

# Empirical Evaluation of SVM for Facial Expression Recognition

Saeeda Saeed<sup>1</sup>, Junaid Baber<sup>1</sup>, Maheen Bakhtyar<sup>1</sup>, Ihsan Ullah<sup>1</sup>, Naveed Sheikh<sup>2</sup>, Imam Dad<sup>1</sup>, Anwar Ali Sanjrani<sup>1</sup>  
<sup>1</sup> Department of Computer Science & IT, University of Balochistan  
<sup>2</sup> Department of Mathematic, University of Balochistan

**Abstract**—Support Vector Machines (SVMs) have shown better generalization and classification capabilities in different applications of computer vision; SVM classifies underlying data by a hyperplane that can separate the two classes by maintaining the maximum margin between the support vectors of the respective classes. An empirical analysis of SVMs on the facial expression recognition task is reported with high intra and low inter class variations by conducting an extensive set of experiments on a large-scale Fer 2013 dataset. Three different kernel functions of SVM are used; linear kernel, quadratic kernel and cubic kernel, whereas, Histogram of Oriented Gradient (HoG) is used as a feature descriptor. Cubic Kernel achieves highest accuracy on Fer 2013 dataset using HoG.

**Keywords**—Facial Expression Recognition; Support Vector Machine (SVM); Histogram of Oriented Gradients (HoG)

## I. INTRODUCTION

Facial expression is one of the most important non-verbal means of communication, enabling human beings to exchange the social information. There is a vast range of applications of facial expression recognition (FER) such as human and man-machine interaction [1], smart healthcare [2], biometrics, medical diagnosis [3], surveillance [4] and mental state identification [5].

However, high intra and low inter-class variations make human FER more challenging task in the field of machine learning and computer vision. Based on emotions and social interactions, facial expressions are usually categorized into six basic emotions, namely, happiness, sadness, anger, fear, surprise, and disgust [6]. The expression *neutral* is also added as another facial expression and widely accepted by the researchers [7]. Figure 1 represents some sample images of these emotions/facial expressions.

In recent years, several interesting solutions have been proposed for FER, relying on the diversified set of techniques and strategies to better represent and classify facial expressions in various application domains [5], [8]. Most of the existing FER frameworks focus on algorithms to extract better features, such as Gabor filter and RBF network [7], CNN [9], hybrid CNN-SIFT aggregator [10], and SVM [11].

In this paper, empirical study of SVMs with different kernels and features on FER is reported. In experiments, the analysis of false positives for each expression are also discussed. The rest of the paper is organized as follows: section II discusses the related work. The detailed description of the methodology adopted for this empirical study is provided in the section III. The section IV reports the experiments, the dataset



Fig. 1. Sample images from the Fer 2013 dataset. Images in the same column show identical expressions. The labels of expressions from left to right are: Anger, Disgust, Fear, Happy, Sad, Surprise and Neutral.

used for the evaluations along with the detailed analysis of the obtained results. Finally, section V provides some concluding remarks.

## II. RELATED WORK

In recent years, the research community has shown great interest in FER. As a result, several interesting solutions have been proposed for FER. For instance, Usman et al. in [12] proposed a three step solution to FER. In the first step, a state-of-the-art Viola-Jones face detection method has been employed to detect faces in the images. Subsequently, HoG features are extracted from the identified regions of interest followed by autoencoder and PCA based features dimensional reduction techniques. Finally, SVMs are used for the classification purposes. A Similar technique has been adopted in [13], where, initially, faces are detected using a Viola-Jones face detector followed by a feature extraction phase where Local Binary Pattern (LBP) features are used for representation purposes.

On the other hand, Anurag De et al. in [14] modeled FER system using an eigenface based approach where hue-saturation values are used for face detection. By calculating Euclidean distance between the test image and mean of training dataset, the expressions are then classified. For FER, Lekdioui et al. in [15] rely on a novel facial decomposition technique. First, the regions of a face are extracted using facial landmarks by using the algorithm IntraFace. For feature extraction, different techniques, such as LBP, CSLBP, LTP and Dynamic LTP, are used. SVMs are then used for classification purposes. Suzan Anwar et al. in [16] proposed an Active Shape Model (ASM) tracker, which takes input from a webcam and tracks 116 facial landmarks. These tracked landmark points are used for extracting the feature of expression from face and Support Vector Machine (SVM) is used for classification.

## III. PROPOSED METHODOLOGY

Figure 2 provides a block diagram of the methodology adopted in this study. The method is mainly composed of two

components, namely (i) feature extraction and (ii) classification. For feature extraction, state-of-the-art feature descriptors, namely HoG, is used, whereas, three different kernels are used of SVM for the classification.

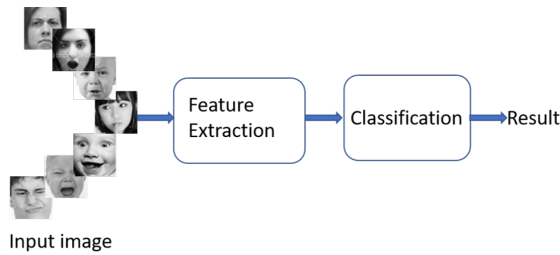


Fig. 2. The flow chart of the proposed model

The Histogram of Oriented Gradient (HoG) features are widely used for classification in different fields. The steps of HoG for extracting features are explained as follows:

- 1) The input image is divided into  $8 \times 8$  pixel grid.
- 2) Then,  $2 \times 2$  cells are created as block by 50% of overlapping, the total number of blocks are  $5 \times 5 = 25$  for  $48 \times 48$  images.
- 3) The direction of the gradient orientation range ( $0 - 180^\circ$ ) is divided into 9 bins. By the equation 1 gradient magnitude and orientation are calculated as follow

$$\begin{aligned}
 dx &= f(x+1, y) - f(x-1, y) \\
 dy &= f(x, y+1) - f(x, y-1) \quad (1) \\
 m(x, y) &= \sqrt{dx^2 + dy^2}
 \end{aligned}$$

In Equation 1,  $f(x, y)$  describes the pixel value at location  $(x, y)$  in the given face patch  $I$ , and  $m$  describes the magnitude.

- 4) The length of the HoG feature vector for given  $I$  is  $N \times C \times P$ , where  $N$  denotes the number of the blocks,  $C$  denotes the number of cells in each block, and  $P$  denotes the number of orientation bins. The orientation are quantized into 9 bins, where makes the HoG features of length  $25 \times 4 \times 9 = 900$  [17], [18]. An example of the HoG process for feature extraction is shown in Figure 3.

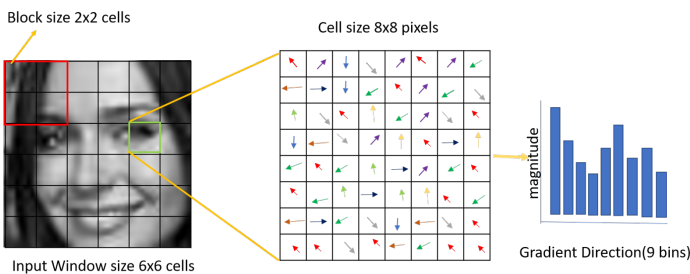


Fig. 3. HoG process for extracting features.

The SVM has shown outstanding performances in different application domains [19], [20]. The core goal of SVM is to

find the best hyperplane to separate binary classes at maximum distance with the minimum number of support vectors. Consider a set of  $l$  training examples  $(r_i, t_i), i = 1, \dots, l$ , where each example is  $n$ -dimensional,  $r_i \in \mathbb{R}^n$ , a class label  $t_i \in \{+1, -1\}$ . A function  $\phi$  is learned that maps given unknown instance  $r_j$  to  $t'_j, t'_j = \phi(r_j)$ .

Inherently, SVM is a binary classifier. To extend SVM for multi-class classification, either one-vs-one (OVO) or one-vs-all (OVA) approach is used. In case of OVA, for  $q$  different classes where  $q > 2$ ,  $q$  different classifiers are trained, for each class  $i$ , it assumes  $i$  as positive and rest all other as negative. The OVA often leads to imbalanced training dataset. In Fer 2013 dataset, it also create huge imbalance for few classes such as facial expression *Disgust*. If binary classifier for *Disgust* expression is trained using OVA, then there will be only 436 positive instances and 28273 negative instances; the positive instances will only be 1.5% of given training instances.

For many applications, OVO has shown better accuracy compared to OVA. The OVO approach trains  $q(q-1)/2$  binary classifiers. In case of FER, there will be 21 different binary classifiers. The ensemble/voting approach is used to decide the label of given instance  $r_j$ , the unknown instance  $r_j$  is given to 21 different binary classifiers, and the label with highest frequency is decided as the expression; OVO is used in this paper.

## IV. EXPERIMENTS AND RESULTS

### A. Dataset

The SVMs are evaluated on a large-scale benchmark dataset, namely, Fer 2013 dataset [21]. The imbalanced nature of the dataset makes it one of the most challenging datasets for FER. In total, the dataset covers 7 different facial expressions, namely *Angry, Disgust, Fear, Happy, Sad, Surprise* and *Neutral*. The dataset is divided into training and test sets. The test set is further divided into private and public testing. The training set contains 28709 instances against all expressions, whereas, the test set contains 7180 instances. The number of instances of each facial expression in test and training set are described in Table I.

It can be observed that some expressions have very few training instances. For instance, *Disgust* has only 436 training instances, which makes dataset imbalanced for the disgust expression. Since classification models are data greedy models, therefore, the recognition heavily depends on the number of training instances per class.

TABLE I. FER 2013 DATASET DESCRIPTION

Facial Expression	Training Set Instances	Test Set Instances
Angry	3995	958
Disgust	436	111
Fear	4097	1024
Happy	7215	1774
Sad	4830	1247
Surprise	3171	831
Neutral	4965	1233

### B. Experimental Results

Table II summarizes the recognition accuracies. It can be seen that linear SVM has limited performance compared to

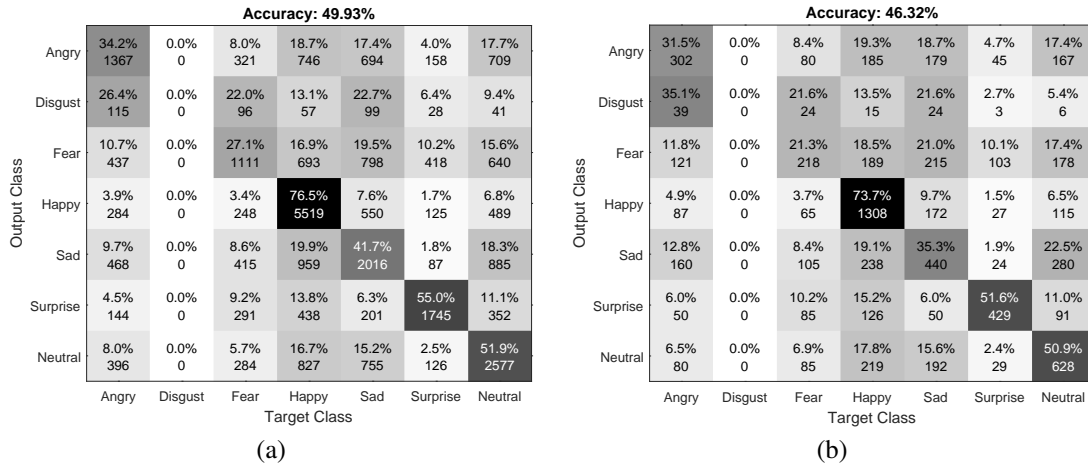


Fig. 4. FER on linear SVM using HoG features. a) Shows the results on the training set, and b) Shows the results on the test set.

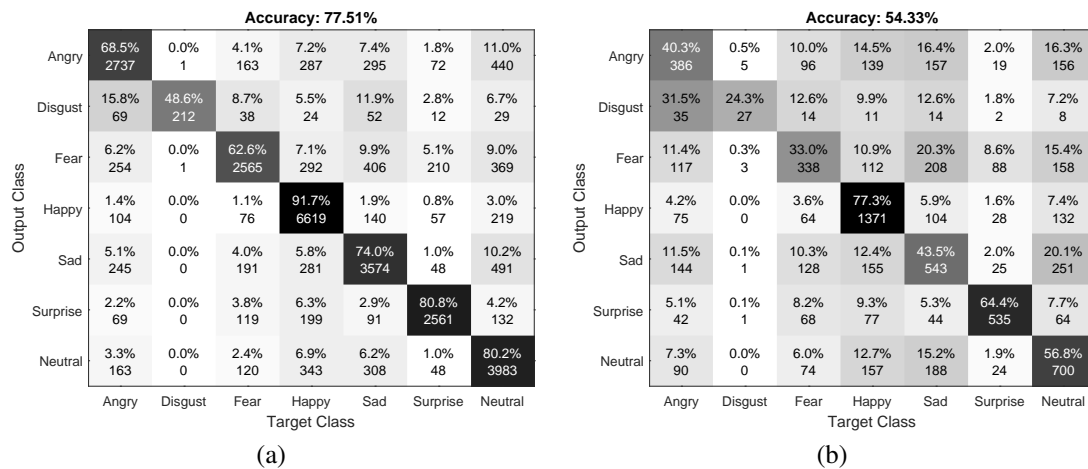


Fig. 5. FER on quadratic SVM using HoG features. a) Shows the results on the training set, and b) Shows the results on the test set.

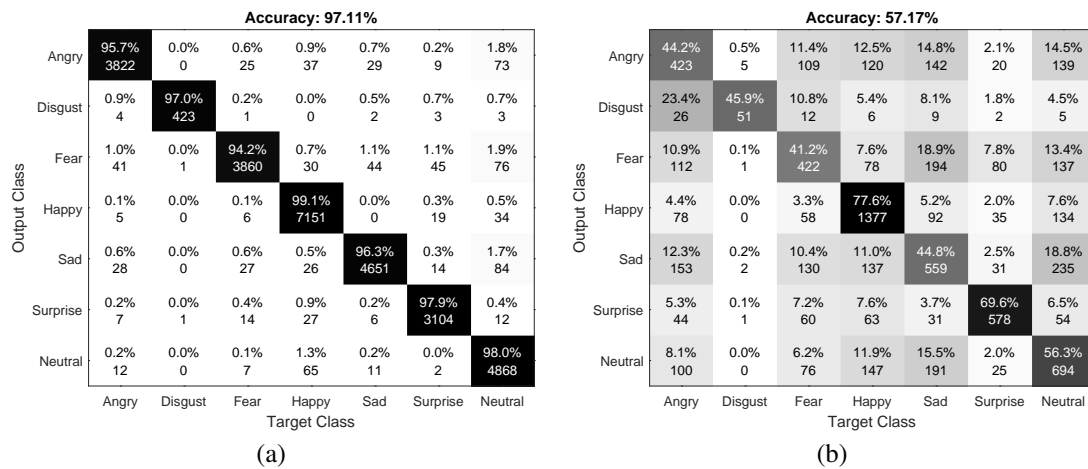


Fig. 6. FER on cubic SVM using HoG features. a) Shows the results on the training set, and b) Shows the results on the test set.

the other kernels. The quadratic SVM gives 77.51% accuracy on training set but 54.33% accuracy on the test set which gives an impression that feature space is quite challenging and HoG feature representation is limited for FER application. Even cubic SVM, which has better performance for non-linear feature space, has only 57.17% accuracy on test set.

Interestingly, cubic SVM over-fits on the training set and achieves 97.11% accuracy.

Moreover, in order to provide more detailed analysis of the results; Figure 4, Figure 5 and Figure 6 show the confusion matrices of linear, quadratic and cubic kernels, respectively, on HoG features.

TABLE II. COMPARISON BETWEEN THE RESULTS OF LINEAR SVM,  
QUADRATIC SVM AND CUBIC SVM ON HOG FEATURES

Feature Vector	Training Set Accuracy	Test Set Accuracy
Linear SVM	49.93%	46.32%
Quadratic SVM	77.51%	54.33%
Cubic SVM	97.11%	57.17%

It is interesting to observe that linear SVM fails to learn hyperplane for *Disgust* expression in training set and also in test set which indicates that *Disgust* expression is not linearly separable. It is also evident from Figure 4 that expression *Happy* achieves better recognition compared to the all other expressions. The happy expression also have similar recognition in quadratic and cubic SVMs test sets. On a leader-board, top three positions are the same for all SVM kernels whereas few expressions keep changing their positions. In case of disgust expression, it is on the last position (7th) on linear and quadratic SVMs, whereas, it improves the position on Cubic SVM that is 4th which indicates that cubic SVM gives better results in imbalance training as well. Almost all expressions improve their accuracies from linear to cubic SVMs except *Neutral* expression; for *Neutral* expression quadratic SVM gives the highest accuracy.

## V. CONCLUSION

In this paper, an empirical study is conducted on SVM for investigating the impact of different kernel functions on the performance of FER. There are seven different facial expressions; the expressions include *Angry*, *Disgust*, *Fear*, *Happy*, *Sad*, *Surprise* and *Neutral*. Cubic SVM achieves highest accuracy on the test set, whereas, it over-fits on the training set. The cubic SVM also learns better hyperplanes under imbalance training instances. In case of *Disgust* expression, the training instances are only 1.52%. The linear SVM fails to learn the hyperplane for *Disgust* expression, the accuracy of *Disgust* is 0%; it neither appeared as false positive for other expressions as well. The cubic SVM has better accuracies for all the expressions except *Neutral*; quadratic SVM gives highest accuracy for the *Neutral*. The linear SVM give poor performance for all the expressions except *Happy* expression; linear SVM gives competitive accuracy for *Happy* expression, that is 73.7%, whereas, the best is 77.6% using cubic SVM.

## REFERENCES

- [1] C. Shan and R. Braspenning, *Recognizing Facial Expressions Automatically from Video*. Springer US, 2010, pp. 479–509.
- [2] M. S. Hossain, *Patient State Recognition System for Healthcare Using Speech and Facial Expressions*. Plenum Press, 12 2016, pp. 251–287.
- [3] L. C. Jain, U. Halici, I. Hayashi, S. Lee, and S. Tsutsui, *Intelligent biometric techniques in fingerprint and face recognition*. CRC press, 1999, vol. 10.
- [4] C. Ilyas, K. Nasrollahi, T. Moeslund, M. Rehm, and M. Haque, “Facial expression recognition for traumatic brain injured patients,” *Imaging and Computer Graphics Theory and Applications*, pp. 259–275, 2018.
- [5] B. Fasel and J. Luetttin, “Automatic facial expression analysis: A survey,” *PATTERN RECOGNITION*, pp. 259–275, 1999.
- [6] F. Ekman, Paul and W. V., “Constants across cultures in the face and emotion,” *Journal of Personality and Social Psychology*, pp. 124–129, 1971.
- [7] J. K. Shruti Karkra, “Compound facial expression recognition through gabor filter and rbf network,” *International Journal of Computer Science and Mobile Computing*, pp. 576–583, 2016.
- [8] D. Astler, H. Chau, K. Hsu, A. Hua, A. Kannan, L. Lei, M. Nathanson, E. Paryavi, M. Rosen, H. Unno, C. Wang, K. Zaidi, X. Zhang, and C.-M. Tang, “Increased accessibility to nonverbal communication through facial and expression recognition technologies for blind/visually impaired subjects,” in *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2011, pp. 259–260.
- [9] Z. Yu and C. Zhang, “Image based static facial expression recognition with multiple deep network learning,” in *ICMI*, 2015.
- [10] M. Al-Shabi, W. P. Cheah, and T. Connie, “Facial expression recognition using a hybrid CNN-SIFT aggregator,” *CoRR*, vol. abs/1608.02833, 2016.
- [11] M. Georgescu, R. T. Ionescu, and M. Popescu, “Local learning with deep and handcrafted features for facial expression recognition,” *CoRR*, 2018.
- [12] M. U. Ilyas, S. Latif, and J. Qadir, “Using deep autoencoders for facial expression recognition,” *CoRR*, 2018.
- [13] N. Kauser and J. Sharma, “Facial expression recognition using lbp template of facial parts and multilayer neural network,” *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 445–449, 02 2017.
- [14] A. S. Anurag Dea and D. M. Palb, “A human facial expression recognition model based on eigen face approach,” *International Conference on Advanced Computing Technologies and Applications (ICACTA2015)*, 2015.
- [15] K. Lekdioui, Y. Ruichek, R. Messoussi, Y. Chaabi, and R. Touahni, “Facial expression recognition using face-regions,” in *Proceedings of the 3rd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2017, pp. 1–6.
- [16] M. M. Suzan Anwar, “Real time intention recognition,” *IEEE International Conference on Advances in Computer Applications*, 2015.
- [17] A. Alsubari and R. J. Ramteke, “Extraction of face and palmprint features based on lbp, hog and zernike moments,” *International Journal of Computer Applications*, vol. 172, pp. 31–34, 08 2017.
- [18] S. Kim and K. Cho, “Fast calculation of histogram of oriented gradient feature by removing redundancy in overlapping block,” *Journal of Information Science and Engineering*, vol. 30, pp. 1719–1731, 11 2014.
- [19] H. Byun and S.-W. Lee, “Applications of support vector machines for pattern recognition: A survey,” in *Pattern recognition with support vector machines*. Springer, 2002, pp. 213–236.
- [20] K. Ahmad, M. L. Mekhalif, N. Conci, F. Melgani, and F. D. Natale, “Ensemble of deep models for event recognition,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 2, p. 51, 2018.
- [21] D. Goodfellow, Ian J. and Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, “Challenges in representation learning: A report on three machine learning contests,” in *Neural Information Processing*, 2013, pp. 117–124.