EPiC Series in Computing

Volume 82, 2022, Pages 150–159

Proceedings of 37th International Conference on Computers and Their Applications

EPiC Computing

Facial Emotion Characterization and Detection using Fourier Transform and Machine Learning

Aishwarya Gouru and Shan Suthaharan*

University of North Carolina at Greensboro, Greensboro, NC 27402, USA
a_gouru@uncg.edu, s_suthah@uncg.edu

Abstract

We present a Fourier-based machine learning technique that characterizes and detects facial emotions. The main challenging task in the development of machine learning (ML) models for classifying facial emotions is the detection of accurate emotional features from a set of training samples, and the generation of feature vectors for constructing a meaningful feature space and building ML models. In this paper, we hypothesis that the emotional features are hidden in the frequency domain; hence, they can be captured by leveraging the frequency domain and masking techniques. We also make use of the conjecture that a facial emotions are convoluted with the normal facial features and the other emotional features; however, they carry linearly separable spatial frequencies (we call computational emotional frequencies). Hence, we propose a technique by leveraging fast Fourier transform (FFT) and rectangular narrow-band frequency kernels, and the widely used Yale-Faces image dataset. We test the hypothesis using the performance scores of the random forest (RF) and the artificial neural network (ANN) classifiers as the measures to validate the effectiveness of the captured emotional frequencies. Our finding is that the computational emotional frequencies discovered by the proposed approach provides meaningful emotional features that help RF and ANN achieve a high precision scores above 93%, on average.

1 Introduction

Facial expressions—such as the happy, sad, and sleepy emotions—are some of the brain's responses to psychological events. Hence, the development of computational and machine learning techniques to characterize and detect facial emotions would be useful for addressing the recently reported psychological problems and mental disorders in [14]. In computer science discipline—especially in the last two decades—many face recognition and emotion detection techniques have been proposed [8, 6, 19]. However, they still suffer from two major drawbacks. The first drawback is the data paucity problem. It means that it is difficult to acquire sufficient data samples for all the possible emotions such that a machine learning classifier can be trained accurately and built. The second drawback is the detection and extraction of emotional feature vectors so that a meaningful feature space can be constructed to develop machine learning models. Therefore, there is an urgent need to study facial emotion detection problems with novel

[*] Corresponding author – s_suthah@uncg.edu

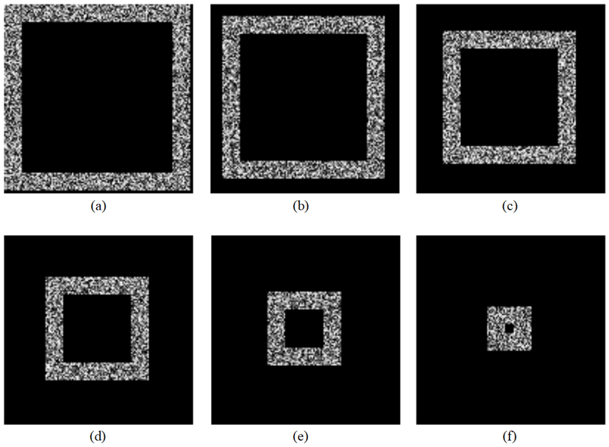

Figure 1: It illustrates some of the narrow-band rectangular frequency kernels. The emotions create distinct horizontal and vertical distortions, in turn, they generate such orthogonal emotional frequencies. Hence, the narrow-band "rectangular" frequency kernels help capture such frequencies with minimum spatial aliasing.

approaches. Hence, this paper introduces a new terminology— the computational emotional frequency—in the Fourier domain and study its contribution to characterize emotional features.

Emotional frequency has been defined and widely used in psychology discipline [7]; however, the computational emotional frequency never been introduced or studied using Fourier transform for facial emotion detection. Hence, to the best of our knowledge, the proposed work brings novelty to the facial emotion recognition research. In simple terms, the computational emotional frequency may be described by the modeling and simulation of emotional frequencies. With the computational emotional frequencies, we have introduced a novel approach to uniquely characterize different emotions with distinct frequency bands by leveraging the discrete Fourier transform [1]. We validated this approach using the classification performance scores of the random forest and artificial neural network classifiers as the qualitative measures.

2 Objectives

We present a Fourier-based machine learning technique that translates the emotional signals—emitted by the facial emotions—into linearized emotional frequencies in vertical and horizontal directions to characterize emotions, construct feature vectors, and detect facial emotions. The first objective is to find the origins of a range of vertical frequencies of the signal emitted by the emotional features, given a narrow-band horizontal frequencies and the same range of horizontal frequencies of the signal, given a narrow-band vertical frequencies. It includes the discovery of these emotional frequencies using the rectangular narrow-band kernels. A subset of rectangular narrow-band kernels applied in the Fourier domain are shown in Figure 1. Our definition of the computational emotional frequencies is defined by spatial origins of the hidden narrow-band frequencies in the Fourier domain, which are masked by these kernels. The second objective is to construct feature vectors (or a feature space) using these analogies and study their effectiveness using performance scores of the RF and ANN classifiers, as the qualitative measures, by applying them on the feature space of the emotional features.

3 Background

The current facial emotion recognition systems are evolved from the fundamental system—the facial cction coding system (FACS)—that helps the grouping of the facial changes based on their descriptions on the surface of the facial regions [5]. The FACS has been later extended by Tian, Kande, and Cohn [17] with the concept of Action Units, which allow the integration of facial landmarks (e.g., lips, eyes, and mouth) into the system to improve its performance in terms of the grouping of the emotions (e.g., happy, anger, and sad). The integration of the landmarks creates a significant amount of uncertainties while increasing computational complexity. Hence, Viola and Jones [18] proposed an algorithm to increase the speed of computation by adapting the AdaBoost learning and cascade approach to speed up the computation along with a new concept called Integral Image that also allows faster computation in a real-time environment. In our approach, we tried to avoid the use of landmarks; however, the frequency domain techniques indirectly include the emotional frequencies emitted in the facial landmark regions.

The random forest models have been extensively studied to understand their suitability for face detection [9, 10, 12]. For example, in [9], authors have conducted an experimental research and studied the performance of random forest and support vector machine (SVM) in terms of detecting faces to make them suitable for mobile applications. In [10], authors mainly focused on the facial textures; hence, they combined the local patterns and random forest learning to develop models for facial recognition. In [12], authors have implemented a technique by combining Gabor Filter and Oriented Gabor Phase Congruency Image with random forest learning. All these approaches resulted in significantly high accuracy for the random forest classification. This is one of the reasons for us to use the performance scores of the random forest as the measure to validate the proposed approach to discover emotional frequencies.

Teo, De Silva and Vadakkepat [16] developed a model for facial expression detection that uses integral projection, statistical computation, a neural network and Kalman filtering. This approach also achieves a very good accuracy. Another emotion recognition recognition technique is proposed in [8] that performs facial expression analysis using the concept of neurofuzzy network that integrates psychological findings in the analysis. Authors of [19] used kernel whitening, support vector data descriptors, and Gaussian based classifiers in their proposed approach. This approach concluded that one-class classification methods can reach a good balance between the labeling and computation overheads, and the recognition performance. Facial expression recognition problem was also studied by Otsuka and Ohya [11], but they used the hidden Markov models (HMM) in a dynamic environment with multiple subjects. Authors first used a two-dimensional Fourier transformation to construct Feature vectors, by image processing techniques, and then applied HMM to represent distinct facial expressions.

Similarly, Cohen et al. [3] studied the facial expression recognition problem, but they focused on video sequences; hence, they developed temporal and static models. They used the multi-level HMM classifiers to recognize facial expression sequences and segment video sequences. In contrast, De Silva, Miyasato, and Nakatsu [4] studied the facial emotion recognition problem using multimodal information of the emotions acquired from the audio, video, and hybrid clips. They assigned weights to both audio and video inputs and generated outputs based on the input information. Similarly, a technological review by Garcia-Garcia, Penichet and Lozano [6] discussed some of the different types of expression analysis models that are available in this research domain. This literature survey suggests none of these techniques addresses the problem of characterizing the emotions, before developing machine learning, without using the landmarks (region of interest) and extracting emotional features at various levels of narrow-band frequencies. Our goal is to fill this gap by presenting a novel approach.

4 Methods

Facial images are generally formed by the convoluted facial and emotional features, and an additive noise. Hence, we can mathematically model this example as follows:

$$s = u \circledast v + \epsilon \tag{1}$$

where u represents a set of convoluted facial features, v represents a set of convoluted emotional features, and the operator $\circledast$ represents the convolution operator. For the purpose of emotion detection, we ignore the noise term, since efficient noise filtering may be applied to remove noise. Hence, the noise-free model for facial emotion detection is:

$$s = u \circledast v \tag{2}$$

To simplify the explanation, we consider two subjects that satisfy the above noise-free model.

$$\texttt{Subject 1}: \qquad s_1 = u_1 \circledast v_1; \quad s_1' = u_1 \circledast v_1', \tag{3}$$

where u_1 represents the first subject's facial features, and v_1 and v_1' represent two emotional features of the first subject.

$$\texttt{Subject 2}: \qquad s_2 = u_2 \circledast v_1; \quad s_2' = u_2 \circledast v_1' \tag{4}$$

where u_2 represents the second subject's facial features, and v_1 and v_1' represent the same two emotional features of the second subject. Now suppose a facial image $s \in \{s_1, s_2\}$ that satisfies the model $s = u * v$ is given, then we could raise two important questions:

- Question 1: Is it possible to extract u from s and determine if $u = u_1$ or $u = u_2$ by using a machine learning model? In other words, if a facial image is given, is it possible to find who is the person? This is facial recognition;
- Question 2: Is it possible to extract v from s and determine if $v = v_1$ or $v = v_1'$ by using a machine learning model? In other words, if a facial image is given, is it possible to find what is the emotion? This is emotion detection.

The goal of this paper is to address the challenges associated with the second question without using a facial recognition solution and propose a ML-based solution to detect facial emotions. In essence, we propose to develop a ML model f such that

$$y = f_{\alpha,\beta}(s) \tag{5}$$

where the parameter set α is the training parameter, β is the hyper-parameter, and y is the response variable which represents the emotion labels (e.g., sleepy, happy, or sad).

4.1 Generalized Model

In our proposed approach, we are mainly interested in grouping image pixels of s with respect to their frequency bands (i.e., computational emotional frequencies) with the hope of revealing the emotional features that are usually convoluted in the spatial domain s, and latent in the frequency spectrum S of s. The revelation of emotional features helps us construct a meaningful feature vectors and a feature space to develop machine learning models to classify facial emotions. Hence, we use FFT to transform the image signal s in equation(2) as follows:

$$S = T(u \circledast v), \tag{6}$$

where $s = u \circledast v$, u is the convoluted facial features, v is the convoluted emotional features, and T represents FFT. Facial landmarks generally play major roles in characterizing and detecting facial emotions in many applications. Our proposed approach does not utilize the landmarks; however, the emotional frequencies of the facial landmarks are automatically included in the frequency spectrum. Using the convolution theorem, we could write the convolutional model in equation(6) by the following product equation:

$$S = T(u) \times T(v) \tag{7}$$

Hence, by applying the frequency masking technique, we could expand the above model to a consolidated parametric model for multiple subjects and emotions as follows:

$$F_i = M_i(S) = M_i(T(u)) + M_i(T(v)), \tag{8}$$

where M_i is the i^{th} narrow-band frequency kernel that is parametrized and linearized with respect to the fixed width, b, of the narrow-band kernels and the number of kernels, p, F_i is considered the i^{th} emotional frequency, and $i = 1, 2, \ldots, p$.

4.2 Computational Emotional Frequencies

The emotional frequencies F_is are captured in the Fourier domain using the narrow-band frequency kernel M_i, while computational emotional frequencies s_i are defined in the spatial domain by applying the inverse FFT (i.e., iFFT), as follows,

$$s_i = T^{-1}(F_i) = T^{-1}(M_i(T(u))) + T^{-1}(M_i(T(v))) \tag{9}$$

where s_i is considered the i^{th} computational emotional frequency that provides the i^{th} feature for the feature space, where the feature space satisfies:

$$y = f_{\alpha,\beta}(s_1, s_2, \ldots, s_p), \tag{10}$$

where $f_{\alpha,\beta}$ is the machine learning model that is parametric with α and β with the goal of optimizing it to develop a model to detect emotions. In this paper, we use RF and ANN for the parametric machine learning model $f_{\alpha,\beta}$. In essence it solves the problem of developing a ML model to detect facial emotions, while the equation (9) characterizes the facial emotions.

We empirically show, through a set of simulations with machine learning, that FFT extracts meaningful emotional features that define separable facial and emotional frequencies. For this purpose we use the performance scores of the RF and ANN classifiers as the measures of quantifying the emotional features. The goal of this paper is to estimate the computational emotional features $s_1, s_2, \ldots, s_p$ using the emotional frequencies captured in frequency spectrum.

5 Feature Learning

We used a narrow-band frequency kernels in Frequency domain to mask out the spatial frequencies to define emotional frequencies by using equation (8) and detect them using equation (9) such that a meaningful feature space can be constructed for training and building a machine learning model. We have presented a few narrow-band frequency kernels in Figure 1. This process acts as a frequency masking (M_i) technique that captures spatially orthogonal emotional

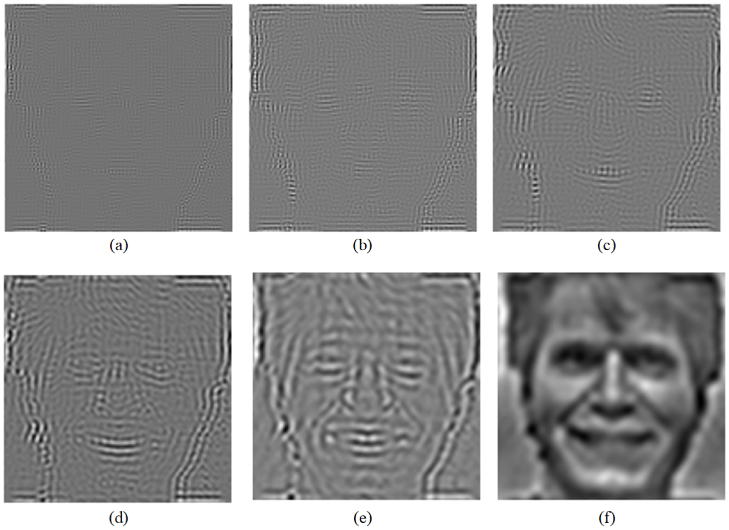

Figure 2: It illustrates some of the emotional frequencies of the happy facial expression.

frequencies, as described in section (2), with respect to given narrow-band frequencies, either vertically or horizontally. The spatial frequency features of the hidden emotions in the Fourier domain are revealed by the kernels are presented in Figures 2 and 3, for the happy and sad emotions, respectively. We can clearly see the effects of these two emotional expressions, in particular, in the frequency-band presented in Figures 2(e) and 3(e).

6 Experimental Results

In our experimental research, we have adapted the Yale-Face dataset [2] to validate our proposed approach and build a machine learning model. We have used the performance scores of the RF and ANN learning model as the measures to validate the computational emotional features detected by the proposed approach. As a result, we have developed an automated machine learning model with a companion feature space to classify facial emotions. In the Yale-Face dataset, there are 9 different facial expressions with 15 subjects; hence, there are a total of 135 grayscale images. We have only used the following 5 expressions to test our hypothesis and develop a machine learning model: 1. Happy, 2. Sad, 3. Sleepy, 4. Surprised, 5. Wink. It means we have 75 images to perform our tasks and build a model. Since we don't have sufficient data to train and build a model, we have used the fast Fourier transformation and the availability of large number of frequency bands (narrow-bands) to extract many spatial frequency features to extend the number of samples with high-dimensional feature vectors.

6.1 Feature Space Generation

As mentioned in the previous section, we have used 75 images for the feature space construction. The first step is the application of a simple dimensionality reduction. In dimensionality reduction, we have eliminated the color channels and converted the images to grayscale. The resulting grayscale images are resized to a standard of 128 x 128 pixels, we used a Bicubic interpolation

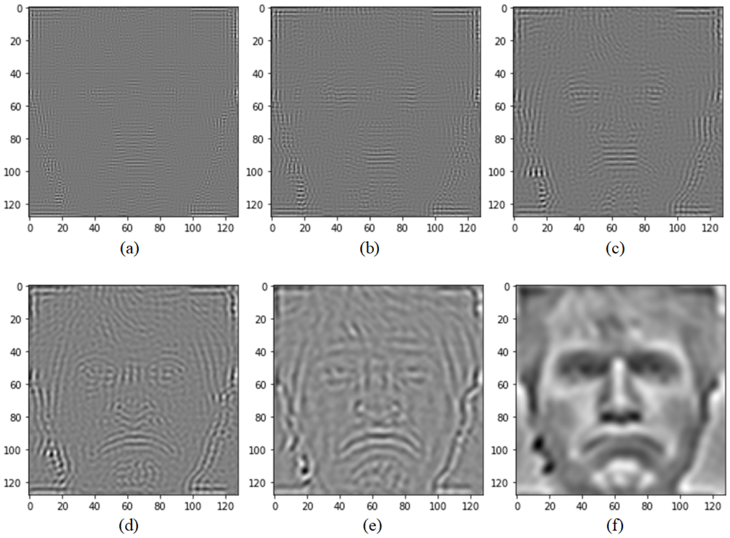

Figure 3: It illustrates some of the emotional frequencies of the sad facial expression.

for this purpose. This step is important to generate a balanced feature space. The built-in module of the Scipy FFT library is used for the generation of Fourier transformation of the images. The algorithm presented in this paper uses FFT and iFFT. FFT converts the spatial domain (input image) to the frequency domain. That means the input image is converted to a spectral image in the Fourier domain. Twenty five narrow-band frequency kernels are applied to the frequency domain as masks and the resulted amplified spectrum are converted back to the spatial domain using iFFT. These images display computational emotional features; hence, they are used to construct a feature space with dimension 25. When an image data is being written to feature space, each expression is labeled with a specific number. Labeled classes are as follows, 1- Happy, 2- Sad, 3- Sleepy, 4-Surprised, 5- Wink, which means, we have built a supervised machine learning model. These labels are manually assigned to each expression.

6.2 Model Training

For the model training, a training and testing set is built and these are used to validate the machine learning model. As mentioned earlier, the feature space generated is based on pixels, and all the image dimensions are 128 × 128; hence, the number of observations for one image is 16,384 vectors. Each image is of size 128 × 128 pixels and we used 25 frequency frames. The dimensions, 25 are achieved by choosing the alternate frames ranging from 14 to 64 that are generated by the narrow-band frequency kernels in the Fourier domain. This implies, for each image $128 \times 128 \times 25 = 409600$ feature vectors are generated. There are 15 subjects and we selected 5 emotions; hence, we have 75 image in total. Therefore, our data domain consists of 30,720,000 feature vectors in total. Some of the frames used to generate the feature space are shown in Figures 2 and 3. To train the RF model with such a huge data domain, it took an average of 30 minutes in an i5 processor when 100 trees are used. Similarly, it took about 5 hours for ANN to be trained with 75 epochs. This computational constraint is the main limitation of the proposed approach while its goal is to alleviate the data paucity problem.

Table 1: Performance scores of Random Forest model

Emotions	Accuracy	Precision	Specificity	Sensitivity
Happy	95.80	85.49	96.07	94.69
Sad	98.26	95.45	98.90	95.64
Sleepy	95.76	94.64	85.31	98.67
Surprised	97.83	95.17	98.82	93.81
Wink	98.75	94.92	98.75	96.31

6.3 Domain Splitting

After generating the feature space, we have implemented the RF and ANN classifiers. For the preparation of the algorithms, we have used domain splitting where a percentage of the data domain is randomized and stored in two data frames, namely, training and testing dataset. These data frames are used to implement he RF and ANN classifiers. The primary step to implementing the algorithm is to split the data into training and testing sets. In this experiment, we have used an 80:20 ratio to split the data domain. 80% of the data domain is randomized and is used for the training of the model. The remaining 20% of the randomized data domain is used for validating the data. The randomization helps the data to evenly populate across the data frame and help develop a better training strategy.

6.4 Model Validation

As we have implemented RF and ANN classifiers on the training and test set, the results obtained are based on the qualitative measures of the model. Since the data domain is a balanced dataset, we can include the accuracy measure if the validation process. We have evaluated the model based on the well-known qualitative measures, these measures are calculated based on the true positive, true negative, false positive, and false negative values. These values are used to generate accuracy, precision, sensitivity, and specificity scores [15]:

$$\texttt{Accuracy} = \frac{1}{1+\frac{FP+FN}{TP+TN}} \qquad \texttt{Precision} = \frac{1}{1+\frac{FP}{TP}} \tag{11}$$

$$\texttt{Sensitivity} = \frac{1}{1+\frac{FP}{TN}} \qquad \texttt{Specificity} = \frac{1}{1+\frac{FN}{TP}} \tag{12}$$

Since the feature space is approximately balanced, we included the accuracy measure, along with other measures, for the model validation, and presented the results in Tables 1 and 2.

6.5 Performance Analysis

As the dataset contains 5 expressions, the validation results of each expression using inbuilt measures are provided in Tables 1 and 2 for the RF and ANN models. Based on the obtained results, the primary expressions, all the expressions are accurately detected as measured by the qualitative measures presented above. The result obtained concludes that the computational emotional frequencies used to construct the feature space are meaningful and the ML developed by using them are highly efficient. For example, the precision value of 95.45% for the sad

Table 2: Performance scores of Artificial Neural Network model

Emotions	Accuracy	Precision	Specificity	Sensitivity
Happy	97.06	93.81	98.44	91.72
Sad	97.45	94.22	98.55	93.09
Sleepy	96.65	90.52	97.64	92.62
Surprised	97.29	92.98	98.25	93.42
Wink	97.58	93.63	98.39	94.34

emotion of the RF results indicate the TP is very high while the FP is low, and the precision value of 94.22% for the same emotion of ANN also explains the same result. We can see the similar patterns for all the emotions and the models. However, one observable difference is the precision value of 85.49% for the happy emotion of the RF results; however, it is not that significant, since the TP is very high while the FP is very low in the case as well. Hence, overall, we can conclude the proposed approach provide a solution to characterize and detect facial emotions using Fourier transform and machine learning. Also, note that we used the following parameters: For RF:- (i) 500 trees in the forest, (ii) Gini impurity for quality split of a tree, (iii) square root of the number of features is used in the best split, and (iv) no constraint is imposed in the maximum tree depth. For ANN:- (i) 4 hidden layers are used, (ii) 800, 600, and 400 neurons are used in the first, second, and third layers respectively, (iii) 5 neurons are used for the fourth layer, (iv) the activation function "relu" is used in the first three layers, and "sigmoid" is used in the fourth layer, (v) the "uniform" kernel initializer is used, and (vi) the stochastic gradient descent is used to minimize the loss that is measured by mean squared error with the "accuracy." The ANN parameters are selected by adapting the ones used in [13].

7 Conclusion

The proposed Fourier-based machine learning technique was able to characterize and detect facial emotions efficiently, which is evidenced by the high qualitative performance scores (accuracy, precision, specificity, and sensitivity) of the RF and ANN models. The proposed approach was also able to extend the data samples to tackle the high-dimensionality of the emotional features, and the limitations in the availability of data for learning. It was also able to identify the hidden emotional frequency features and extract them to linearized convoluted emotional features. However, it still suffers from the computational drawback because of the very high dimensional system that it creates. Our future research will focus on the development of low-dimensional structures from the high-dimensional system to build machine learning classifiers.

8 Author Contributions

Gouru contributed to the literature survey, computer simulation of the design and the models, and the development of the results; writing/revising of manuscript (particularly to the background and the experimental results sections). **Suthaharan** contributed to the development of the ideas that include the computational emotional frequencies, their spatial correspondence to feature vectors, and their applications to machine learning; design, modeling, and development of the experimental studies; interpretation of ML results; writing/revising of manuscript.

References

[1] Normand Beaudoin and Steven S Beauchemin. An accurate discrete Fourier transform for image processing. In *Object recognition supported by user interaction for service robots*, volume 3, pages 935–939. IEEE, 2002.

[2] Peter N. Belhumeur, Joao P Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[3] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S Chen, and Thomas S Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and image understanding*, 91(1-2):160–187, 2003.

[4] Liyanage C De Silva, Tsutomu Miyasato, and Ryohei Nakatsu. Facial emotion recognition using multi-modal information. In *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat.*, volume 1, pages 397–401. IEEE, 1997.

[5] Paul Ekman and Wallace V Friesen. Facial action coding system consulting psychologists press. *Palo Alto, CA*, 1978.

[6] Jose Maria Garcia-Garcia, Victor MR Penichet, and Maria D Lozano. Emotion detection: a technology review. In *Proceedings of the XVIII international conference on human computer interaction*, pages 1–8, 2017.

[7] E Tory Higgins, James Shah, and Ronald Friedman. Emotional responses to goal attainment: strength of regulatory focus as moderator. *Journal of personality and social psychology*, 72(3):515, 1997.

[8] Spiros V Ioannou, Amaryllis T Raouzaiou, Vasilis A Tzouvaras, Theofilos P Mailis, Kostas C Karpouzis, and Stefanos D Kollias. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks*, 18(4):423–435, 2005.

[9] Emir Kremic and Abdulhamit Subasi. Performance of random forest and svm in face recognition. *Int. Arab J. Inf. Technol.*, 13(2):287–293, 2016.

[10] Brian O'Connor and Kaushik Roy. Facial recognition using modified local binary pattern and random forest. *International Journal of Artificial Intelligence & Applications*, 4(6):25, 2013.

[11] Takahiro Otsuka and Jun Ohya. Recognizing multiple persons' facial expressions using hmm based on automatic extraction of significant frames from image sequences. In *Proceedings of International Conference on Image Processing*, volume 2, pages 546–549. IEEE, 1997.

[12] YC See, NM Noor, JL Low, and Eugene Liew. Investigation of face recognition using gabor filter with random forest as learning framework. In *TENCON 2017-2017 IEEE Region 10 Conference*, pages 1153–1158. IEEE, 2017.

[13] Himanshu Singh. *Practical Machine Learning and Image Processing.* Springer, 2019.

[14] Praveen Suthaharan, Erin J Reed, Pantelis Leptourgos, Joshua G Kenney, Stefan Uddenberg, Christoph D Mathys, Leib Litman, Jonathan Robinson, Aaron J Moss, et al. Paranoia and belief updating during the covid-19 crisis. *Nature Human Behaviour*, 5(9):1190–1202, 2021.

[15] Shan Suthaharan. Supervised learning algorithms. In *Machine learning models and algorithms for big data classification*, pages 183–206. Springer, 2016.

[16] WK Teo, Liyanage C De Silva, and Prahlad Vadakkepat. Facial expression detection and recognition system. *Journal of The Institution of Engineers, Singapore*, 44(3), 2004.

[17] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 23(2):97–115, 2001.

[18] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[19] Zhihong Zeng, Yun Fu, Glenn I Roisman, Zhen Wen, Yuxiao Hu, and Thomas S Huang. Spontaneous emotional facial expression detection. *J. Multim.*, 1(5):1–8, 2006.