What does a typical CNN “see” in an emotional facial image?

Conference or Workshop Item

Published Version

Identification Number/DOI: https://doi.org/10.11159/mvml23.114 <https://doi.org/10.11159/mvml23.114>

Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. MVML 114
DOI: 10.11159/mvml23.114

What Does A Typical CNN "See" In An Emotional Facial Image?

Mathy Vandhana Sannasi[1], Dr. Markos Kyritsis[2], Dr. Katie L. H. Gray[3]
[1,2]Department of Business Informatics, Systems and Accounting, Henley Business School,
University of Reading, Whiteknights Rd, Reading RG6 6UD, United Kingdom;
[3]School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6UR, UK
[1]m.sannasi@pgr.reading.ac.uk; [2]m.kyritsis@henley.ac.uk; [3]k.l.h.gray@reading.ac.uk

Abstract **-** The objective of this research is to understand the current capabilities of artificial neural network algorithms and contrast them to the human visual system, in order to identify the most effective features to support affective automation. This can, in turn, aid in optimisation of resources used for storage and transmission by understanding which level of information can be used to augment and potentially accelerate accurate identification of emotional facial expressions. For the first part of our experiment, which we present in this paper, we focused on evaluating feature selection of facial expression images using machine learning. 70 (10 examples of each basic emotion) randomly selected from the NIMSTIM dataset images were used, which were split into train (56) and test (14) sets. The testing images were then processed using Singular Vector Decomposition to vary the levels of information shown in the image. Next, the training dataset was used to train a Convolutional Neural Network algorithm with 18 layers (with convolutional, max pooling, dropout, flattening and activation layers) and 66,884,615 trainable parameters. The validation accuracy was 45% and the confusion matrix showed that the emotion disgust was predicted at almost 100% accuracy, surprise at 55%, and sorrow/happiness/neutral at 46-47%. As expected, the granularity level of the test images had an effect on the successful predictions. A feature map visualisation was performed to demonstrate what the CNN "sees" (i.e., the feature selection) in the image in order to accurately predict the emotional expression type. For the next phase of our experiment, we plan on contrasting the features and performance to that of the human visual system using an experimental design with eye tracking.

Keywords**:** artificial facial emotion recognition, convolutional neural networks, information loss, singular value decomposition, feature visualisation, affective computing, human-computer interaction

1. Introduction

The need for artificially identifying emotional expressions from faces, especially fear, is increasingly important (e.g., for road safety, security (national and home), healthcare, etc) [1, 2]. However, state-of-the-art facial emotion recognition algorithms require large datasets to train, which are then decomposed into eigenspace, right from the first task of detecting face and its features [3]. This in turn leads to large storage and transmission costs as well as transfer rates [4]. In contrast, evidence suggests that detecting emotional faces is effortless for the typical human visual system, as shown by experiments involving masking, where an emotional face is rapidly succeeded by a neutral expression. For humans, studies have shown that there is an objectively measurable psychophysiological response (e.g., a galvanic response, or dilated pupils) even when the emotional face stimulus does not enter the participants' subjective awareness [5, 6]. Therefore, amongst the various limitations of facial emotion and expression analysis, use of artificial neural networks has not managed to mimic the human visual system, neither in terms of process, nor in terms of prediction performance.

2. Background

Artificial emotion recognition, also called affective computing [7], is an area of research that aims to improve user experience by enabling machines to identify, understand and react appropriately to human emotional states [8]. Emotions have been known to play a vital role in human socialisation and civilisation [9], and thus, recognising human emotions is considered, by many, to be important in increasing efficiency and comfort of human interactions with computers and computerised systems. Human emotions, an integral part of personal and social spheres, can be artificially recognized using several ways. Recent research has seen a lot of tractive force towards machine learning and artificial intelligence, and their scope includes both statistical models and neural network models. Some of these models have been inspired by the biological substrates of learning (e.g., perceptron networks and evolutionary algorithms), with a particularly high interest seen in areas of facial and speech expression recognition [10]. In the field of facial expression recognition, the six basic human emotions (sorrow, happiness, anger, disgust, fear and surprise) have been considered the benchmark for its widely occurring research [10, 11].

2.1. Facial expression recognition

Facial expressions enable emotional communication [12]. Facial action can illustrate a person's positive and/or negative experience, with the most enjoyed emotion usually getting displayed on one's face [13]. Artificial facial expression recognition involves two types of approaches: feature detection and template matching [12, 14]. Both of these phases directly or indirectly start with facial detection, which is usually performed using the well-tested and efficient Viola Jones method combining support vectors, boosting and cascading classifier [15]. This method, proposed by Paul Viola and Michael Jones in 2001, is known for its good and concurrent detection rates and is based on object recognition [16]. Feature selection approaches were succeeded by model-based algorithms such as Convolutional Neural Networks (CNN) [17] since they were more efficient in solving non-linear, pattern recognition problems such as emotional expression recognition.

2.2. Convolutional Neural Networks (CNN)

Neural networks, inspired from the way human neurons interact with each other [18], are regarded as the main constituents of artificial intelligence frameworks, and are being predominantly used for classification problems, right since their inception in the forms of Perceptron and ADALINE [19-21]. CNNs, a variant of artificial neural networks, have recently been used for many pattern detection tasks and thus were considered the framework of choice for several applications such as face detection [22], facial expression recognition [23], and image classification [24]. These networks have one or more of the following layers: convolutional layer, pooling layer, activation layer and fully connected layer and are known to have high performance in machine learning problems [25]. The deep and highly connected layers with feed-forward propagation, along with the profound architecture comprising of local receptive fields, weight sharing and spatial subsampling makes CNNs computationally less intensive and better capable of generalising, and thus, learning highly abstract features and objects in an effective and efficient manner [26, 27]. However, still being heavily data-driven in terms of the need for sizable image volume for better performance and the tendency to be overfitted to the training data [28], CNN has its pitfalls and can benefit greatly from being benchmarked against a superior system [29, 30] – i.e., the human visual system.

3. Materials and methods

3.1. Dataset

NIMSTIM set of facial expressions [31] is a large and multi-racial (African, Asian, European and Latino-American) image dataset of facial expressions (672 images by 43 professional actors demonstrating 16 different facial poses including neutral). The reliability and validity of the dataset was established using validation sessions by third party participants for precise identification of expressions and high intra-participant agreement. These aspects have been valuable for the current research and hence this dataset was selected. For this study, a 70-image subset was randomly chosen from the original dataset using GraphPad web service (https://www.graphpad.com/quickcalcs/randomSelect1/) using stratified random sampling i.e. 10 images of each emotion category were chosen to enable a balanced dataset and were processed further.

3.2. Image pre-processing

A set of preparatory steps were carried out on the selected NIMSTIM images using Python libraries such as OpenCV. As the first step, since presence of colour is known to hinder salient feature identification such as edges [32], the images were converted from colour- to grayscale. This can also aid in decreasing the computational requirements since colour can introduce more information into the algorithm, which in most cases, might not be necessary, and thus will amplify the training of the model to achieve a good performance [33]. Secondly, the images were subject to Histogram equalisation, which equalises the density graph of the image [34]. This technique is claimed to boost recognition rates by reducing noise (inequalities) in the brightness of the images and thus, enhancing the features of the image [35]. The third step was face detection and cropping. A well-known frontal face detection algorithm proposed by Paul Viola and Michael Jones in 2001, uses Haar-like features for face detection [36]. The faces in the images were detected and an oval mask (with average value of the grayscale of the pictures) was placed to isolate the facial area. Finally, the images were standardised to a resolution of 226 x 364.

The prepared images were then randomly split, using the GraphPad web service, into a training set (56) and test set (14), with the same number of images for each emotion type (eight in the training set and two in the test set for each emotion type). The test images were then subject to dimensionality reduction using the matrix factorisation technique called Singular Value Decomposition (SVD). SVD was chosen since it has been used for several years for image processing in order to reduce storage requirements and to meet stringent image-size constraints [37] for large computing requirements [38]. In SVD, the image A (dimensions m x n) is decomposed into a product of three matrices, U, S, V which can be written as:

$$A = USV^T \tag{1}$$

where U and V are orthonormal matrices with dimensions m x m and n x n respectively. U contains the left singular vectors of A, V contains the right singular vectors of A and S is a diagonal matrix with the singular values of matrix A [38]. Rank k of the matrix A is used to denote the number of its non-zero singular values [37]. The above expression can further be decomposed into:

$$A = \sum_{i=1}^{r} u_i \sigma_i v_i^T \tag{2}$$

where u is the value in the i^{th} column of U, v is the value in the i^{th} row of V and σ is the singular value in the ith row of S [39]. This summation exists for all the k values of σ and thus, SVD can be used to compress an image to a lower rank by performing the summation till the desired level of k [38, 40]. This technique was used to dimensionally reduce the test images to k = 2 till k = 20 and the reduced images were individually exported into a .png format. The rank 20 was ascertained by deciding to keep 99.75% of the original image information in the rank 20 image. Thus, the test dataset had 266 images (14 original images with 19 levels of granularity) and these images were then used to test the CNN model.

3.2.6. Convolutional Neural Network Model

A CNN model was built using 56 training images belonging to seven emotion classes (anger, disgust, fear, happiness, sorrow, surprise, neutral).

Layer	Type	Output Shape	No. of Parameters
1	Conv2D	(None, 362, 224, 128)	1280
2	MaxPooling2D	(None, 181, 112, 128)	0
3	Dropout	(None, 181, 112, 128)	0
4	Conv2D	(None, 179, 110, 256)	295168
5	MaxPooling2D	(None, 89, 55, 256)	0
6	Dropout	(None, 89, 55, 256)	0
7	Conv2D	(None, 87, 53, 512)	1180160
8	MaxPooling2D	(None, 43, 26, 512)	0
9	Dropout	(None, 43, 26, 512)	0
10	Conv2D	(None, 41, 24, 512)	2359808
11	MaxPooling2D	(None, 20, 12, 512)	0
12	Dropout	(None, 20, 12, 512)	0
13	Flatten	(None, 122880)	0
14	Dense	(None, 512)	62915072
15	Dropout	(None, 512)	0
16	Dense	(None, 256)	131328
17	Dropout	(None, 256)	0
18	Dense	(None, 7)	1799
Total parameters			**66,884,615**

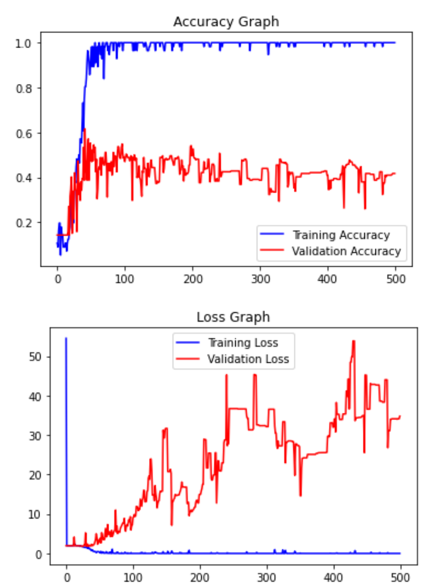

Fig. 1: Model summary of the built CNN model (left) and Accuracy and loss graphs of train/test data (right)

The model used convolutional, max pooling and RELU activation layers with the final activation layer as a SoftMax layer suitable for multiclass classification (of 7 classes). The parameters used are batch size 5, epochs 100 and input shape 364 x 226 x 1. The compiler used was RMSProp and the loss function, categorical cross entropy. Max pooling function was used to localise the features better [24]. Dropout layers were added to avoid overfitting of the model. The model generated 66,884,615 trainable parameters generated (Fig. 1 (right)).

4. Results and discussion

4.1. Results

The overall accuracy of the model was 100% on the training set, yet 42% on the test set (Fig. 1 (left)). A confusion matrix was built to check the class-wise accuracy for the emotion type variable. The model predicts disgust emotion type with the highest accuracy of 100% irrespective of the granularity level of the test image, followed by surprise facial images at 55%, then neutral and sad facial images at 47%. The model fails to predict fear images and most angry images (accuracy 3%) (Table 1), indicating a large variance in model accuracy.

Table 1: Confusion matrix for the validation data (multiclass dependent variable, emotion type).

Actual emotion \ Predicted emotion	A	D	F	H	N	S	Sr
A	3%	0%	0%	0%	0%	97%	0%
D	0%	100%	0%	0%	0%	0%	0%
F	0%	0%	0%	0%	50%	0%	50%
H	0%	8%	38%	46%	8%	0%	0%
N	3%	0%	42%	0%	47%	8%	0%
S	0%	0%	47%	0%	3%	47%	3%
Sr	0%	0%	0%	0%	45%	0%	55%

A = Anger; D = Disgust, F = Fear, H = Happy, N = Neutral, S = Sad, Sr = Surprise

4.2. Further analysis using Generalised Mixed Models (GMMs)

The test data output of the CNNs was used to build generalised mixed models to check for fixed and random effects on the success of the prediction. The first GMM model was built using Granularity levels as the random variable and the model returned a singular fit. Two more GMMs were built with emotion type (actual) and File ID as random variables respectively (comparison in Table 2). In both these models, granularity level was considered the fixed variable.

Table 2: Comparison of the GMM models.

	GMM1	GMM2
Random variables	Label	File ID
Fixed variables	Gran (not significant)	Gran (levels 3 and 4 were significant)
AIC	244.9	113
BIC	312.2	180.3
Deviance	204.9	73.0
Loglik	-102.4	-36.5
Variance of random effects	8.66 (Label)	70.47 (File ID)
Std deviation of random effects	2.94 (Label)	8.40 (File ID)
Bayes factor	1	4.27×10^{28}

In addition, a Bayes factor estimation was performed on the data using Bayesian Information Criteria (BIC) which compares the fit of the data under null hypothesis as opposed to that of the alternate hypothesis [41]. It suggested that the second model, with the random variable, File ID, is 4.27×10^{28} more likely to fit the data than the first model. Hence this model was considered for further investigation. The normality of the residuals was checked, and it was found to be mostly normal. This logistic mixed model (estimated using ML and optimx optimizer) was built to predict Success with Gran (formula:Success ~ Gran) and included File ID as a random effect (formula: ~1 | File ID). The model's total explanatory power is substantial (conditional R2 = 0.96), and the part related to the fixed effects alone (marginal R2) is of 0.02. The model's intercept, corresponding to Gran = 2, is at -0.57 (95% CI [-7.78, 6.64], p = 0.876).

Within this model: The effect of Gran level 3 is statistically significant and negative (beta = -5.04, 95% CI [-9.42, -0.66], p = 0.024; Std. beta = -5.04, 95% CI [-9.42, -0.66]) and the effect of Gran level 4 is statistically significant and negative (beta = -5.14, 95% CI [-9.59, -0.69], p = 0.024; Std. beta = -5.14, 95% CI [-9.59, -0.69]). The effects of the rest of the levels of granularity are statistically non-significant. A simple logistic regression with Gran and File ID as the independent variables and success as the dependent variable was built. The File ID and Granularity variables were not found to be statistically significant in this model and the GMM model displayed better performance (GMM accuracy 98% against Logistic regression accuracy 94%; GMM balanced accuracy 98% vs. Logistic regression 95% on the validation set) and hence the importance of the random effect of File ID variable was illustrated. Considering the high validity measures of the original NIMSTIM dataset (mean overall proportion correct = 0.81 (SD = 0.19); mean kappa = 0.79 (SD = 0.17)) [31], the random effect of the file ID might imply that there could be individual actor's variations in emotional expressions and/or features of emotional faces, that needs further evaluation.

4.3. Feature visualisation

Activation visualisations was performed in Python, and it showed the features captured in the deeper layers of the CNNs for each emotion type. For example, the surprise emotion type image (Fig. 2) showed that the features, eyebrows, open mouth and wide eyes were captured by Layer 10.

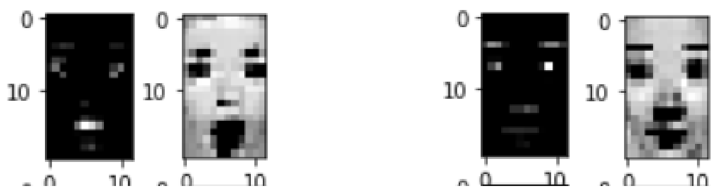

Fig. 2: Activation visualisation for surprise emotion. From left to right (excerpts of visualisations for surprise emotion type): Deep layer of masked_61F_SP_O; Deep layer of masked_62F_SP_C

Investigating one of the images of disgust emotion type (Fig. 3) for all layers from 1 till 10, we can see that the regions of the face around the eyes and mouth were focussed upon as the layers went deeper. Since this network is almost 100% accurate in recognising the disgust emotion type, another image (with a male actor) was chosen to visualise the activations. A similar feature detection was noticed with the eyes and mouth being the primary focus (Fig. 3).

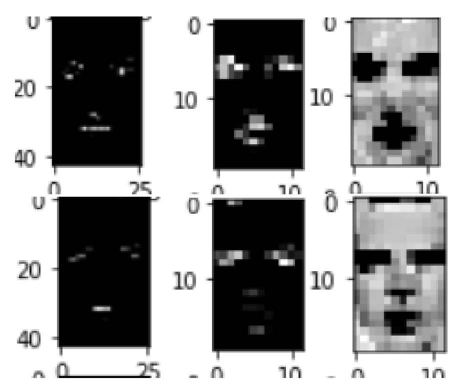

Fig 3: Activation visualisation. Top: From left to right (disgust emotion type): excerpts from deep layers of masked_11f_di_c; Below: From left to right (disgust emotion type (male actor)): excerpts from deep layers of masked_15M_DI_C

It was, however, interesting to note from the confusion matrix that the misclassifications for happy, neutral and sad emotions were mostly fear. Hence the visualisations of these emotion types were compared with the fear emotion activation visualisation. In these visualisations (Fig. 4), there is a major focus on the eyes, and it is also surprising to note that there is lesser focus on the mouth (smile) for happy emotion.

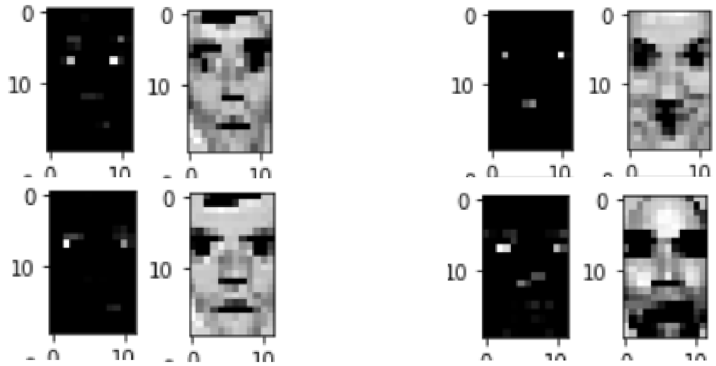

Fig 4: Activation visualisation excerpts for misclassified emotions with Fear. From top to bottom, left to right: deep layers of masked_28M_FE_C (Fear), masked_40M_HA_C (Happy), masked_48M_NE_C (Neutral) and masked_58M_SA_C (Sad)

An anger emotion which was recognised as a sad emotion in 97% of the test anger images was compared in terms of visualisation with the sad emotion type. It showed (Fig. 5) that in both the cases, a furrowed eyebrow was detected by the network, and it could have been the reason behind the misclassification.

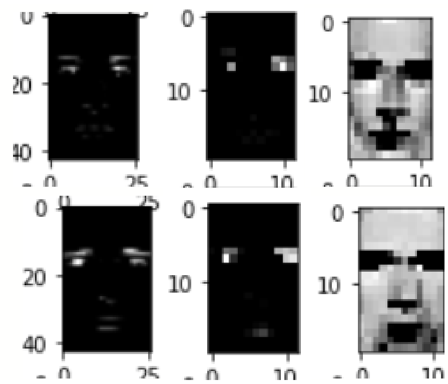

Fig. 5: Activation visualisation excerpts for misclassified anger with Sad emotion. From top to bottom: deep layers of masked_01F_AN_C (anger) and Layer 7 and layer 10 of masked_56M_SA_C (sad)

5. Conclusion

In this paper, we have performed the machine learning phase of our research which aims to compare both human- and machine-based predictors of emotion. A set of facial images showing basic emotional expressions were used and it was seen that the convolutional neural network model was good at predicting the disgust expression, followed by surprise, sad, and neutral expressions. The GMMs built on the test results of the model showed that there was a significant influence of granularity levels on the success of the model's prediction. Moreover, the random effects of File IDs were evident in the model's exploratory power and that the model assuming random effects performed better than the GMM model that did not, as well as a simple logistic regression with fixed variables. This implies that there are more factors (e.g. individual variations in expressions, facial features etc.) than just the emotion type that could have influenced the model's decisions, needing further investigation. Furthermore, the feature visualisation of the activation layers showed that the eyes were the major features visualised during most emotions, with the exception of disgust, that had a focus on the features of the nose and mouth. This gives a good basis for the next phase of our research, which aims to investigate the features used by the human visual system, as well as to compare accuracy of facial expression recognition to the CNN under varying levels of information loss. Therefore, for the next phase of our research, we are planning an event-related experiment using human participants on the same image set, while feature selection will be determined with the help of an eye-tracker.

References

[1] J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, M. Portela-García, and C. López-Ongil, "Fear Recognition for Women Using a Reduced Set of Physiological Signals," *Sensors*, vol. 21, no. 5, p. 1587, 2021.

[2] A. Leone, A. Caroppo, A. Manni, and P. Siciliano, "Vision-Based Road Rage Detection Framework in Automotive Safety Applications," (in eng), *Sensors (Basel, Switzerland),* vol. 21, no. 9, p. 2942, 2021, doi: 10.3390/s21092942.

[3] S. Lawrence, C. L. Giles, T. Ah Chung, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks,* vol. 8, no. 1, pp. 98-113, 1997, doi: 10.1109/72.554195.

[4] A. M. Rufai, G. Anbarjafari, and H. Demirel, "Lossy image compression using singular value decomposition and wavelet difference reduction," *Digital Signal Processing,* vol. 24, pp. 117-123, 2014/01/01/ 2014, doi: https://doi.org/10.1016/j.dsp.2013.09.008.

[5] M. Tamietto and B. de Gelder, "Neural bases of the non-conscious perception of emotional signals," *Nature Reviews Neuroscience,* vol. 11, no. 10, pp. 697-709, 2010/10/01 2010, doi: 10.1038/nrn2889.

[6] D. D. Vlastos, M. Kyritsis, V. A. Varela, S. R. Gulliver, and A. Papaioannou-Spiroulia, "Can a Low-Cost Eye Tracker Assess the Impact of a Valent Stimulus? A Study Replicating the Visual Backward Masking Paradigm," *Interacting with Computers,* vol. 32, no. 1, pp. 132-141, 2020, doi: 10.1093/iwc/iwaa010.

[7] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies,* vol. 59, no. 1, pp. 55-64, 2003/07/01/ 2003, doi: https://doi.org/10.1016/S1071-5819(03)00052-1.

[8] I. Lopatovska and I. Arapakis, "Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction," *Information Processing & Management,* vol. 47, no. 4, pp. 575-592, 2011/07/01/ 2011, doi: https://doi.org/10.1016/j.ipm.2010.09.001.

[9] D. Sander, D. Grandjean, G. Pourtois, S. Schwartz, M. L. Seghier, K. R. Scherer, and P. Vuilleumier, "Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody," *NeuroImage,* vol. 28, no. 4, pp. 848-858, 2005/12/01/ 2005, doi: https://doi.org/10.1016/j.neuroimage.2005.06.023.

[10] A. Saxena, A. Khanna, and D. Gupta, "Emotion recognition and detection methods: A comprehensive survey," *Journal of Artificial Intelligence and Systems,* vol. 2, no. 1, pp. 53-79, 2020.

[11] S. B. Daily, M. T. James, D. Cherry, J. J. Porter, S. S. Darnell, J. Isaac, and T. Roy, "Chapter 9 - Affective Computing: Historical Foundations, Current Applications, and Future Trends," in *Emotions and Affect in Human Factors and Human-Computer Interaction*, M. Jeon Ed. San Diego: Academic Press, 2017, pp. 213-231.

[12] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Procedia Computer Science,* vol. 108, pp. 1175-1184, 2017/01/01/ 2017, doi: https://doi.org/10.1016/j.procs.2017.05.025.

[13] P. Ekman, W. V. Freisen, and S. Ancoli, "Facial signs of emotional experience," *J Pers Soc Psychol,* vol. 39, no. 6, p. 1125, 1980.

[14] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image and Vision Computing,* vol. 31, no. 2, pp. 164-174, 2013/02/01/ 2013, doi: https://doi.org/10.1016/j.imavis.2012.10.002.

[15] M. D. Putro, T. B. Adji, and B. Winduratna, "Adult image classifiers based on face detection using Viola-Jones method," in *2015 1st International Conference on Wireless and Telematics (ICWT)*, 17-18 Nov. 2015 2015, pp. 1-6, doi: 10.1109/ICWT.2015.7449208.

[16] K. Vikram and S. Padmavathi, "Facial parts detection using Viola Jones algorithm," in *2017 4th international conference on advanced computing and communication systems (ICACCS)*, 2017: IEEE, pp. 1-4.

[17] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A survey of 3d face recognition methods," in *International Conference on Audio-and Video-based Biometric Person Authentication*, 2005: Springer, pp. 891-899.

[18] L. Noriega, "Multilayer perceptron tutorial," *School of Computing. Staffordshire University,* 2005.

[19] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing,* vol. 37, no. 1, pp. 54-115, 1987/01/01/ 1987, doi: https://doi.org/10.1016/S0734-189X(87)80014-2.

[20] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological review,* vol. 65, no. 6, p. 386, 1958.

[21] B. Widrow and M. E. Hoff, "Adaptive switching circuits," Stanford Univ Ca Stanford Electronics Labs, 1960.
[22] D. Timoshenko and V. Grishkin, "Composite face detection method for automatic moderation of user avatars," *Computer Science and Information Technologies (CSIT'13)*, 2013.
[23] A. Uçar, "Deep Convolutional Neural Networks for facial expression recognition," in *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 3-5 July 2017 2017, pp. 371-375, doi: 10.1109/INISTA.2017.8001188.
[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
[25] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017: Ieee, pp. 1-6.
[26] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015/05/01 2015, doi: 10.1038/nature14539.
[28] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, 2021/03/31 2021, doi: 10.1186/s40537-021-00444-8.
[29] J. M. Garcia-Garcia, V. M. Penichet, and M. D. Lozano, "Emotion detection: a technology review," in *Proceedings of the XVIII international conference on human computer interaction*, 2017, pp. 1-8.
[30] A. P. Anokhin, S. Golosheykin, E. Sirevaag, S. Kristjansson, J. W. Rohrbaugh, and A. C. Heath, "Rapid discrimination of visual scene content in the human brain," *Brain Research*, vol. 1093, no. 1, pp. 167-177, 2006/06/06/ 2006, doi: https://doi.org/10.1016/j.brainres.2006.03.108.
[31] N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, and C. Nelson, "The NimStim set of facial expressions: judgments from untrained research participants," (in eng), *Psychiatry Res*, vol. 168, no. 3, pp. 242-9, Aug 15 2009, doi: 10.1016/j.psychres.2008.05.006.
[32] O. E. Hammoumi, F. Benmarrakchi, N. Ouherrou, J. E. Kafi, and A. E. Hore, "Emotion Recognition in E-learning Systems," in *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*, 10-12 May 2018 2018, pp. 1-6, doi: 10.1109/ICMCS.2018.8525872.
[33] C. Kanan and G. W. Cottrell, "Color-to-Grayscale: Does the Method Matter in Image Recognition?," *PLOS ONE*, vol. 7, no. 1, p. e29740, 2012, doi: 10.1371/journal.pone.0029740.
[34] T. Celebi, I. Shayea, A. A. El-Saleh, S. Ali, and M. Roslee, "Histogram Equalization for Grayscale Images and Comparison with OpenCV Library," in *2021 IEEE 15th Malaysia International Conference on Communication (MICC)*, 1-2 Dec. 2021 2021, pp. 92-97, doi: 10.1109/MICC53484.2021.9642082.
[35] L. Jing, X. Liu, and G. Yin, "The research and implementation of the method of pretreating the face images based on OpenCV machine visual library," in *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, 12-14 Aug. 2011 2011, vol. 5, pp. 2719-2721, doi: 10.1109/EMEIT.2011.6023595.
[36] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, 2001, vol. 1: Ieee, pp. I-I.
[37] H. Andrews and C. Patterson, "Singular value decompositions and digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 1, pp. 26-53, 1976, doi: 10.1109/TASSP.1976.1162766.
[38] L. Cao, "Singular value decomposition applied to digital image processing," *Division of Computing Studies, Arizona State University Polytechnic Campus, Mesa, Arizona State University polytechnic Campus*, pp. 1-15, 2006.
[39] K. M. Aishwarya, R. Ramesh, P. M. Sobarad, and V. Singh, "Lossy image compression using SVD coding algorithm," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 23-25 March 2016 2016, pp. 1384-1389, doi: 10.1109/WiSPNET.2016.7566363.

[40] S. Kahu and R. Rahate, "Image compression using singular value decomposition," *International Journal of Advancements in Research & Technology,* vol. 2, no. 8, pp. 244-248, 2013.
[41] E.-J. Wagenmakers, "A practical solution to the pervasive problems ofp values," *Psychonomic Bulletin & Review,* vol. 14, no. 5, pp. 779-804, 2007/10/01 2007, doi: 10.3758/BF03194105.