



Analysis of Micro Facial Expression by Machine and Deep Learning Methods: Haar, CNN, and RNN

Dibakar Raj Pant ^a, Rolisha Sthapit ^b

Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal

Corresponding Authors: ^a drpant@ioe.edu.np, ^b rolishasthapit@gmail.com

Received: 2020-08-27

Revised: 2021-03-21

Accepted: 2021-03-22

Abstract:

Facial expressions are due to the actions of the facial muscles located at different facial regions. These expressions are two types: Macro and Micro expressions. The second one is more important in computer vision. Analysis of micro expressions categorized by disgust, happiness, anger, sadness, surprise, contempt, and fear are challenging because of very fast and subtle facial movements. This article presents one machine learning method: Haar and two deep learning methods: Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) to perform recognition of micro-facial expression analysis. First, Haar Cascade Classifier is used to detect the face as a pre-image-processing step. Secondly, those detected faces are passed through series of Convolutional Neural Network (CNN) layers for the features extraction. Thirdly, the Recurrent Neural Network (RNN) classifies micro facial expressions. Two types of data sets are used for training and testing of the proposed method: Chinese Academy of Sciences Micro-Expression II (CASME II) and Spontaneous Actions and Micro-Movements (SAMM) database. The test accuracy of SAMM and CASME II are obtained as 84.76%, and 87% respectively. In addition, the distinction between micro facial expressions and non- micro facial expressions are analyzed by the ROC curve.

Keywords: Micro expression, Convolutional Neural Network, Long Short-term Memory

1. Introduction

Facial expression is an important means of emotional expression generated by the actions of the facial muscles. They are categorized into two types: macro and micro expressions. The first one is longer facial expression (about 1 to 4 sec) whereas the second one changes rapidly for about less than half of a second [12]. There are seven universal micro expressions, which are categorized as Disgust, Anger, Fear, Sadness, Happiness, Surprise and Contempt [2].

Micro facial expressions are important in computer vision. Such expressions have applied in crime prediction, clinical therapy, emotion analysis, and human-computer interaction. However, analysis of micro facial expressions is still challenging due to the very fast, and subtle facial movements [6].

In other words, micro expressions are involuntary movements that occur only in high-stake situations and

it makes hard to generate training and testing dataset [6].

Over the past few decades, many researchers have made their efforts to help computer better understand facial expressions and the form of emotional communications among humans. Pfister et al. [8] projected an outline for recognizing spontaneous facial micro-expressions using LBP-TOP (Local Binary Pattern-Three Orthogonal Planes). Song et al. [10] used a Harris3D detector with combination of HOG and the Histograms of Oriented Optical Flow (HOOF) features, and used codebook to encode features in a sparse manner of micro-expressions. Kim et al. [7] projected a new feature representation for micro-expressions by using CNN. However, no approaches are capable to detect and recognize the all seven universal micro expressions. In this article, the author has implemented one machine learning method: Haar Cascade Classifier [6] and two deep learning methods, the Convolution Neural Network (CNN) [1] in conjunction with the Recurrent



Figure 1: Example of seven different facial micro expressions (Taken from Eilert Akademie) [5]

Neural Network (RNN) [5] to analyze micro-facial expressions. The Haar Cascade Classifier is used to detect the face, the CNN is used for feature extraction and the RNN classifies different types of micro facial expressions. We use confusion matrix and ROC curve for the validation of our method.

2. Methodology

The proposed method for the classification of micro facial expression is illustrated in the Figure 2 that describes the role of three learning methods: the Haar Cascade Classifier, the Convolution Neural Network (CNN) and the Recurrent Neural Network (RNN).

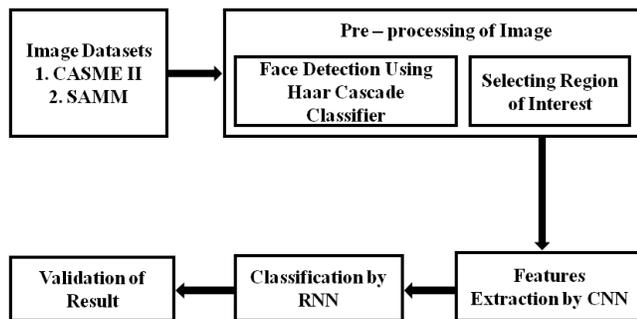


Figure 2: Block diagram for the classification micro facial expressions

2.1 Image Datasets

CASME II (Chinese Academy of Sciences Micro-Expression II) [12] and SAMM (Spontaneous Actions and Micro-Movements) [2] are two standard datasets that are used to train and test the performance of the proposed model as given in the Figure 2. The detail descriptions of data sets are given in table 1.

Table 1: Characteristics of CASME II and SAMM Dataset

Dataset	Number of Training Images	Number of Test Images	Emotions
SAMM	1330	120	Anger, Contempt, Happiness, Sadness, Disgust, Fear, Surprise and Others
CASME II	2550	210	Repression, Happiness, Sadness, Disgust, Fear, Surprise and Others

2.2 Preprocessing

There are two steps in preprocessing: detection of the face and selection of the region of interest. The Haar Cascade Classifier employs integral image concept to form a ‘cascade’ that discards non-face regions in an image [6]. The learning algorithm is based on AdaBoost which selects a small number of important features from a large set and gives highly competent classifiers and is used to create strong classifier from combining a collection of weak classification functions. The complex classifiers are combined to form a ‘cascade’ which discard any non-face regions in an image [4].

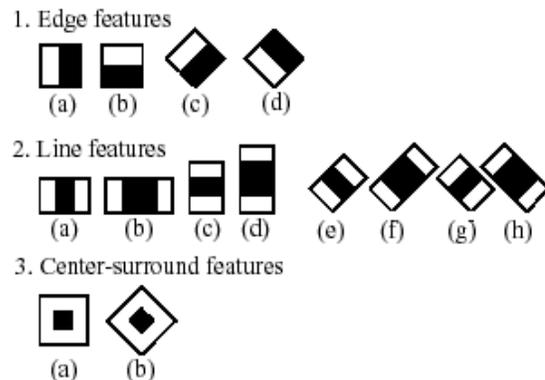


Figure 3: Examples of Haar Features

The image (matrix) I is transformed to an integral image representing a certain matrix L whose sizes coincide with sizes of the initial image as given in the Equation (1).

5	2	3	4	1	0	0	0	0	0	0
1	5	4	2	3	0	5	7	10	14	15
2	2	1	3	4	0	6	13	20	26	30
3	5	6	4	5	0	8	17	25	34	42
4	1	3	2	6	0	11	25	39	52	65
					0	15	30	47	62	81

Original image

Integral image

Figure 4: Integral Image

In this case, each matrix L element is the sum of elements I arranged in the above and to the left of this matrix element L ,

$$L(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j), \quad 0 \leq x \leq N, \quad 0 \leq y \leq M \quad (1)$$

Where $L(x, y)$ is the output image and $I(i, j)$ is the intensity of corresponding pixels of the initial image. Likewise, each matrix element can be calculated using the following relation

$$L(x, y) = I(x, y) - L(x-1, y-1) + L(x, y-1) + L(x-1, y) \quad (2)$$

After detection of the face, a rectangular bounding box is created for each image with the coordinates (x, y) and (w, h) where (x, y) are the coordinates of the image and (w, h) represents the height and width of the image. Haar Cascade Classifier groups the features into separate sub-windows and the classifier at each stage determines whether or not the sub-window is a face. In case it is not, the sub-window is discarded along with the features in that window. If the sub-window moves past the classifier, it continues to the next stage where the second stage of features is applied which is shown in figure 5.

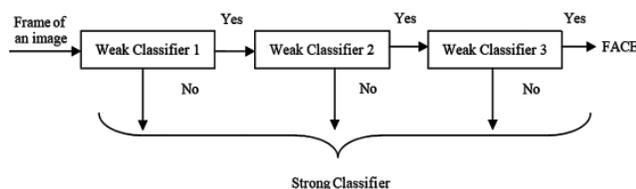


Figure 5: Cascade Classifier

2.3 Features Extraction by the Convolution Neural Network (CNN)

For the features extraction from the input images, VGGFACE [11], a modified Convolution Neural Network architecture, where, seven convolution layers each followed by the activation function ReLU (rectified linear unit) and 3 max pooling layers with the stride of value 2 is used. The kernel size is 2×2 . The mathematical form of the convolution layer is given in Equation (3):

$$\begin{aligned} o[m, n] &= f[m, n] \times g[m, n] \\ &= \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} f[u, v] g[m-u, n-v] \end{aligned} \quad (3)$$

Where $o[m, n]$ is the convolved feature map, $f[m, n]$ is the input image and $g[m, n]$ is a kernel.

The activation function ReLU performs element wise operation to reconstitute all negative values in the feature map by zero. It is defined in Equation 4.

$$f(x) = \max(0, x) \quad (4)$$

Where, x is the neuron input.

The last convolution layer of a two-dimension array of size 32 is reshaped as one dimension array and this array is fed as the input to RNN.

2.4 Classification by Recurrent Neural Network (RNN)

RNN remember their inputs over a long period of time as it contains the information in a memory. The output of CNN is reshaped to feed into the RNN using 32 layers followed by the activation function ReLU. The output from RNN is passed through the series of dropout to flatten with a dropout value of about 0.2 to create the feature map for the classification of micro facial expressions which is then fed to softmax for classification.

2.5 Softmax Classification

This layer is the final layer which classifies the different types of micro facial expressions. The output of the softmax function is equivalent to a categorical

probability distribution. Softmax assigns decimal probabilities to each class in a multi-class problem. Mathematically, the softmax function is shown in Equation (5), where z is a vector of the inputs to the output layer and again, j indexes the output units, so $j = 1, 2, \dots, K$.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^N e^{z_k}} \quad (5)$$

2.6 Validation by the Confusion Matrix

Confusion matrix has been used for the validation and performance evaluation of the model. The precision, recall and F1- score has been calculated.

Table 2: Confusion Matrix for binary classification

	Predicted True	Predicted False
Actual True	TP	FN
Actual False	FP	TN

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Precision (Exactness)} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall (Completeness)} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

3. Results and Discussion

The result of the preprocessing using the Haar cascade classifier to detect the face and selecting the region of interest is shown in figure 6.

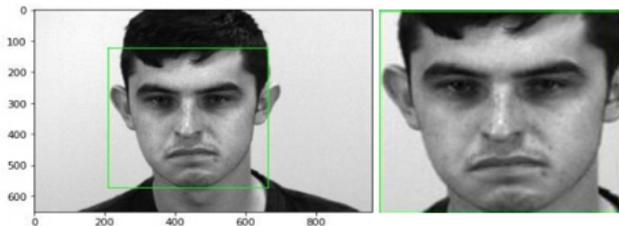


Figure 6: Face Detected by Haar cascade Classifier and Selecting Region of Interest

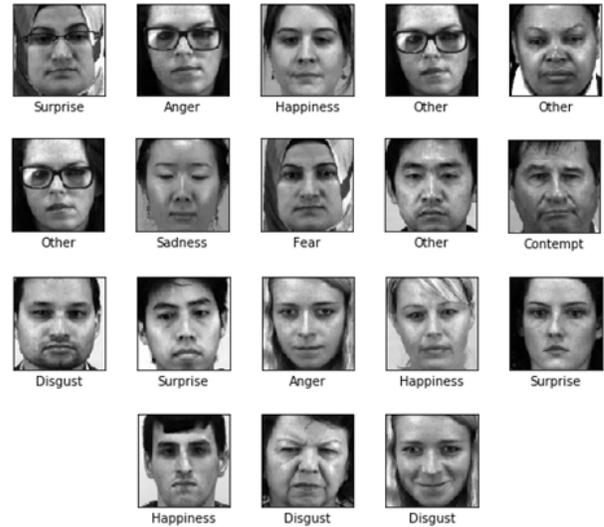


Figure 7: Results of Different Expression of SAMM Dataset

Figure 7 shows the classification of different micro expressions like anger, sadness, happiness, contempt, disgust, fear, surprise and others (non-micro expression) of SAMM dataset.

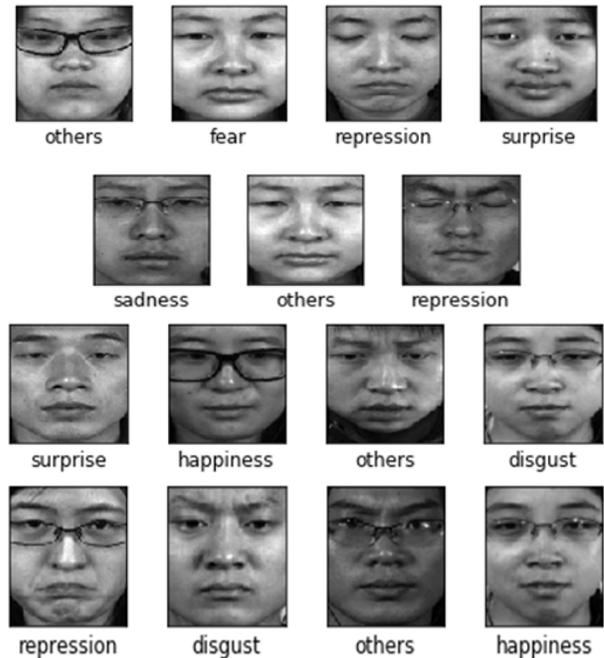


Figure 8: Results of Different Expression of CASME II Dataset

Figure 8 shows the classification of different micro expressions of CASME II dataset. The CASME II dataset has also defined that its dataset contains eight classes but the standard dataset contains no anger in it. Instead of contempt it contains the emotion repression.

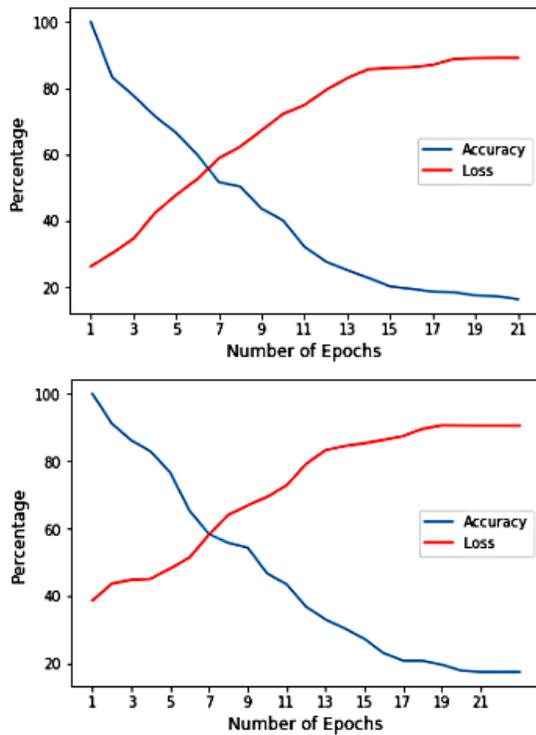


Figure 9: Loss/Accuracy vs Epochs in Recognition of micro expression analysis of SAMM Dataset (top) and CASME II Dataset (bottom)

Figure 9 shows check the accuracy/loss versus number of epochs plot in the classification of micro facial expressions from SAMM and CASME II dataset. In the both data sets, optimization of loss functions occurs at 21 epochs.

The confusion matrix evaluates the performance of the proposed model and its parameters for multi class classification are calculated and shown in the table 3.

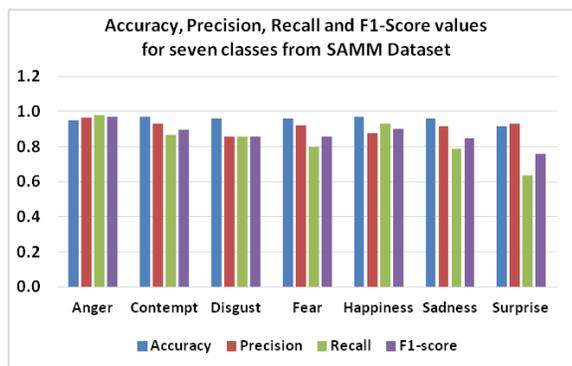


Figure 10: Accuracy, Precision, Recall and F1-Score values for six classes from SAMM Dataset

The accuracy, precision, recall and F1-score for the classification of micro expression and non-micro

expression of SAMM dataset are 0.95, 0.9626, 0.981 and 0.9687 respectively.

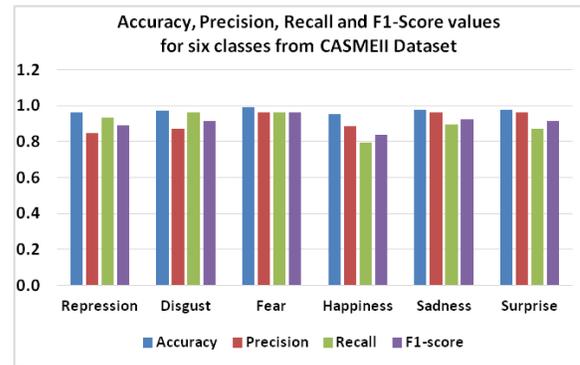
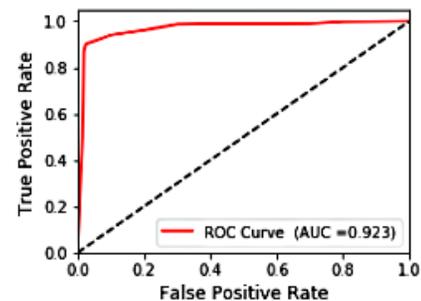


Figure 11: Accuracy, Precision, Recall and F1-Score values for six classes from CASMEII Dataset

The accuracy, precision, recall and F1-score for the classification of micro expression and non-micro expression of CASME II dataset are 0.9524, 0.983, 0.9611 and 0.97193 respectively.

ROC for Micro-expressions Vs Non-Micro-expressions in SAMM dataset



ROC for Micro-expressions Vs Non-Micro-expressions in CASME II dataset

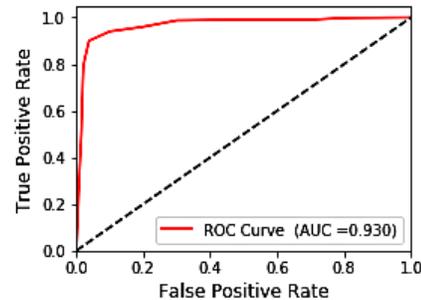


Figure 12: ROC plot between micro-expressions and non-micro-expressions of SAMM (top) and CASME II (bottom)

The area under the curve value of the SAMM and CASME II dataset for distinction between micro expression and non- micro expression is 0.923 and 0.930 respectively.

Table 3: Confusion Matrix for multi class classification of SAMM dataset [9]

		Predicted Class							
		Other	Anger	Contempt	Disgust	Fear	Happiness	Sadness	Surprise
Actual Class	Other	11	1	0	0	0	1	0	2
	Anger	0	13	0	1	1	0	0	0
	Contempt	1	0	11	0	0	1	0	2
	Disgust	0	0	1	12	0	1	0	1
	Fear	0	0	0	0	12	0	1	2
	Happiness	0	1	0	0	0	14	0	0
	Sadness	1	0	0	0	0	0	11	3
	Surprise	0	0	0	1	0	0	0	14

Table 4: Confusion Matrix as a binary classification for micro expression and non-micro expression from SAMM Dataset [9]

		Predicted Class	
		Micro expression	Non- Micro expression
Actual Class	Micro expression	103	2
	Non- Micro expression	4	11

Table 5: Confusion Matrix for multi class classification of CASME II dataset [9]

		Predicted Class							
		Other	Anger	Repression	Disgust	Fear	Happiness	Sadness	Surprise
Actual Class	Other	27	0	1	1	0	1	0	0
	Anger	0	0	0	0	0	0	0	0
	Repression	0	0	28	0	1	1	0	0
	Disgust	2	0	0	27	0	1	0	0
	Fear	3	0	0	1	25	1	0	0
	Happiness	1	0	4	2	0	23	1	0
	Sadness	1	0	1	1	0	0	25	1
	Surprise	0	0	0	0	0	3	0	27

Table 6: Confusion Matrix as a binary classification for micro expression and non-micro expression from CASME II Dataset [9]

		Predicted Class	
		Micro expression	Non- Micro expression
Actual Class	Micro expression	173	7
	Non- Micro expression	3	27

4. Conclusion

The seven different types of micro facial expressions are classified using machine and deep learning methods. Haar Cascade classifier is used to detect the face, convolutional neural network for the feature extraction and recurrent neural network for the classification of the micro expressions. The CASME II dataset provides the better result than that of the SAMM in terms of accuracy. The CASME II dataset has 87% accuracy and that of SAMM dataset has 84.76% accuracy. In SAMM dataset, among all the micro expression that is classified, happiness and contempt have more accuracy. Similarly, the micro expression fear is classified with greater accuracy in CASME II dataset. For each type of micro expressions accuracy, precision, recall and f1 score values are also calculated. In SAMM dataset, the micro expression anger has the maximum values of precision, recall and F1-score, that are obtained 96%, 98% and 97% respectively. Likewise, in CASME II dataset the micro expression fear has the maximum values of precision, recall and F1-score that are obtained 96%, 96% and 96% respectively. Further, the evaluation using ROC plot shows that the classifier is also able to differentiate between micro facial expression and non- micro facial expression with the accuracy of 92.2% and 93% for SAMM and CASME II dataset respectively.

References

- [1] "Convolutional Neural Networks (LeNet)," <http://deeplearning.net/tutorial/lenet.html>, (Accessed on August 8, 2020).
- [2] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "Samm: A spontaneous micro-facial movement dataset," *IEEE transactions on affective computing*, vol. 9, no. 1, pp. 116–129, 2016.
- [3] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.
- [4] A. Egorov, A. Shtanko, and P. Minin, "Selection of viola-jones algorithm parameters for specific conditions," *Bulletin of the Lebedev Physics Institute*, vol. 42, no. 8, pp. 244–248, 2015.
- [5] "Eilert-akademie fur emotionale intelligenz. 7 basic emotions." <https://eilert-akademie.de/>, (Accessed on August 8, 2020).
- [6] P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. WW Norton & Company, 2009.
- [7] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 382–386.
- [8] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *2011 international conference on computer vision*. IEEE, 2011, pp. 1449–1456.
- [9] A. R. Shrestha, "Micro expression analysis using recurrent neural network," Master's thesis, Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Kathmandu, Nepal, 2019.
- [10] Y. Song, L. P. Morency, and R. Davis, "Learning a sparse codebook of facial and body microexpressions for emotion recognition," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 237–244.
- [11] Y. Wang, J. See, R. C. W. Phan, and Y. H. Oh, "LBP with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition," in *Asian conference on computer vision*. Springer, 2014, pp. 525–537.
- [12] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, and X. Fu, "Casme ii: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS one*, vol. 9, no. 1, 2014.

