

Emotion Detection Using Facial Expressions



Shivani Nandani, Rohin Nanavati, and Manish Khare

Abstract Emotion detection has been an area of interest for researchers for several decades. Since emotions can influence our decisions in life, they have been defined as an essential part of every human behaviour predictor. With the increasing need to understand human behaviour, facial expressions are vital in recognising emotions via non-verbal communication. For this study, two methods were used—CNN and PCA + SVM—with two data sets—the CK+ data set and the JAFFE data set. This study intends to compare some of the widely accepted machine learning algorithms for the given problem and a given data set.

Keywords Emotion detection · Facial expressions · CNN · PCA · SVM

1 Introduction

Emotions help us convey our intentions and form an essential part of communication that requires no language. They are universal. Our facial expressions often betray the emotions we are feeling. According to Mehrabian [12], 55% of a message pertaining to feeling and attitude is in the facial expression.

Initially studied by Ekman [3], facial features can be classified into six basic emotions—happiness, sadness, fear, disgust, surprise and anger. Therefore, emotion detection is possible via facial expressions, called facial emotion recognition (FER). FER has significant academic and commercial potential.

Our objective is to detect emotions using facial expressions to predict emotions. The scope of this paper remains limited to FER, whilst we continue to work on trans-

S. Nandani · R. Nanavati · M. Khare (✉)

Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India

e-mail: manish_khare@daiict.ac.in

S. Nandani

e-mail: 201801076@daiict.ac.in

R. Nanavati

e-mail: 201801108@daiict.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
P. K. Singh et al. (eds.), *Futuristic Trends in Networks and Computing Technologies*,
Lecture Notes in Electrical Engineering 936,
https://doi.org/10.1007/978-981-19-5037-7_45

627

lating the results into real-time applications. Significant research has been conducted in the field of FER. After reviewing the current state of the art methods, we chose to experiment using two different approaches. We attempted FER using a convolution neural network (CNN) as our choice of unsupervised learning method. For supervised learning, we chose to use a support vector machine (SVM) to solve the classification problem. The two data sets that we used were the JAFFE data set [10, 11] and the CK+ data set [9].

2 Related Work

Since emotions play an essential role in the choices that humans make in everyday life (such as what to eat or whom to talk to) as well as plan the future course of action (such as what career to pursue or whom to marry), the ability to correctly recognise emotions, or FER, is an area of interest for several researchers working in the domain of computer vision. The reason for this interest is the application of FER in several disciplines:

- In social interaction systems, facial expressions are an essential form of communication since they allow individuals to communicate their feelings in a non-verbal manner.
- Emotion processing is a crucial part of normal brain development in humans. Thus, tracking facial expressions from childhood to adolescence may prove helpful in understanding a child's growth [4].
- With the increase in AI-oriented technology and robotics, FER will be essential in these automated systems, especially when used to provide services that require a good analysis of the service user's emotional state. Examples of such scenarios would be security, entertainment, household and medicine [1].
- For distance learning programmes, which have become increasingly popular due to the COVID-19 pandemic, FER can identify students' understanding during the study process and, subsequently, change teaching strategies based on the data obtained [15].

2.1 Convolution Neural Network

Krizhevsky and Hinton gave a landmark publication that showed how a deep neural network works and its resemblance to the human visual cortex's functionality [7]. With the recent advances in deep learning, CNN models have been employed for several machine learning problems and have outperformed the existing models. In the case of FER, a typical model consists of three different stages—facial recognition, feature extraction and classification. CNN models are biologically inspired and thus

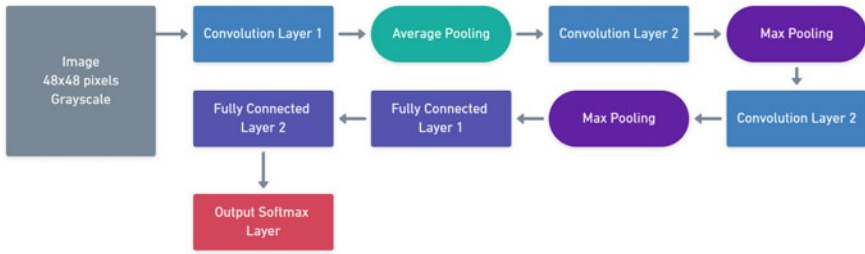


Fig. 1 The architecture proposed by Dachapally [2]

combine the feature extraction and classification steps. The input to a CNN model is a localised face image, and the output is classified labels. Several models have been proposed earlier—Liu et al. [8] proposed a model that uses three CNN subnets and achieved an accuracy of 65.03%; Dachapally [2] presented a model that used multiple convolution layers and achieved an accuracy of 86.38%. Shin et al. [14] proposed several other architectures that achieve $\approx 60\%$ accuracy. The architecture used by Dachapally [2] is shown in Fig. 1.

2.2 Support Vector Machines

SVM is one of the oldest machine learning models and is still relevant in the industry. This can be attributed to the simple interpretation of an SVM model—it finds a hyperplane in an N-dimensional space that distinctly classifies data points.

SVM has been employed for the FER problem in the past. In the model proposed by Abdulrahman and Eleyan [1], the PCA + SVM model has been used on JAFFE and MUFÉ data sets and produces an average accuracy of 87% and 77%, respectively.

3 Data Sets

The data sets used for this study include the extended CK+ data set [9] and the JAFFE data set [10, 11]. Both data sets were divided into an 80% training set and a 20% validation set.

The **Extended Cohn-Kanade Data set (CK+)** [9] is a public benchmark data set for emotion recognition. It comprises a total of 327 labelled grey-scale images which have similar backgrounds. A sample of the same is shown in Fig. 2.

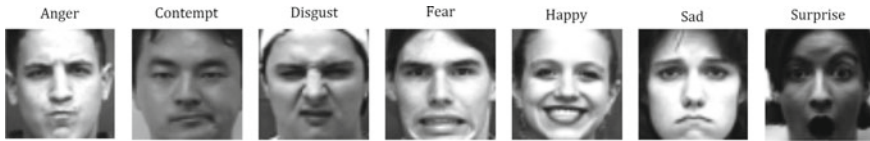


Fig. 2 Sample images—CK+ data set [9]



Fig. 3 Sample images—JAFFE data set [10, 11]

The **Japanese Female Facial Expression Data set (JAFFE)** [10, 11] has 213 labelled images. As shown in sample images in Fig. 3, all images are 8-bit grey scale and 256×256 pixels. 10 Japanese female expressers make up the data set with seven posed facial expressions (6 basic facial expressions + 1 neutral). This data set does not portray ‘contempt’.

4 Proposed Method

Two different methods were used—convolution neural networks and principal component analysis combined with support vector machine.

4.1 Convolution Neural Networks

Based on the success of the previous publications with CNN-based models, we decided to build our model from scratch with four convolution layers, one pooling layer and one fully connected layer. The architecture is represented in Fig. 4.

Here, the input layer takes grey-scale images in 48×48 pixel format. The first, second, third and fourth convolution layers use 6×6 kernel size, 2×2 kernel size, 2×2 kernel size and 2×2 kernel size, respectively. A max-pooling layer follows the third convolution layer. For the max-pooling layers, we have used a 2×2 window. Finally, we have the fully connected layer and the output layer, which uses the ‘softmax’ function.

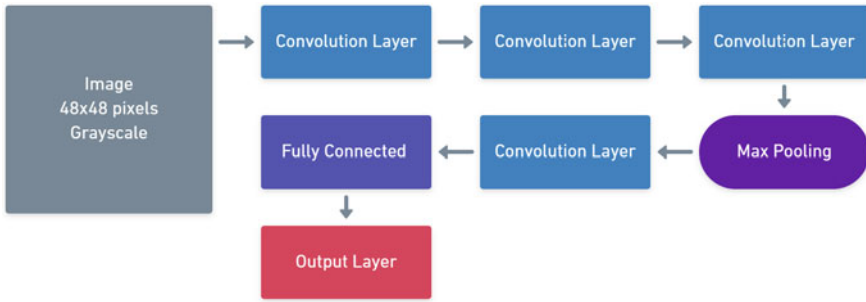


Fig. 4 Proposed CNN model



Fig. 5 Proposed PCA + SVM model

4.2 Support Vector Machine (SVM)

As seen in Fig. 5, we used principal component analysis (PCA) using randomised singular value decomposition for feature extraction. PCA reduces the image matrix into its principal vectors, sorted by the eigenvalues. We used the first 120–130 principal components for the SVM.

5 Experimental Results and Analysis

All the experiments were performed on the system with the hardware architecture shown in Table 1.

Table 1 Hardware architecture

Architecture	x86_64
Model name	Intel(R) Core(TM) i7-8750H CPU@2.20 GHz
Socket(s)	1
L1d cache	192 KiB
L2 cache	1.5 MiB
L3 cache	9 MiB

For both the methods mentioned in Sect. 4, we divided the data sets (CK+ [9] and JAFFE [10, 11]) into 80% training and 20% testing set. To evaluate the models, we have used accuracy as the performance metric.

5.1 CK+ Data Set

Both the models achieve an accuracy of above 80%. However, since the data set is skewed, some emotions are identified more accurately than others.

Convolution Neural Network For the CK+ data set, the CNN model gives 83.84% accuracy. The training was done for 80 epochs.

As seen in Fig. 6, the model is able to identify ‘happy’, ‘surprise’, ‘fear’ and ‘disgust’ with maximum precision, which is in accordance with human-level emotion detection. Some examples of the same are shown in Fig. 6. The model misclassifies ‘sadness’ the most. This misclassification results in an accuracy of 83.84%. The model loss and accuracy for each epoch can be seen in the graphs given in Fig. 7.

Support Vector Machine For the PCA + SVM model, we get an accuracy of 81.81%. We used the first 120 principal components.

As can be seen from Fig. 8, the significance of the principal components decreases rapidly. In fact, after the 14th component, the significance dropped below 1%. As we wanted to retain at least 95% of the image information, we needed to use more than 80 components, as can be observed from Fig. 8. With this constraint, we then chose 120 principal components as that gave the best accuracy.

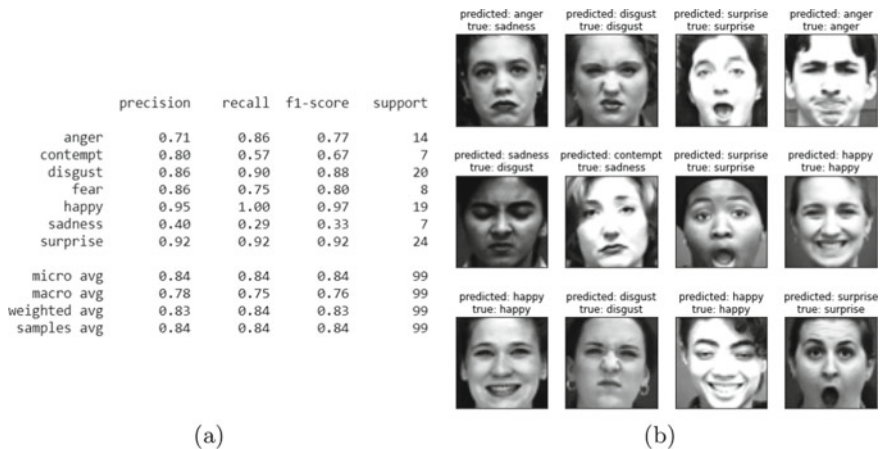


Fig. 6 Results for CNN model for CK+ data set [9] **a** classification report, **b** sample predictions

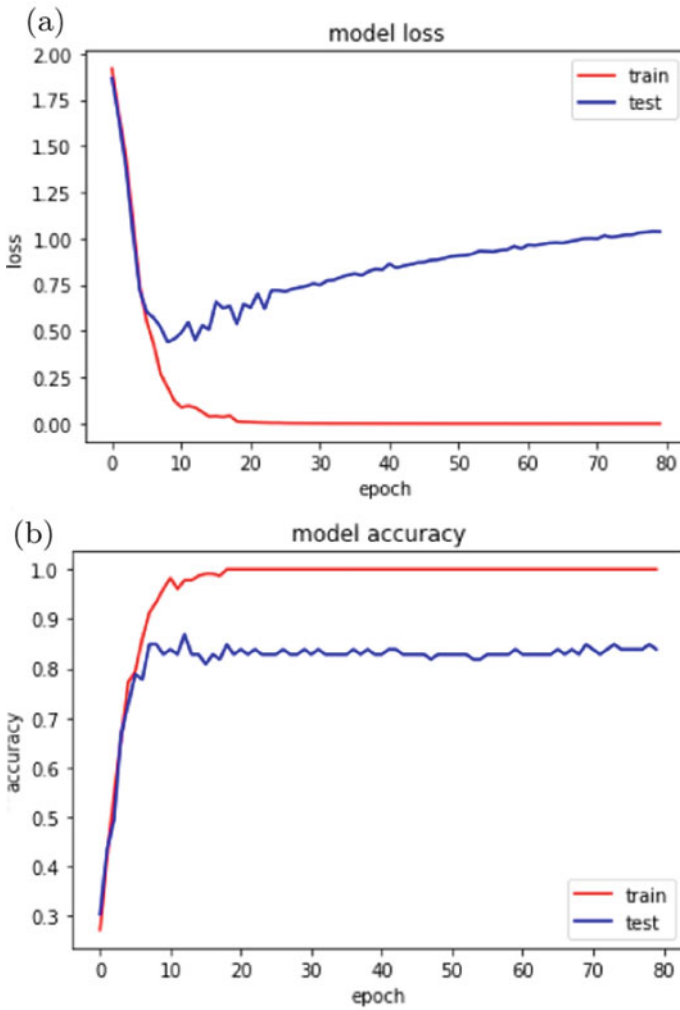


Fig. 7 a Model loss versus number of epochs, b model accuracy versus number of epochs

From the classification report given in Fig. 9, it is evident that the model, same as humans, can predict emotions such as ‘happy’, ‘fear’ and ‘surprise’ with greater accuracy than the other emotions. Some examples are shown in Fig. 9. ‘Sadness’ and ‘anger’ emotions are again misclassified the most, as was the case with the CNN model.

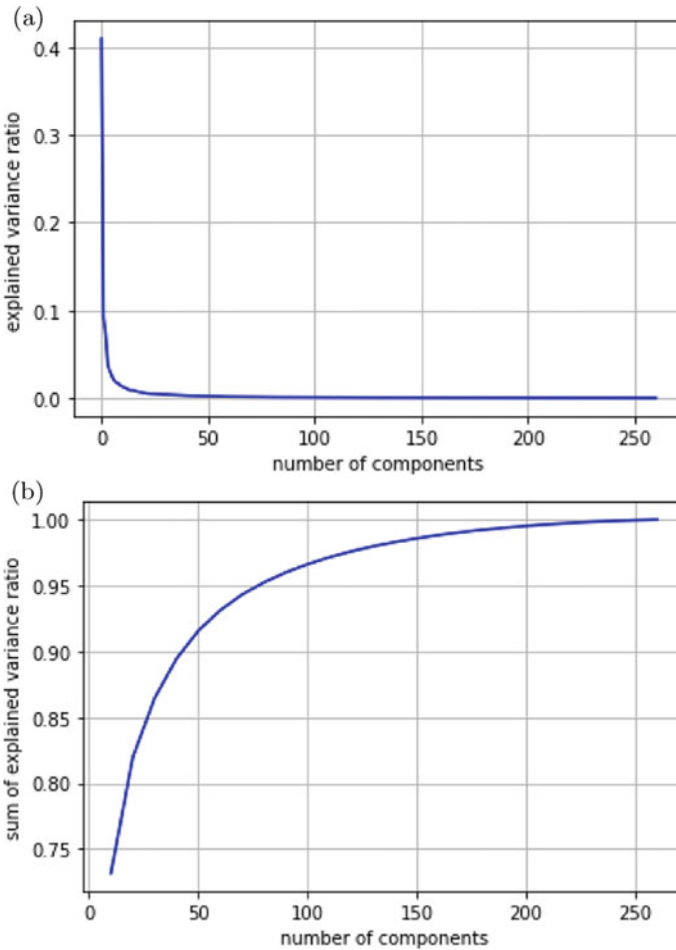


Fig. 8 **a** Explained variance ratio, **b** sum of explained variance ratio

5.2 JAFFE Data Set

Both the models give accuracy above 85%; however, the PCA + SVM model beats the CNN model on the JAFFE data set [10, 11].

Convolution Neural Network When CNN is performed on the JAFFE data set, we get an accuracy of 87.50%. The training was done for 80 epochs.

As seen in Fig. 10, the model can identify ‘happiness’, ‘anger’ and ‘fear’ with extremely high precision, whilst ‘disgust’ has the least precision value. Examples of the model’s predictions are shown in Fig. 10.

The model loss and accuracy for each epoch can be seen in the graphs shown in Fig. 11.

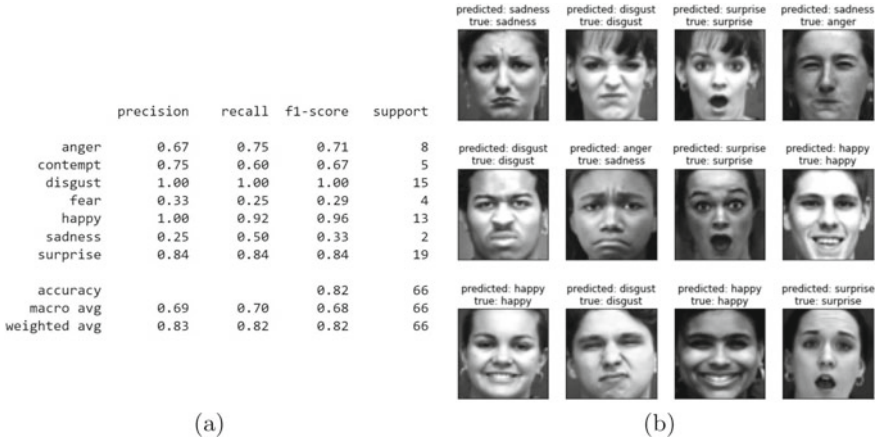


Fig. 9 Results for PCA + SVM model for CK+ data set [9] **a** classification report, **b** sample predictions

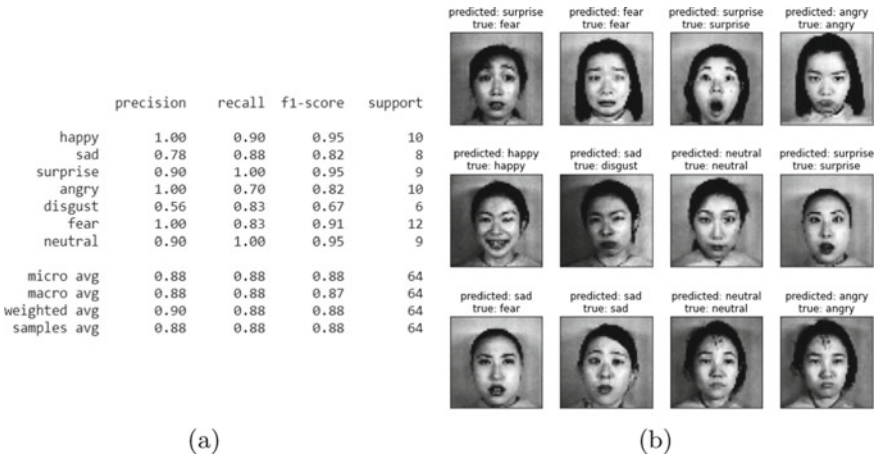


Fig. 10 Results for CNN model for JAFFE data set [10, 11] **a** Classification report **b** Sample predictions

Support Vector Machine For the PCA + SVM model, we get an accuracy of 95.35%. We used the first 130 principal components.

As can be seen from Fig. 12, the significance of the principal components decreases rapidly. In fact, after the 15th component, the significance dropped below 1%. We wanted to retain at least 95% of the image information; therefore, we needed to use more than 80 components, as can be observed from Fig. 12. With this constraint, we then chose 130 principal components as that gave the best accuracy.

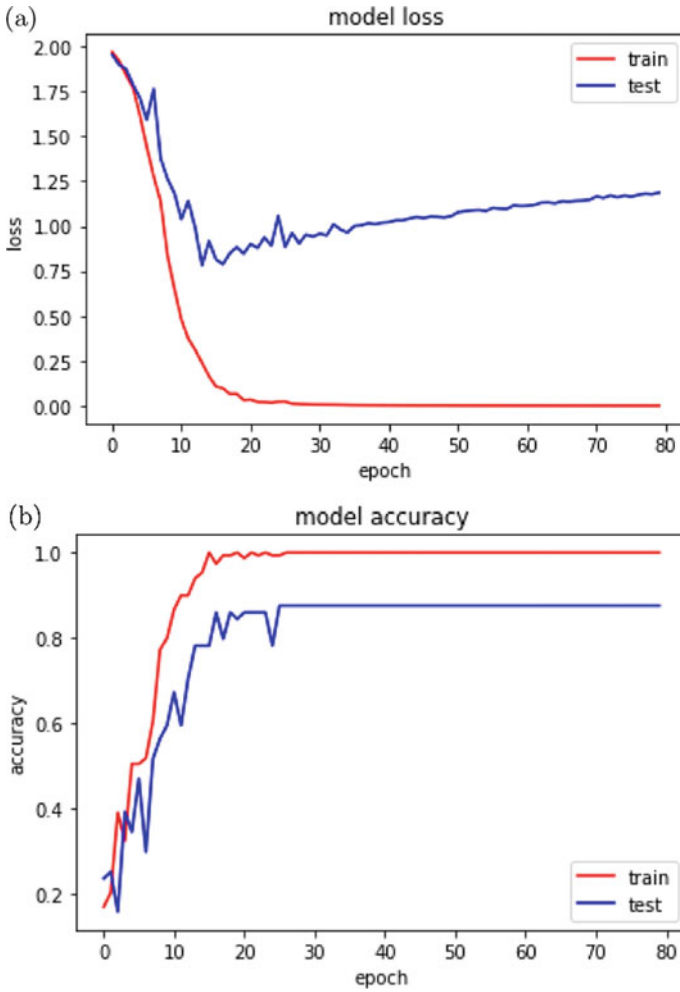


Fig. 11 a Model loss versus number of epochs, b model accuracy versus number of epochs

The model only struggled with ‘neutral’ and ‘fear’ emotions, as can be interpreted from Fig. 13. The high accuracy of the model is evidenced by the few examples shown in Fig. 13.

5.3 Against Other Comparable Architecture

Our proposed PCA + SVM model performs best on the JAFFE data set [10, 11] and beats several other proposed methods. As can be seen from Table 2, many methods

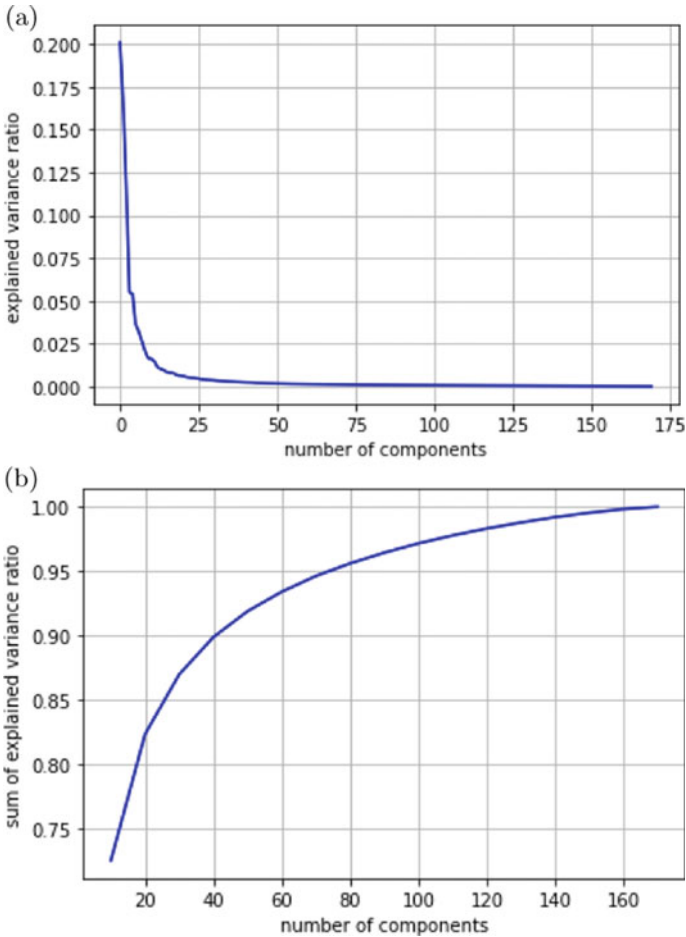


Fig. 12 **a** Explained variance ratio, **b** sum of explained variance ratio

have been proposed to extract features from the images, namely SNE [6], GPLVM [5], NMF [16] and LDA [13], after which SVM is used to classify the images. However, using PCA followed by SVM gives the best results with 95.35% accuracy. Abdulrahman and Eleyan [1] also used PCA followed by SVM; however, our model beats their performance. Our PCA + SVM model also beats the CNN model for the JAFFE data set [10, 11]. For the CK+ data set [9], the CNN model beats the PCA + SVM model marginally. Due to PCA + SVM model being much less complex than the CNN model, we believe that the PCA + SVM model will be more efficient in real-time computing.

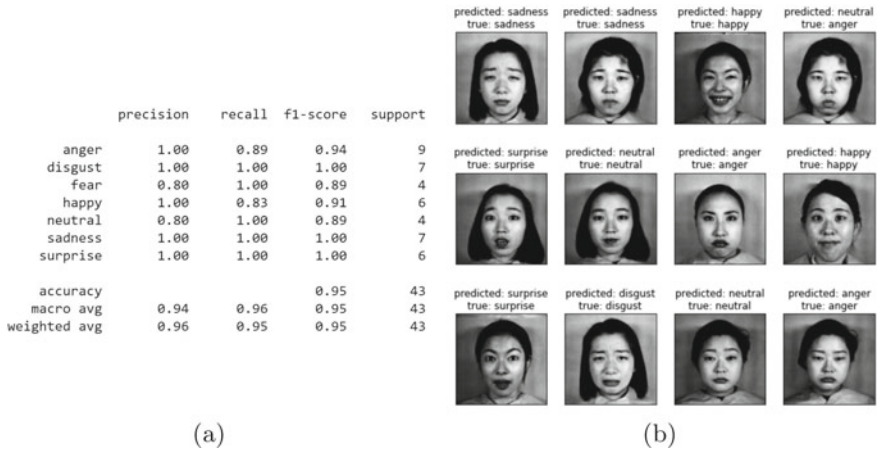


Fig. 13 Results for PCA + SVM model for JAFFE data set [10, 11] **a** classification report, **b** sample predictions

Table 2 Proposed models versus other comparable models

Method	Data set	Accuracy (%)
SNE + SVM [6]	JAFFE	65.7
GPLVM + SVM [5]	JAFFE	65.24
NMF + SVM [16]	JAFFE	66.19
PCA + SVM [1]	JAFFE	87
LDA + SVM [13]	JAFFE	93.6
CNN [2]	JAFFE	86.38
CNN [14]	JAFFE	50.61
CNN [14]	CK+	65.54
PCA + SVM (proposed)	JAFFE	95.35
CNN (proposed)	JAFFE	87.50
PCA + SVM (proposed)	CK+	81.81
CNN (proposed)	CK+	83.84

6 Conclusions

This paper tried to address the facial emotion recognition problem using two popular methods—CNN and PCA with SVM. The proposed models were then tested on two different data sets—the CK+ data set and the JAFFE data set. The accuracy of the models is higher than the human level and can easily detect everyday emotions. With the increase in computer vision tasks, FER models will play an important role in

evaluating user engagement for various commercial products. The source code of this work has been made publicly available at GitHub.¹

Future work for this paper shall include employing these models for real-time emotion detection, which can be used by e-learning platforms to understand students' engagement. Another extension of the task attempted in this paper is to use multiple CNN networks instead of one and compare the accuracy of the same to a single CNN model.

References

1. Abdulrahman M, Eleyan A (2015) Facial expression recognition using support vector machines. In: 2015 23rd signal processing and communications applications conference (SIU), pp 276–279. <https://doi.org/10.1109/SIU.2015.7129813>
2. Dachapally PR (2017) Facial emotion detection using convolutional neural networks and representational autoencoder units
3. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. *J Personal Soc Psychol* 17(2):124–129
4. Herba C, Phillips M (2004) Annotation: development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. *J Child Psychol Psychiatry Allied Disciplines* 45:1185–1198. <https://doi.org/10.1111/j.1469-7610.2004.00316.x>
5. Huang MW, Wang ZW, Ying ZL (2010) A novel method of facial expression recognition based on GPLVM plus SVM. In: IEEE 10th international conference on signal processing proceedings, pp 916–919. <https://doi.org/10.1109/ICOSP.2010.5655729>
6. Huang M, Wang Z, Ying Z (2011) Facial expression recognition using stochastic neighbor embedding and svms. In: Proceedings 2011 international conference on system science and engineering, pp 671–674. <https://doi.org/10.1109/ICSSE.2011.5961987>
7. Krizhevsky A, Hinton G (2009) Learning multiple layers of features from tiny images. Technical report 0, University of Toronto, Toronto, Ontario
8. Liu K, Zhang M, Pan Z (2016) Facial expression recognition with cnn ensemble. In: 2016 international conference on cyberworlds (CW), pp 163–166. <https://doi.org/10.1109/CW.2016.34>
9. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE computer society conference on computer vision and pattern recognition—workshops, pp 94–101. <https://doi.org/10.1109/CVPRW.2010.5543262>
10. Lyons M, Kamachi M, Gyoba J (1998) The Japanese female facial expression (JAFFE) dataset, Apr 1998. <https://doi.org/10.5281/zenodo.3451524>
11. Lyons MJ, Kamachi M, Gyoba J (2020) Coding facial expressions with Gabor wavelets (IVC special issue). *CoRR ArXiv:2009.05938*, <https://doi.org/10.48550/arXiv.2009.05938>
12. Mehrabian A (1968) Some referents and measures of nonverbal behavior. *Behavior Res Methods Instrument* 1(6):203–207. <https://doi.org/10.3758/BF03208096>
13. Shah J, Sharif M, Yasmin M, Fernandes S (2017) Facial expressions classification and false label reduction using LDA and threefold SVM. *Pattern Recogn Lett* 139. <https://doi.org/10.1016/j.patrec.2017.06.021>
14. Shin M, Kim M, Kwon DS (2016) Baseline CNN structure analysis for facial expression recognition. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 724–729. <https://doi.org/10.1109/ROMAN.2016.7745199>

¹ <https://github.com/shivani-nandani/emotion-detection.git>.

15. Yang D, Alsadoon A, Prasad P, Singh A, Elchouemi A (2018) An emotion recognition model based on facial recognition in virtual learning environment. Proc Comput Sci 125:2–10; The 6th international conference on smart computing and communications. <https://doi.org/10.1016/j.procs.2017.12.003>
16. Zilu Y, Guoyi Z (2009) Facial expression recognition based on NMF and SVM. In: 2009 international forum on information technology and applications, vol 3, pp 612–615. <https://doi.org/10.1109/IFITA.2009.279>