Identification and Recognition of Facial expression using CNN Algorithm

[1]Dara Deepthi & [2]Dr.B.V.SeshuKumari

[1]M-Tech (CNIS), Dept of IT, Vallurupalli Nageswara Rao Vignana Jyothi Institution of Engineering &Technology, Mail Id: - daradeepthi044@gmail.com

[2]Associate Professor, Dept of IT, Vallurupalli Nageswara Rao Vignana Jyothi Institution of Engineering &Technology, Mail Id: - seshukumari_bv@vnrvjiet.in

_____

Abstract

Now a day's artificial intelligence (AI) has become most efficient and perfect algorithm for the non-linear data. In this paper we are developing Facial expression recognition system (FER) by using deep learning technique. Every person will be interested to know about the emotion of a person by his face gestures, to know the emotions, so many techniques are developed in the existing, and here we are using AFBN and GFBN from the DNN. In the AFBN the given image will be converted into LBP image after performing the pre-processing and then CNN will be applied to the output image. In GFBN the action units (AU) calculated when there is change in facial expressions. In this paper the proposed method mainly focuses on combining these two results by Softmax function to finding the error. The Top-2 error will be considered, and it is called as second highest emotion. Whenever to calculate all the emotions of a person there must be image with neutral emotions and it will be generated by using auto-encoder technique. After getting the data of neutral image we can easily get. The output results are compared with previous methods to know the efficiency of proposed method. For CK+ dataset we got 96% accuracy and for JAFFE dataset 91% accuracy. By using the DNN structure we are getting 1.5-3% improvement, when we compare to older methods.

Keywords: - LDA, LDP, SVM, FER, CNN

_____

1. INTRODUCTION

In the improved Artificial intelligence, there are voice recognition and face recognition methods. By these methods we can get the recognition of voice and languages. the FER system there are majorly four steps are those face detections, face registration, feature extraction, and classification. The face detection will be done based on some algorithms like AdaBoost, HOG and so many. The second step is performed by doing face rotation and differentiation of muscles, the expressions are captured perfectly and there must be landmarks. The differentiation in facial expressions like eyebrows, eyes, nose, and mouth are known as action units (AUs). In the feature extraction the feature points are extracted by using GFBN and AFBN. In the dimension reduction we can use Principal component analysis (PCA) and linear Discriminant analysis (LDA). For the binary features we can use LBP and local directional pattern (LDP). The classification will be performed based on the support vector machine (SVM) and hidden markov model.

In this system, we explore the practical investigation to estimate the importance of different facial regions of a human in the task of expression classification. For this motivation

_____

we implement CNN classifier with help of face images for facial expression classification such as Happy, Sad, Neutral, etc. Based to the experimental results, the upper regions of the face demonstrate to be the most important task for facial expression classification. Here for facial expression classification, we acquire facial expression classification dataset from Kaggle web resources and this dataset contains many images of seven expressions.

2. RELATED WORK

The FER system will be divided two types classical feature extraction (CFE) and DNN. In the CFE, we can consider Appearance feature-based network extraction, in this completion face will be taken into consideration. In the Geometric feature extraction method only the muscle movements and structure of the face is considered.

The comparison methodologies are divided into two types: existing database and published reports. In the first method we are going to have constrained and unconstrained data. In the second one we have training and testing sets, landmarks, error metrics and so many. Here we are taking 300 faces and applying precious techniques which are semi-automatic and the obtained results are compared with existing methods to know the accuracy of the proposed method.

3. PROPOSED METHOD

Here deep learning neural network is proposed for the FER and we focus in this paper is error detection and restructuring the result. The error will be most happening error it is called as Top-2 error. In this method we are using two networks those are appearance feature-based network and geometric feature-based network. In the first network CNN is used highly and features are extracted based on LBP feature. In the second network geometric changes are extracted for all the six emotions. These two features are combined by using the weighting function.
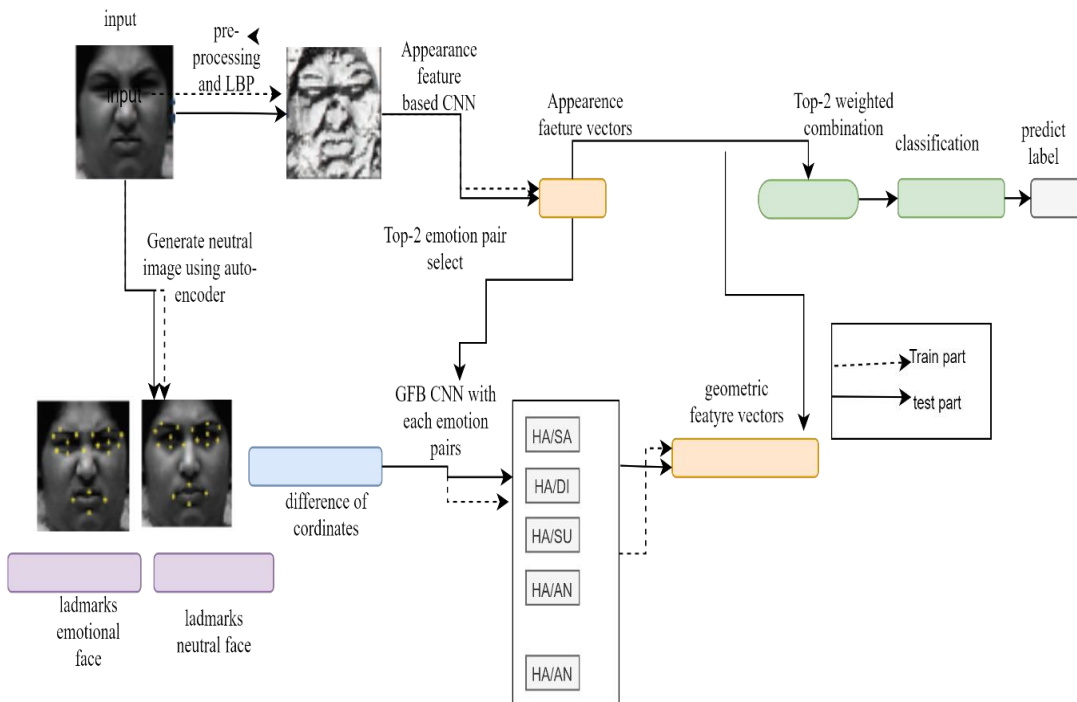

Fig.1. The proposed method for FER algorithm

By using the six-length Softmax results are measured to find the correct answer. This method is used to recognize the facial errors and reason for errors are detected by using a network. Here 150 images are taken with two datasets with 10-fold cross validation method. In this method the total dataset is divided into 10 parts and nine are used for training and remaining one is used for verification. All the results are arranged in descending order and rank was given and in every fold validation point is calculated.

Pre-processing:

Whenever we are entering into the main process, first we must select the image and then it will be cropped based on our requirement and noise everything will be removed. Here face is the most important and highly used object. The focus will be on all the parts in the face and remaining portion is not required. By using HOG algorithm, we can detect the face part, here we are using SVM, it will divide the face and non-face part with positive and negative symbols with sliding window. For removing the noise, we will apply blurring process. After pre-processing step all the images are converted into LBP image.
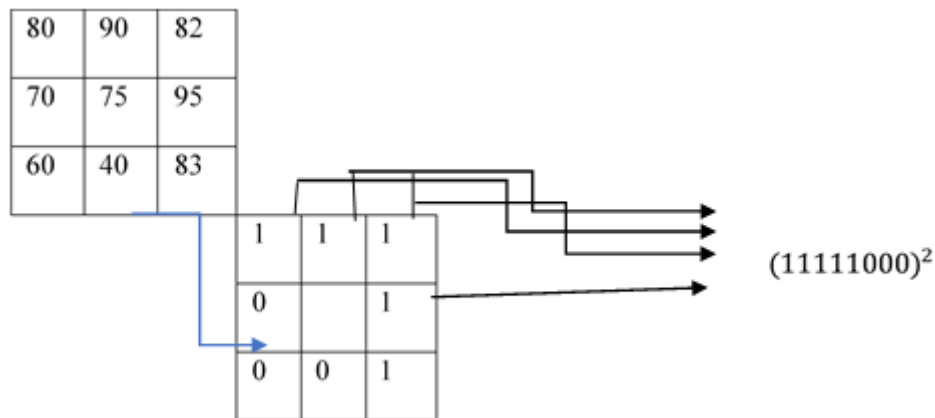

Fig:2 Encoding by using LBP

Appearance Feature-Based Network (AFBN) In this method the holistic features are extracted, here LBP are taken into consideration for the efficient process. From the image 3*3 pixels are taken into consideration and then each pixel is compared with the centre pixel if the value is greater than the centre pixel it is calculated as one otherwise zero., Finally, this process will be performed on entire process in clockwise direction. After completion of the process the entire image is converted as LBP.
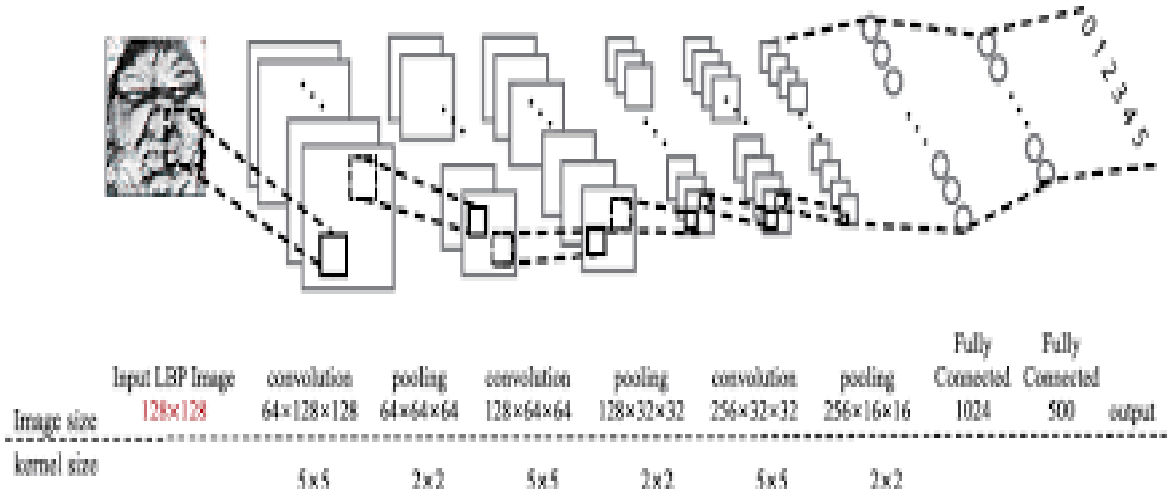
Image size	input LBP Image	convolution	pooling	convolution	pooling	convolution	pooling	Fully Connected	Fully Connected	
	128×128	64×128×128	64×64×64	128×64×64	128×32×32	256×32×32	256×16×16	1024	500	output
kernel size		5×5	2×2	5×5	2×2	5×5	2×2			

Fig:3 CNN structure for Appearance Feature-Based Network

$$LBP_{C,D}(x_a, y_a) = \sum_{C=0}^{C} S(i_c - i_a)\, 2^C \qquad (1)$$

$LBP_{C,D}$ means c= neighbouring pixels, D= centre pixel, i_c, i_a are converted Gray scale values, S(x) is binary number.

The main reason and advantage of LBP process is high speed and easy for the calculation. In the final output of LBP process the binary code will be converted into a decimal format. The main reason to convert the image into decimal format is to reduce the computational complexity of the system. In the AFBN method two methods are used one is LBP and another one is CNN. The first process is completed then we entered the second step, here the output of LBP image is given as input to the CNN.   The convolutional operation is performed on the given data with a kernel, feature maps will be generated by making the convolutional layers and pooling layers for every component the input and output shapes are extracted from the face image.

$$x_{ab}^{c,f} = \sigma\left(g_{ab} + \sum_{k=0}^{s-1}\sum_{l=0}^{s-1} w_{op} x_{a+kb+b}^{c,f} \right), \qquad (2)$$

In the figure the input image we applied convolution operation, the image size is 128*128 and convolution layer and pooling layer is applied to the image with 5*5 kernel with 4 number. Then pooling is applied to 2*2 block then 3*3 kernal is applied to the result then we got 64*64 input, then the result is changed to 64*64*64.
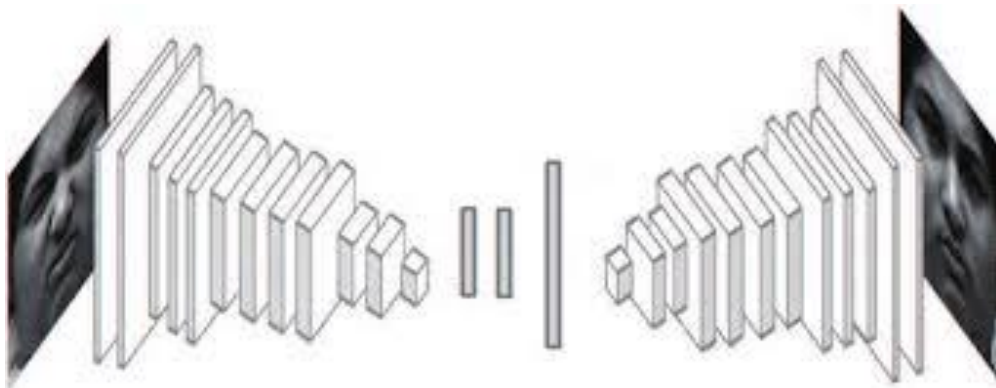

Fig:4   Neutral image with auto encoder structure   its not clear what is in this image

The Geometric Feature-Based Network (GFBN)

The emotion recognition will become very difficult by using only one feature, the difficulties like rotation, illumination, emotional difference and so many, and to get the more accurate results and error finding nature we are going for GFBN. The GFBN mainly used to capture different movements like emotions of different images. The error occurs mostly in AFBN for second highest probability, so from all 6 probabilities top-2 is selected from the softmax function. For the geometric we need a neutral face to detect the dynamic features, but in the dataset there may or may not the presence of neutral images. To detect the dynamic features of an image there has to be a presence of neutral image, so here we are applying auto encoder to find the neutral image. Image dynamic features are extracted based the difference between the coordinates.
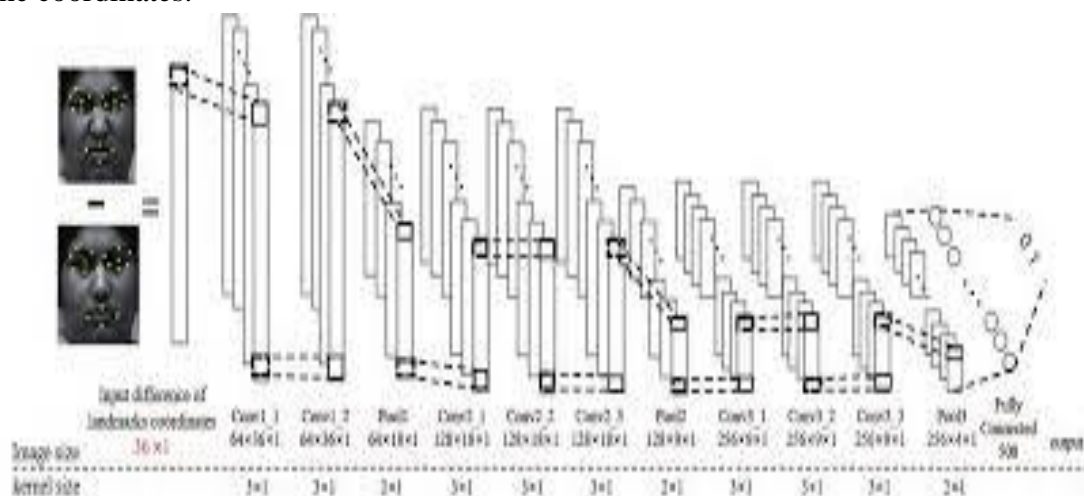

Fig:5 CNN structure for Geometric-based feature network

The neutral face image will be generated by doing encoding and decoding. In the encoding the image is given to the pool 3 layer, then convolution and max pooling layers are applied, then 4096 nodes are used to compress the layers. In the decoding process 4096 codes are extracted then up sampling and convolution process is performed. Here input will be equal to the output image, but the error function is derived from the emotion input facial image and neutral image differences are calculated.

Weighting Function for Top-2 Emotions

The two highest result is taken to perform the next process, the results of the AFBN and GFBN are taken into consideration to get more accurate and exact features. The anger, sad and all emotions are calculated with AFBN and GFBN, the AUs will be taken, and results are shown.

Evaluations of Sigmoid activation function:

Table.1 Training and validation accuracy

Epochs	Training accuracy	Validation accuracy
1	0.2394	0.2545
2	0.2419	0.2545
3	0.2481	0.2545
4	0.2525	0.2643
5	0.2687	0.2643

Table.2 Training and validation loss

Epochs	Training loss	Validation loss
1	2.7469	1.9603
2	1.8348	1.8797
3	1.8194	1.8847
4	1.8227	1.8933
5	1.8104	1.8857

Evaluations of ReLU activation function:

Table.3 Training and validation accuracy

Epochs	Training accuracy	Validation accuracy
1	0.9094	0.9173
2	0.9145	0.9242
3	0.9190	0.9255
4	0.9224	0.9286
5	0.9265	0.9298

Table.4 Training and validation loss

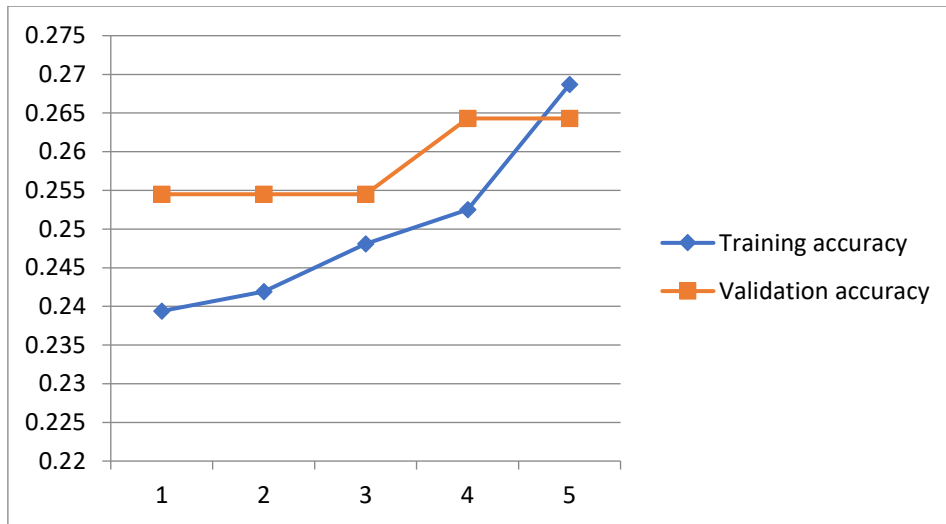| Epochs	Training loss	Validation loss
1	0.3236	0.2959
2	0.3026	0.2780
3	0.2859	0.2652
4	0.2722	0.2542
5	0.2604	0.2458

4. EXPERIMENTAL RESULTS

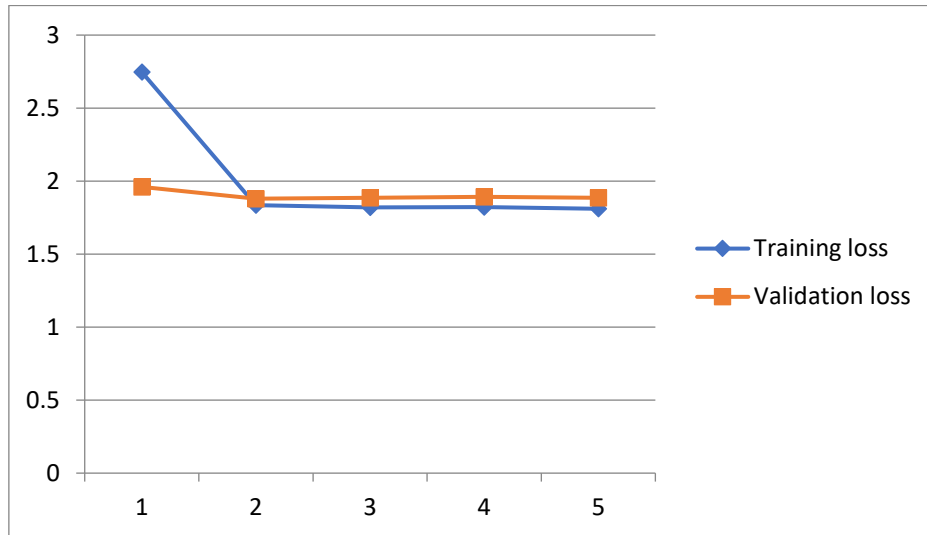Figure.6. Training and validation accuracy of Sigmoid
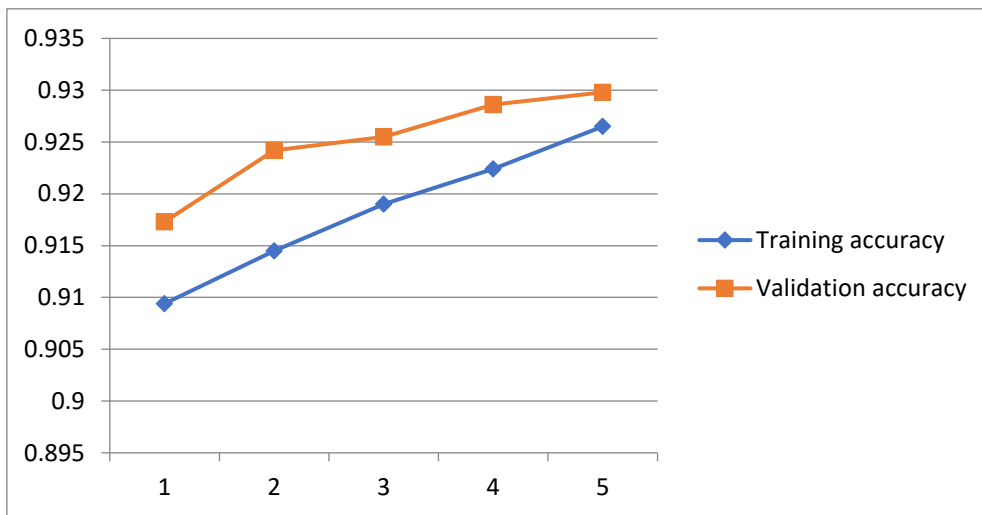
Figure.7. Training and validation loss of Sigmoid

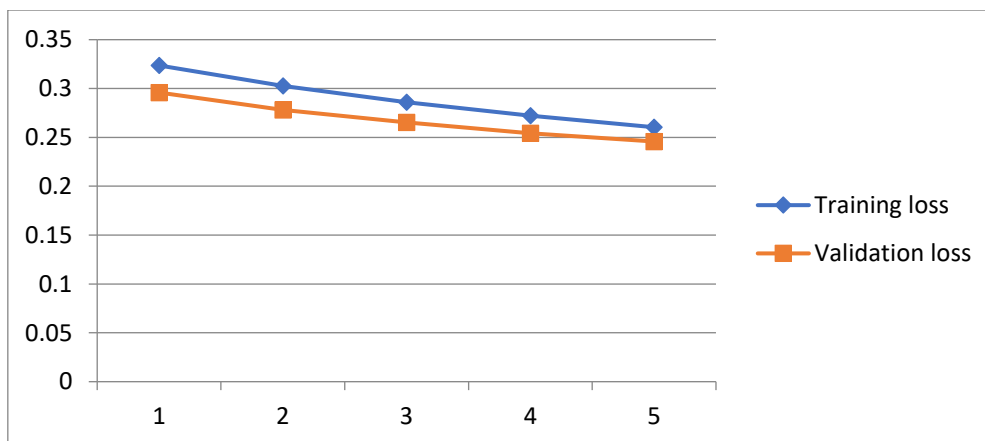

Figure.8. Training and validation accuracy of ReLU

Figure.9. Training and validation loss of ReLU

_____

Application Testing Results:
Happy:

Angry:

Neutral image

5. CONCLUSION

Neural network is the most advantages and efficient network, in this paper we are using DNN based AFBN and GFBN. By using AFBN we got the static features and by using GFBN we got the dynamic features. The static and dynamic features are combined to get the perfect facial expressions. In the CFBN the LBP features are extracted then the results are given to the CNN to get the AU. In the GFBN the neutral face and expressive face is considered and

_____

the dynamic features, the static and dynamic features are extracted from AFBN and GFBN. From the results Top-2 error is taken into consideration with 82% appearance-based feature extraction and 97% accuracy of with geometric based feature extraction.

REFERENCES

[1] C. Sagonas, "300 faces in-the-wild challenge: Database and results."

[2] V. Kazemi "One millisecond face alignment with an ensemble of regression trees."

[3] K. Simonyan" Very deep convolutional networks for large-scale image recognition."

[4] P. Lucey, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression."

[5] M. Lyons "Coding facial expressions with Gabor wavelets".

[6] S. Elaiwat "A spatio-temporal RBM based model for facial expression recognition."

[7] Y. Liu, "Facial expression recognition with fusion features extracted from salient facial areas."

[8] M. Goyani "Multi-level haar-wavelet based facial expression recognition using logistic regression."

[9] C. Shan "Facial expression recognition based on local binary patterns

[10] Y. LeCun, "Deep learning," Nature, vol. 521, pp. 436444, May 2015.

[11] N. Srivastava "Dropout: A simple way to prevent neural networks from overfitting,"

[12] I. Goodfellow "Deep Learning, Cambridge, MA, USA: MIT Press, 2016.