Open **A**ccess
Full-Length Article

Machine Learning Based Statistical Analysis of Emotion Recognition using Facial Expression

Aqib Ali[1], Jamal Abdul Nasir[2], Muhammad Munawar Ahmed[3], Samreen Naeem[1], Sania Anam[4], Farrukh Jamal[5,*], Christophe Chesneau[6], Muhammad Zubair[3], Muhammad Saqib Anees[7]

[1]Department of Computer Science and IT, Glim Institute of Modern Studies, Bahawalpur, 61300, Pakistan
[2]Department of Statistics, GC University Lahore, 54000, Pakistan
[3]Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[4]Department of Computer Science, Govt Degree College for Women Ahmadpur East, Bahawalpur 63350, Pakistan
[5]Department of Statistics, Govt. S.A Postgraduate College Dera Nawab Sahib, Bahawalpur, Punjab 63100, Pakistan
[6]Department of Mathematics, Université de Caen, LMNO, Campus II, Science 3, 14032 Caen, France
[7]Govt. Elementary School 7/BC, Bahawalpur, 61300, Pakistan

ABSTRACT

Background: Humans can deliver many emotions during a conversation. Facial expressions show information about emotions.

Objectives: This study proposed a Machine Learning (ML) approach based on a statistical analysis of emotion recognition using facial expression through a digital image.

Methodology: A total of 600 digital image datasets divided into 6 classes (Anger, Happy, Fear, Surprise, Sad, and Normal) was collected from the publicly available Taiwan Facial Expression Images Database. In the first step, all images are converted into a grey level format and 4 Regions of Interest (ROIs) are created on each image, so the total image dataset gets divided into 2400 (600 x 4) sub-images. In the second step, 3 types of statistical features named texture, histogram, and binary feature are extracted from each ROIs. The third step is statistical feature optimization using the best-first search algorithm. Lastly, an optimized statistical feature dataset is deployed on various ML classifiers.

Results: The analysis part was divided into two phases: firstly boosting algorithms-based ML classifiers (named as LogitBoost, AdaboostM1, and Stacking) which obtained 94.11%, 92.15%, and 89.21% accuracy, respectively. Secondly, decision tree algorithms named J48, Random Forest, and Random Committee were obtained with 97.05%, 93.14%, and 92.15% accuracy, respectively.

Conclusion: It was observed that decision tree-based J48 classifiers gave 97.05% classification accuracy.

INTRODUCTION

The message has always been given importance in human life. A complete message can be conveyed in two different ways; verbally and non-verbally. Non-verbal messages are exchanged in various ways, such as facial expressions and body movements. During communication, people make different facial expressions and these expressions are an important part of non-verbal communication with importance in explaining human

actions[1]. To obtain the required information, facial expression data are divided into several areas which are implemented in different approaches to artificial vision. The area of the chin, nose, eye, and lips changes during the conversation.

Dynamic facial expressions have caused many problems with facial expressions. Therefore, many problems were encountered when collecting facial datasets such as eyes, nose, lips, and chin. In this regard, a total of 90% accuracy was obtained by using K-Nearest Neighbor (KNN) algorithm[2]. Various artificial vision methods have been used for facial recognition. It also helps to understand human emotions, mentalities, and behavior. Facial muscles allow us to obtain useful information on facial expressions. In general, Human-Computer Interaction (HCI) and computer interaction have been used in computational models to obtain achievable 94% results[3]. In the face detection process, automatic facial expression methods faced several challenges. The facial expressions of the individual community are easier to explain than the difficulty of classifying datasets when compared to multicast data sets. The architecture of the multiple databases of the Deep Neural Network (DNN) get better results, e.g. Multi-PIE 94.80%, MMI 56.05%, CK+ 92.21%, DISFA 56.15%, FERA 77.44%, SFEW 48.62% and FER-2013 61.11%[4].

Emotion-based expression recognition uses a computer vision approach. Locally collected datasets have been divided into 3 classes; happy, angry, and sad, and generate a region of interest for feature extraction. Multi texture features are used for classification and obtain 96% overall accuracy using random forest classifiers[5]. Statistical and knowledge-based edge detection methods will help in facial expression recognition. During human communication, different conditions are observed in the area of the eyes and lips. Canny Edge Detection (CED) provides 96.6% of results on facial expressions[6]. Transactional Parameter Transfer (TPT) was used to develop a customized classification model. The main goal of the TPT model is to use the pre-trained regression function and the computational cost, which is much lower than other algorithms[7]. Human-Computer Interaction (HCI) is based on facial expressions. Signal recognition technologies are evaluated using various classification methods. Based on PAINFUL 78.33% and CK+ 92.72% database[8], TPT will perform better.

Facial expressions have several salient features of emotion. The patch comparison operation technique is useful for extracting the characteristics of these distinct areas. Gabbar's patch-based features produced better results based on the JAFEE and CK+ databases, with an overall accuracy of 92.9% and 94.4%[9]. The facial recognition system is easy to use. In particular, the person is identified through facial recognition techniques through access to the network, automatic gender, and FER review[10].

The idea is to divide the image into 6 parts and label it. The random decision for classification gets forest classification and SVM classification, and 96.25% accuracy[11]. Men and women's emotions were associated with behavioral disorders. The use of synchronous tracking was related to the performance of the classification to allocate attention to participants[12]. In the FER field, processing images are configured in real-time from a mobile device, which is healthcare. The recommended approach is based on privileged neural network and better rating results[13].

MATERIALS AND METHODS

In this study, 6 types of human emotions have been used named Anger, Happy, Fear, Surprise, Sad, and Normal, collected from publicly available Taiwan Facial Expression Images Database[14] as shown in Fig. **1**. A total of 600 (350 females, and 250 males) digital images are collected for experimentation of size 512 x 512 and 24-bit (JPEG) format.

Figure 1. Facial expression digital image dataset[14].

Proposed Methodology

The proposed methodology has been discussed with the proposed algorithm. In this case, the first step was a digital image collection from a publicly available human emotion database. The second step was image preprocessing. In this step, all 600 images were resized into 512 x 512 size for data standardization and converted into an 8-bit grey-level format. The third step was a Region of Interest (ROI) selection. In this step, we draw 5 ROI on each image as shown in Fig. **2**. So, the final size of the sub-image dataset is 3000 (600 x 5). The fourth step was the statistical features extraction. In this step, 3 types of statistical features were extracted, namely texture, spectral, and histogram features. The fifth step was the statistical feature selection. In this step, 7 optimized statistical features were selected using a correlation-based feature selection approach. In the last step, 6 Machine Learning (ML) classifiers optimized statistical features dataset was used for obtaining classification results. The proposed algorithm is discussed below:
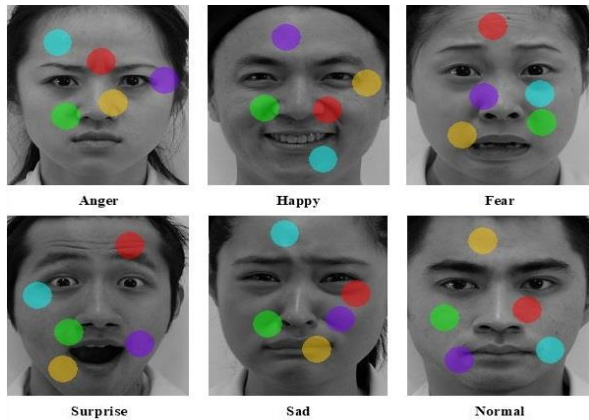

Figure 2. Gray level images with non-overlapping ROIs[14].

Proposed Algorithm

Start main ()

  {

    Input: Facial emotion image dataset

  For

  {

    Stage 1 to Stage 5

    Stage 1: Six types of images dataset.

    Stage 2: Image pre-processing.

    Stage 3: 5 Non overlapping ROI selection.

    Stage 4: Extract statistical features.

    Stage 5: Statistical feature optimization.

  End For

  }

    Stage 6: Classification based on ML approaches.

    Output: Emotion Recognition

End main

  }

The proposed framework explained all necessary steps as shown in Fig. **3.**
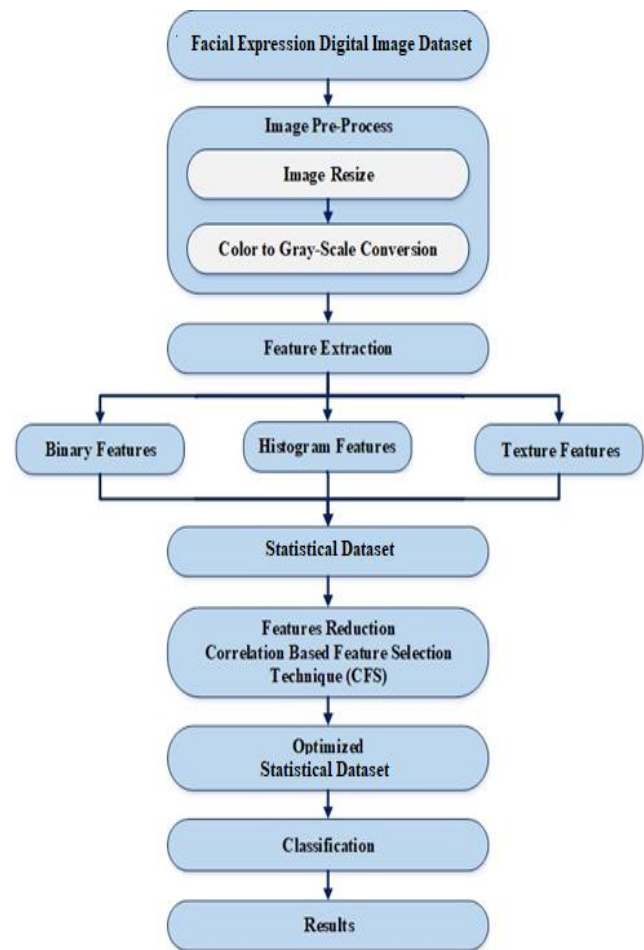

Figure 3. Proposed framework for emotion recognition using facial expression.

Feature Selection

Many scholars have applied several techniques to extract hidden features[15-17]. The accuracy of the classification work of a large dataset depends on the choice of the

feature, so the method of selecting the feature plays a very important role in image processing.

Binary Feature

The binary properties indicate the shape of the object in image processing based on the area (at least on the other axis of the moment), Eller number, area, and center of projection.

Histogram Feature

A histogram consists of a set of adjacent rectangles with bases along the x-axis, with centers for class marks and regions proportional to the frequency of the class[18]. The histogram shapes are calculated based on the gray level and the number of pixels. The first order in Equation 1 is the following feasibility histogram of the order P(h).

$$P(h) = \frac{K(h)}{N} \quad (1)$$

Here, N is a Total Pixel and the Gray Level of K(h) has a Total Pixel.

The probability of the first order histogram is used using the following methods for calculating the data: Mean deviation, standard deviation, applique, energy, and entropy.

The averages are average values, with luminous symbols (medium-high) and black symbols (medium-low). The meaning is the following in Equation 2.

$$\bar{h} = \sum_{h=0}^{P-1} hP(h) = \sum_{x}\sum_{y} \frac{I(x,y)}{N}, \quad (2)$$

Here, P is the Total Number of Gray Degrees from 0 to 255 and I (x,y) represents x (rows) and y (columns) of the pixels.

Standard deviation (SD) describes the opposite of the image, as define in Equation 3.

$$\sigma_h = \sqrt{\sum_{h=0}^{P-1}(h - \bar{h})^2 P(h)} \quad (3)$$

Skewness (Skew), lack of balance in the distribution around some central values (mean, median, format). The frequency curve is divided into a long tail on the right (positive) and a long tail on the left (negative). Compression is explained in Equation 4 as follows.

$$Skew = \frac{1}{\sigma_h^3}\sum_{h=0}^{P-1}(h - \bar{h})^3 P(h) \quad (4)$$

Energy measures the distribution of the brown surface, which is determined in Equation 5.

$$Energy = \sum_{h=0}^{P-1}[P(h)]^2 \quad (5)$$

Entropy is measured by the total number of bits required to code which is represented in the data. It is given in Equation 6.

$$Entropy = -\sum_{h=0}^{P-1} P(h)\log_2[P(h)] \quad (6)$$

Texture Feature

Selecting the purpose of the image in the text by the row and column coordinates. The texture is calculated in five different ways: energy, communication, entropy, inverse difference, and inertia[19].

In Equation 7, energy is measured smoothly or evenly by calculating the distribution between the gray surfaces.

$$Energy = \sum_{m}\sum_{n}(C_{mn})^2 \quad (7)$$

Here, C_{mn} is the value in the co-event matrix in terms of pixel distribution values.

The method of communication defines the similarity of pixels to the default pixel distance. The method of assistance is as follows in Equation 8.

$$Correlation = \frac{1}{\sigma_a\sigma_b}\sum_{m}\sum_{n}(m - \mu_a)(n - \mu_b)C_{mn} \quad (8)$$

Here μ_a and μ_b are the means of a and b, respectively.

$$\mu_a = \sum_m m \sum_n C_{mn} \quad (8.1)$$

$$\mu_b = \sum_n n \sum_m C_{mn} \quad (8.2)$$

$$\sigma_a^2 = \sum_m (m - \mu_a)^2 \sum_n C_{mn} \qquad (8.3)$$

$$\sigma_b^2 = \sum_n (n - \mu_b)^2 \sum_m C_{mn} \qquad (8.4)$$

Gray level-dependent matrix is measured by another histogram method with distance and angle parameters.

Entropy is measuring image content information. The entropy is explained in Equation 9 as follows.

$$Entropy = -\sum_m \sum_n C_{mn} \log_2 C_{mn} \qquad (9)$$

The inverse difference method is measuring the local coherence of the image which explains the following in Equation 10.

$$Inverse\ Difference = \sum_m \sum_n \frac{C_{mn}}{|m - n|} \qquad (10)$$

The method of inertia is to measure the opposite, which explains the following in Equation 11.

$$Inertia = \sum_m \sum_n (m - n)^2 C_{mm} \qquad (11)$$

Classification

For the classification approach, this study was divided into 2 categories: the first category is ML classification based on boosting algorithms (LogitBoost, AdaboostM1, Stacking) and the second category is ML classification based on decision tree algorithms (J48, Random Forest, Random Committee)[20].

RESULTS

For this study, the facial emotional dataset was divided into 6 classes namely: Anger, Happy, Fear, Surprise, Sad, and Normal. Various tinning parameters have been calculated to measure the performance of these image datasets e.g. "receiver operating characteristics" (ROC), "mean absolute error" (MEA), "false positive" (FP), "true positive" (TP), "root mean square error" (RMSE), and time (T)[21, 22]. In the first step, the statistical features dataset was deployed on ML classifiers based on Boosting algorithms[23, 24] named as LogitBoost, AdaboostM1, and Stacking and obtain 94.11%, 92.15%, and 89.21% accuracy, respectively as shown in Table **1.**

In the second step, statistical features dataset was deployed on ML classifiers based on Decision Tree Algorithms named as J48, Random Forest, and Random Committee were obtained 97.05%, 93.14%, and 92.15% accuracy, respectively as shown in Table **2.** It has been observed that decision tree-based[25-26] ML classifiers showed promising results as compared to Boosting algorithms.

Table 1. Boosting Algorithm-Based ML Classification.

S. No.	Classifiers	Kappa Statistics	TP Rate	FP Rate	ROC	MAE	RMSE	T (sec)	Accuracy
01	LogitBoost	0.9292	0.941	0.011	0.995	0.0232	0.1353	0.17	94.11%
02	AdaboostM1	0.9056	0.922	0.015	0.980	0.0548	0.1643	0.09	92.15%
03	Stacking	0.8703	0.892	0.019	0.962	0.2255	0.3153	0.11	89.21%

Table 2. Decision Tree Algorithm-Based ML Classification.

S. No.	Classifiers	Kappa Statistics	TP Rate	FP Rate	ROC	MAE	RMSE	T (sec)	Accuracy
01	J48	0.9646	0.971	0.006	0.998	0.0504	0.1154	0.07	97.05%
02	Random Forest	0.9174	0.931	0.014	0.973	0.0328	0.1493	0.11	93.14%
03	Random Committee	0.9056	0.922	0.015	0.998	0.0416	0.1362	0.03	92.15%

Table 3. The Confusion Matrix Table for J48 Classifier.

S. No.	Classes	Anger	Happy	Fear	Surprise	Sad	Normal	Accuracy
01	**Anger**	**2320**	55	5	10	0	10	**96.66 %**
02	**Happy**	6	**2330**	0	40	16	10	**97.08 %**
03	**Fear**	55	5	**2340**	0	0	0	**97.75 %**
04	**Surprise**	20	10	0	**2346**	6	20	**97.50 %**
05	**Sad**	9	10	1	20	**2320**	40	**96.66 %**
06	**Normal**	5	50	5	0	4	**2336**	**97.33 %**

The J48 classifier based on the decision tree algorithm was performed most promising and provided higher accuracy insisted on other implemented ML classifiers. The confusion matrix shows correctly predicted values paced diagonally, others are un-corrected for the J48 classifier as shown in Table **3** and detail accuracy of the J48 classifier is shown in Fig. **4.**
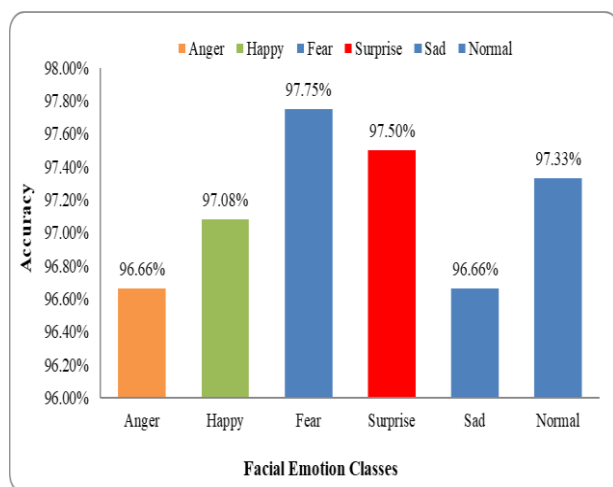

Figure 4. Confusion matrix detail accuracy graph using decision tree J48 classifier.

Emotion recognition using facial expression comparison graph applying boosting based algorithms and decision tree-based algorithms as shown in Fig. **5.**
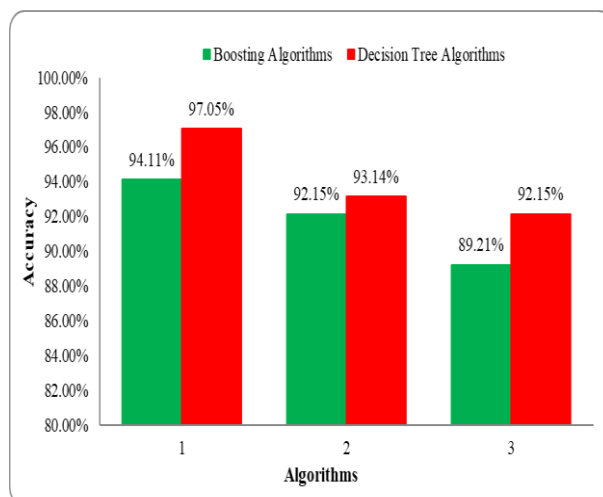

Figure 5. Comparative analysis of employed boosting and decision tree-based classifiers.

CONCLUSION

In this study, we proposed a Machine Learning (ML) approach based on a statistical analysis of emotion recognition using facial expression through a digital image. A total of 600 digital image datasets were divided into 6 classes (Anger, Happy, Fear, Surprise, Sad, and Normal) collected from the publicly available Taiwan Facial Expression Images Database. Three types of statistical features named texture, histogram, and binary feature were extracted for analysis. The analysis part was

divided into 2 phases, firstly boosting algorithms-based ML classifiers named as LogitBoost, AdaboostM1, and Stacking and obtain 94.11%, 92.15%, and 89.21% accuracy, respectively. Secondly, decision tree algorithms named J48, Random Forest, and Random Committee were obtained with 97.05%, 93.14%, and 92.15% accuracy, respectively. It has been observed that decision tree algorithms based on J48 ML classifiers performed better as compared to other deployed classifiers.

ACKNOWLEDGEMENTS

LIST OF ABBREVIATION

DNN	Deep Neural Network
FP	False Positive
HCI	Human-Computer Interaction
KNN	K-Nearest Neighbor
MEA	Mean Absolute Error
ML	Machine Learning
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristics
ROI	Region of Interest
T	Time
TP	True Positive
TPT	Transactional Parameter Transfer

REFERENCES

1. Ekman P, Friesen WV. Unmasking the face Cambridge. MA: Malor Books. 2003.
2. Perveen N, Ahmad N, Khan MA, Khalid R, Qadri S. Facial expression recognition through machine learning. Int J Sci Technol Res. 2016; 5(03):91-7.
3. Martinez A, Du S. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. J Mach Learn Res. 2012; 13:1589-608.
4. Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. In 2016 IEEE Winter conference on applications of computer vision (WACV) 2016; 1-10.
5. Jamal F. Emotion based facial expression detection using machine learning approach. J Appl Emerg Sci. 2020; 10(1):35-40.
6. Chen X, Cheng W. Facial expression recognition based on edge detection. Int J Comp Sci Eng Sur. 2015; 6(2):1-9.
7. Valstar MF, Almaev T, Girard JM, McKeown G, Mehu M, Yin L, Pantic M, Cohn JF. Fera 2015-second facial expression recognition and analysis challenge. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) 2015; 6:1-8.
8. Zen G, Porzi L, Sangineto E, Ricci E, Sebe N. Learning personalized models for facial expression analysis and gesture recognition. IEEE transactions on multimedia. 2016; 18(4):775-88.
9. Zhang L, Tjondronegoro D. Facial expression recognition using facial movement features. IEEE transactions on affective computing. 2011; 2(4):219-29.
10. Lin SH. An introduction to face recognition technology. Info Sci Int J Emerg Transdiscipl. 2000; 3:1-7.
11. Khan K, Mauro M, Migliorati P, Leonardi R. Gender and expression analysis based on semantic face segmentation. International conference on image analysis and processing. 2017; 37-47.
12. Martin-Key NA, Graf EW, Adams WJ, Fairchild G. Facial emotion recognition and eye movement behaviour in conduct disorder. J Child Psychol Psych. 2018; 59(3):247-57.
13. Gervasi O, Franzoni V, Riganelli M, Tasso S. Automating facial emotion recognition. In Web Intelligence. 2019; 17(1):17-27.
14. Brain Mapping Laboratory and Brain Research Unit. TFEID (Taiwanese Facial Expression Image Database). [Online]. http://bml.ym.edu.tw/tfeid/ 2020.
15. Ali A, Qadri S, Khan Mashwani W, Kumam W, Kumam P, Naeem S, Goktas A, *et al*. Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image. Entropy. 2020; 22(5):567-73.
16. Naeem S, Ali A, Qadri S, Mashwani WK, Tairan N, Shah H, *et al*. Machine-learning based hybrid-feature

analysis for liver cancer classification using fused (MR and CT) images. App Sci. 2020; 10(9):31-4.

17. Ali A, Qadri S, Mashwani WK, Brahim Belhaouari S, Naeem S, Rafique S, *et al*. Machine learning approach for the classification of corn seed using hybrid features. Int J Food Prop. 2020; 23(1):1097-111.

18. Liu ZT, Sui GT, Li DY, Tan GZ. A novel facial expression recognition method based on extreme learning machine. In 2015 34th Chinese Control Conference (CCC) 2015; 12:3852-7.

19. Zhang Z. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. Int J Patt Recog Artif Intel. 1999; 13(06):893-911.

20. Nelson CA. The development and neural bases of face recognition. Infant and child development. Int J Res Pract. 2001; 10(2):3-18.

21. Umbaugh SE. Digital image processing and analysis: Human and computer vision applications with CVIPtools. CRC press; 2010.

22. Khan K, Mauro M, Migliorati P, Leonardi R. Gender and expression analysis based on semantic face segmentation. International Conference on Image Analysis and Processing 2017; 11(5):37-47.

23. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics. 1973; (6):610-21.

24. Lopes AT, De Aguiar E, Oliveira-Santos T. A facial expression recognition system using convolutional networks. In 2015 28th SIBGRAPI Conference on Graphics, Patterns, and Images 2015; (2):273-80.

25. Mirmehdi M. Handbook of texture analysis. Imperial College Press; 2008.

26. DE Almeida Freitas F, Peres SM, DE Moraes Lima CA, Barbosa FV. Grammatical facial expressions recognition with machine learning. In The Twenty-Seventh International Flairs Conference 2014.