Soft Computing (2023) 27:8721–8737
https://doi.org/10.1007/s00500-023-08230-9

DATA ANALYTICS AND MACHINE LEARNING

Automated facial expression recognition using exemplar hybrid deep feature generation technique

Mehmet Baygin[1] · Ilknur Tuncer[2] · Sengul Dogan[3] · Prabal Datta Barua[4,5] · Turker Tuncer[3] · Kang Hao Cheong[6] · U. Rajendra Acharya[7]

Accepted: 8 April 2023 / Published online: 28 April 2023

Abstract

The perception and recognition of emotional expressions provide essential information about individuals' social behavior. Therefore, decoding emotional expressions is very important. Facial expression recognition (FER) is one of the most frequently studied topics. An accurate FER model has four prime phases. (i) Facial areas are segmented from the face images. (ii) An exemplar deep feature-based model is proposed. Two pretrained deep models (AlexNet and MobileNetV2) are utilized as feature generators. By merging both pretrained networks, a feature generation function is presented. (iii) The most valuable 1000 features are selected by neighborhood component analysis (NCA). (iv) These 1000 features are selected on a support vector machine (SVM). We have developed our model using five FER corpora: TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA. Our developed model is able to yield an accuracy of 97.01, 98.59, 96.54, 100, and 100%, using TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA, respectively. The results obtained in this study showed that the proposed exemplar deep feature extraction approach has obtained high success rates in the automatic FER method using various databases.

Keywords Facial expression recognition · Exemplar deep feature · Neighbor component analysis · Emotion detection

✉ Sengul Dogan
sdogan@firat.edu.tr

Mehmet Baygin
mehmetbaygin@ardahan.edu.tr

Ilknur Tuncer
ilknur.tuncer@icisleri.gov.tr

Prabal Datta Barua
prabal.barua@usq.edu.au

Turker Tuncer
turkertuncer@firat.edu.tr

Kang Hao Cheong
kanghao_cheong@sutd.edu.sg

U. Rajendra Acharya
Rajendra.Acharya@usq.edu.au

1 Department of Computer Engineering, Faculty of Engineering, Ardahan University, Ardahan, Turkey

2 Interior Ministry, Elazig, Turkey

3 Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig, Turkey

4 School of Business (Information System), University of Southern Queensland, Toowoomba, QLD 4350, Australia

5 Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

6 Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, Singapore 487372, Singapore

7 School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia

1 Introduction

1.1 Background

Facial expressions are one of the most effective tools human beings use to communicate. People show their constantly changing emotional states with these facial expressions (Othman et al. 2023; Geiger and Wilhelm 2023). In this way, they can even communicate independently of the languages (Nikolaus and Fourtassi 2023). Facial expression recognition (FER) is one of the most recently studied hot topics in the literature (Canal et al. 2022). FER systems can use various signals such as electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG) as input (Celniak and Augustyniak 2022; Cha and Im 2022; Wang et al. 2022). However, the most effective input for FER systems is images taken through a camera (Shen et al. 2022). Nowadays, the increase in human–computer interaction has made automatic emotion recognition even more interesting (Febrian et al. 2023; Zhen et al. 2023). Automated emotion recognition is already used in many fields, such as virtual reality, robotics, interactive games, and smartphones, and these areas are growing day by day (Ahmed et al. 2023; Arul Vinayakam Rajasimman et al. 2023; Dzedzickis et al. 2020). Thus, providing correct, fast, and effective solutions in FER systems is vital. When the studies in the literature are examined, it is seen that six different emotions are categorized. These are anger, disgust, fear, happiness, sadness, and surprise (Liu et al. 2023; Wani and Hashmy 2023). These feelings are the basic emotions (Ghosh et al. 2023; Gil and Bigot 2023). In addition, it has been accepted that these feelings are universal expressions regardless of language, ethnic origin, and culture (Ekman 1973). In addition to these emotions, a neutral mood has also been added. FER approaches performed in the literature are generally based on machine vision and machine learning (Arul Vinayakam Rajasimman et al. 2023; Kavitha and RajivKannan 2023). These methods generally focus on some basic points in a face image, such as eyes, nose, and lips. A wide variety of optimization algorithms exist in the literature (Liu et al. 2019; Zhao et al. 2022a, b). These optimization algorithms are also used to identify these basic points (Zhou et al. 2019). They obtain the highest performance by extracting features from these important points. However, the complex structure of the datasets used for automatic emotion recognition makes it difficult to find the optimal solution (Foggia et al. 2023). In particular, the similarity of some emotional states reduces the success of classification in automatic FER methods (Gao et al. 2023). In this study, a new model has been developed for automatic FER. Five well-known public datasets were used in this exemplar deep feature extraction work. Our model basically consists of preprocessing, feature generation, feature selection, and classification stage.

1.2 Literature review

There are several automated FER approaches proposed in the literature. Some of these approaches use classical machine learning methods, while others have used deep learning systems. Some of these studies are summarized in Table 1.

1.3 Motivation and our method

Computer vision solves various problems, for instance, object detection and face recognition, but one of the most challenging problems is FER (Jupalli et al. 2023). To solve the FER problems, many computer vision-based models have been presented in the literature (Porcu et al. 2020). However, most of these studies use deep learning networks for feature extraction and classification. Particularly, these models have used CNNs, but these models cannot reach satisfactory results and are costly. Nowadays, patch-based models are trendy since they can extract features from the local areas. The patch-based models generally use fixed-size patches, and this situation increases the time burden of this model. In order to use the advantages of the patch-based models with a lower time complexity, we have used dynamic-size patch division.

Moreover, we have used a transfer learning hybrid deep feature extractor. There are many CNNs in the literature, and they have individual performances. To use their advantage together, we have presented a new deep feature extractor. Also, we have designed a deep feature engineering model to get high classification performances. In this aspect, deep learning and feature engineering have been used together to attain high classification capability. In the literature, FER models generally use one or a few small datasets to test their classification performances. In this research, we have used five datasets (both small and big datasets) for testing our model.

In this paper, a deep feature extraction-based exemplar FER method is proposed. We have presented a deep feature engineering model. Therefore, MobileNetV2 and AlexNet deep learning network architectures were used in the study. The fully connected layers of these deep learning networks are used for feature generation, and the Logits layer in MobileNetV2 and the fc6 layer in AlexNet are used for meaningful feature generation. In addition, the NCA feature selection algorithm was used in the proposed method, and the most weighted 1000 features obtained were classified using SVM. Five different datasets available in the literature were used to justify the robustness of the

Table 1 Summary of few recent studies conducted in the literature

References	Year	Method	Dataset	Evaluation criteria	Classification method	Performance
Farajzadeh and Hashemzadeh (2018)	2018	Exemplar-based LBP and HOG features	CK, CK+, JAFFE, TFEID, MMI	Tenfold cross-validation	SVM	CK = 97.14% JAFFE = 92.53% TFEID = 98.9% CK+ = 97.94% MMI = 83.12%
Liu et al. (2018)	2018	Multi-channel pose-aware CNN	BU-3DFE, KDEF,	Fivefold cross-validation	–	BU-3DFE = 91.22% KDEF = 86.9%
Eng et al. (2019)	2019	HOG features	JAFFE, KDEF	Holdout validation (70%/30%)	SVM	JAFFE = 76.19% KDEF = 80.95%
Kas et al. (2021)	2020	Orthogonal parallel-based directions generic quad map binary patterns	CK+, KDEF, JAFFE, Oulu-CASIA, RAFD	LOSO cross-validation	SVM	JAFFE = 77.62% KDEF = 90.2% CK+ = 96.48% Oulu-CASIA = 77.32% RaFD = 97.2%
Sun et al. (2020)	2020	Self-adaptive feature learning approach based on a priori knowledge	JAFFE, CK+, KDEF	LOSO cross-validation	Active Feature Learning	JAFFE = 78.64% CK+ = 92.52% KDEF = 81.22%
Porcu et al. (2020)	2020	Data augmentation based on CNN	KDEF, CK+, ExpW	–	Softmax	KDEF = 83.30% (for training) CK+ = 83.3% ExpW = 40.4% (for testing)
Chowdary et al. (2021)	2021	CNN (Inception V3)	CK+	Holdout validation (84%/16%)	Softmax	CK+ = 98.5%
Akhand et al. (2021)	2021	Deep CNN with DenseNet161	KDEF, JAFFE	Holdout validation (90%/10%) and tenfold CV	–	KDEF = 98.78% JAFFE = 100% (for holdout validation) KDEF = 96.51% JAFFE = 99.52% (for tenfold CV)
Tang et al. (2020)	2021	Frequency neural network	CK+, Oulu-CASIA, KDEF, FER2013	Tenfold cross-validation	Softmax	CK+ = 98.97% Oulu-CASIA = 86.67% KDEF = 90.07% FER2013 = 64.41%
Khattak et al. (2022)	2022	Custom-designed CNN	JAFFE	Holdout validation (90%/10%)	Softmax	JAFFE = 95.65%

*All Datasets: CK (Chowdary et al. 2021), CK+ (Akhand et al. 2021), MMI (Tang et al. 2020), RAFD (Khattak et al. 2022), BU-3DFE (Jupalli et al. 2023), FER2013 (Krizhevsky et al. 2012), ExpW (Deng et al. 2009), TFEID (Sandler et al. 2018), JAFFE (Goldberger et al. 2004), KDEF (Vapnik 1998), Oulu-CASIA (Robnik-Šikonja and Kononenko 2003)

proposed model. These datasets tested on the developed model showed the highest classification performances.

1.4 Novelties and contributions

Novelties of this paper are as follows:

- Patch-based feature extraction is a popular model and transformer. Herein, we have presented a new fused deep feature extractor.
- Using the proposed fused deep feature extractor, a new-generation exemplar deep feature engineering model

has been presented and our proposal has been tested on the five FER corpora.

Contributions of the deep feature generation-based exemplar FER model are as follows:

- In this study, feature generation phases of deep learning approaches were used. In addition, the images were divided into blocks by the exemplar method, and the features of both the main image and the exemplars were extracted. Thus, more details/features are extracted from the face image during the feature generation phase.
- Automated FER methods are one of the most complex topics studied in the literature for a long time. Therefore, machine learning or deep learning approaches are generally used in this field. This study aims to develop a simple, effective, and high-accuracy model. For this purpose, a deep feature generation-based exemplar model, NCA feature selection, and SVM classifier were used together.
- We have obtained an accuracy of 97.01, 98.59, 96.54, 100, and 100% with TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA datasets, respectively. In this aspect, we have proposed a highly accurate FER model and this model attained over 97% classification accuracy on the five FER datasets.

1.5 Organization

The organization of the remainder of this paper is given as follows. The details of our proposed exemplar fused deep feature engineering model are presented in Sect. 2, Sect. 3 presents the calculated classification performances of our proposal, the results and findings are discussed in Sect. 4, and Sect. 5 gives conclusions.

2 The proposed exemplar deep feature generation-based FER method

Nowadays, many variable-effective computer vision models and patch-based models are trendy since they used local areas of the images to get valuable features. In this section, we explain the details of the presented exemplar deep feature engineering model. The main objective of this model is to present a patch-based effective deep feature engineering model to get high classification performance on the used five FER image dataset corpora. The proposed method in this paper basically consists of five essential phases: preprocessing, exemplar-based deep feature generation, feature concatenation, NCA-based feature selection, and classification. The methods used in the proposed model are briefly summarized as follows:

1. Exemplar division: The main images are divided into 3×3 segments in this phase. This process aims to apply a detailed feature extraction process by dividing the image into patches. This way, features can be generated from all the details in the image.
2. AlexNet (Krizhevsky et al. 2012): AlexNet is one of the popular CNN architectures in the machine learning society. This architecture consists of 25 layers and contains more than 60 million parameters. It has three fully connected layers named fc6, fc7 and fc8. It is one of the champions of the ImageNet competition. In this research, we have used a pretrained model of this network and the used AlexNet was trained on the ImageNet1K (Deng et al. 2009) and the fc6 layer is used for feature extraction. Using this layer, 4096 features have been extracted.
3. MobileNetV2 (Sandler et al. 2018): MobileNetV2 uses bottleneck blocks and is one of the famous lightweight CNNs in the literature. Therefore, MobileNetV2 is a very popular CNN in the computer vision society. This CNN (MobileNetV2) contains 154 layers and 3.4 million parameters. Here, we have used a pretrained version of the MobileNetV2. To create a deep feature extractor, we have used the "*logits*" layer of this network. This network was trained on the ImageNet1K (this dataset contains about 1.3 million images with 1000 object categories). Therefore, 1000 features have been extracted by deploying this feature extractor.
4. NCA (Goldberger et al. 2004): NCA is a feature selection method and is a feature selection version of the kNN since it uses a distance-based fitness function. NCA generates weights and these generated weights defines classification effects of the features. In this paper, NCA is used to eliminate the features generated by AlexNet and MobileNetV2. This way, it aims to exclude irrelevant features from the feature vector.
5. SVM (Vapnik 1998): This method is a well-known shallow classification approach in the literature. SVM has variable kernels. In order to show high classification performances of the generated features, we have used SVM classifier. The main objective of this method is to classify the features selected by NCA to determine to which class the images belong.

A block diagram summarizing the method is shown in Fig. 1. In addition, the procedure that performs this process is given in Algorithm 1, and the details of the methods used in the developed model are presented in the following sections.

Algorithm 1. Procedure used for exemplar deep feature generation-based automatic FER method.

Algorithm 1. Procedure used for exemplar deep feature generation based automatic FER method.

Input: Face images

Output: Predicted values

0: Load image.

1: Segment facial areas of the image.

2: Resize the segmented image to 227 × 227 and 224 × 224 sized images.

3: Extract 4096 features from AlexNet/fc6 layer and 1000 features from MobileNetV2/Logits layer

4: Divide resized image into nine exemplars using 3 × 3 sized non-overlapping blocks.

5: Extract 4096 and 1000 features from each block, like step 3, by using AlexNet/fc6 layer and MobileNetV2/Logits layers

6: Concatenate features and obtain 50,960 features.

7: Apply NCA to concatenated 50,960 features, and 1000 most discriminative features are selected.

8: Classify reduced features and obtain predicted values.

2.1 Preprocessing

Using only facial area is the best way to get meaningful features for the FER problem; hence, an effective preprocessing method must be used to solve these problems. The preprocessing step is used to improve the performance of the proposed method and has a crucial role. At this stage, the segmentation process was applied to all datasets first. In this way, unnecessary areas in the image are cropped. Afterward, the resizing process was applied to the images with different dimensions to extract deep features. The raw image data have been resized to 224 $\times$ 224 and 227 $\times$ 227 sizes for MobileNetV2 and AlexNet deep network architectures, respectively. The procedure for the preprocessing step of the application is presented in Algorithm 2.

Algorithm 2. Preprocessing steps for datasets.

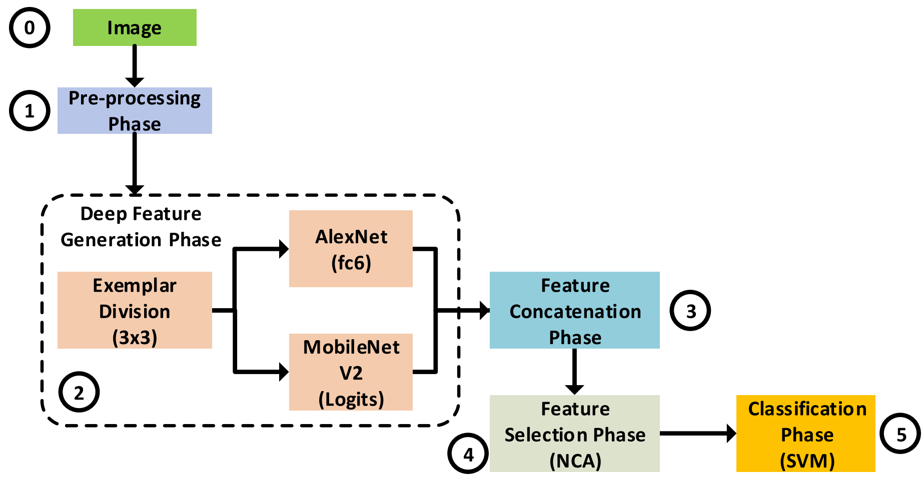
Algorithm 2. Preprocessing steps for datasets.

Input: The used raw FER image (I) with a size of M × N, where M and N are the height and weight of the used image.

Output: Preprocessed image (PI)

```
00: Load FER corpus
01: for i=1 to D do // Herein, D shows the number of images
02:     Read each FER image
03:     Segment facial area.
04:     Resize the image to generate features using deep networks.
05: end for i
```

Fig. 1 Flowchart of the proposed method



2.2 Feature generation

For a feature engineering model, feature extraction is one of the most important phases since the distinctive features bring high classification performance. To generate meaningful features, we have used two pretrained CNNs (CNNs are the flagships of computer vision) and patch division. The main objective of the presented feature extraction method is to generate local (by using patches) and global deep features by deploying the used hybrid deep feature extractor. Each CNN has individual classification performances. Therefore, we have presented a hybrid feature extraction function by using the most used two most well-known CNNs to use the advantages of both CNNs. This study uses MobileNetV2 and AlexNet deep learning models for feature extraction. These deep networks take the preprocessed and resized images as input parameters. Then, weights that were previously trained and obtained by the transfer learning process were applied to the datasets used in this study. MobileNetV2 deep network architecture has 154 layers, including the input layer (Sandler et al. 2018). AlexNet deep network architecture consists of 25 layers (Krizhevsky et al. 2012). In the feature generation process, the "*Logits*" layer, the fully connected layer of MobileNetV2, and the "*fc6*" layer, which is the fully connected layer in the AlexNet architecture, have been preferred. In addition, in this study, the images in the datasets are divided into 3 × 3 non-overlapping blocks. This process is called exemplar division. In this way, more features are obtained from preprocessed images, and the success rate is further increased. A block diagram of the feature generation process is given in Fig. 2.

As shown in Fig. 2, the recommended method primarily takes the resized images as input. Then, the application extracts feature from MobileNetV2, "*Logits*" layers and "*fc6*" layers from AlexNet. In the application, 1000 features are generated from the Logits layer and 4096 features

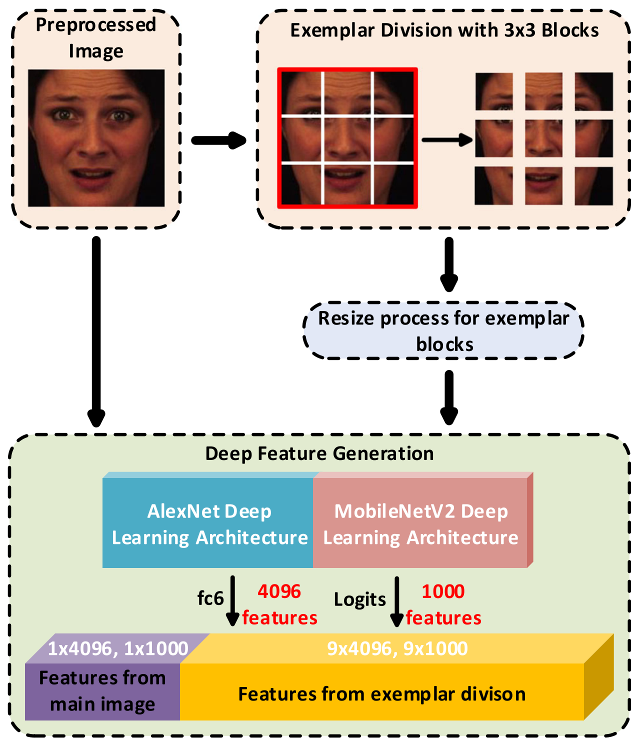


Fig. 2 Architecture of exemplar-based deep feature generation

from the fc6 layer. In this way, 5096 features are obtained from the main image. In the second stage of feature generation, exemplar division is applied to the main image, and the image is divided into M × M windows. Here, the "M" value is 3, so nine new patches are obtained, revealing the outline of the face (eyes, nose, lips, etc.). In the next step, the deep feature generation process applied to the main image is also applied to these images. In this way, 5096 features have been extracted from each input. Finally, a pseudo-code for the feature generation process is presented in Algorithm 3.

Algorithm 3. Exemplar deep feature generation.

Algorithm 3. Exemplar deep feature generation.

Input: Preprocessed images (PI)

Output: Generated deep features

Load PI
for i=1 to D **do**
    Read PI
    Deploy AlexNet to PI and extract 4096 features using the fc6 layer. This feature vector named f^1.
    Deploy MobileNetV2 to PI extract 1000 feature using the Logits layer. This feature vector named f^2.
    $X(i, 1{:}5096) = conc(f^1, f^2)$; // Concatenate the generated features. where $conc(.,.)$ defines concatenation function.
    $cnt = 1$;
    for m=1 to 3 **do**
        for n=1 to 3 **do**
            Create exemplars by dividing the image into 3 × 3 sized non-overlapping blocks.
            Generate features from each exemplar using AlexNet and MobileNetV2 and update f^1 and f^2
            $X(i, (cnt) \times 5096 + 1{:}(cnt + 1) \times 5096) = conc(f^1, f^2)$;
// Concatenate generated features. Herein, X defines the final feature vector with a length of 50,960 and $conc()$ function is the merging function.
            $cnt = cnt + 1$;
        end for n
    end for m
end for i

In Algorithm 3, both feature generation and feature concatenation are given.

2.3 Feature selection

One of the most important steps in machine learning is feature selection. This phase has two important advantages: (i) it increases the classification performance and (ii) it reduces the computational complexity in the classification process. Neighborhood component analysis (NCA) is one of the most well-known and frequently used methods in the literature (Goldberger et al. 2004). The NCA algorithm selects the essential features from the obtained features and performs this process by weighting the features. It is a nonparametric method. In addition, unlike the feature selection algorithms, the weights produced in this method have positive values. In the developed model, after all the features are weighted with the NCA algorithm, the features are sorted from large to small according to their weights. This process aims to eliminate the low-weight features in the classification process. After the sorting process, the last 1000 features with the highest weight are selected and given as input to the classification process. The feature selection process reduces the feature matrix of 1 × 50,960 size extracted for a single image to 1 × 1000 size. A pseudo-code for this process is given in Algorithm 4.

Algorithm 4. Pseudo-code used for NCA selector.

Algorithm 4. Pseudo-code used for NCA selector.

Input: The generated feature vector (X) with a size of D x 50960, actual output (y) with a length of D.

Output: The selected features (sf) with a size of D x 1000.

```
00: Load X
01: weights = NCA(X, y); // Calculate NCA weights.
02: idx = sort(weights, 'desc'); // Sort NCA weights by descending.
03: for i=1 to D do
04:     for j=1 to 1000 do
05:         sf(i, j) = X(i, idx(j));
06:     end for j
07: end for i
```

2.4 Classification

The last stage of implementation is classification; in this stage, SVM classifier is used. This section uses a shallow/conventional classifier to demonstrate the high classification capability of the generated and selected features. SVM is a supervised learning-based classification technique (Vapnik 1998). We have used MATLAB Classification Learner Toolbox (MCLT) to perform the classification. We have obtained the highest classification performance using the cubic kernel. In addition, box constraint level "1," kernel scale mode "auto"," and multiclass method "One-vs-All" are selected. Finally, we have performed a tenfold cross-validation technique to develop the model.

3 Experimental results

In this section, we present datasets and results. The details of these datasets are clarified in the following.

3.1 Materials

Five FER corpora were used to test the presented exemplar deep model. These public datasets are TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA. Sample images of these datasets used in experimental studies are shown in Fig. 3. Furthermore, details of these FER corpora are listed in Table 2.

This paper is about facial expression recognition (FER), and we have presented a new deep feature engineering model in this work. The used FER corpora are explained in this section. In this paper, five (Taiwanese Facial Expression Image Database) TFEID (Sandler et al. 2018), (Japanese Female Facial Expression) JAFFE (Goldberger et al. 2004), (The Karolinska Directed Emotional Faces) KDEF (Vapnik 1998), CK+ (Extended Cohn–Kanade) (Akhand et al. 2021), and Oulu-CASIA (Robnik-Šikonja and Kononenko 2003) datasets are used. Detailed explanations of these datasets are presented in the following sections.

3.1.1 TFEID Dataset

In this dataset, facial emotion images from 40 subjects were collected. Seven basic emotional images were taken for each of these subjects. Detailed features of this dataset are presented in Table 3.

3.1.2 JAFFE dataset

This dataset contains the facial expressions of 10 Japanese women. There are seven emotional states in this dataset, six of which are basic and one neutral. Detailed features of this dataset used in the study are given in Table 4.

3.1.3 KDEF dataset

The third dataset used in the study is KDEF. Images in this dataset were obtained from the Department of Clinical Neuroscience, Psychology, Karolinska Institutet. The features of this dataset, which includes 980 images and seven emotional expressions, are given in Table 5.

Fig. 3 Sample images used in this study from different datasets (row 1—TFEID, row 2—JAFFE, row 3—KDEF, row 4—CK+, row 5—Oulu-CASIA)

Table 2 Number of images used in the datasets

Class	Datasets				
	TFEID	JAFFE	KDEF	CK+	Oulu-CASIA
Anger	34	30	140	135	1796
Disgust	40	29	140	177	1790
Fear	40	32	140	75	1633
Happy	40	31	140	207	1791
Sad	39	31	140	84	1668
Surprise	36	30	140	249	1671
Neutral	39	30	140	–	–
Contempt	–	–	–	54	–
Total	268	213	980	981	10,349

3.1.4 CK+ dataset

The images in this dataset were obtained from the Department of Clinical Neuroscience, Psychology, Karolinska Institutet. The features of this dataset, which includes 980 images and seven emotional expressions, are given in Table 6.

3.1.5 Oulu-CASIA dataset

The final dataset used to verify the method proposed in this paper is Oulu-CASIA. This dataset includes six basic emotional expressions. This dataset consists of 10,349 images. In addition, the dataset contains visible (VIS) and near-infrared (NIR) images. VIS images were used in this study, and the features of the Oulu-CASIA dataset are presented in Table 7.

3.2 Results

As stated at the beginning of the section, five face corpora were used to develop this model. In addition, a tenfold cross-validation technique was applied in the classification process, and the confusion matrix obtained for each dataset is shown in Fig. 4.

The performance metrics used to calculate the classification including accuracy, recall, specificity, precision, F-

Table 3 Attributes of the TFEID dataset (Sandler et al. 2018)

Feature	Details
Gender	20 males, 20 females
	40 subjects in total
Facial expressions	7 facial expressions:
	Angry, disgust, fear, happy, neutral, sad, surprise
Images from	Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan
Database release year	2007
Total images for datasets	268
Images per expression	Average 38 images for each expression
Image format	JPG
Image color palette	8-bit Grayscale

Table 4 Attributes of the JAFFE dataset (Goldberger et al. 2004)

Feature	Details
Gender	10 females
	10 subjects in total
Facial expressions	7 facial expressions:
	Angry, disgust, fear, happy, neutral, sad, surprise
Images from	Kyushu University, Fukuoka, Japan
Database release year	1998
Total images for datasets	213
Images per expression	Average 30 images for each expression
Image format	TIFF
Image color palette	8-bit Grayscale

Table 5 Attributes of the KDEF dataset (Vapnik 1998)

Feature	Details
Gender	35 males, 35 females
	70 subjects in total
Facial expressions	7 facial expressions:
	Angry, disgust, fear, happy, neutral, sad, surprise
Images from	Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
Database release year	1998
Total images for datasets	980
Images per expression	Average 140 images for each expression
Image format	JPG
Image color palette	24-bit RGB

measure, and geometric mean are given in the following. The mathematical equivalents of these values are given in Eqs. (1)–(6) (Powers 2020).

$$\text{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \tag{3}$$

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$F_{\text{Measure}} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

$$\text{Geometric}_{\text{mean}} = \sqrt{\text{recall} \times \text{specificity}} \tag{6}$$

where TP, FN, FP, and TN are the numbers of true positives, false negatives, false positives, and true negatives, respectively. The performance results obtained for each corpora used in this study using our proposed model are given in Table 8. It can be seen from Table 8 that an accuracy of 97.01, 98.59, 94.69, 100, and 100% was obtained for TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA, respectively.

Table 6 Attributes of the CK+ dataset (Akhand et al. 2021)

Feature	Details
Gender	65 males, 145 females
	210 subjects in total
Facial expressions	7 facial expressions:
	Angry, disgust, fear, happy, contempt, sad, surprise
Images from	Carnegie Mellon University
Database release year	2000, 2010
Total images for datasets	981
Images per expression	Average 140 images for each expression
Image format	PNG
Image color palette	8-bit Grayscale

Table 7 Attributes of the Oulu-CASIA dataset (Robnik-Šikonja and Kononenko 2003)

Feature	Details
Gender	59 males, 21 females
	80 subjects in total
Facial expressions	6 facial expressions:
	Angry, disgust, fear, happy, sad, surprise
Images from	Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland
Database release year	2008, 2009
Total images for datasets	10,349
Images per expression	Average 1724 images for each expression
Image format	JPEG
Image color palette	24-bit RGB

4 Discussion

The results and findings are discussed in this section. This paper proposes exemplar-based deep feature generation method for automatic FER. To select the most appropriate deep feature generator, 11 deep features were tested on the KDEF dataset.

Figure 5 shows the graph of accuracy versus various deep feature generators. It may be noted that the most appropriate deep networks for feature generation are AlexNet (fc6) and MobileNetV2. Therefore, these pre-trained networks are utilized as feature generators. After the deep feature generators were determined, four different feature selection algorithms were tested on the KDEF dataset. These algorithms are ReliefF (Robnik-Šikonja and Kononenko 2003), NCA (Yang et al. 2012), mRMR (Ding and Peng 2005), and Chi2 (Liu and Setiono 1995). The feature vectors obtained from the deep feature extractors have been tested with these algorithms, and the test results are given in Fig. 6.

The given results in Fig. 6 were tested using the KDEF dataset, and these results were obtained using deep feature extractors (AlexNet and MobileNetV2), feature selection algorithms (ReliefF, NCA, mRMR, Chi2), and classification (SVM). The results given in Fig. 6 show that the best feature selection method is the NCA algorithm for the KDEF dataset. After that, decision tree (Tree), discriminant analysis (DA), k-nearest neighbor (KNN), and SVM classifiers are used for automated classification. The accuracy value for each classifier was obtained with KDEF dataset using the deep feature vector and the NCA-selected features. Finally, the obtained results of classification are given in Fig. 7.

It can be noted from Fig. 7 that SVM classifier is most suitable for classification. Hence, we have used it in our developed model. The obtained results during the model development process showed that the best performance would be obtained with AlexNet (fc6) and MobileNetV2-based deep feature extraction, NCA-based feature selection, and SVM-based classification.

This paper used five commonly used datasets to develop the model with tenfold cross-validation. Our developed model has obtained 97.01% accuracy using TFEID, 98.59% with JAFFE, 94.69% for KDEF, 100% using CK+, and 100% success rate using Oulu-CASIA datasets. The comparison of the results obtained in this paper with the studies carried out in the literature is given in Table 9.

The papers tested using these corpora are given in Table 9. In this table, previous studies in the field of FER are compared with the results obtained by our model. It can be seen from Table 9 that the model developed in this study provided the highest success rates using JAFFE, CK+, and Oulu-CASIA datasets. According to Table 9, the highest success obtained with the KDEF dataset is 97.93% (Vedantham and Reddy 2020). A deep belief network is used as a classifier in their study. In addition, the parameters were optimized by applying an optimization

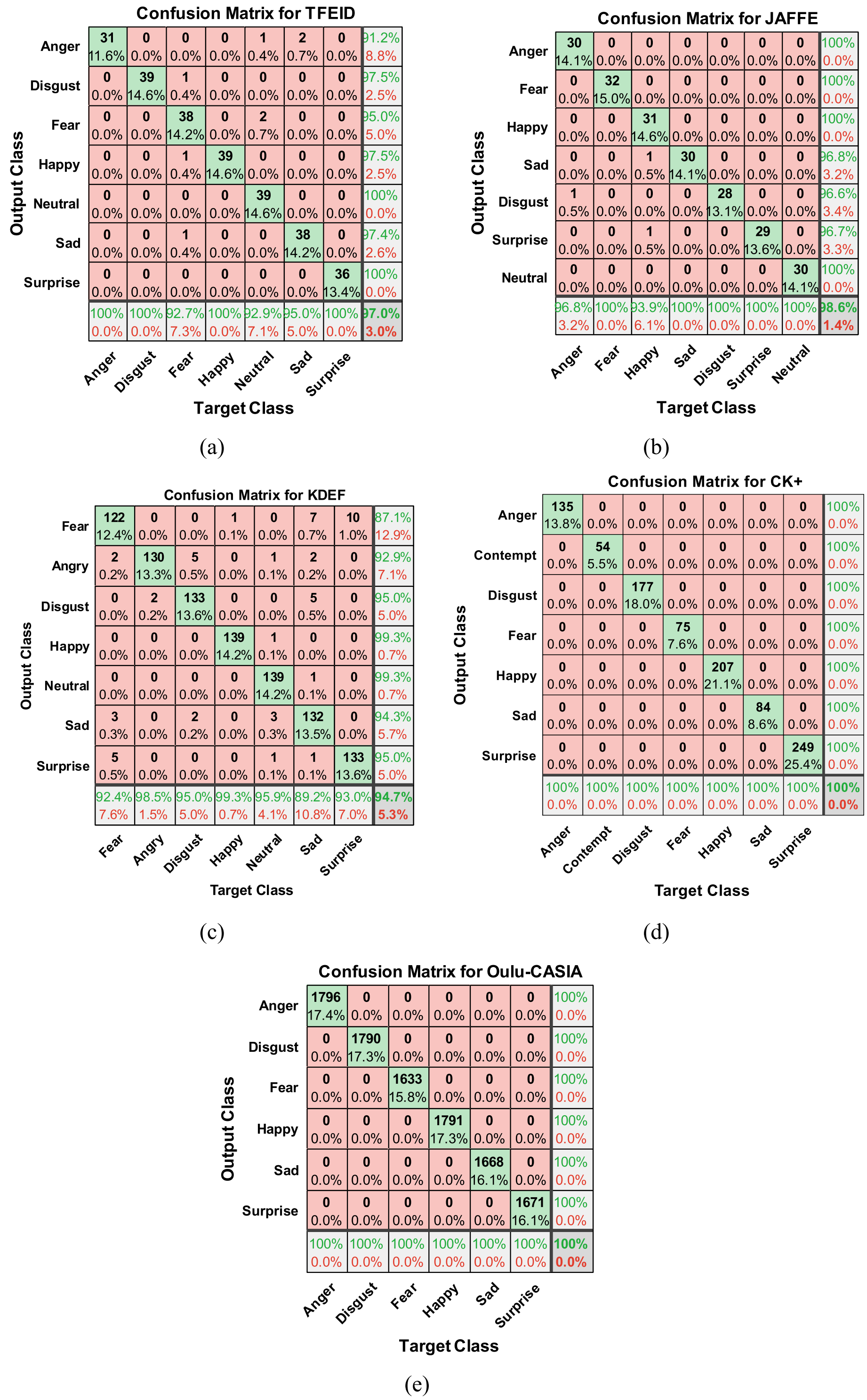


Fig. 4 Confusion matrices obtained for various face datasets

algorithm during the classification process. This process has increased the performance of the classifier. However, similarly, the highest performance could not be obtained for the Oulu-CASIA dataset. The highest success rate obtained with TFEID dataset is 98.9% (Zhao et al. 2022a). An exemplar-based method is proposed in their work. An accuracy rate close to this result was obtained in our method. Our proposed FER model generally obtained high performance using all five public datasets.

Automatic emotion detection is a very important and popular topic in the literature. In addition, there are various datasets in the literature for automatic FER systems. In this

Table 8 Performance results obtained for five corpora using our proposed model

Characteristics of corpora		Performance metrics (%)					
Datasets	Dataset size	Acc.	Rec.	Spe.	Pre.	FMea.	Geo.
TFEID	268	97.01	91.18	97.86	86.11	88.57	94.46
JAFFE	213	98.59	100	98.36	90.91	95.24	99.18
KDEF	980	94.69	87.14	95.95	78.21	82.43	91.44
CK+	981	100	100	100	100	100	100
Oulu-CASIA	10,349	100	100	100	100	100	100

**Acc.* accuracy, *Rec.* recall, *Spe.* specificity, *Pre.* precision, *FMea.* F-measure, *Geo.* geometric mean

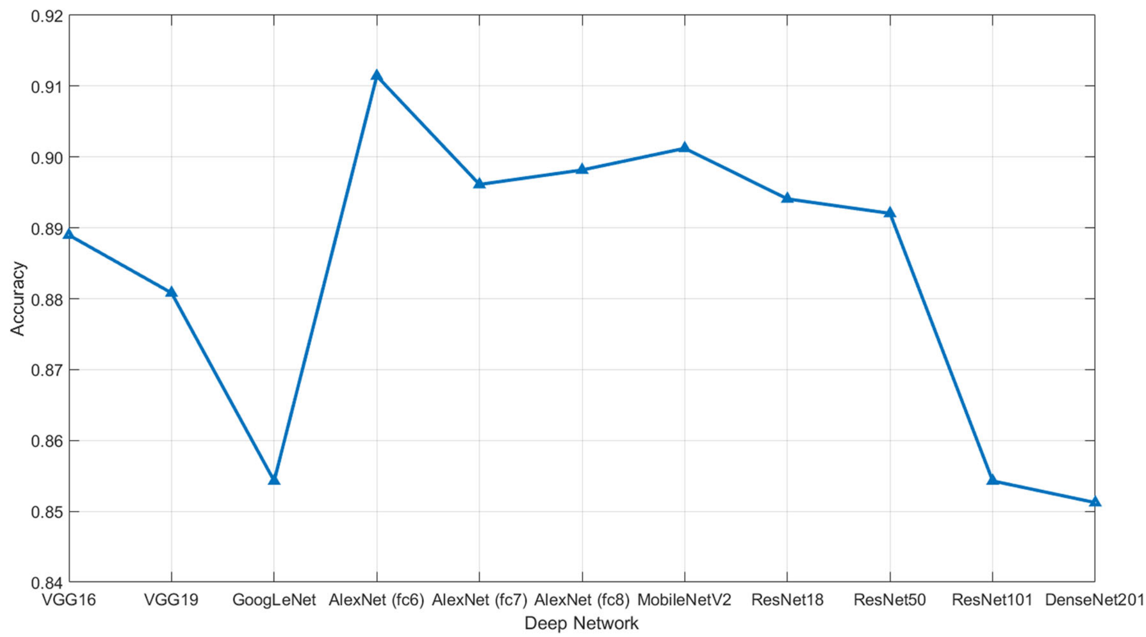


Fig. 5 Graph of accuracy versus the various deep feature generators

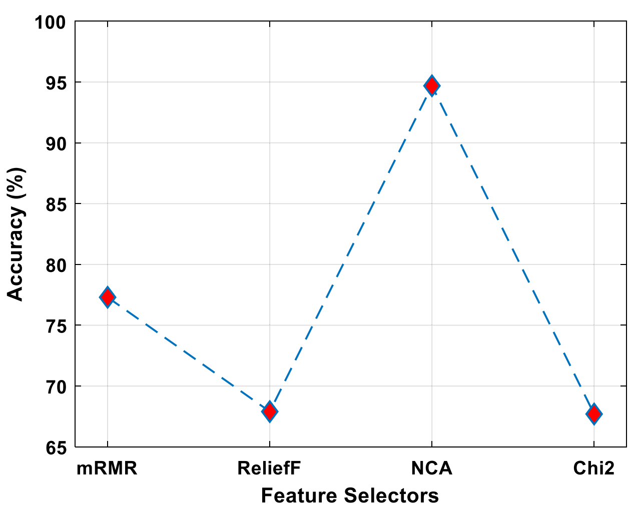


Fig. 6 Performance of feature selection algorithms

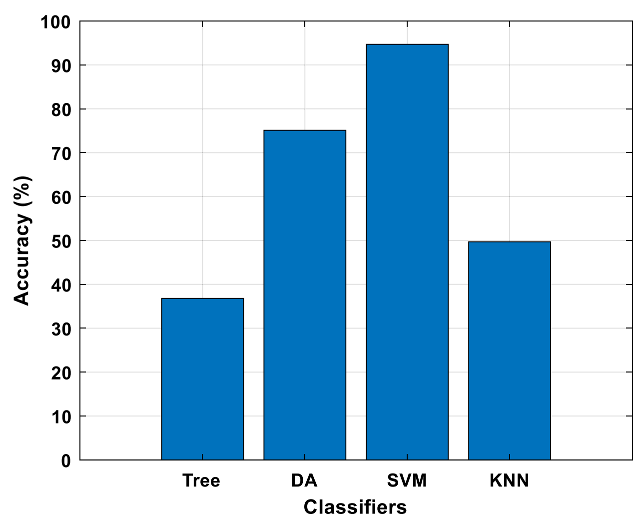


Fig. 7 Performance of classification algorithms

Table 9 Comparison of accuracy (%) with state-of-the-art methods and different datasets

Refs.	Year	Method	Evaluation criteria	TFEID	JAFFE	KDEF	CK+	Oulu-CASIA
Zhao et al. (2022a)	2018	Exemplar-based LBP and HOG features	Tenfold cross-validation	98.9	92.53	–	97.94	–
Eng et al. (2019)	2018	Multi-channel pose-aware CNN	Fivefold cross-validation	–	–	86.9	–	–
Porcu et al. (2020)	2019	HOG features	Holdout validation (70:30)	–	76.19	80.95	–	–
Zhao et al. (2022b)	2020	Orthogonal parallel-based directions generic quad map binary patterns	LOSO cross-validation	–	77.62	90.2	97.53	77.32
Foggia et al. 2023)	2020	Self-adaptive feature learning approach based on a priori knowledge	LOSO cross-validation	–	78.64	81.22	92.52	–
Sun et al. (2020)	2020	Data augmentation based on CNN	–	–	–	–	83.3	–
Vedantham and Reddy (2020)	2020	LPQ, WLD, and LBP-based feature extraction and deep belief network classification	–	–	–	97.93	97.58	92.38
Kas et al. (2021)	2021	Frequency neural network	Tenfold cross-validation	–	–	90.07	98.41	86.67
Khattak et al. (2022)	2022	Custom-designed CNN	Holdout validation (90:10)	–	95.65	–	–	–
Kumari and Bhatia (2022)	2022	Contrast-limited adaptive histogram equalization modified joint trilateral filter and CNN	–	–	–	–	98.01	–
Our method			Tenfold cross-validation	97.01	98.59	94.69	100	100

paper, a new automatic FER model has been developed. We have discussed the advantages and drawbacks of this model in the following.

Advantages:

- In this study, an exemplar-based feature extraction approach was used. In addition, deep feature extraction was performed using AlexNet and MobileNetV2 deep network architectures. The division process has been performed on facial images using the exemplar method, and high success has been achieved in the classification process. The division process helps in detecting important features such as eyes, nose, and lips.
- A fast, simple and accurate model is presented. This model has a low time burden since we have used transfer learning to extract features. The proposed deep feature extractor is very simple since we have only used the well-known two CNNs and other methods are very simple. Our model reached over 97% on the five FER corpora.
- We have used shallow methods. In this aspect, a cognitive model has been presented since any metaheuristic model was not used to increase classification performance.
- Our proposed fused deep feature engineering model outperformed.

Limitation:

- Bigger datasets can be used, but these are the commonly used datasets to get comparative results.

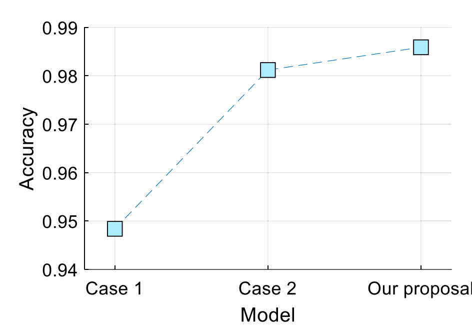


Fig. 8 Performances of the proposed

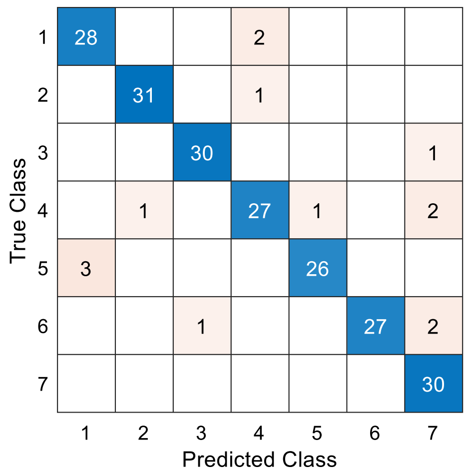


Fig. 9 Confusion matrix of the proposed model by deploying LOSO CV on the JAFFE dataset. The enumerated classes are explained as follows: 1: anger, 2: fear, 3: happy, 4: sad, 5: disgust, 6: surprise, 7: neutral

4.1 Ablation

In order to show the effectiveness of the methods used in our proposal, we have defined cases and these cases are given in the following:

Case 1 We have used only MobileNetV2 to get features.

Case 2 We have used only AlexNet to get features.

The classification performances of Case 1 and Case 2 are compared to our model, and the calculated classification accuracies are demonstrated in Fig. 8.

Figure 8 shows that our proposed methodology is the best and uses the advantages of MobileNetV2 and AlexNet together. These tests were applied to the JAFFE dataset.

To show the high classification capability of this model, we have tested this model using LOSO CV on the JAFFE dataset. The calculated confusion matrix by deploying the proposed model is demonstrated in Fig. 9.

As can be seen from Fig. 9, our model reached 93.43% classification accuracy on the JAFFE dataset by deploying LOSO CV. This result demonstrated the robust classification performance of the proposed model.

5 Conclusion

This study proposes a novel FER method using an exemplar hybrid deep feature generation technique. To generate comprehensive features in a short execution time, two pretrained deep networks are used as feature generation functions, and exemplar feature extraction is done using the presented hybrid deep feature generator. NCA is employed to select the most informative 1000 features, and finally, these features are classified by the SVM classifier. Our presented exemplar deep feature extractor and NCA selector-based model attained 97.01, 98.59, 94.69, 100, and 100.0% accuracies using TFEID, JAFFE, KDEF, CK+, and Oulu-CASIA datasets, respectively. In future, our proposed FER model can be extended for other applications such as biometric verification, security (fight terrorism, find fugitives), and protect personal information (ATM, home access without keys/passwords).

Author contributions MB, IT, SD, PDB, TT, KHC, and URA were involved in conceptualization, methodology, and writing—review and editing; TT and SD were involved in software; MB, IT, and SD were involved in validation and investigation; MB and IT were involved in formal analysis and visualization; MB, IT, and PDB were involved in resources; MB, IT, PDB, SD, and TT were involved in data curation; MB, SD, and TT were involved in writing—original draft preparation; URA was involved in supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Funding The authors state that this work has not received any funding.

Data availability Enquiries about data availability should be directed to the corresponding author.

Declarations

Conflict of interest The authors of this manuscript declare no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals.

Informed consent Not applicable.

References

Ahmed N, Al Aghbari Z, Girija S (2023) A systematic survey on multimodal emotion recognition using learning algorithms. Intell Syst Appl 17:200171

Akhand MAH, Roy S, Siddique N, Kamal MAS, Shimamura T (2021) Facial emotion recognition using transfer learning in the deep CNN. Electronics 10:1036

Arul Vinayakam Rajasimman M, Manoharan RK, Subramani N, Aridoss M, Galety MG (2023) Robust facial expression recognition using an evolutionary algorithm with a deep learning model. Appl Sci 13:468

Canal FZ, Müller TR, Matias JC, Scotton GG, de Sa Junior AR, Pozzebon E et al (2022) A survey on facial emotion recognition techniques: a state-of-the-art literature review. Inf Sci 582:593–617

Celniak W, Augustyniak P (2022) Eye-tracking as a component of multimodal emotion recognition systems. In: International conference on information technologies in biomedicine. Springer, pp 66–75

Cha H-S, Im C-H (2022) Performance enhancement of facial electromyogram-based facial-expression recognition for social virtual reality applications using linear discriminant analysis adaptation. Virtual Real 26:385–398

Chowdary MK, Nguyen TN, Hemanth DJ (2021) Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput Appl. https://doi.org/10.1007/S00521-021-06012-8

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248–255. IEEE

Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 3:185–205

Dzedzickis A, Kaklauskas A, Bucinskas V (2020) Human emotion recognition: review of sensors and methods. Sensors 20:592

Ekman P (1973) Cross-cultural studies of facial expression. Darwin and facial expression: a century of research in review. Academic Press, New York, pp 169–222

Eng S, Ali H, Cheah A, Chong Y (2091) Facial expression recognition in JAFFE and KDEF Datasets using histogram of oriented gradients and support vector machine. In: IOP conference series: materials science and engineering. IOP Publishing, p 012031

Farajzadeh N, Hashemzadeh M (2018) Exemplar-based facial expression recognition. Inf Sci 460:318–330

Febrian R, Halim BM, Christina M, Ramdhan D, Chowanda A (2023) Facial expression recognition using bidirectional LSTM-CNN. Procedia Comput Sci 216:39–47

Foggia P, Greco A, Saggese A, Vento M (2023) Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition. Eng Appl Artif Intell 118:105651

Gao H, Wu M, Chen Z, Li Y, Wang X, An S et al (2023) SSA-ICL: multi-domain adaptive attention with intra-dataset continual learning for Facial expression recognition. Neural Netw 158:228–238

Geiger M, Wilhelm O (2023) Computerized facial emotion expression recognition. Digital phenotyping and mobile sensing: new developments in psychoinformatics. Springer, Cham, pp 43–56

Ghosh S, Priyankar A, Ekbal A, Bhattacharyya P (2023) Multitasking of sentiment detection and emotion recognition in code-mixed Hinglish data. Knowl Based Syst 260:110182

Gil S, Le Bigot L (2023) Emotional face recognition when a colored mask is worn: a cross-sectional study. Sci Rep 13:1–15

Goldberger J, Hinton GE, Roweis S, Salakhutdinov RR (2004) Neighbourhood components analysis. Adv Neural Inf Process Syst 17:513–520

Jupalli TK, Reddy MST, Kondaveeti HK (2023) Artificial intelligence in higher education. Mobile and sensor-based technologies in higher education. IGI Global, pp 1–30

Kas M, Ruichek Y, Messoussi R (2021) New framework for person-independent facial expression recognition combining textural and shape analysis through new feature extraction approach. Inf Sci 549:200–220

Kavitha M, RajivKannan A (2023) Hybrid convolutional neural network and long short-term memory approach for facial expression recognition. Intell Autom Soft Comput 35:689–704

Khattak A, Asghar MZ, Ali M, Batool U (2022) An efficient deep learning technique for facial emotion recognition. Multimed Tools Appl 81:1649–1683

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105

Kumari N, Bhatia R (2022) Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter. Soft Comput. https://doi.org/10.21203/rs.3.rs-866042/v1

Liu W-L, Gong Y-J, Chen W-N, Liu Z, Wang H, Zhang J (2019) Coordinated charging scheduling of electric vehicles: a mixed-variable differential evolution approach. IEEE Trans Intell Transp Syst 21:5094–5109

Liu S, Gao P, Li Y, Fu W, Ding W (2023) Multi-modal fusion network with complementarity and importance for emotion recognition. Inf Sci 619:679–694

Liu H, Setiono R (1995) Chi2: feature selection and discretization of numeric attributes. In: 7th IEEE international conference on tools with artificial intelligence, pp 388–391. IEEE

Liu Y, Zeng J, Shan S, Zheng Z (2018) Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pp 458–465. IEEE

Nikolaus M, Fourtassi A (2023) Communicative feedback in language acquisition. New Ideas Psychol 68:100985

Othman E, Werner P, Saxen F, Al-Hamadi A, Gruss S, Walter S (2023) Classification networks for continuous automatic pain intensity monitoring in video using facial expression on the X-ITE Pain Database. J vis Commun Image Represent 91:103743

Porcu S, Floris A, Atzori L (2020) Evaluation of data augmentation techniques for facial expression recognition systems. Electronics 9:1892

Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn 53:23–69

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–20

Shen J, Yang H, Li J, Cheng Z (2022) Assessing learning engagement based on facial expression recognition in MOOC's scenario. Multimed Syst 28:469–478

Sun Z, Chiong R, Hu Z-P (2020) Self-adaptive feature learning based on a priori knowledge for facial expression recognition. Knowl Based Syst 204:106124

Tang Y, Zhang X, Hu X, Wang S, Wang H (2020) Facial expression recognition using frequency neural network. IEEE Trans Image Process 30:444–457

Vapnik V (1998) The support vector method of function estimation. Nonlinear Modeling: advanced black-box techniques. Springer, pp 55–85

Vedantham R, Reddy ES (2020) A robust feature extraction with optimized DBN-SMO for facial expression recognition. Multimed Tools Appl 79:21487–21512

Wang Y, Song W, Tao W, Liotta A, Yang D, Li X et al (2022) A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fus. https://doi.org/10.48550/arXiv.2203.06935

Wani AH, Hashmy R (2023) A supervised multinomial classification framework for emotion recognition in textual social data. Int J Adv Intell Paradig 24:173–189

Yang W, Wang K, Zuo W (2012) Neighborhood component feature selection for high-dimensional data. J Comput 7:161–168

Zhao F, Hu X, Wang L, Zhao J, Tang J (2022b) A reinforcement learning brain storm optimization algorithm (BSO) with learning mechanism. Knowl Based Syst 235:107645

Zhao F, Di S, Wang L (2022a) A hyperheuristic with q-learning for the multiobjective energy-efficient distributed blocking flow shop scheduling problem. IEEE Trans Cybern

Zhen R, Song W, He Q, Cao J, Shi L, Luo J (2023) Human-computer interaction system: a survey of talking-head generation. Electronics 12:218

Zhou S, Xing L, Zheng X, Du N, Wang L, Zhang Q (2019) A self-adaptive differential evolution algorithm for scheduling a single batch-processing machine with arbitrary job sizes and release times. IEEE Trans Cybern 51:1430–1442