted or genuine) which will be used as our class representation.

Sections 3.2 and 3.3 will describe in more details how each of these stages were performed in our experiments.

3.2 Face detection and facial landmarks

As the first step of the image processing stage of this experimentation, we need, initially, to detect the face on the images and then define a boundary block. In order to do that, we used the Cascade Classifier method from OpenCV library that is based on the Haar-like features algorithm, a Haar feature-based cascade classifier developed and trained to properly locate faces or objects[2].

We have decided to use this specific technique because it was the most popular used in the current literature, as it was discussed in Section 2.

Subsequently, we did the facial landmarks with dlib support that extracts keypoints from regions with pre-trained models and estimates the location of 68 coordinates that maps salient regions of the face, such as eyes, eyebrows, nose, mouth, jawline [19].

Figure 1 illustrates one of the images from IMPA-FACE3D database with the detected face boundary block and marked landmarks and it was implemented using the same as the work presented in [7].

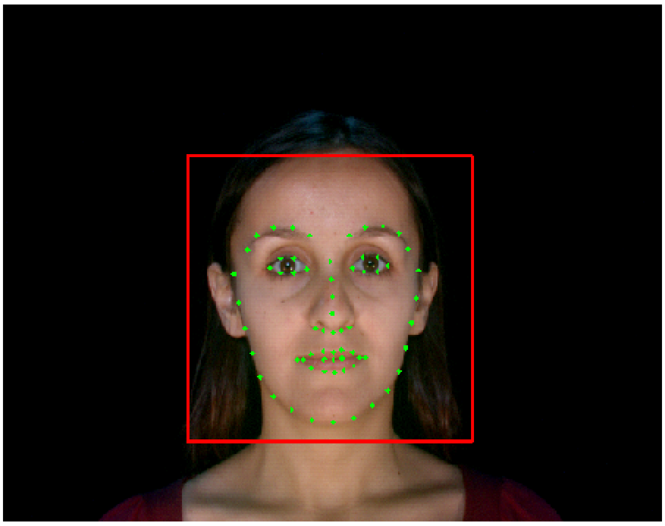

Figure 1. Face detection and facial landmarks with opencv and dlib libraries.

[2]https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier

3.3 Feature extraction

After the facial landmarks detection, we needed to calculate a measure for facial deformations. These feature values were calculated by identifying geometric features on the face. For example, the inner eyebrow height, eye and mouth openness and so on. Each distance on the mouth, eye, eyebrow and nose represents one feature value.

The considered model was reformulated by [7] from the model proposed by [20], because they presented three new features to solve some problems with the same nature from this work.

4 Differentiating fake from natural emotions from face

Our main goal is to show the importance of using datasets with naturally generated emotions from face images and from that we can use very simple classification techniques for identifying the ones that are fake from the ones that are not.

Thus, three algorithms were implemented to analyse the data. The k-means algorithm was used to analyse the database, the k-Nearest Neighbors and Backpropagation Neural Network were used to classify. In this section, we will present their specifications.

4.1 k-means

Clustering algorithms aim to automatically group the n data from database by unsupervised learning into k-groups, called clusters. Our database already had the correct labels for genuine and acted emotions, so the objective of using k-means was to classify the entire database into two clusters based on feature values differences without having access to the labels. Thus, this makes it possible to analyse the reasons that led to the incorrect classifications and try to improve the classification algorithm through relevant metrics.

The data from our initial was separated into happiness and sadness emotions and then processed separately with the k-means. The data considered for this algorithm were the columns containing the differences of feature values between neutral and expressive face, resulting in points with 14 dimensions.

The first step was to specify the number of clusters k and the stop criteria. So, we defined clusters being the type of emotion acted or genuine, thus we have k=2. As for the stop criteria, it was completed after two interactions without any changing on the clusters. After, the centroids were initialised randomly among dataset points. Also, two versions of k-means were implemented, each using a different distance: Euclidean and Manhattan. Finally, the criteria to define the new centroids of clusters was to calculate the average of all data points from each dimension.

The metric used to analyse the results was to compute the number of classifications made correctly and to analyse the data that was classified wrongly.

4.2 K-Nearest Neighbors (KNN)

Our problem is to predict results in discrete output (genuine or acted), so we needed to choose a simple benchmark classification algorithm. Thus, the k-Nearest Neighbors classification algorithm [21] was implemented, in which learning is based on the similarity (distance) between the data.

Similar to what was done for k-means, the database was divided into happiness and sadness emotions to analyse the classifications within each one separately. Firstly, we have defined the classes considered for classification: genuine and acted. In addition, two versions of this algorithm were implemented using the two distances: Euclidean and Manhattan. In both, the five shortest distances were used to classify the data (k=5).

4.3 Backpropagation Neural Network

Artificial Neural Networks are models inspired on human brain simulating neurons and connections between them (synapses). For this work, it was implemented a Multilayer Perceptron (MLP) with three layers, with only one hidden layer, because it is a simple feedforward network [21].

Our final model of MLP had 14 neurons in input layer that represents the feature values extracted from images. We tried training our network with 3 to 6 neurons on the hidden layer. In the output layer, we considered just one neuron that represents the output genuine or acted, 0 or 1 respectively. In the test stage, we have considered 0.5 as the threshold that separates the two classes.

The weights were initially set to a random number between 0 and 1. The activation function used was the Sigmoid and the method used to weight adjust was the backpropagation, a supervised learning technique. We trained our network with a learning rate of 0.01 and 0.001. The iteration of the network terminates when the average error values of an epoch are smaller than 0.1.

The parameters used in our neural network were chosen through experimentation. We have applied different metric values to train our network and the values presented were the best obtained. Due to the lack of space, we were unable to add all the details in the experimentation.

5 Results and Discussions

In order to evaluate the results, we have extracted the true-positives (tp), true-negatives (tn), false-positives (fp) and false-negatives (fn) values and calculated the precision, recall and accuracy of the k-Means, k-NN and Neural Network for happiness and sadness emotions.

We initially implemented k-Means algorithm with Euclidean distance, but we obtained the worst result. The Manhattan distance had a similar purpose, since both calculated the geographic distance between two points. So we have also tried running the algorithm with Manhattan distance and obtained a better result. The comparison between the two approaches are presented in Table 2.

For this reason, we have also used Manhattan distance on k-NN. In addition, we have analysed the feature values used on

Table 2. Comparison between different distances on k-Means algorithm

Distance	Emotion	Accuracy	Precision	Recall
Euclidean	Happiness	0.78	0.95	0.77
Euclidean	Sadness	0.53	0.74	0.61
Manhattan	Happiness	0.81	0.96	0.8
Manhattan	Sadness	0.52	0.74	0.59

k-Means and verified that some features have a very detailed description about the face detected from images. So we disregarded the feature values ieb_height, oeb_height, lc_height, and obtained the results shown in Table 3 for the k-Means and k-NN algorithms.

Table 3. Comparison between k-Means and kNN algorithms

Algorithm	Emotion	Accuracy	Precision	Recall
k-Means	Happiness	0.84	0.97	0.84
k-Means	Sadness	0.58	0.77	0.66
kNN	Happiness	0.988	0.986	1.0
kNN	Sadness	0.73	0.9	0.74

The neural network was implemented based on modifications made to k-Means and k-NN. In order to analyse our algorithm efficiency, we randomly split the data into a sample of 70 percent for training and 30 percent for testing. Also, we have tried to train and test our network with different parameter values, the results with the best configurations was presented in Table 4.

Table 4. Backpropagation Neural Network results

Emotion	Accuracy	Precision	Recall	Hidden Layers	Learning Rate
Happiness	0.89	0.93	0.93	6	0.001
Happiness	0.83	0.9	0.9	5	0.01
Sadness	0.81	0.89	0.89	4	0.01
Sadness	0.78	0.92	0.82	7	0.01

Therefore, we have obtained good results on classifying the happiness and sadness emotions as fake or genuine, similar to the results presented in the literature. We have obtained these results with simpler algorithms, thus indicating the importance of using a well collected dataset.

It is not possible to perform a direct comparative analysis with the state-of-art works because there is no consistency in the way their results are presented. For instance, [8] analysed the results through cross-validation tests with four emotions, on the other hand [9] presented the results through the overall accuracy with six emotions. In [10], the authors presented the accuracy with each dataset separately. The others works do not present results about fake emotions, they focused on a classification on the six universal emotions. In our work, we have focused on the emotions happiness and sadness and of using all the datasets to train and test.

In addition, we have shown that by using acted emotions datasets to classify 'real' emotions does not necessarily reflect a real scenario, even with good results. Thus, our main conclusion is that datasets with elicited emotions can create a better mapping of the user features.

6 Conclusion

In this work, we have explored the possibility of classifying emotions from facial expressions as acted and genuine. There are many approaches on literature to solve the problem of detecting emotions and, recently, to classify an emotion in acted or genuine. We propose to use k-Means to analyse the behavior of our dataset, and a k-Nearest Neighbor and Neural Network to classify the emotions in acted or genuine.

The results obtained showed that the best classifier for happiness emotion is the k-Nearest Neighbor with 98.6% of precision and 100% of recall and for sadness emotion is the Backpropagation Neural Network with 92% of precision and 82% of recall.

Finally, as future work, some modifications on image processing will be made, analysing Feature Values considered, which also can be done with features extraction based on intensity of muscle contractions. Also the neural network can be improved to try to obtain a better result. In addition, it can increase the quantity of genuine and acted emotions images and also can be done an analysis of other emotions.

References

[1] F. Mandal, "Nonverbal communication in humans," *Journal of Human Behaviour in the Social Environment*, vol. 24, pp. 417–421, 05 2014.

[2] F. Salmam, A. Madani, and M. Kissi, "Emotion recognition from facial expression based on fiducial points detection and using neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, p. 52, 02 2018.

[3] X. Wang, X. Wang, and Y. Ni, "Unsupervised domain adaptation for facial expression recognition using generative adversarial networks," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–10, 07 2018.

[4] I. Tautkute and T. Trzciński, "Classifying and visualizing emotions with emotional dan," *Fundam. Inform.*, vol. 168, pp. 269–285, 2019.

[5] D. Duncan, G. Shine, and C. English, "Facial emotion recognition in real time." Available in: http://cs231n.stanford.edu/reports/2016/pdfs/022_Report.pdf., 2016.

[6] G. Sharma, L. Singh, and S. Gautam, "Automatic facial expression recognition using combined geometric features," *3D Research*, vol. 10, 06 2019.

[7] M. Costa-Abreu and G. Bezerra, "Famos: a framework for investigating the use of face features to identify spontaneous emotions," *Pattern Analysis and Applications*, vol. 22, pp. 683–701, 12 2017.

[8] Z. Qin, T. Gedeon, and S. Caldwell, "Neural networks assist crowd predictions in discerning the veracity of emotional expressions," in *Neural Information Processing*, pp. 205–216, Springer International Publishing, 2018.

[9] Y. Kim and X. Huynh, "Discrimination between genuine versus fake emotion using long-short term memory with parametric bias and facial landmarks," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 3065–3072, Oct 2017.

[10] K. Kulkarni, C. Corneanu, I. Ofodile, S. Escalera, X. Baró, S. Hyniewska, J. Allik, and G. Anbarjafari, "Automatic recognition of facial displays of unfelt emotions," *IEEE Transactions on Affective Computing*, pp. 1–1, 2018.

[11] J. Walczyk, D. Griffith, R. Yates, S. Visconte, B. Simoneaux, and L. Harris, "Lie detection by inducing cognitive load: Eye movements and other cues to the false answers of "witnesses" to crimes," *Criminal Justice and Behavior - CRIM JUSTICE BEHAV*, vol. 39, pp. 887–909, 06 2012.

[12] J. H. Goldberg, "Relating perceived web page complexity to emotional valence and eye movement metrics," *Proceedings Of The Human Factors And Ergonomics Society Annual Meeting*, vol. 56, pp. 501–505, 2012.

[13] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 154–154, June 2004.

[14] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.

[15] M. Zanarini, F. Frankenburg, D. Reich, K. Silk, J. Hudson, and L. McSweeney, "The subsyndromal phenomenology of borderline personality disorder: A 10-year follow-up study," *The American journal of psychiatry*, vol. 164, pp. 929–35, 07 2007.

[16] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. Lecompte, T. Pitcairn, and P. E. e. a. Ricci-Bitti, "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of personality and social psychology*, vol. 53, pp. 712–717, 1987.

[17] M. Tavakolian, C. G. Bermudez Cruces, and A. Hadid, "Learning to detect genuine versus posed pain from facial expressions using residual generative adversarial networks," in *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*, pp. 1–8, May 2019.

[18] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," pp. 94 – 101, 07 2010.

[19] C. Sagonas and S. Zafeiriou, "Facial point annotations." Available in: https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/. Access in: oct 2019.

[20] E. Jongh, "Fed: an online facial expression dictionary as a first step in the creation of a complete nonverbal dictionary," Master's thesis, Delft University of Technology, 6 2002.

[21] B. Coppin, *Artificial Intelligence Illuminated*. Sudbury: Jones And Bartlett Publishers, 2004.