FACIAL EMOTION RECOGNITION USING NEURAL NETWORK

M Anusha[1] G Phani Kumar[2]
[1]*M.tech Scholar, Dept. of ECE, Ganapathi Engineering College, Visakhapatnam, A.P, India*
[2]*Associate Professor, Dept. of ECE, Ganapathi Engineering College, Visakhapatnam, A.P, India*

Abstract: Facial emotion recognition is an essential aspect in human-machine interaction. In the real-world conditions, it faces many challenges, i.e., illumination changes, large pose variations and partial or full occlusions, which cause different facial areas with different sharpness and completeness. Inspired by this fact, we focus on facial expression recognition based on partial faces in this paper. We compare contribution of seven facial areas of low- resolution images, including nose areas, mouse areas, eyes areas, nose to mouse areas, and nose to eyes areas, mouth to eyes areas and the whole face areas. Through analysis on the confusion matrix and the class activation map, we find that mouth regions contain much emotional information compared with nose areas and eyes areas. In the meantime, considering larger facial areas is helpful to judge the expression more precisely. To sum up, contributions of this paper are two-fold: (1) we reveal concerned areas of human in emotion recognition. (2) We quantify the contribution of different facial parts.

Index Terms—facial emotion recognition, facial areas, class activation map, confusion

I. INTRODUCTION

Ever since September 11, 2001, people's need for security. This is due to the increase in sales for safety [1]. By outwitting the security system of the airport it could only come to such a disaster. Politicians and professionals are looking for since then, for new or better security systems to such a disaster in to be able to prevent the future.

An important role is played by the identification of persons who are already dangerous are classified. Among others, biometrics (bios = life and metron = measure). Biometrics is the doctrine of the application mathematical methods [2] on the measurement and number ratios of living things and their body parts defined. The "characteristics" of these systems become biometric Called feature. In addition to identification, there is also the verification in biometrics.

In contrast, the verification verifies the already known identity for its truth content. The identification and the verification take a broad spectrum. Already at the Door of a private individual can be a simple Identification system [3], the front door lock, being found. In the Federal Republic of Germany one thinks about the introduction of biometric features in ID cards to make identification easier and safer.

In biometrics the person identifiable by a key, chip or the like must verifizieren by a biometric feature. This means that they are e.g. through a Fingerprint as owner of the key. Among the most famous features next to the fingerprint is the iris. Man himself uses the face as identification aid. This is the acceptance of facial recognition systems greater than that of fingerprint or iris recognition systems at the user, which again and again an association in the direction of prison and surveillance state cause. A big advantage of face recognition is non-contact identification and verification. The maintenance of the face recognition [4] station is therefore lower. Grease and other traces of dirt through touching arise cannot interfere with the system. In addition, the non-contact offers Identification further applications.

There are different algorithms for face recognition: template matching, facial recognition with geometric features, facial recognition with two-dimensional Fourier transformation, the Hierarchical graph matching and facial recognition with facial faces, the face recognition with neural networks.

In this papers, different proposed neural network algorithms are Convolution Neural Network, AlexNet, and ResNet.

II. LITERATURE SURVEY

Kobayashi et al [5-6] proposed geometric feature based methods, facial components or key facial points are marked on the face and extracted to form the feature vector which represents the face geometry. Kobayashi and Hara proposed a geometric face model with 30 facial characteristic points and later they developed a real time system which works with real time images where subjects did not have facial hair or glasses.

Mehmood et al [7] have discussed ASM and AAM use both shape and texture based methods for emotion recognition. After analyzing various shape models Cootes developed ASM where an iterative refinement algorithm is adopted to fit the data in consistent with the training set. A highly effective technique for facial feature localization is AAM developed by Cootes that can detect the contours of face structures and describes the facial shape and appearance by a set of parameters. Later Batur presented Adaptive Appearance Model that can extract facial feature points under illumination variation. PCA, a classical statistical method is used for feature extraction [Andrew, 2001] and it has the ability to classify data; therefore it is often used as a classifier and a dimension reduction technique at the same time. Using

late positive component of the brain Mehmood proposed a method for emotion recognition. The next section outlines some of the commonly used classifiers for emotion recognition.

Yurtkan et al [8] have carried the Facial features that affect changes on the face are common in 3D space than 2D surface. Also, several emotions include skin wrinkles. Due to the limitations in describing facial surface deformations in 2D, there is a need for 3D space features in order to represent 3D motions of the face successfully. When compared with 2D, 3D image represents information such as depth and head rotation which is robust to light and head pose variations. In this context, 3D database BU3DFE has been developed by Binghamton University and 3D geometrical feature point data has been studied.

Mukherjee et al [9] have studied fuzzy classifier is one of the most powerful classifier to solve classification problem with vague input. The fuzzy classifier can be described as set of fuzzy rules. The RF classifier was proposed by Breima and defined as a meta-learner comprised of many individual trees. It is designed to operate quickly over large datasets and more importantly to be diverse by using random samples to build each tree in the forest. Adaboost is chosen for real time classification as it provides an added value of choosing features that are most informative to test at real time. AdaBoost is a kind of self-adaptation boosting algorithm. Using this algorithm, the multi weak learner is boosted into a strong one; it works on the fundamental philosophy that when the classifier classifies samples correctly, the weight of these samples will be reduced.

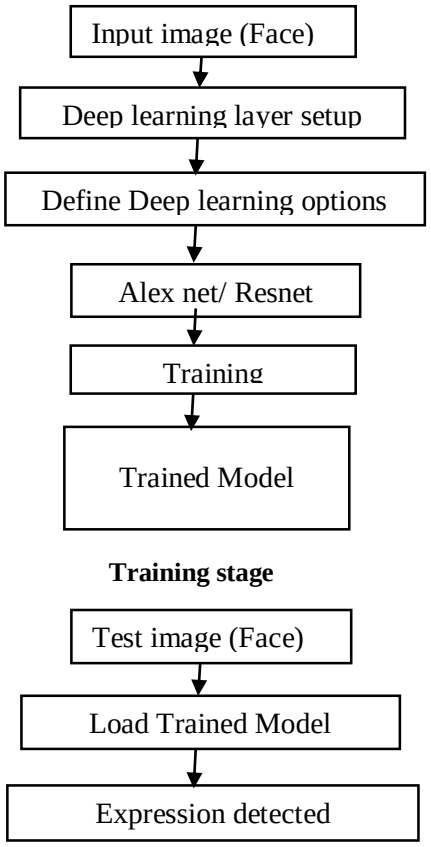
Sohail et al [10] have proposed Multiclass SVM classifier is used for classifying six emotions. Range of kernel functions like Linear, Gaussian Radial Basis Function (RBF), and Polynomial, Sigmoid etc., are used for classification. Sohail has applied SVM for an automated system for emotion recognition and found that RBF outperforms other kernels.

Wilson et al [11] have discussed most common classifier is NN. The NN is capable of dealing with features on the face which can carry out independent muscle movements: brow/forehead, eyes/lids and base of nose. The hidden layer in NN is connected with each part of the face and with each output unit. As stated by Barnard and Botha, large sample size and training set with equal number of samples from each group shows improved performance in this work, and balanced training sample given to NN yields better predictive performance as mentioned by Wilson and Sharda. Recent research has concentrated on techniques other than NN with an aspiration of gaining higher accuracy.

Dongcheng et al [12] have carried Naïve Bayes classifier performed well when labelled data is used for training and performed poorly with a mix of unlabelled data in the training set. In KNN classifier, all the training samples are considered as "representative point". The distance between test samples and the "representative point" is used to classify the emotion's decision-making.

III. METHODOLOGY

3.1 Proposed Method



Proposed system block diagram

We can observe the block diagram of the proposed method in the above figure. 2. The training phase consisted of reading the input images, defining the layers followed by the options to the layers. Once the network is defined, the training of the images starts. This produces the model of the Deep learning. This model is used for recognition of the images in the testing face.

3.2 CNN Architectures

Some specific kinds of CNN Architectures are used in this paper in cutting edge applications and research and some of the most commonly used architectures for these that are winners of ImageNet classification benchmarks are discussed in depth and are named in chronological order AlexNet, and ResNet. Some other architectures that are not as prominently used these days, but are interesting either from a historical perspective, or as recent areas of research are discussed briefly as a quick review.

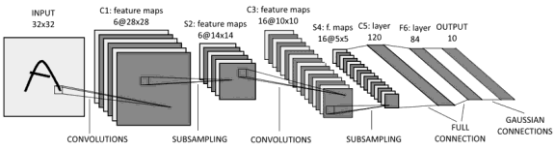


Fig.3. 2: Architecture of LeNet-5

LeNet, which is one of the first instantiations of a com Net that was successfully used in fact, is presented in detail. In addition to it, this was the com Net that took an input image, used com filters five by five filters applied at stride one and had a couple of conv layers, a few pooling layers and then some fully connected layers at the end. This fairly simple comNet was very successfully applied to digit recognition.

So AlexNet from 2012 which is also presented already before, was the first large scale convolutional neural network that was able to do well on the ImageNet classification task so in 2012 AlexNet was entered in the competition, and was able to outperform all previous non deep learning based models by a significant margin, and so this was the comNet that started the spree of comNet research and usage afterwards. So the basic comNet AlexNet architecture is a conv layer followed by pooling layer, normalization, com pool norm, and then a few more conv layers, a pooling layer, and then several fully connected layers afterwards.

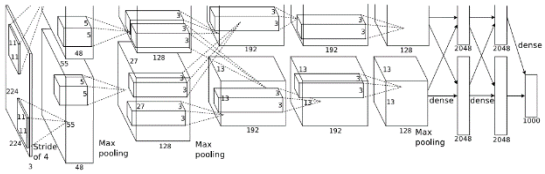


Fig.3. 3: Architecture of AlexNet

If the first layer which is a conv layer for the AlexNet, it's 11 by 11 filters, 96 of these applied at stride 4 is considered. What would be the output volume size of this first layer? And a hint is given. If input size is remembered, it is clear that convolutional filters are present, ray. A formula, which is the hint over here that gives you the size of the output dimensions after applying com. The formula is given as the full image, minus the filter size, divided by the stride, plus one and 55 is given in the figure. The spatial dimensions at the output are going to be 55 in each dimension and then 96 total filters are presented here and the depth after the conv layer is 96, which is the output volume. Now the total number of parameters in this layer are given using 96 11 by 11 filters.

Now consider at the 2015 winner, which is the ResNet network and so here this idea is really, this revolution of depth net. Depth in 2014 is started to increase, and here this hugely deeper model at 152 layers is only present which the ResNet architecture was. Now let's look at that in a little bit more detail.

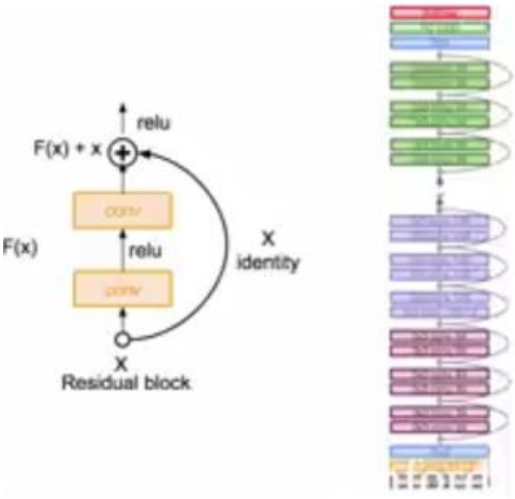


Fig.3. 4: ResNet Very deep networks using residual connection

The full ResNet architecture and it's very simple and elegant just stacking up all of the ResNet blocks on top of each other, and total depths of up to 34, 50, 100, are present and up to 152 for ImageNet are tried up.

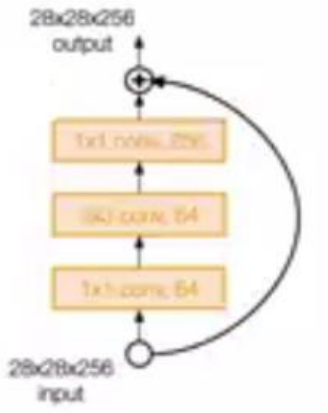


Fig.3. 5: ResNet block

One additional thing to be discussed for a very deep network is that the ones are more than 50 layers deep, they also use bottleneck layers similar to what GoogleNet did in order to improve efficiency and within each block now and what is did is going to be performed, have this 1x1 conv filter, that first projects it down to a smaller depth. So again if it is observed that let it be approximately 28x28x256 implant, this 1x1 conv is done by taking it's depth down projection and 28x28x64 is obtained. Now the convolution of 3x3 conv, that only one is present here and is operating over this reduced step so it's going to be less expensive, and then afterwards they have another 1x1 conv that projects the depth back up to 256, and so, this is the actual block that is observed in deeper networks. So in practice the ResNet also uses batch normalization after every conv layer, and use Xavier

initialization with an extra scaling factor that helped the introduction to improve the initialization trained with SGD + momentum. The learning rate uses a similar learning rate type of schedule where the learning rate is decayed when the validation error plateaus. Mini batch size 256, a little bit of weight decay and there is no drop out.

IV. RESULTS

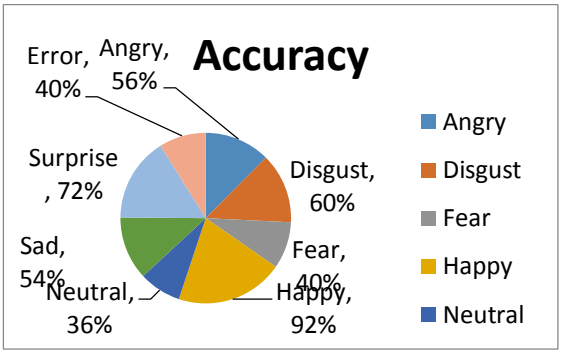
The experiments are performed on the FER 13 and Japanese face lady expression dataset. The FER 13 had images of size 48x48 and the Japanese lady dataset had images of size 256x256. Different sets of images have been considered for training and testing. The training phase consisted of reading the input images, defining the layers followed by the options to the layers. Once the network is defined, the training of the images starts. This produces the model of the Deep learning. This model is used for recognition of the images in the testing face.

Table 4.1: input taring images

angry	
disgust	
Sad	
happy	
neutral	
surprise	

a) Sad b) happy

Fig.4.1. results of expression recognition



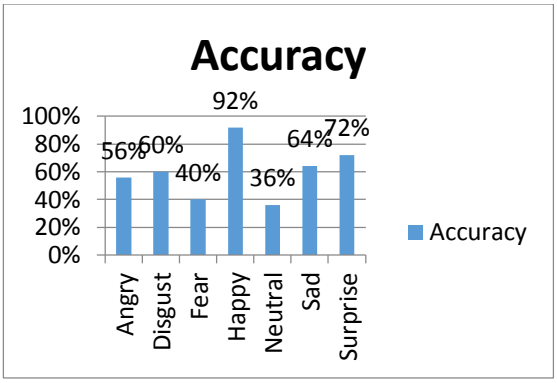


Fig.4. 2: CNN pie chart

In the above fig. 4.2 The accuracy pie chart represents the percentage of the facial expressions of the human face like Angry, fear, happy, Disgust et.,

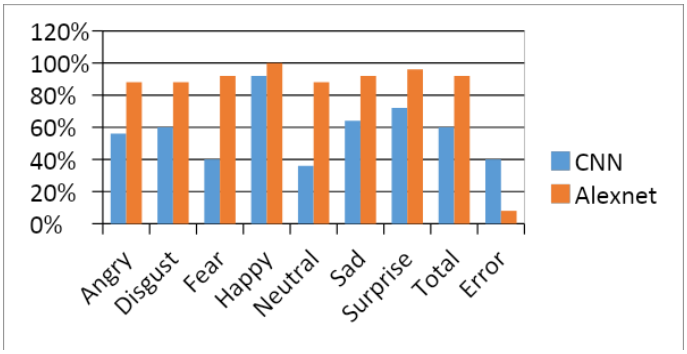


Fig.4. 3: Comparison Between CNN and Alexnet

In the above fig.4.3. The chart represents comparison between the CNN and AlexNet percentages. Blue one indicates the CNN percentage and another one is the AlexNet percentage. In this chart AlexNet results are better than CNN.

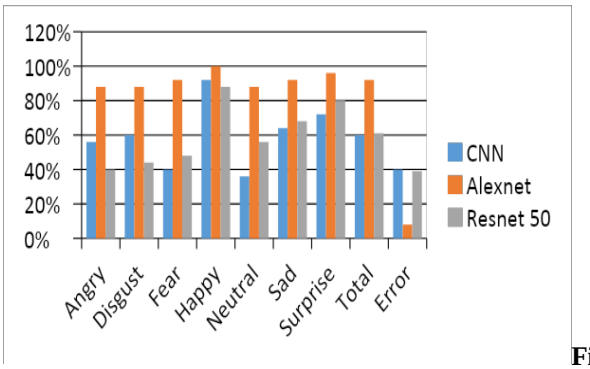


Fig.4. 4: Comparison Table

In the Fig.4.4. this figure shows that comparisons of the different face expression in CNN, AlexNet, and Resnet 50. The AlexNet results are better than CNN results.

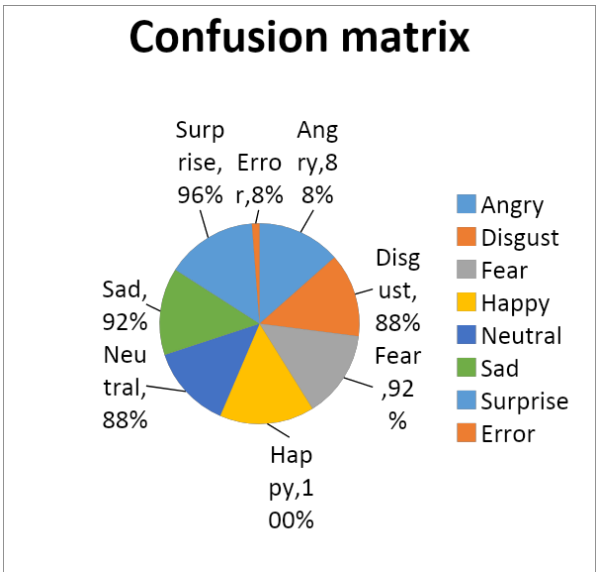


Fig.4. 5: Confusion matrix

In the above figure. 4.5 The pie chart represents the Confusion matrix of the AlexNet to different face Expressions percentages.

Table 4. 1 Confusion matrix-Alexnet

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Total
Angry	22 88%	0 0.0%	0 0.0%	0 0.0%	2 8%	1 4%	0 0.0%	88% 12%
Disgust	1 4%	22 88%	2 8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	88% 12%
Fear	0 0.0%	0 0.0%	23 92%	0 0.0%	1 4%	1 4%	0 0.0%	92% 8%
Happy	0 0.0%	0 0.0%	0 0.0%	25 100%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
Neutral	1 4%	0 0.0%	1 4%	0 0.0%	22 88%	1 4%	0 0.0%	88% 12%
Sad	0 0.0%	1 4%	0 0.0%	0 0.0%	1 4%	23 92%	0 0.0%	92% 8%
Surprise	0 0.0%	0 0.0%	0 0.0%	1 4%	0 0.0%	0 0.0%	24 96%	96% 4%
								92% 8%

In the below table 4.1. This table represents the Confusion Matrix of the Alexnet. To different face expressions.

Table 4. 2: Confusion matrix-CNN

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Total
Angry	14 56%	0 0.0%	3 12%	1 4%	2 8%	2 8%	3 12%	56% 44%
Disgust	2	15	5	1	1	1	0	60%

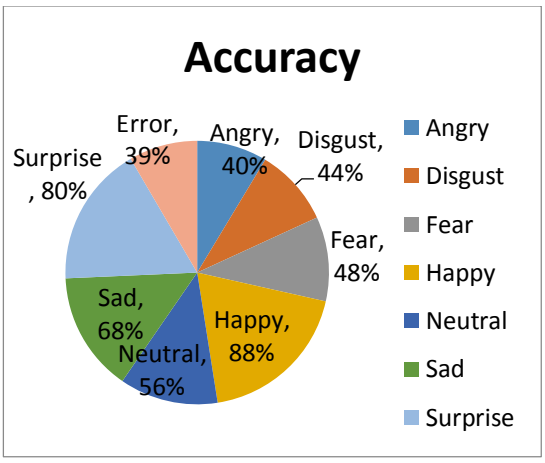
	8%	60%	20%	4%	4%	4%	0.0%	40%
Fear	3 12%	2 8%	10 40%	3 12%	1 4%	4 16%	2 8%	40% 60%
Happy	0 0.0%	1 4%	1 4%	23 92%	0 0.0%	0 0.0%	0 0.0%	92% 8%
Neutral	2 8%	3 12%	3 12%	3 12%	9 36%	4 16%	1 4%	36% 64%
Sad	2 8%	0 0.0%	4 16%	3 12%	0 0.0%	16 64%	0 0.0%	64% 36%
Surprise	0 0.0%	0 0.0%	5 20%	2 8%	0 0.0%	0 0.0%	18 72%	72% 28%
								60% 40%

In the Table 4.2 the above table shows the confusion matrix of the CNN to the different face expressions. In this table input class values and output class values.

Table 4. 3: Confusion matrix-Resnet 50

	Angry	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**	**Total**
Angry	10 40%	1 4%	3 12%	1 4%	3 12%	4 16%	3 12%	40% 60%
Disgust	2 8%	11 44%	5 20%	2 8%	2 8%	2 8%	1 4%	44% 56%
Fear	3 12%	1 4%	12 48%	2 8%	2 8%	1 4%	4 16%	48% 52%
Happy	1 4%	1 0.0%	0 0.0%	22 88%	0 0.0%	0 0.0%	1 4%	88% 12%
Neutral	2 8%	2 8%	1 4%	1 4%	14 56%	4 16%	1 4%	56% 44%
Sad	0 0.0%	2 8%	1 4%	1 4%	3 12%	17 68%	1 4%	68% 32%
Surprise	0 0.0%	0 0.0%	3 12%	2 8%	0 0.0%	0 0.0%	20 80%	80% 20%
								61% 39%

In the above table 4.3 represents the input class values and output class values of the ResNet50 network to different face Expressions.



In the Above figure 4.6. It shows that Accuracy of the confusion matrix. In different face Expressions and different values.

V. CONCLUSION

This paper concluded a brief review of FER approaches. As we described, such approaches can be divided into two main streams: conventional FER approaches consisting of three steps, namely, face and facial component detection, feature extraction, and expression classification. The classification algorithms used in conventional FER include Adaboost, random forest; by contrast, deep-learning-based FER approaches highly reduce the dependence on face-physics-based models and other pre-processing techniques by enabling "end-to-end" learning in the pipeline directly from the input images. As a particular type of deep learning, a CNN visualizes the input images to help understand the model learned through various FER datasets, and demonstrates the capability of networks trained on emotion detection, across both the datasets and various FER related tasks. A few recent studies have provided an analysis of a CNN architecture for facial expressions that can outperform previously applied CNN approaches using temporal averaging for aggregation. However, deep-learning-based FER approaches still have a number of limitations, including the need for large-scale datasets, massive computing power, and large amounts of memory, and are time consuming for both the training and testing phases.

VI. REFERENCES

[1] Y.Tian, T.Kanade, and J.F. Cohn, "Facial Expression Analysis," Handbook of Face Recognition, Springer, October 2003.

[2] M.S.Hossain and G.Muhammad, "An Emotion Recognition System for Mobile Applications," IEEE Access Special Section on Emotion-Aware Mobile Computing", vol. 5, pp. 2281- 2287, 2017.

[3] A. Konar, A. Chakraborty, "Emotion Recognition: A Pattern Analysis Approach", Wiley, 2015.

[4] T. Ojala, M. Pietikainen, and D Harwood, "Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions," Proc. 12th Int'l Conf. Pattern Recognition, vol. 1, Jersusalem, pp. 582 – 585, 9-13 October 1994.

[5] H. Kobayashi and F. Hara, "Recognition of Mixed Facial Expressions by Neural Network," *Proc. Int'l Workshop Robot and Human Comm*, Tokyo, Japan pp. 387- 391, 1-3 September 1992.

[6] H. Kobayashi and F. Hara, "Recognition of Six Basic Facial Expression and Their Strength by Neural Network," *Proc. Int'l Workshop Robot and Human Comm*, Tokyo, Japan, pp. 381-386, 1-3 September,1992.

[7] R. M. Mehmood and H. J. Lee, "A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns," Comput. Elect. Eng., vol. 53, pp. 444-457, 2016.

[8] K.Yurtkan, and H. Demirel, "Feature Selection for Improved 3D Facial Expression Recognition," Pattern Recognition Letters, vol. 38, pp. 26-33, 2014.

[9] I. Mukherjee, and R.E. Schapire, "A Theory of Multiclass Boosting," J. Machine Learning Research, vol. 14, pp. 437-497, 2013.

[10] A.S.M. Sohail, and P. Bhattacharya, "Classifying facial expressions using point-based analytic face model and Support Vector Machines", Proc. Int'l Conf. Systems, Man and Cybernetics, ISIC, Montreal, Que, Canada, pp.1008-1013, October 2007.

[11] RL.Wilson and R. Sharda, "Bankruptcy Prediction using Neural Networks," Decision Support Systems, vol. 11, pp. 545–557, 1994.

[12] S. Dongcheng, and J.Jieqing, "The Method of Facial Expression Recognition Based on DWT-PCA/LDA", 3rd Int'l Congress Image and Signal Processing (CISP2010), China, pp. 1970-1974, 2010.