Journal of Current Science and Technology, September-December 2021 Vol. 11 No. 3, 402-410

ISSN 2630-0656 (Online)

Cite this article: Naveen, P., & Sivakumar, P. (2021, September). A deep convolution neural network for facial expression recognition. *Journal of Current Science and Technology, 11(3),* 402-410.
DOI: 10.14456/jcst.2021.40

Journal of Current Science and Technology
Journal homepage: https://jcst.rsu.ac.th

A deep convolution neural network for facial expression recognition

Palanichamy Naveen[1*] and Pothiraj Sivakumar[2]

[1]Department of ECE, Sri Eshwar College of Engineering, Coimbatore 641202, India
[2]Kalasalingam Academy of Research and Education, Krishnankoil -626126, India

*Corresponding author; E-mail: naveenamp88@gmail.com

Received 26 May 2021; Revised 2 July 2021; Accepted 10 July 2021;
Published online 28 September 2021

Abstract

Facial expressions play an important role in non-verbal communication processes to identify people and recognize emotions. This article proposes a deep convolution neural network (CNN) for recognizing seven basic emotional states (anger, disgust, fear, happiness, neutral, sadness, and surprise) and presents results comparing the accuracy of the proposed and existing methods. A support vector machine (SVM) was used in the convolutional layer to classify the input images. Then a rectified linear unit (ReLU) was introduced to train the inputs more easily and to improve performance. To reduce the tendency of overfitting, global average pooling was employed, and the rescaled hinge loss function was introduced to decrease noise in the classification system. Finally, to reduce the memory and time requirements, Nesterov-accelerated adaptive moment estimation was introduced. When implemented with different data sets, the proposed method demonstrated increased recognition accuracy compared to existing methods. The recognition accuracy of the proposed method improved significantly to 78.76% and 82.58%, respectively, for the Radbound Faces Database (RaFD) and the Karolinska Directed Emotional Faces (KDEF) database and BU-3DFE data sets while the overall average accuracy was 62.3% for randomly downloaded images from the Internet.

Keywords: *convolution neural network; facial expression; global average pooling; hinge loss function; ReLU; SVM.*

1. Introduction

With the demand for secure access and human-computer communication, efforts to improve and refine biometric technologies have drawn increasing interest. While fingerprint identification remains the predominant method in everyday security applications, other biometric technologies, such as iris scans and voice recognition, are being adopted for regular real-world applications. However, all these identification techniques require the cooperation of the identified person (Rahulamathavan, Phan, Chambers, & Parish, 2013; Watson, 2012).

Human facial identification, also termed facial recognition, is a technique that does not require the cooperation of an individual. They need not to involve the targeted individual facilitates law enforcement agencies' rapid identification of suspects in criminal cases. Facial recognition is not, however, restricted to law enforcement applications. Rather, it has been applied in various fields to restrict known problem makers' access to large public gatherings, such as sporting events. Despite its widespread use, however, proper facial recognition implementation requires additional efforts to address issues, including variations in poses, occlusion, illumination, and expressions (Naveen & Sivakumar, 2021a).

Facial expressions are important ways for people to express emotions and intentions in non-verbal communication (Bonaccio, O'Reilly, O'Sullivan, & Chiocchio, 2016). Facial expressions thus offer a rich source of information in day-to-day life. Research on automatic facial expression analysis has made significant achievements in past two decades (Pantic & Rothkrantz, 2000; Zeng, Pantic, Roisman, & Huang, 2009; Zhang et al., 2016). The enhanced performance of facial expression technology has also relied on developments in artificial intelligence, cognitive science, and neuroscience.

The aim of any facial expression analysis is to establish a system that can automatically classify various expressions. The basic facial expressions are anger, disgust, fear, happiness, neutral, sadness and surprise surprise (Mao, Rao, Yu, & Dong, 2017). The aim of our proposed system is to classify these seven-basic expressions.

Recently, visual recognition functions have exhibited significant improvements due to the development of deep learning. In fact, performance features, such as robustness, generalization, scalability, and a universal learning approach, make deep learning the most successful method. The convolution neural network (CNN) is a type of deep learning that automatically identifies the relevant facial features without any human supervision (Abbas, Ibrahim, & Jaffar, 2019).

CNN architecture has several layers. The most important layer is the convolution layer. In this paper, we present a deep CNN capable of automatic facial expression recognition. We employ a support vector machine (SVM) for the binary classification of input images. Then we train a two-layer ReLU network on linearly separable data. Recognizing that linearly separable data increase the power of random noise, we propose a simple algorithm to converge the global minima in finite number of non-zeros and to discharge local minima and saddle points effectively. Neural networks with rectified linear unit (ReLU) activation functions have demonstrated great success in various fields. To reduce the offline storage of CNN on the parameter level, we compressed the first fully connected layer with global average pooling. Compression reduces the spatial requirements, and hence, the use of global average pooling correspondence between feature maps and categories is more native to CNN. We avoid overfitting at this stage since no parameter is required to optimize global average pooling. Next, to tackle the imbalanced noisy classification problem, we introduce rescaled hinge loss function and its properties. Finally, to enhance both training speed and accuracy, we employ Nestorov-accelerated adaptive moment estimation. We utilise several standard data sets along with some randomly selected images downloaded from the Internet to compare the facial expression recognition accuracy of our proposed method and existing methods.

The rest of the paper is organized as follows. Section 2 outlines the existing research in the field. Section 3 describes the proposed methodology while Section 4 presents the experimental results. Finally, in Section 5, we discuss avenues for future research and conclude the paper.

2. Related works

Tang et al. (2019) proposed the kernel nearest-farthest subspace (KNFS) classifier for facial recognition (Tang, Li, Su, J. et al., 2019). By utilizing the kernel function, KNFS creates sample points that are linearly separable and thus maps linear inseparable sample points in low-dimensional kernel space to high-dimensional kernel space. The KNFS technique improves the dimensions of the sample points and considers the relationship between the sample points. The proposed approach achieved an average recognition rate of 99.58% for basic facial expressions, such as amaze, anger, neutral, and smiling.

He et al. (2016) proposed divide-and-conquer SVMs to enable big data training (He, Chen, Ji, Rho, & Kung, 2016). SVMs are also used as classifiers. They grouped the system into two broad classifications: local-descriptor classification and image classification. In the experimental results, the image classification accuracy reached 90.5%. In addition, most of the pixels were assigned successfully with exact colors.

Dubey and Chakraborty (2021) proposed that the discriminative capability of deep image representation using the trained model can be enhanced by the average biased ReLU (AB-ReLU) in the last few layers (Dubey & Chakraborty, 2021). To introduce non-linearity, ReLU ignores some values. AB-ReLU enhances discriminatory capability in two ways. First, it rejects the inappropriate and positive information in ReLU, and second, it uses limited discriminative and discarded negative information. Employed with many databases, the proposed method demonstrated excellent performance in its experimental results (Li et al., 2020).

Lin, Chen, and Yan (2014) proposed using global average pooling to replace the conventional fully connected layers in CNN and create one feature map for each class of the classification task in the last multilayer perceptron convolution layer (Lin, Chen, & Yan, 2014; Calvo, & Lundqvist, 2008). Irrespective of the fully connected layers stacked atop the feature maps, their method takes the average value of each feature map and feeds the resulting vector directly into the hinge loss function.

Xu et al. (2017) proposed the rescaled hinge loss function (Xu, Cao, Hu, & Principe, 2017). This proposed method is monotonic, bounded, and non-convex loss. Utilizing the rescaled hinge loss function, which is based on the correntropy-induced loss (C-loss) function, the authors developed a new robust SVM. In their experimental results, the sparseness property of the proposed method exhibited state-of-art performance.

To enhance the performance of the deep learning system, Dozat proposed Nesterov momentum (Dozat, 2016). The proposed Nestorov-accelerated adaptive moment algorithm (Nadam) is better than adam.

Mao et al. (2017) proposed a pose-based hierarchal Bayesian (PBHB) model to address challenging multi-pose facial expressions (Mao et al., 2017). This method incorporates local appearance features and global geometric information to learn intermediate facial representations before recognizing expressions. The proposed method effectively identifies facial expressions by sharing a set of features with different poses and thus does not require additional training and parameter adjustment for each pose (Yang, 2002). In this study, we compared our proposed method's recognition rates with the results of the PBHB model (Dhall, Goecke, Lucey, & Gedeon, 2011).

Wang et al. (2018) devised an intelligent emotion recognition system to overcome deterioration in facial recognition performance due to facial image translation. It utilizes stationary wavelet entropy for feature extraction and a feed-forward single hidden layer as a network classifier to increase the method's robustness by preventing the classifier from stagnating at the local optimum points. Employing the powerful Jaya training algorithm, the composed network and training exhibited better overall performance results than state-of-the-art methods (Wang, Phillips, Dong, & Zhang, 2018).

Zhang, Y.-D. et al. (2016) proposed a novel facial emotion recognition system with biorthogonal wavelet entropy for feature detection and fuzzy multiclass SVM. The performance of the devised method showed better performance than the state-of-the-art methods (Zhang, Y. D. et al., 2016).

3. Proposed methodology

3.1 Convolution layer

Figure 1 depicts the proposed system. The most significant component in the CNN architecture is the convolution layer, which comprises a collection of convolution filters called kernels. The kernel method is used to map input patterns to a feature area with a dot product and to classify the patterns of the feature area using a well-understood algorithm in which all functions in the future space can be expressed by dot products. The aim is to efficiently calculate these internal products at the input space using the kernel function (Zhang, Zhang, Chao, & Tseng, 2018). To design a kernel classifier, accurate kernels must be chosen with the correct feature space (Francis & Raimond, 2021).

3.2 Support vector machines

This approach is used to minimize the norms of the kernel weight (Gu & Wu, 2009). The aim of SVMs is to map the input patterns to a particular feature space and to separate the transformed data. Here, (x_i, y_i), $i \leq 1 \leq l$ are the standard training examples, where $y_i \in \{-1,1\}$ is the label for the input pattern $x_i \in X$

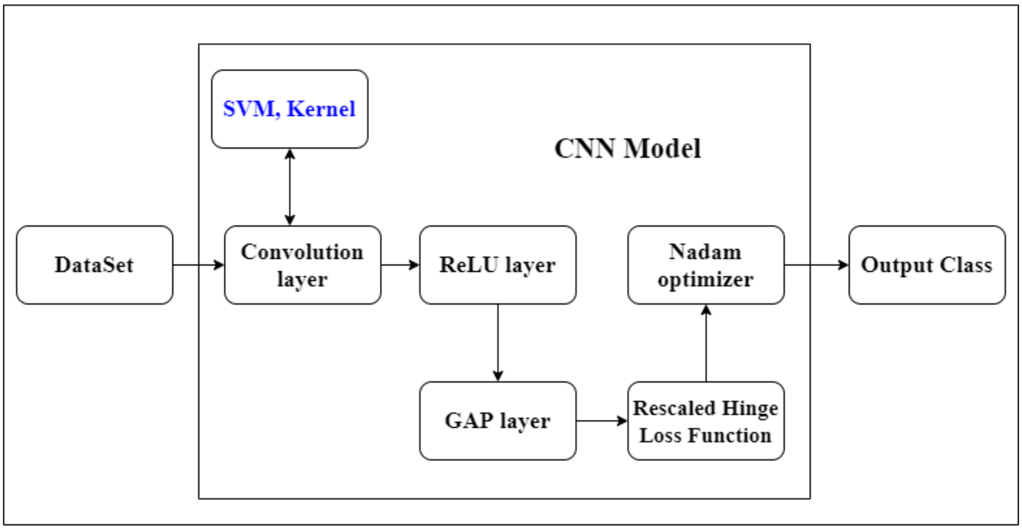

Figure 1 Proposed System Architecture

Patterns are characterized by the identification of the function of the form.

$$f(x)=\langle w,\phi(x)\rangle+b=\sum_{i=1}^{l} \alpha_i^{*} y_i k(x_i,x)+b \quad (1)$$

The coefficient α_i^* defining the kernel weight (Burges, 1998) is computed by solving the optimization problem and maximizing the dual optimization problem as follows:

$$E[\alpha]=\sum_{i=1}^{l} \alpha_i-\frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i,x_j) \quad (2)$$

3.3 Parametric linear units

A simple algorithm (Naveen & Sivakumar, 2021b) escapes from local minima and saddle points to effectively train any two-layer ReLU network and achieve global optimization.

The input-output relationship of a single hidden layer with scalar inputs, a single output and hidden neurons $k > 0$ is expressed as follows:

$$x\mapsto f(x):=\sum_{j=1}^{k} v_j \sigma\left(w_j^{T} x\right) \quad (3)$$

The equation above maps the input vector $x \in \Re^d$ to a scalar output by merging k nonlinear maps of nonlinear predictable interpretations of x through the ReLU activation $\sigma(z) := \max\{0, z\}$. To denote positive and negative classes, the output $f(\cdot)$ can take both positive and negative values and, therefore, $k \geq 2$ hidden layers due to the non-negativity of ReLU outputs. Here, v_j is the weight of the link from the j-th hidden neuron to the output and $w_j \in R^d$ heaps the weight of the links connecting the input x to the j-th hidden neuron (Zou, Cao, Zhou, & Gu, 2020).

3.4 Global average pooling

To reduce memory consumption, we must compress the weight matrix by replacing the fully connected layer using global average pooling (GAP). In the absence of the fully connected layer, however, it is difficult to conduct feature transfer learning on the trained set (Ni, Moulin, & Yan, 2015). To overcome this challenge, we cascaded the GAP layer before cascading the fully connected layer. It is, moreover, necessary to adjust the weight of the matrix.

$$\mathbf{W}'_{i,j}=\sum_{k=(j-1)fm^2+1}^{j\times fm^2} \mathbf{W}'_{i,k} \quad (4)$$

Where $\mathbf{W}'$ is the modified weight matrix, f_m is the size of the input feature map, i is the index of the output neurons and j is the index of the input feature.

3.5 Rescaled hinge loss function

The presence of SVMs results in a noisy classification system, and this is especially true for facial recognition tasks due to the limitless measures in their losses. The effective solution for noisy classification in facial recognition is the rescaled hinge loss function, which combines the hinge loss function and C-loss function (Xu et al., 2017). The rescaled hinge loss is represented as follows:

$$l_{rhinge}(y, f(x)) = \beta[1 - exp(-\eta l_{hinge}(y, f(x)))] \quad (5)$$

where $\eta = (2\sigma^2)^{-1} > 0$ is the rescaled parameter, $\beta = [1 - \exp(-\eta)]^{-1} > 0$ is the normalizing constant, and $l_{\text{rhinge}}(y, f(x))$ has the same architecture as the C-loss function. The C-loss is constructed as follows:

$$l_C(y,f(x))=\frac{1}{1-\exp\left(-\frac{1}{2\sigma^2}\right)}\left[1-\exp\left(-\frac{(1-yf(x))^2}{2\sigma^2}\right)\right] \quad (6)$$

Where, $\sigma > 0$ is the window width.

3.6 Nestorov-accelerated adaptive moment estimation

Nestorov-accelerated adaptive moment estimation can be used to optimize the CNN and improve the training speed and accuracy. The main advantage of this estimation is its lower computational power and greater efficiency in memory. It is represented as follows:

$$\theta_t \leftarrow \theta_{t-1} - \eta \frac{\mu \mathbf{m}_{t-1}}{\sqrt{\mathbf{n}_{t-1}} + \varepsilon} - \eta \frac{(1-\mu)\mathbf{g}_t}{\sqrt{v\mathbf{n}_{t-1} + (1-v)\mathbf{g}_t^2} + \varepsilon} \quad (7)$$

Where, $\mathbf{g}_t$ is current gradient, η is portion of the gradient, m and n are initialization bias correction terms. Therefore, Nadam is not dependent on $\mathbf{g}_t$, and we obtain $\overline{\mathbf{m}_t}$ and θ_t as follows:

$$\overline{\mathbf{m}_t} \leftarrow (1-\mu_t)\mathbf{g}_t + \mu_{t+1}\mathbf{m}_t \quad (8)$$

$$\theta_t \leftarrow \theta_{t-1} - \eta \frac{\overline{\mathbf{m}_t}}{\sqrt{\mathbf{v}_t} + \varepsilon} \quad (9)$$

Therefore, $\mathbf{m}_t$ is obtained from consequent timestep while $\mathbf{g}_t$ is obtained from current timestep.

4. Experimental results

We evaluated our proposed method and the existing approaches to subject-dependent experimenting in the following databases. The CMU Multi-PIE face database contains more than 750,000 images of 337 people recorded in four sessions over a five-month period. Subjects were filmed while showing a range of facial expressions under 15 vision points and 19 illumination conditions (Gross, Matthews, Cohna, Kanade, & Baker, 2010; Naveen & Sivakumar, 2021b). In the Static Facial Expression in the Wild (SFEW) data set, which contains 698 images of 95 subjects extracted from movies, the images are assigned to one of seven types of expressions: anger, disgust, fear, neutral, happiness, sadness, and surprise. The three-dimensional facial expression, or BU-3DFE, database, contains 100 subjects exhibiting 2,500 facial expressions (Yin, Wei, Sun, Wang, & Rosato, 2006). The Radbound Faces Database (RaFD) is a collection of images of 67 models (including Caucasian men and women, Caucasian children, boys and girls and Moroccan Dutch men) showing eight emotional expressions (Langner et al., 2010). The Karolinska Directed Emotional Faces (KDEF) is a collection of 4,900 images of human facial expressions. Finally, we randomly selected 465 images from the Internet, which included 364 facial expressions. Other than the random images, the data set remains the same as those used to test the PBHB (Mao et al., 2017), which facilitates our comparisons and analyses.

To estimate our proposed model, we compared its performance with state-of-art methods, such as VGG-Face (Parkhi, Vedaldi, & Zisserman, 2015), Multi-SVM (Hesse, Gehrig, Gao, & Ekenel, 2012), sLDA (Lade, 2015; Blei & McAuliffe, 2008), DHMM (Ojo & Adeniran, 2010), and PBHB. The VGG-Face method shows that deep CNN can achieve results comparable to the state-of-the-art methods without any added feature but with proper training. Table 1 presents the overall recognition accuracies along with the standard deviations for each method. Based on those values, the proposed method outperformed the existing methods. Table 2 shows the accuracies of the methods in recognizing the seven basic facial expressions along with the standard deviations. Except for two facial expressions—disgust and neutral, the proposed method was more accurate than the existing methods.

Table 1 Overall performance analysis of data set across proposed method and existing methods

Database	Proposed	PBHB	Multi-SVM	sLDA	DHMM
RaFD & KDEF	78.76 ± 1.96	74.96 ± 2.28	67.84 ± 2.31	59.06 ± 2.96	61.86 ± 2.32
BU-3DFE	82.58 ± 2.35	80.47 ± 1.79	74.07 ± 2.01	64.67 ± 2.66	62.72 ± 2.35

Table 2 Recognition accuracy in terms of basic facial expressions across proposed method and existing methods

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Multi-SVM	78.56 ± 0.86	63.38 ± 0.74	56.02 ± 1.05	82.14 ± 0.53	46.50 ± 1.08	68.26 ± 0.95	80.04 ± 0.98
sLDA	72.81 ± 0.62	68.75 ± 0.68	31.00 ± 0.78	70.60 ± 0.33	65.21 ± 1.12	51.01 ± 0.51	53.01 ± 0.89
DHMM	74.68 ± 0.77	70.27 ± 0.94	34.79 ± 1.71	78.10 ± 0.85	52.01 ± 1.13	63.92 ± 1.23	59.23 ± 1.99
PBHB	80.41 ± 0.56	82.70 ± 0.65	55.65 ± 0.71	88.25 ± 0.23	65.35 ± 0.91	71.41 ± 0.43	80.65 ± 0.88
Proposed	82.36 ± 0.84	80.36 ± 1.65	58.73 ± 0.23	89.36 ± 0.84	65.23 ± 0.84	73.26 ± 0.23	82.34 ± 0.33

Table 3 records the performance of the VGG-Face, PBHB, and our proposed method with the SFEW data set. Our method exhibited extremely consistent performance in mean recognition accuracy with different facial expressions and surpassed the VGG-face and PBHB methods. Finally, we compared the proposed method's recognition accuracy with the recognition accuracy of the PBHB, Multi-SVM, sLDA, and DHMM methods for the randomly selected images downloaded from the Internet. The results are available in Table 4. The average success rate of the proposed method exceeded those of the other methods, as illustrated in Figure 2.

Table 3 Performance on SFEW across VGG-face, PBHB and proposed method

Methods	Disgust	Happiness	Neutral	Surprise	Mean
VGG-Face	63.59	17.24	28.92	56.95	41.68
PBHB	43.59	43.01	41.67	50.6	44.72
Proposed	41.36	53.78	44.76	54.26	**48.54**

Table 4 Performance on random images across proposed method and existing methods

Parameters		**Proposed**		**PBHB**		**Multi-SVM**		**SLDA**		**DHMM**	
Expressions	**No. of Facial Images**	**Success %**	**No. of Facial Images**	**Success %**	**No. of Facial Images**	**Success %**	**No. of Facial Images**	**Success %**	**No. of Facial Images**	**Success %**	**No. of Facial Images**
Anger	**56**	50.00	28	51.79	29	44.64	25	37.50	21	33.93	19
Disgust	**55**	65.45	36	76.36	42	63.64	35	29.09	16	47.27	26
Fear	**43**	58.14	25	39.53	17	32.56	14	27.91	12	32.56	14
Happiness	**62**	67.74	42	62.90	39	54.84	34	41.94	26	33.87	21
Neutral	**46**	63.04	29	47.83	22	45.65	21	39.13	18	47.83	22
Sadness	**52**	53.85	28	50.00	26	42.31	22	38.46	20	32.69	17
Surprise	**50**	76.00	38	72.00	36	62.00	31	48.00	24	24.00	12
Average	364	62.03		57.20		49.38		37.43		36.02	

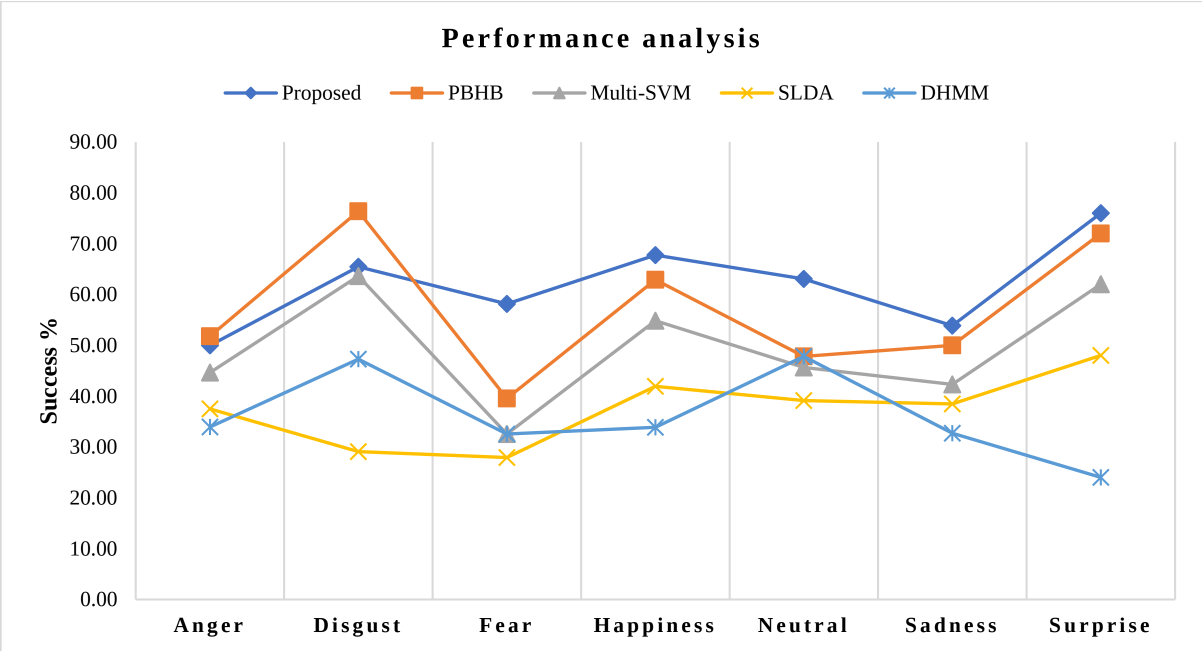

Figure 2 Performance comparison of proposed method and existing methods for basic facial expressions

5. Conclusion

This paper proposed a deep CNN method for facial expression recognition. Several poses were drawn from various data sets and randomly selected images downloaded from the Internet. The recognition accuracy of our proposed method improved significantly to 78.76% and 82.58%, respectively, for the RaFD, KDEF, and BU-3DFE data sets. The proposed model thus recognized the basic facial expressions, outperformed the existing method in five out of seven facial expressions, and achieved an overall average accuracy of 62.3% for randomly selected Internet images. Further improvement of the proposed method can be achieved for 3D data sets and real-time videos by enhancing the CNN architecture.

6. References

Abbas, Q., Ibrahim, M. E. A., & Jaffar, M. A. (2019). A comprehensive review of recent advances on deep vision systems. *Artificial Intelligence Review*, *52*(1), 39-76. DOI: https://doi.org/10.1007/s10462-018-9633-3

Blei, D., & McAuliffe, J. (2008). Supervised topic models. *Proceedings of Advanced Neural Information Processing System*, *20*, 121-128.

Bonaccio, S., O'Reilly, J., O'Sullivan, S. L., & Chiocchio, F. (2016). Nonverbal behavior and communication in the workplace: A review and an agenda for research. *Journal of Management*, *42*(5), 1044-1074. DOI: 10.1177/0149206315621146

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* *2*(2), 121-167. DOI: 10.1023/A:1009715923555

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior research methods*, *40*(1), 109-115.

Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. (2011). Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. *Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2106-2112. DOI:10.1109/ICCVW.2011.6130508

Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. *ICLR Workshop*, (1), 2013-2016. https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ

Dubey, S. R., & Chakraborty, S. (2021). Average biased ReLU based CNN descriptor for improved face retrieval. *Multimedia Tools and Applications, 80*, 23181-23206. DOI: https://doi.org/10.1007/s11042-020-10269-x

Francis, D. P., & Raimond, K. (2021). Major advancements in kernel function approximation. *Artificial Intelligence Review*, *54*(2), 843-876. DOI: https://doi.org/10.1007/s10462-020-09880-z

Gross, R., Matthews, I., Cohna, J., Kanade, T., & Baker, S. (2010). Multi-pie. *Proceedings of the International Conference on Automatic Face and Gesture Recognition, 28*(5), 807-813. DOI: 10.1016/j.imavis.2009.08.002

Gu, L., & Wu, H. Z. (2009). Applying a novel decision rule to the sphere-structured support vector machines algorithm. *Neural Computing and Applications*, *18*(3), 275-282. DOI: https://doi.org/10.1007/s00521-008-0179-1

He, X., Chen, B. W., Ji, W., Rho, S., & Kung, S. Y. (2016). Erratum to: Large-scale image colorization based on divide-and-conquer support vector machines. *The Journal of Supercomputing*, *72*(4), 1678-1678. DOI: https://doi.org//10.1007/s11227-015-1414-z

Hesse, N., Gehrig, T., Gao, H., & Ekenel, H. K. (2012, November). Multi-view facial expression recognition using local appearance features. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 3533-3536). IEEE.

Lade, P. (2015). *Probabilistic topic models for human emotion analysis*. Ph.D. dissertation, Arizona State University, USA. Retrieved form https://repository.asu.edu/attachments/146460/content/Lade_asu_0010E_14644.pdf

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377-1388. DOI: https://doi.org/10.1080/02699930903485076

Li, X., Zhang, X., Yang, H., Duan, W., Dai, W., & Yin, L. (2020). An EEG-based multi-modal emotion database with both posed and authentic facial actions for emotion analysis. *15th IEEE International Conference on Automatic Face and Gesture Recognition (FG).* DOI: 10.1109/FG47880.2020.00050

Lin, M., Chen, Q., & Yan, S. (2014). *Network In Network*. Retrieved form https://arxiv.org/abs/1312.4400

Mao, Q., Rao, Q., Yu, Y., & Dong, M. (2017). Hierarchical Bayesian theme models for multipose facial expression recognition. *IEEE Transactions on Multimedia*, *19*(4), 861-873. DOI: 10.1109/TMM.2016.2629282

Naveen, P., & Sivakumar, P. (2021a). Adaptive morphological and bilateral filtering with ensemble convolutional neural network for pose-invariant face recognition. *Journal of Ambient Intelligence and Humanized Computing, 12*, 10023–10033. DOI: https://doi.org/10.1007/s12652-020-02753-x

Naveen, P., & Sivakumar, P. (2021b). Pose and head orientation invariant face detection based on optimised aggregate channel feature. *Annals of the Romanian Society for Cell Biology, 25*(5), 4368-4390. https://www.annalsofrscb.ro/index.php/journal/article/view/5455

Ni, B., Moulin, P., & Yan, S. (2015). Pose adaptive motion feature pooling for human action analysis. *International Journal of Computer Vision*, *111*, 229-248. DOI: https://doi.org/10.1007/s11263-014-0742-4

Ojo, J. A., & Adeniran, S. A. (2010). One-sample face recognition using HMM model of fiducial areas. *International Journal of Image Processing, 5*(1): 58-68.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1424-1445.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *Xianghua Xie, Mark W. Jones, and Gary K. L. Tam, editors, Proceedings of the British Machine Vision Conference (BMVC)*, 41.1-41.12. BMVA Press. DOI: 10.5244/C.29.41

Rahulamathavan, Y., Phan, R. C.-W., Chambers, J. A., & Parish, D. J. (2013). Facial expression recognition in the encrypted domain based on local Fisher Discriminant Analysis. *IEEE Transactions on Affective Computing, 4*(1), 83-92. DOI: 10.1109/T-AFFC.2012.33

Wang, S.-H, Phillips, P., Dong, Z.-C., & Zhang, Y.-D. (2018). Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing, 272*(C), 668-676. DOI: https://doi.org/10.1016/j.neucom.2017.08.015

Watson, P. (2012). A multi-level security model for partitioning workflows over federated clouds. *Journal of Cloud Computing: Advances, Systems and Applications*, *1*(1), 1-15. DOI: https://doi.org/10.1186/2192-113X-1-15

Xu, G., Cao, Z., Hu, B.-G., Principe, J. C. (2017). Robust support vector machines based on the rescaled hinge loss function, *Pattern Recognition, 63*(C), 139-148, ISSN 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2016.09.045

Yang, M. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods. *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, 21-21 May 2002*, 215-220. DOI: 10.1109/AFGR.2002.4527207

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3D facial expression database for facial behavior research. *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06),* 211-216. DOI: 10.1109/FGR.2006.6

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39-58. DOI: 10.1109/TPAMI.2008.52

Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., & Yan, K. (2016). A deep neural network-driven feature learning method for multi-view facial expression recognition. *IEEE Transactions on Multimedia*, *18*(12), 2528-2536.

Zhang, Y.-D., Yang, Z.-J., Lu, H.-M., Zhou, X.-X., Phillips, P., Liu, Q.-M., & Wang, S. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access, 4,* 8375-8385. DOI: 10.1109/ACCESS.2016.2628407

Zhang, W., Zhang, Z., Chao, H.-C., & Tseng, F.-H. (2018). Kernel mixture model for probability density estimation in Bayesian classifiers. *Data Mining and Knowledge Discovery volume, 32*, 675-707. DOI: https://doi.org/10.1007/s10618-018-0550-5

Zou, D., Cao, Y., Zhou, D., & Gu, Q. (2020). Gradient descent optimizes over-parameterized deep ReLU networks. *Machine Learning*, *109*(3), 467-492. DOI: https://doi.org/10.1007/s10994-019-05839-6