Facial Expression Recognition: A Review of Methods, Performances and Limitations

Olufisayo Ekundayo
School of Maths, Statistics & Computer Science
University KwaZulu-Nata
Durban, South Africa
Email: 218085734@stu.ukzn.ac.za

Serestina Viriri
School of Maths, Statistics & Computer Science
University KwaZulu-Nata
Durban, South Africa
Email: viriris@ukzn.ac.za

Abstract—**Facial expression is one of the profound nonverbal channels through which human emotion state is communicated, its automation involves analysis and recognition of facial features. Facial Expression Recognition (FER) is categorized as behavioral biometrics, and also applicable in the field of computer vision and human computer interaction. Variations in the nature of the images to be processed; head pose, image background, light intensity and occlusion are some of the sources of the challenges with facial expression recognition system. Achieving a robust automatic facial expression recognition system invariant to the aforementioned challenges, is the goal of this research area. This paper presents an analysis of major feature extraction and classification methods, their performances in terms of accuracy and their respective limitations.**

Keywords: FER, feature extraction, Classification

I. INTRODUCTION

Facial Expression Recognition (FER) simply means inference deduced from facial deformation or facial components movement which is based on the analysis of facial muscles activation. Fasel and Luettin [1] warned that facial expression recognition should not be misplaced for human emotional detection. But since emotion state is involved in activating the facial muscles hence facial expression analysis could lead to human emotion detection. This ability is embedded in man. The reproduction of this special human intelligence in machine is the goal of Facial expression recognition system. Fig 1 is the image of six major affective states (happy, sad, fear, disgust, surprise and anger) in FER. FER has gained popularity in affective computing and this anchors on the fact that human face conveys more information in nonverbal communication channels than its counterparts; speech and body gesture [2] [3]. FER could be categorized as a behavioral biometrics and it is also prominent in several areas like; Computer Vision (CV), Human Computer Interaction (HCI), Games, medical and learning environments.

The general architecture of FER contains three phases which are; the pre-processing, the feature extraction and the classification phase. In Pre-processing, the quality of input data (image) is enhanced and redundancy is minimized or removed. In feature extraction phase, input data from the pre-processing is transformed so as to get the best representative features. It is necessary that the extracted data contain useful information that could guide the classification technique towards a right prediction. This implied that the outcome of the extraction phase has influence on the performance of the system. Just like every other biometric features, extracted features must have large between class variability and small within class variability. Classification is the last phase of FER system, the actual mapping of the action units to the labelled emotion takes place here. Classification is achieved by some special classification algorithms such as Support Vector Machine (SVM), AdaBoost algorithm, Artificial Neural Network (ANN) to mention a few. As the feature extraction is important for good system performance so also is the classification algorithm implements.

FER could be processed with static or dynamic images and the processing is challenged with non-generality of emotion, but [4] proved that the manners of expressing six basic emotion states in man are similar. Other challenges include; head pose, light intensity, background and occlusion. Dynamic images have additional temporal information which is an advantage for better recognition in FER system, yet it suffers from inconsistence frame duration of extracted dynamic features and temporal scale information loss. The goal of every research in this field is to develop a FER system that is invariant to the stated challenges. This paper presents an up to date trend of major methods employed towards achieving the dreamed FER system, their performance and respective limitation.

This paper is limited to some major methods for feature extraction and feature classification in FER. The arrangement of the paper is as follows: Section I is the introduction, Section II comprises of the feature extraction methods; this shall be viewed from the perspective of hand crafted methods and learned method. the pros and cons of each of the method involved shall be elucidated. Section III is feature classification methods; some of the conventional methods

ISBN 978-1-5386-7365-2/19/$31.00 ©2019 IEEE

fig 1 is the image of six major affective states (happy, surprise, fear, angry, sad and disgust) [5].

will be considered so also the state-of-the art methods (Deep learning). Section IV is the review discussion and V is the conclusion.

II. FEATURE EXTRACTION METHODS

The major challenge to consider in feature extraction phase is getting representative features that have small within class variance and large between class variance. The feature extracted are categorised into hand crafted features and the learned features. The hand crafted features are the features extracted using the conventional methods of pre-processing and the likes, and these can also be further classified into appearance based features, geometric based features and the hybrid based features. the appearance based feature capture the global or subsectional part of facial image, its major concern is the colour and the texture of the image. Most often employed features that fall under this category are; Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG), and Gabor Texture (GT). These methods need to take into account inherent challenges which include head poses and illumination variants.

A. Local Binary Pattern

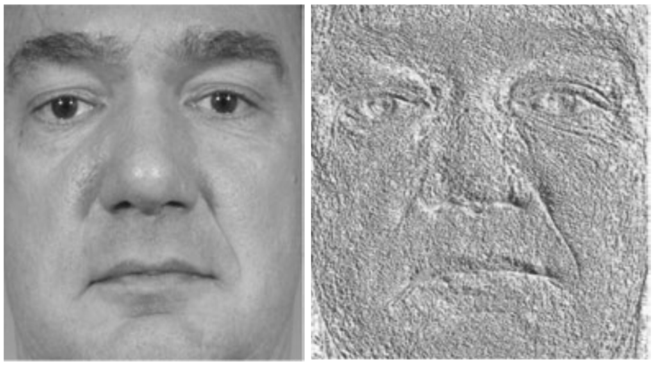
LBP is good for texture analysis, the motive behind LBP operator is that image texture can be represented by the local spatial and the gray scale contrast which are regarded as complementary measures [6]. LBP uses 3x3 pixels that contain gray scale values and threshold every neighbour pixel P(0,…,7) with the centre pixel R(1) to generate a binary sequence using a binary thresholding function given in (1) and then compute the decimal equivalent for the centre pixel with (2).

If g_p is the grey scale value in pixel P(0, 1,….7) and g_c is the grey scale value of the centre pixel R(1) then, the threshold function is given as:

$$s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad \text{where } x = g_p - g_c \tag{1}$$

$$LBP_{R,P} = \sum_{p=0}^{p-1} s(x) 2^p \tag{2}$$

Fig 2 shows the LBP version of the input image**.** LBP strength lies in its discriminating prowess, computational simplicity, high tolerance for low image resolution and invariant to illumination changes. Shahreen et al. [7] claimed that using LBP feature on SVM classifier for seven expressions in JAFFE database improved the system performance by 22%. Although the result could not be generalised because it was only experimented on a single database. Likewise, [8] carried out a study with linear programming on LBP features using JAFFE database, the result showed that an average accuracy of 93.8% was achieved. LBP is challenged with some factors



(a) (b)

Fig. 2 (a) is the input image and (b) is the corresponding LBP image [9].

like; rotation, increases in computational complexity as the size of features increases in terms of time and space, small sample size and limited information representation, because it does not consider magnitude information but pixel difference. These demerits led to its variants, [10] employed one of LBP variants called CLBP which considered the sign and magnitude information of the differences between the centre and the neighbour gray values. It achieved a better result when compared with some other appearances based methods on Cohn Kanade and JAFFE database using SVM as classifier. Details of the LBP variants is available in a survey by[9] [11]**.**

B. Histogram of Oriented Gradient

Histogram of Oriented Gradients (HOG) is a feature descriptor employed in several fields where characterization of objects through their shapes is essential. Local object appearance and shape can often be described by the distribution of local intensity gradients and edge directions. The motivation behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients and corresponding edge directions [12]. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is computed. The descriptor is the concatenation of these histograms. Since different images may have different contrast hence, contrast normalization is necessary to improve performance. This normalization results in better invariance to changes in illumination and shadowing. Kumar et al. [13] extracted HOG features from the active facial patches and fed it to SVM for classification into neutral or six universal expressions, the result shows an average accuracy of 95% with 5 folds cross-validation in extended Cohn Kanade (CK+) dataset. HOG gives better performance than many other methods. However, it is computationally intensive and hence increase the time of computation.

C. Gabor Filter Texture

Gabor texture is popularly use for feature extraction, it is a wavelet transform algorithm with great ability for good directional selectivity, shift variance sensitivity, spatial locality and maximization of information in both space and frequency domain. A Garbor filter is a function obtained from amplitude modulation of a sinusoid with Gaussian function in a spatial domain and also capture the relevant frequency spectrum in all directions. It is formally defined in equation (3) and (4). Assuming the following parameters: (x,y) to be the pixel position in the spatial domain, λ to be the wavelength in pixel, θ to be orientation of the Garbor filter and S_x,S_y to be standard deviation along the x and y direction; then:

$$G(x,y) = \frac{1}{2\pi s_x s_y} exp\left[-1/2\left(\frac{x'^2}{s_x^2}+\frac{y'^2}{s_y^2}\right)\right] exp\left[j\,\frac{2\pi x'}{\lambda}\right] \quad (3)$$

Where $x' = xcos\theta + ysin\theta, y' = -xsin\theta + ycos\theta$ (4)

Gu et al. [14] use Gabor filter for feature extraction and then the resulting Gabor decompositions were encoded with radial grids and claimed that it improved expression recognition with local classifiers. Nevertheless, Gabor filter in its quest to maximise available information incurs computational complexity and large space utilization.

D. Geometric Features

Geometric features are features extracted statistically from facial landmark displacement. The theory behind this approach is that there are subsets of face components that are more pronounced in facial expression analysis, getting the metrics relationship of these components could effectively represent the feature vectors [2]. Geometric features are not affected by lighting condition, they are not difficult to register and perform well for some Action Units. But they are not good to represent action unit that do not cause landmark displacement. Ghimire and Lee [3] extracted geometric feature automatically from video frames by using displacements based on elastic bunch graph matching displacement estimation.

E. Hybrid Features

Hybrid features give room for the research question of how best can features be combined for utmost performance. Liu et al. [15] proposed an algorithm that fuse LBP and HOG features extracted from CK+ and JAFFE database and reduced the extracted features dimensionality with PCA, after permutated the fusion on several classifiers, he found that the fused features on softmax classifier produced 98.3% on CK+ and 90% on JAFFE database. The result is an evidence that a proper hybrid features could significantly improve system performance.

F. Learned Features

Learned features, are attributed to Artificial Neural Network (ANN). Here, ANN learned the features from the input that could represent without any assistant from man. Learned features could have been the best because it is not affected by illumination, rotation, head pose and so on. But the major problem is lack of sufficient data for network to learn, which could cause overfitting. Breuer and Kimmel [16] use visualization techniques in deep learning to see the kind of feature that Convolution Neural Network (CNN) is usingfor classification, he observed that the features at the low level resembled low level Gabor filters. Khorrami et al. [17] showed that CNN learned features correspond to Facial Action Units (FACs) [4]. In order to improve the performance of CNN, both LBP and SIFT have been introduced into CNN of 1D layer [18], [19]. They recorded a better performance compared to the performance of only hand crafted features or learned features in the CNN network. Complementing learned features with hand crafted feature is still a research focus in FEA that demands consideration.

III. CLASSIFICATION METHODS

Classification algorithms are grouped based on the type of image they take has input for classification; Support Vector Machine, Adaboost algorithm, Linear Discriminant Analysis, Principal Component Analysis, Artificial Neural Network are used for static or still images. Sometimes they are also considered for an independently treated frame from sequence of frames [20]. while Hidden Markov Model and Recurrent Neural Network are used for dynamic images because classification of dynamic images requires the temporal information of the image sequences which is the strength of HMM and RNN.

A. Support Vector Machine

SVM has proved to be successful in recognizing facial expression, based on its generality especially when the labels are properly defined [21]. SVM is characterized with High performance in terms of accuracy and data size flexibility. It was purposefully model for binary problem classification with the aim of obtaining a hyperplane that maximizes the margin between two classes.

Given two class labelled training samples set {(x_i, t_i)} when I = 1, 2, ……………, N then SVM tends to find a hyperplane with the function:

$$f(x) = W.x + b \quad (5)$$

And W(weights) is computed by linear combination of vectors for the training samples as:

$$W = \sum_{i=1}^{N} \alpha_i t_i x_i \quad (6)$$

Where α is obtained by solving constrained quadratic programming problem.

FER is a multiclass problem, then the application of SVM to FER is a research focus in this field. Several multiclass approaches of SVM include; one versus all, pairwise, Simultaneous classification, and the use of loss functions [22][23].

SVM accuracy is challenged by the presence of much coarse noise in the label [21]. This affect the parallelism of the hyperplanes and also its tangential to the support vectors, which are the basics for SVM classifiers. To address the tangential condition, [24] proposed Proximal Support Vector Machine (PSVM). Also issues of hyperplanes parallelism and tangent to the support vectors have been considered in[25].

B. Adaboost

Adaboost has its name coined from Adaptive Boosting. It is a classifier that sequentially generates a strong classifier from set of week classifiers. The following are the stepwise operations of AdaBoost algorithm:

Given a set of training samples $\{(x_1,y_1),\ldots\ldots,(x_N,y_N)\}$ such that $x_i \in X$ and $y_i \in \{-1,1\}$, average weight of each samples is given as:

$$W_i = \frac{1}{N} \quad (7)$$

For some numbers of classifier h_t $t = 1, \ldots\ldots T$. there is need to get a $h_t : X \rightarrow \{-1,1\}$ with minimum error rate ε_t = prob. Of $h_t(xi) \neq yi$ and compute its updated weight as:

$$w_{t+1} = (w_i \exp(-\alpha_t y_i h_t(x_i)))\,/z_t \quad (8)$$

Where is given as:

$$\alpha = \frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t} \quad (9)$$

The esulting classifier output is then obtained as:

$$H(x) sign(\textstyle\sum_{t=1}^{T} \alpha_t h_t \ (x)) \quad (10)$$

Basically, AdaBoost is a binary classifier and it is employed in solving binary related problems like face detection [13] and gender classification [26] . FER is a multiclass problem, in order to apply Adaboost to FEA, researchers in the field adopts the multi-class version [27].

PCA and LDA are mostly use for dimensionality reduction which ensures reduction in the computational complexity cost. In most cases they are used to complement SVM and Adaboost. PCA and SVM in [28], [29] has reportedly said to have produced better result. Ghimire et al. [3] proposed two methods for FER, one is the combination of Multiclass Adaboost and dynamic time warping and the other is SVM classifier on a boosted feature vector. The result showed that SVM had better performance of 97.35% compare to the formal of 95.17%.

ANN is another classifier that is thriving in developing FER system. It was almost forgotten despite its predictive prowess as a result of optimization difficulty and its affinity for large set of data for scalability which is not available. Recently, the discovery of Deep learning for Deep Neural Networks has made neural networks a compelling tool for FER system.

C. Deep Learning

The current research and the state-of-the-art is the use of deep learning, Deep learning looks so promising as it gives better performance and it requires little or no data engineering. The most relevant type of deep learning for image processing generally is Convolution Neural Network (CNN). Application of CNN to FER involves the following processes; (1) Normalization: which is similar to preprocessing task in handcrafted approach. (2) Image cropping: this is where redundancies like image background and other parts of the image that has little or no contribution to expression representation are removed. (3) Down sampling: this ensures alignment of facial components so that CNN could learn the region specific to each expression. (4) Convolutional network: this phase has three layers which are the convolution layer, the subsampling layer and the classification layer. Convolution layer gets the pixel format of the input image using convolution kernel. Subsampling perform maximum pooling operation on the output of the convolution layer to cause dimensional reduction of the extracted features. Classification layer employs softmax classifier for prediction purposes.

Several works have been carried out with CNN on Emotion detection from facial expression [19], [16], [18]. The compelling challenges with CNN are insufficient facial expression data for proper learning and prediction and also computational cost. The issue of computational cost has been addressed from the hardware perspective. Presently, there are graphic processing devices (GPU) majorly meant for image processing. This has reduced time of computation to the bearable minimum, but it is costly to acquire for personal use most especially for researcher in the developing countries. Furthermore, overfitting occurs when the network does not have enough data to learn from. Some optimization techniques like the use of pretrained network such as; VGG16, VGG19, ResNET50, MobileNET, Xception and so on, for fine-tuning network and/or for feature extraction has been employed. Other methods include; reducing network size, adding weight regularization and adding dropout. Also the use of hand crafted features has been introduced [30], [18]. Anggraeni et al. [31] carried out experiment by engineering feature data with preprocessing techniques to improve CNN performance and found that the system performance increase from 86.08% to 97.06%. Yang et al. [19] proposed a method called weighted mixture deep neural network (WMDNN), he employed LBP feature and gray scale feature which were extracted from shallow CNN and by fine-tuning VGG16 on ImageNet respectively. Their respective outputs were fused in a weighted manner for softmax classification. When tested on CK+, JAFFE and Oulu-CASIA database, he claimed that the result outperforms FER methods that are based on the hand-crafted features or deep learning methods with one channel.

D. Hidden Markov Model

HMM is a generative model with ability to model temporal dynamics in sequenced based data. Its strength lies in modelling temporal segmentation and alignment. Sandbach et al. [32] employed Gentle AdaBoost to selectively extract feature from 3D motion based feature in BU-4DFE database and then model the extracted feature with HMM, he showed that the system has high performance of 81.93% for expression classification. Although HMM is capable of modelling variable length expression time series [33], yet as a generative model, it has a problem of being trapped in local optimal solution [34]. The disadvantage was improvised for by hybridizing a discriminant classifier and HMM model. Valstar and Pantic [35] hybridized HMM and SVM with the objective of using HMM to capture the temporal dynamics of a facial action and then employed SVM as a multiclass classifier of the features on a frame by frame basis. He claimed performance improved by 7%. Likewise, SikKa et al.

TABLE I.
SUMMARY OF THE FER METHODS, PERFORMANCE AND LIMITATION

AUTHOR	METHOD (FEATURE and CLASSIFIER)	PERFORMANCE	DATABASE/TYPE	LIMITATION
G. Sandbach et al. [32]	3D motion based feature + HMM	81.93%	BU-4DFE (Spontaneous)	Being a generative model it experience Local optimal Solution trap.
K. SikKa et al [33].	HMM+SVM	75.62% 93.89%	OULU-CASIA VIS (Spontaneous) CK+ (Posed)	SVM performs greatly with binary problem than given a multiclass problem like expression recognition.
Feng et al [8].	LBP + Linear Programming	93.8%	JAFFE(posed)	Posed dataset for system training will mislead the system in real time environment.
D. Huang et al [11].	LBP + AdaBoost	84.6%	CK+ (posed)	Loss of relevant information is possible.
Eleyan et al [29].	PCA+SVM	87%	MUFE	SVM is susceptible to noise as it affects the hyperplane parallelism and hyperplane tangent to support vector.
D.Anggraeni et al [31]	Hand crafted Preprocessing stages + CNN	97.06%	CK+ (posed), JAFFE (posed) and MUG (posed)	Tendency to overfit because of Limited data for training.
P. Kumar et al [13].	HOG+SVM	95%	CK+(posed)	It is computationally intensive because more time is spent in extracting HOG features
Y. Liu et al [16].	Hybrid (LBP&HOG) +Softmax	98.3% 90%	CK+(posed) JAFFE(posed)	Research is still finding best combinations of features.
B. Yang et al [19].	LBP + CNN (fine tune with VGG16)	Better performance than either methods	CK+, JAFFE & Oulu-CASIA	Generalization problem
T Zhang et al[18].	SHIFT + CNN	State of the art performance	BU-3DFE and Multi-PIE datasets	Generalization problem

[33] hybridized SVM and HMM as exemplar-HMM model and use it on both pose and spontaneous database, he showed exemplar-HMMs achieved 75.62% accuracy on OULU-CASIA VIS dataset and 93.89% on CK+.

IV. DISCUSSION

From table 1 the use of different classifiers; AdaBoost [11] and Linear programming [8] to classify LBP produced 84.6% and 93.8% accuracy respectively. Likewise, SVM was employed to classify PCA [29] and HOG [13], the results showed that PCA had 87% accuracy while HOG had 95% accuracy. Furthermore, it was observed that spontaneous databases [32], [33] recorded lower accuracy compared to the posed databases [16]. Even the hybrid features (LBP and HOG) generate different accuracy on different posed databases [16]. Both Hybrid features and combination of the hand crafted feature and deep learning generate a promising performance, then more focus should be given to finding the best combination of either hand crafted features or combination of both hand crafted feature and learned features.

V. CONCLUSION

The application of deep learning to FER system ease people of the stress from feature extraction and classification problems, because these are properly considered. Nevertheless, deep learning is challenged with size of dataset but this could be accounted for using some data engineering techniques on the available dataset. Also hand crafted methods can be employed as additional information that could guide a network in prediction.

REFERENCES

[1] B. F. and J. Luettin, "(PDF) Automatic Facial Expression Analysis: A Survey," *IDIAP Research Report.* [Online]. Available:https://www.researchgate.net/publication/283477585_Automatic_Facial_Expression_Analysis_A_Survey. [Accessed: 24-Sep-2018].

[2] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic

Analysis of Facial Actions: A Survey," *IEEE Trans. Affect. Comput.*, vol. 3045, no. c, 2017.

[3] D. Ghimire and J. Lee, "Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines," *Sensor*, vol. 13, pp. 7714–7734, 2013.

[4] P. Ekman, "Cognition & Emotion An argument for basic emotions An Argument for Basic Emotions," *COGNlTfON Emot.*, vol. 6, p. 169, 1992.

[5] A. Jan, "Deep Learning Based Facial Expression Recognition and Its Application," Brunel University London, 2017.

[6] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns," Springer, Berlin, Heidelberg, 2000, pp. 404–420.

[7] S. Kasim, R. Hassan, N. H. Zaini, A. S. Ahmad, and A. A. Ramli, "A Study on Facial Expression Recognition Using Local Binary Pattern," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1621–1626, 2017.

[8] X. Feng, M. Pietikäinen, and A. Hadid, "Facial Expression Recognition Based on Local Binary Patterns 1," *Pattern Recognit. Image Anal.*, vol. 17, no. 4, pp. 592–598, 2007.

[9] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Local Binary Patterns for Still Images*, 2011 editi. Springer US, 2011.

[10] F. Ahmed, H. Bari, and E. Hossain, "Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)," *Int. Arab J. Inf. Technol.*, vol. 11, no. 2, pp. 195–203, 2014.

[11] D. Huang, C. Shan, M. Ardebilian, Y. Wang, and L. Chen, "Local Binary Patterns and Its Application to Facial Image Analysis : A Survey," *IEEE Trans. Syst. Man, Cybern. Part C*, vol. 41, no. 6, pp. 765–781, 2011.

[12] M. Dahmane and J. Meunier, "Emotion Recognition using Dynamic Grid-based HoG Features," *IEEE Xplore*, pp. 884–888, 2011.

[13] P. Kumar, "A Real-time Robust Facial Expression Recognition System using HOG Features," *Int. Conf. Comput. Anal. Secur. Trends*, pp. 289–293, 2016.

[14] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognit.*, vol. 45, no. 1, pp. 80–91, Jan. 2012.

[15] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial Expression Recognition with Fusion Features," *Sensor*, vol. 712, no. 17, pp. 1–18, 2017.

[16] R. Breuer and R. Kimmel, "A Deep Learning Perspective on the Origin of Facial Expressions," *arXiv*, vol. 5, no. 5, pp. 1–16, 2017.

[17] P. Khorrami, T. Le Paine, and T. S. Huang, "Do Deep Neural Networks Learn Facial Action Units When Doing Expression Recognition ?," *IEEE Xplore*, pp. 19–27.

[18] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition," *IEEE Trans. Multimed.*, vol. 18, no. 12, pp. 2528–2536, Dec. 2016.

[19] B. Yang, J. Cao, R. Ni, and Y. Zhang, "Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images," *IEEE Access*, vol. 6, pp. 4630–4640, 2018.

[20] Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," *Handb. Face Recognit.*, pp. 247–275, 2005.

[21] S. W. Chew, S. Lucey, P. Lucey, S. Sridharan, and J. . Conn, "Improved Facial Expression Recognition via Uni-Hyperplane Classification," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2554–2561, 2012.

[22] H. H. Zhang, Y. Liu, Y. Wu, and J. Zhu, "Variable selection for the multicategory SVM via adaptive sup-norm regularization," *Electron. J. Stat.*, vol. 2, pp. 149–167, 2008.

[23] Z. Wang and X. Xue, "Multi-Class Support Vector Machine," in *Support Vector Machines Applications*, Cham: Springer International Publishing, 2014, pp. 23–48.

[24] G. Fung and O. L. Mangasarian, "Proximal support vector machine classifiers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, 2001, pp. 77–86.

[25] G. M. Fung and O. L. Mangasarian, "Multicategory Proximal Support Vector Machine Classifiers," *Mach. Learn.*, vol. 59, no. 1–2, pp. 77–97, 2005.

[26] B. Wu, H. Ai, and C. Huang, "LUT-Based Adaboost for Gender Classification," Springer, Berlin, Heidelberg, 2003, pp. 104–110.

[27] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Stat. Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[28] A. Eleyan, "Comparative Study on Facial Expression Recognition using Gabor and Dual-Tree Complex Wavelet Transforms," *Int. J. Eng. Appl. Sci.*, vol. 9, no. 1, pp. 1–13, 2017.

[29] M. Abdulrahman, A. Eleyan, A. Kelimeler, İ. Yüz, and Y. İ. Örüntü, "Facial Expression Recognition Using Support Vector Machines Destek Vektör Makineleri ile Yüz İfade Tanıma," pp. 14–17, 2015.

[30] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, 2017.

[31] D. Anggraeni, A. Wulandari, T. Barasaruddin, and D. Yanti, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition Recognition," in *Procedia Computer Science*, 2017, vol. 116, pp. 523–529.

[32] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A Dynamic Approach to the Recognition of 3D Facial Expressions and Their Temporal Models," *Face Gesture*, pp. 406–413, 2011.

[33] K. Sikka, A. Dhall, and M. Bartlett, "Exemplar Hidden Markov Models for classification of facial expressions in videos," *Comput. Vis. Pattern Recognit. Work. (CVPRW), 2015 IEEE Conf.*, pp. 18–25, 2015.

[34] K. Ghanem, A. Draa, and E. Vyumvuhore, "Differential Evolution to Optimize Hidden Markov Models Training : Application to Facial Expression Recognition," *J. Comput. Inf. Technol.*, vol. 23, no. 2, pp. 157–170, 2015.

[35] M. F. Valstar, M. Pantic, and S. Member, "Fully Automatic Recognition of the Temporal Phases of Facial Actions," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 42, no. 1, pp. 28–43, 2012.