# GCV-Based Regularized Extreme Learning Machine for Facial Expression Recognition

**Shraddha Naik and Ravi Prasad K. Jagannath**

**Abstract** Extreme learning machine (ELM) with a single-layer feed-forward network (SLFN) has acquired overwhelming attention. The structure of ELM has to be optimized through the incorporation of regularization to gain convenient results, and the Tikhonov regularization is frequently used. Regularization benefits in improving the generalized performance than traditional ELM. The estimation of regularization parameter mainly follows heuristic approaches or some empirical analysis through prior experience. When such a choice is not possible, the generalized cross-validation (GCV) method is one of the most popular choices for obtaining optimal regularization parameter. In this work, a new method of facial expression recognition is introduced where histogram of oriented gradients (HOG) feature extraction and GCV-based regularized ELM are applied. Experimental results on facial expression database JAFFE demonstrate promising performance which outperforms the other two classifiers, namely support vector machine (SVM) and k-nearest neighbor (KNN).

**Keywords** Facial expression recognition (FER) · Extreme learning machine (ELM) · Regularization · Generalized cross-validation (GCV) · Classification

## 1 Introduction

The inner mind is represented by variations in the gesture, gait, voice and slightest variation in eye and face. Important facial variations are due to face expressions which are categorized into six basic universal expressions which reflect intentional and emotional state of a person. The six basic expressions are happiness, sadness, surprise, fear, anger, and disgust according to Ekman [5]. Extensive research has

S. Naik (✉) · R. P. K. Jagannath
National Institute of Technology Goa, Ponda, Goa, India
e-mail: shraddha@nitgoa.ac.in

R. P. K. Jagannath
e-mail: k.j.raviprasad@nitgoa.ac.in

been carried out so far in the field of facial expression recognition (FER) in numerous domains [3, 6, 18].

In FER, face analysis is broadly carried out in two ways. One way is to describe facial expression measurements using action unit (AU). An AU is an observable component of facial muscle movement. Ekman and Friesen proposed Facial Action Coding System (FACS) in 1978 [4]. All expressions are broken down using different combinations of the AUs [17]. Using another way, the analysis of face was carried out directly based on the facial effect produced or emotions. In the literature, plenty of algorithms have been discussed to solve this problem. In this work, more focus is on the expression classification phase. The "extreme learning machine (ELM)" classifier is used for classification. ELM has a single hidden layer in its structure. Therefore, the whole system boils down to solving linear systems of equations which makes it easy to obtain global optimum value in cost function [10].

Regularized ELM has higher efficiency and a better generalization property. Different types of regularization techniques are possible. $\ell_2$-regularization also known as Tikhonov regularization or ridge regression is one of the commonly used regularization techniques [14]. The best part of $\ell_2$-regularization is that prediction performance increases compared to the traditional ELM. The choice of an appropriate regularization parameter is one of the most daunting tasks. Usually, this choice is accomplished by empirical or heuristic methods.

In this work, FER experiment was performed on JAFFE dataset. Face detection was carried out using Viola–Jones face detection algorithm [19]. "Histogram of oriented gradients (HOG)" technique was applied to obtain feature descriptors. Obtained features were classified using regularized ELM, where regularization parameter ($\lambda$) was calculated using "generalized cross-validation (GCV)" method.

The rest of this paper has been organized as follows: In Sect. 2, we elaborate discussion on the FER, formation of regularized ELM. We also discuss GCV in brief. Proposed methodology is explained briefly in the Sect. 3. Section 4 sheds light on experimental results and discussion. Section 5 concludes the discussion.

## 2  Methods

### 2.1  Facial Expression Recognition

"Facial expression recognition (FER)" system begins with the acquisition of face in image or video frames, followed by feature extraction and expression classification.

Face acquisition phase also allows to detect and locate faces in the complex background images. Many techniques are available for automatic face detection [19, 20]. Viola–Jones face detection algorithm is one of the robust and rapid visual detectors. Haar-like features are used for face detection. Features handled by detectors are computed very quickly using integral image representation. Algorithm trains a cascade

of classifiers with AdaBoost, where AdaBoosting is used to eliminate the redundant features [19].

Next phase of FER is facial feature extraction. It is commonly categorized into geometry-based method and appearance-based method. In the geometry-based method, facial configuration is captured and a set of fiducial points are used to categorize the face into different categories. In the appearance-based method, feature descriptors encode important visual information from the facial images [15].

In HOG feature descriptor, distribution of intensity gradient or edge directions is used to describe local object appearance and shape within an image. Practically, it is accomplished by dividing an image into small cells and gradient is calculated over every pixel of the cell. Combining all the histogram entries together forms the facial feature descriptor [2]. Expression classification is the last phase of FER. Several classifiers are available to classify feature descriptor of images like support vector machine (SVM), artificial neural network (ANN), linear discriminant analysis (LDA) [1, 6].

## 2.2 Extreme Learning Machine

"Extreme learning machine (ELM)" was devised by Huang and Chen [10]. Input weight matrix and biases are being randomly initialized. The time needed for the recursive parameter optimization is excluded. Therefore, computational time of ELM is less. Transcendent characteristics of ELM provide efficient results. The classic ELM model and its few variants are discussed briefly in [10]. ELM training basically consists of two stages. In stage one, random feature map is generated and the second stage solves linear equations to obtain hidden weight parameters. General architecture of ELM is shown in Fig. 1; it is visible that ELM is a feed-forward network with a single hidden layer but input neurons may not act like other neurons [9, 11, 12]. Predicted target matrix $\hat{T}$ is calculated as

$$\hat{T} = f_L(x) = \sum_{j=1}^{L} \beta_i H_i(x) = H\beta, \tag{1}$$

where $d$ is the total number of features, $L$ is the total number of hidden units, $c$ is the total number of classes, $\beta = [\beta_1, \beta_2, \ldots, \beta_L]^T$ is the output weight vector/matrix between the hidden layer and output layer, $T$ is the given training data target matrix, and $H(x) = [H_1(x), H_2(x), \ldots, H_L(x)]$ is ELM nonlinear feature mapping.

Sigmoid, hyperbolic tangent function, Gaussian function, multi-quadratic function, radial basis function, hard limit function, etc., are the commonly used activation functions [9, 11]. For a particular ELM structure, the activation function of hidden units may not be unique. In our experiments, the activation function used is "hard limit function." The output weight matrix $\beta$ can be obtained by minimizing the least square cost function which is given as

**Fig. 1** General architecture of ELM
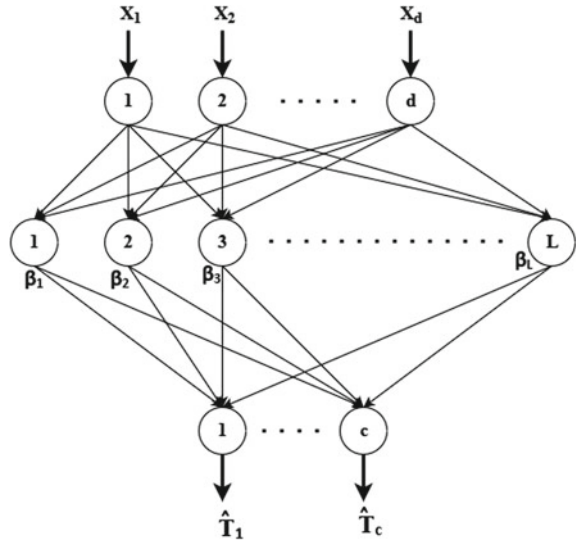


$$\min_{\beta \in R^{L \times c}} ||H\beta - T||^2. \tag{2}$$

The optimal solution can be obtained by $\beta^* = H^\dagger T$, where $H^\dagger$ is Moore–Penrose generalized inverse of matrix $H$ and $\beta^*$ is an optimal $\beta$ solution.

## 2.3  Regularized ELM

In case of overdetermined or underdetermined system of linear equations, optimal solutions may not be obtained using the above method. In such cases, regularized ELM enhances the stability and generalization performance of ELM [10, 12]. Selection of regularization parameter plays a vital role. Basically, regularization parameter is calculated empirically or using some heuristic methods based on the prior experience. There are many variants of Tikhonov-regularized ELM which make accuracy computation more accurate. In this paper, the analysis is done using $\ell_2$-regularization which is as follows:

$$\min_{\beta \in R^{L \times c}} ||H\beta - T||^2 + \lambda ||\beta||^2. \tag{3}$$

Now, after differentiating (3) we get an optimal $\beta^*$ as displayed

$$\beta^* = (\lambda I + H^T H)^{-1} H^T T. \tag{4}$$

The preeminent task is to predict regularization strength that is also known as regularization parameter $\lambda$ in the above (3). Basically, the value of $\lambda$ is chosen heuristically. Heuristic method follows trial and error technique, and therefore, the possibility of finding minima is low. There are many other methods to predict $\lambda$. It takes some range of values for $\lambda$, and after substituting these values in (3), the better value $\lambda$ will be estimated which gives better accuracy. This is called coordinate descent method [7, 12].

### 2.4 Generalized Cross-Validation

"Generalized cross-validation (GCV)" estimates $\lambda$ of $\ell_2$-regularization [8]. The fundamental thing used in GCV is a concept of singular value decomposition (SVD). For any given problem, the value of $\lambda > 0$ for which expected error is less than Gauss Markov estimate. Regularization parameter $\lambda$ minimizes loss function which is nontrivial quadratic depends on variance and $\beta$. However, in GCV there is no need of estimating the variance of errors. Here, $\hat{\lambda}$ which is the optimal value for $\lambda$ is predicted by using

$$G(\lambda) = \frac{\frac{1}{n}||(I - A(\lambda))T||^2}{\left[\frac{1}{n} \operatorname{tr} (I - A(\lambda))\right]^2},$$ (5)

where $A(\lambda) = H(H^T H + n\lambda I)^{-1} H^T$.

Using the above formula, we get value of $\hat{\lambda}$, and substituting this $\hat{\lambda}$ value into the $\ell_2$-regularization, $\beta$ is obtained. GCV method is computationally expensive because $A(\lambda)$ calculation requires finding of an inverse, multiple times.

## 3 Proposed Methodology

The basic FER system is divided into three subsystems as seen in Sect. 2. In the proposed work, neutral images from the dataset are subtracted from other expression images. Face was detected using Jones and Viola Face detection algorithm. Once face was extracted from the given images, HOG features were extracted from the subtracted images.

Furthermore, extracted features are classified using GCV-based regularized ELM. The problem of facial expression analysis is an ill-posed problem. The number of hidden layers used is more than the number of features which is nothing but an input vector. The FER system has become linear, and $\ell_2$-regularization is used to improve generalization. In $\ell_2$-regularization, the choice of regularization parameter plays an important role. Usually, the value of regularization parameters $\lambda$ is chosen randomly

or with some heuristic method. In the proposed method, GCV-based regularization of ELM is used to improvise the performance as shown in Sect. 4.

## 4 Experimental Results and Discussion

In this work, the effectiveness of GCV-based regularized ELM classifier on FER is demonstrated. In the experiments, Japanese Female Facial Expression (JAFFE) [13] database was used which contains a total of 213 images with $256 \times 256$ pixel. Ten Japanese female models posed for six basic expressions (anger, fear, disgust, happy, sad, and surprise), and one was neutral. From training dataset images, neutral images are subtracted. HOG features are extracted from obtained images. Regularized ELM is used to classify. The above process has been explained pictorially in Fig. 2.

GCV-based regularized ELM improves the generalization property. And due to this, recognition accuracy gets enhanced. The selection of hidden number of units (L) is done by choosing different values and based on that accuracies are calculated. The plot of ELM accuracy with the varying hidden number of units has been shown in Fig. 3. Finally, results are compared with non-regularized ELM. The confusion matrices of traditional ELM, GCV-based regularized ELM, SVM and KNN are demonstrated in Table 1. This demonstrates that GCV-based regularized ELM gives better results than other compared classifiers.
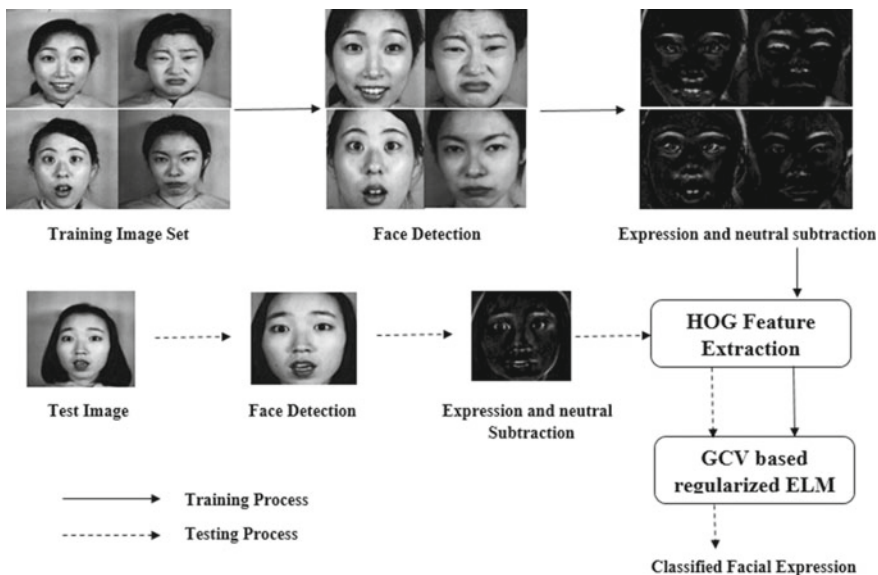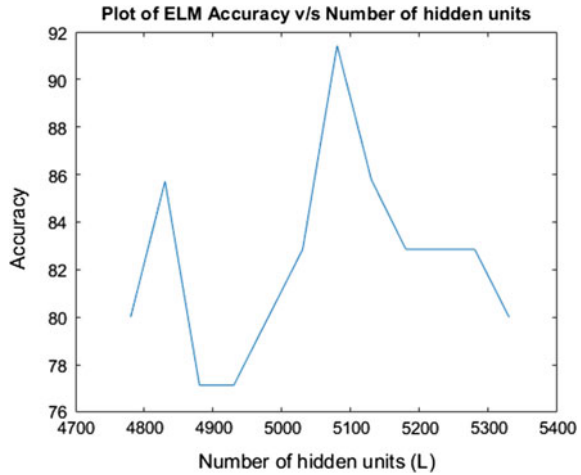


**Fig. 2** Pictorial diagram of experiments using GCV-based regularized ELM

**Fig. 3** The plot depicts ELM testing accuracy (%) versus number of hidden units (L) for underdetermined FER



Experiments are executed using different classifiers like classic ELM, GCV-based regularized ELM, SVM, and KNN. Various performance measures are available to measure the performance of the classification task [16]. The important measures used here are accuracy, precision, recall, and G-measure to demonstrate how the performance of different classifiers varies, and this is given in Table 2.

The solution for FER system can be devised in multiple ways. The main motive of this work is to demonstrate that regularized ELM can be applied to ill-posed problems like FER. ELM can be a good choice as it is easy to design, has a property of online learning, has strong tolerance to input noise, and has good generalization capability. Performance results can be obtained if we obtain correct value of $\lambda$. The correct value of $\lambda$ can be obtained using GCV method. The proposed architecture for FER is one of the simplest and fastest methods. Better techniques can be devised to obtain the regularization parameter which is the strength of regularization. Existing work can be extended on different types of ill-posed problems. This regularization methodology can also be tested on other types of classifiers. Further improvement can also be obtained by pruning the existing architecture. The accuracy of FER system can be further improved by using different better feature extraction techniques, and also by using different feature selection methods. Our main focus in this work is applying automatic computation of regularization parameter for ELM.

## 5 Conclusion

"Extreme learning machine (ELM)" is a prominent feed-forward neural network with three layers. Regularization improves the generalization capability and strong tolerance to input noise which is suitable for ill-posed problems like "facial expression recognition (FER)." The selection of regularization parameter plays a vital role, and

**Table 1** Confusion matrix for facial expression recognition using HOG for traditional ELM, GCV-based regularized ELM, SVM, and KNN

| | Anger | Fear | Disgust | Happy | Sad | Surprise |
|---|---|---|---|---|---|---|
| **Traditional ELM** | | | | | | |
| Anger | **100** | 0 | 0 | 0 | 0 | 0 |
| Fear | 4 | **96** | 0 | 0 | 0 | 0 |
| Disgust | 0 | 36.67 | **63.33** | 0 | 0 | 0 |
| Happy | 0 | 0 | 0 | **100** | 0 | 0 |
| Sad | 0 | 0 | 33.33 | 16.67 | **50** | 0 |
| Surprise | 0 | 0 | 0 | 0 | 0 | **100** |
| Accuracy | 84.89 | | | | | |
| **SVM** | | | | | | |
| Anger | **81.48** | 7.40 | 3.70 | 0 | 3.70 | 0 |
| Fear | 0 | **70.83** | 4.16 | 0 | 8.33 | 16.66 |
| Disgust | 0 | 5 | **55** | 0 | 15 | 25 |
| Happy | 0 | 0 | 0 | **86.66** | 0 | 13.33 |
| Sad | 0 | 8 | 0 | 4 | **52** | 36 |
| Surprise | 0 | 0 | 0 | 0 | 0 | **100** |
| Accuracy | 74.79 | | | | | |
| **GCV-based regularized ELM** | | | | | | |
| Anger | **100** | 0 | 0 | 0 | 0 | 0 |
| Fear | 4 | **96** | 0 | 0 | 0 | 0 |
| Disgust | 3.33 | 6.67 | **90** | 0 | 0 | 0 |
| Happy | 0 | 0 | 0 | **100** | 0 | 0 |
| Sad | 0 | 0 | 26.67 | 16.67 | **56.67** | 0 |
| Surprise | 0 | 0 | 0 | 0 | 0 | **100** |
| Accuracy | 90.44 | | | | | |
| **KNN** | | | | | | |
| Anger | **77.27** | 13.63 | 0 | 0 | 4.54 | 4.54 |
| Fear | 17.64 | **52.94** | 17.64 | 0 | 11.76 | 0 |
| Disgust | 0 | 14.81 | **81.48** | 0 | 3.76 | 0 |
| Happy | 15.38 | 0 | 0 | **80.76** | 0 | 3.86 |
| Sad | 6.66 | 3.33 | 10 | 6.66 | **63.33** | 10 |
| Surprise | 0 | 0 | 0 | 0 | 0 | **100** |
| Accuracy | 75.96 | | | | | |

**Table 2** The performance of different classifiers are listed using various measures

| Classifiers | Accuracy | Precision | Recall | G-measure |
| --- | --- | --- | --- | --- |
| Traditional ELM | 84.89 | 47.57 | 100 | 68.97 |
| GCV-based regularized ELM | **90.44** | **60.68** | **100** | **77.90** |
| SVM | 74.79 | 80.70 | 45.80 | 60.81 |
| KNN | 75.96 | 50 | 72.17 | 60.07 |

it is achieved by "generalized cross-validation (GCV)" method. GCV-based regularization parameter is one of the ways which is applied to avoid heuristics. Using this methodology in FER, recognition rate can be improved. This paper is aimed at proper classification of extracted HOG features in FER which is an ill-posed problem.

# References

1. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 2, pp. 568–573. IEEE (2005)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
3. Deshmukh, S., Patwardhan, M., Mahajan, A.: Survey on real-time facial expression recognition techniques. IET Biom. **5**(3), 155–163 (2016)
4. Ekman, P., Friesen, W.V.: Facial action coding system (1977)
5. Ekman, P., Oster, H.: Facial expressions of emotion. Annu. Rev. Psychol. **30**(1), 527–554 (1979)
6. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recognit. **36**(1), 259–275 (2003)
7. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Statist. Softw. **33**(1), 1 (2010)
8. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), 215–223 (1979)
9. Huang, G.B., Chen, L.: Convex incremental extreme learning machine. Neurocomputing **70**(16), 3056–3062 (2007)
10. Huang, G.B., Chen, L.: Enhanced random search based incremental extreme learning machine. Neurocomputing **71**(16), 3460–3468 (2008)
11. Huang, G., Huang, G.B., Song, S., You, K.: Trends in extreme learning machines: a review. Neural Netw. **61**, 32–48 (2015)
12. Huang, G.B., Zhou, H., Ding, X., Zhang, R.: Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **42**(2), 513–529 (2012)
13. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings, pp. 200–205. IEEE (1998)

14. MartíNez-MartíNez, J.M., Escandell-Montero, P., Soria-Olivas, E., MartíN-Guerrero, J.D., Magdalena-Benedito, R., GóMez-Sanchis, J.: Regularized extreme learning machine for regression problems. Neurocomputing **74**(17), 3716–3721 (2011)
15. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
16. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. **45**(4), 427–437 (2009)
17. Tian, Y.I., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 97–115 (2001)
18. Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Handbook of Face Recognition, pp. 487–519. Springer (2011)
19. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004)
20. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)