IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 39, No. 6, December, 2022, pp. 2165-2172
Journal homepage: http://iieta.org/journals/ts

Multimodal Human Facial Emotion Recognition Using DenseNet-161 and Image Feature Stabilization Algorithm

Angeline R, Alice Nithya A*

Department of Computational Intelligence, SRMIST, Kattankulathur, Chengelpet District-603203, Tamilnadu, India

Corresponding Author Email: alicenia@srmist.edu.in

https://doi.org/10.18280/ts.390630

Received: 20 June 2022
Accepted: 18 December 2022

Keywords:
facial emotion recognition, convolutional neural network (CNN), deep CNN, transfer learning, image feature stabilization algorithm, DenseNet-161, CK+

ABSTRACT

Human Facial Emotion Recognition (FER) is the technology to predict listener's emotion of static images and videos to uncover data on one's enthusiastic states like happy, sad, frustration, anxiety, surprise, hate and neutral states. It's a part of the affective computing technology, which may be a collaborative area of research on listener's emotions. The ability of a computer to recognise and understand and frequently explain human emotions and emotional states which relies on computing (AI) technology. The most difficult part of the FER is to plan the various looks with the separate enthusiastic states. The standard steps of Facial Emotion Recognition are i) Face RoI identification ii) Feature Extraction and iii) Emotion Recognition. Convolutional Neural Network models are the most commonly utilised to detect listener emotions. In this paper, an Image Feature Stabilization Algorithm (IFSA) is proposed to improve the efficiency of facial emotion recognition by implementing Deep Convolutional Neural Network (DCNN) model using the Transfer Learning (TL) technique. The architecture entails employing a FER-compatible pre-trained Densenet-161 based DCNN model and then fine-tuning the model for face emotion data. Initially, the dense layer(s) is/are trained, followed by the fine-tuning of each of the pre-trained DCNN blocks, resulting in an improvement in FER accuracy, particularly for difficult front face views like partial view. Experiments when performed on the CK+ dataset by employing a 10-fold cross validation method using various pre-train models like VGG 16, VGG19, ResNet-18, 34, 50, 152, Inception-V3 and DenseNet-161 showed accuracy of 85.9%, 94.6%, 91%, 91%, 91.1%, 95.1%, 97.1% and 98.7% respectively. Thus, Facial Emotion Recognition performed using the finetuned DenseNet-161 architecture demonstrated exceptionally improved accuracy compared to other pretrained models, along with the proposed image feature stabilization algorithm. The proposed architecture using Densenet-161 showed improved accuracy of 98.78% and 97.52% in other challenging FER datasets like KDEF and JAFEE.

1. INTRODUCTION

Emotions are basic human characteristics that play a significant role in emotional regulation [1, 2]. Humans express their feelings in several ways, like facial features [3], speech, visual communication [4]. Facial emotional analysis is the most prominent and well-researched of the factors connected to emotion identification. Ekman [2] and Suja and Tripathi [5], studied human facial expressions extensively and found broadly accepted facial emotions such as happy, sad, frustration, anxiety, surprise, hate, and neutral states. Recognizing facial emotion expressions has recently been a popular study issue in psychology, psychiatry, and mental health [6]. Smart living technology [7], healthcare systems [8], emotional disorder or mental health diagnosis in autism spectrum disorder [9], computer interaction with humans [10] and robot interaction with humans in social development process [11] all require automated facial emotion detection from human facial expressions. Thus, research in this field has a large number of possible applications.

In FER, the primary objective is to track multiple facial expressions to their corresponding emotional states. This is divided into two parts: first, extracting facial features; and secondly, recognising facial emotions. To achieve these, image preparation is required, which includes face identification, cropping, scaling, and normalising. Face detection crops the facial region after removing the backdrop and non-face items. In the existing FER systems, feature extraction can be done with techniques like Discrete Wavelet Transform (DWT), linear regression approaches [12, 13]. Finally, the retrieved characteristics are used to categorise emotions with the help of Neural Networks (NN) and other suitable machine learning approaches.

Due to their built-in feature extraction method for images, Convolutional Neural Networks (CNNs) have increasingly attracted a lot of interest in Facial Emotion Recognition [14, 15]. A few methods proposed using the CNN to tackle FER difficulties have been described [16-21]. There are a lot of discrete convolutional layers in Deep CNN. It becomes challenging to train a large number of hidden layers in a CNN. The Deep CNN architecture and the training approach are enhanced using efficient ways to boost accuracy [22-25]. VGG-16 [22], Resnet-152 [26], Inception-v3 [24], and DenseNet-161 [27] are some of the most extensively used pre-trained DCNN models. Training a Deep CNN model, on the other hand, needs a large quantity of data and a lot of computer resources. Many previous studies simply looked at frontal views, while others used datasets containing profile views but didn't include the profile view images in the experiment since it was more convenient [12, 28]. As a result, an even more effective FER system that can recognise emotions from both the facial and side perspectives are required [29-32].

The proposed FER system in this paper has an Image Feature Stabilization Algorithm (IFSA), combined with DCNN and Transfer Learning (TL) [29]. By using the proposed IFSA computationally efficient models are created by learning from previously learnt patterns. The reuse of pre-trained models reduces the time and effort of training from the images, which takes a lot of data. In another perspective, transfer learning reuses information by using models that have already been trained [30] that were trained on a large validation set of data for a comparable issue.

The main research findings may be summarised as follows:

(1) Design of a cost-effective FER technique based on DCNN models that addresses the problems using Transfer Learning.

(2) The use of a pipeline training approach to fine-tune the model over time until it produces extremely accurate results.

(3) Facial images in frontal and profile perspectives were used to compare the framework to commonly used pre-trained DCNN models.

(4) The accuracy of the proposed technique in recognising emotions is compared to that of existing ways, as well as a review of the method's competency, particularly with profile images, which is crucial for practical application.

The remaining paper is organised as follows: The present FER approaches are briefly reviewed in Section 2. The suggested Image Feature Stabilization Algorithm (IFSA) is presented in Section 3 with a brief explanation of Transfer Learning and the suggested TL-based FER approach. The results and discussions are presented in Section 4. The study comes to a close with a discussion of future research prospects in Section 5.

2. RELATED WORKS

Traditional FER techniques use features from a face image to determine emotion, and a value will be assigned to it. Recent approaches based on deep learning accomplish the FER problem by merging both processes into a single composite operating process. A number of publications [12, 13] examined and compared current FER approaches, with the most recent ones [30] including deep learning-based methods. The approaches used in the most popular FER methods are briefly described in the subsections below.

2.1 FER methodologies based on machine learning

In Artificial Intelligence, automatic FER is a complicated problem, especially in the machine learning subsystem. To advance the FER issue, many common FER Methodologies Based on Machine Learning have been used in the Table 1.

The fundamental drawback is to examine only frontal views for FER, despite the fact that classic feature extraction methods distinguish between frontal and profile views.

Table 1. Machine learning based FER approaches

Author and Reference	Method / Techniques used for FER
Liew and Yairi's [11]	Worked for categorization on features using several techniques like Gabor, Haar, and LBP
Lee et al. [31]	Augmenting approach for facial image classification and an extended wavelet transform for 2D called Contourlet Transform (CT) for feature extraction
Joseph and Geetha [33]	On their suggested face geometry-based feature extraction, evaluated various classification algorithms, including logistic regression, LDA, KNN, classification and regression trees, naive Bayes, and SVM
Jabid et al. [34]	For feature extraction, studied an appearance-based approach termed Local Directional Pattern (LDP)
Zhi and Ruan [35]	Used 2D classifier-based image projections to extract facial behaviours

Table 2. Deep learning based FER approaches

Author and Reference	Method / Techniques used for FER
Mollahosseini et al. [15]	Looked for more complex design that included four inception layers and two convolutional-pooling layers
Zhao and Zhang [16]	Combined a Deep Belief Network (DBN) used for unsupervised feature learning with a Neural Network (NN) for FER used for feature emotion categorization
Li et al. [17]	Used transfer learning to expand FaceNet2ExpNet
Bendjillali et al. [18]	Employed CNN to FER DWT extracted features
Ngoc et al. [19]	Used graph-based CNN
Pranav et al. [20]	On self-collected face emotional images, used a typical CNN architecture with two convolutional-pooling layers
Ruiz-Garcia et al. [28]	CNN initialization outperforms CNN with random initialization
Ding et al. [36]	Introduced the FaceNet2ExpNet architecture, which extends deep facial recognition architecture to the FER
Jain et al. [37]	Used hybrid deep learning architecture
Shaees et al. [38]	Hybrid architecture with TL, in which features from pre-trained AlexNet are identified using SVM
Liliana [39]	Used a deep CNN architecture with 18 convolutional layers and four subsampling layers
Shi et al. [40]	Suggested a clustering strategy for FER using CNN
Jin et al. [41]	Evaluated both labelled and unlabelled datasets
Porcu et al. [42]	Tested several data preprocessing strategies, incorporating generated images to train the deep CNN, and found that a mixture of new images
Pons and Masip [43]	Individual CNNs were trained with varied thicknesses of filters in convolutional layers or variable numbers of neurons in fully connected layers
Wen et al. [44]	Examined the ensemble of CNNs

2.2 FER methodologies based on deep learning

Deep learning is a relatively new machine learning technique for FER, and CNN-based review is tabulated in the Table 2.

Existing deep learning-based algorithms have taken frontal view images into account, and most research have even eliminated the dataset's profile view images from the research to make the work easier [11, 28, 45-50].

3. PROPOSED SYSTEM

The key contribution of this work is FER employing a pre-trained DCNN model with acceptable TL using proposed Image Feature Stabilization Algorithm (IFSA). The CNN layers first layer catches fundamental information; the next layer recognises more complicated properties and the top layer learns more complex patterns. Constructing DCNN model from the beginning is challenging, so the Transfer Learning approach for emotion recognition uses an already trained model for fine tuning. The image gradients information of horizontal, vertical and diagonal is encoded with Center-Symmetric LBP (CS-LBP) operator for image feature description. Binary codes are used to obtain the CS-LBP descriptor, which is calculated from pixel patches by comparing center-symmetric pairs of pixels. Histogram of Oriented Gradient (HOG) feature descriptor finds the pixels per cell to give the details between the number of dimensions and the number of details in the images. CS-LBP develops the order space of the descriptors used to build the histogram of orders, where the orders are used to compute the entire patch. Based on the Histogram of Orders (HOO), intervals are formed and based on intensities, intervals are formed. The HOG and HOO give information about the overall distribution of pixel patch intensities. The CS-LBP encodes local gradient information.

For image classification, a DCNN model (e.g., VGG, ResNet, Inception-v3 and DenseNet-161) is pre-trained using a dataset (e.g., CK+).

The proposed Image Feature Stabilization Algorithm (IFSA) will extract the feature from the gradients of the original image. The CS-LBP is clearly specified in such a manner that the difference between neighbouring pixels that are located adjacent to a particular pixel is taken into account in order to produce a 4-bit code, which is seen in Figure 1. Histograms of size 16 are generated for each spatial bin, since only four comparisons are used in the analysis. In the course of this research, we have altered the CS-LBP descriptor in a number of different ways, with the goal of enhancing the performance of this descriptor.

The CS-LBP operator can be represented as:

$$\text{CS-LBP}_{R,N}(x,y) = \sum_{i=0}^{\left(\frac{N}{2}\right)-1} s\left(grayimg\ n_i - grayimg\ n_{i+\left(\frac{N}{2}\right)} \right) 2^i$$

$$S(x) = 1 \text{ for } x>1$$
$$0 \text{ for others}$$

where, R-radius, N–equally spaced pixels, x, y are the center-symmetric pairs of pixel and grayimg n_i and grayimg $n_{i+\left(\frac{N}{2}\right)}$ are the gray image values of center symmetric pair of pixels.

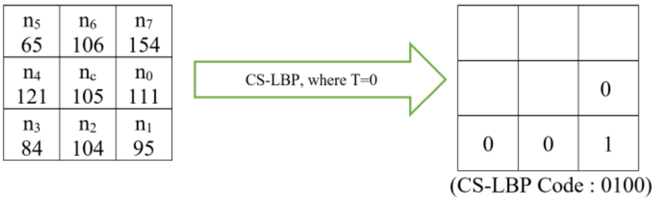

Figure 1. Illustrative diagram for CS-LBP Operator

For one dimensional image function,

$$\frac{d\varphi}{dx} = \lim_{\in\to 0} \frac{\left(\varphi(x\pm\in) - \varphi(x)\right)}{\in} \tag{1}$$

For two-dimensional image function,

$$\frac{\partial\varphi(x,y)}{\partial x} = \lim_{\varepsilon\to 0} \frac{\varphi(x\pm\varepsilon,y) - \varphi(x,y)}{\varepsilon} \tag{2}$$

$$\frac{\partial\varphi(x,y)}{\partial y} = \lim_{\varepsilon\to 0} \frac{\varphi(x,y\pm\varepsilon) - \varphi(x,y)}{\varepsilon} \tag{3}$$

Considering for smallest ∈=1 in the equation 2 and 3, it can be reduced to:

$$\frac{\partial\varphi(x,y)}{\partial x} = \varphi(x\pm 1,y) - \varphi(x,y) = gx \tag{4}$$

$$\frac{\partial\varphi(x,\ y)}{\partial x} = \varphi(x,y\pm 1) - \varphi(x,y) = gy \tag{5}$$

where, *gx* stands for horizontal gradient and *gy* stands for vertical gradient of a point (*x*, *y*) when *ε*=1.

Feature stabilisation is classified into positive and negative image stabilization. If an image feature operator is added to the input image, this is called positive image feature stabilization and subtracted from the input image, this is called negative image feature stabilization. For feature stabilization, gradients of x and y axis (center-symmetric pairs of pixels) are combined with right, left, up, and down gradients of images, respectively. This process will undergo two steps: (i) Image Fusion and (ii) Differential Image Fusion.

Step (i)

$$G_{imgfus} = G_{rx} + G_{lx} + G_{uy} + G_{dy} \tag{6}$$

where,

$$G_{rx} = |\varphi(x+1,y) - \varphi(x,y)|;$$
$$G_{lx} = |\varphi(x,y) - \varphi(x-1,y)|$$
$$G_{uy} = |\varphi(x,y+1) - \varphi(x,y)|$$
$$G_{dy} = |\varphi(x,y) - \varphi(x,y-1)|$$

Step (ii)

$$G_{imgsubfus} = \left(G_{rx} - G_{ly}\right) + \left(G_{uy} - G_{dy}\right) \tag{7}$$

IFSA Steps

Step 1: Input image will be converted to output as gray scale image.

Step 2: Estimate the result of image fusion by adding the top, bottom, left and right gradients of the image captured.

Step 3: Estimate the result of differential image fusion by adding the subtracted value of the top and bottom gradients with left and right gradients.

Step 4: Calculate the pixel value p(a,b) of the pixel point p0(a,b).

Step 5: Repeat the Step 1 – 4 to convert the image into a matrix P of (a-2)*(b-2).

Step 6: Using adjacent value, find the P' transformation matrix of P.

Step 7: PCA is one of the best statistical techniques to maximize the variance and minimize the error which is used for reducing dimensions in order to use the important feature information.

where, p(m,n)=p0(m,n)+Sm*gimgfus(m,n)-Sn*gimgsubfus(m,n), Sm and Sn are the stabilization cooefficient of image fusion and differential image fusion.

The suggested FER system is shown in Figure 2 with a well-known pretrained DCNN model using DenseNet-161.

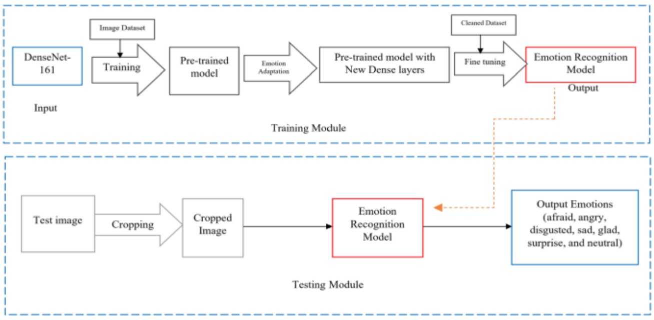
Image Dataset
DenseNet-161
Training
Pre-trained model
Emotion Adaptation
Pre-trained model with New Dense layers
Cleaned Dataset
Fine tuning
Emotion Recognition Model
Output
Input
Training Module
Test image
Cropping
Cropped Image
Emotion Recognition Model
Output Emotions (afraid, angry, disgusted, sad, glad, surprise, and neutral)
Testing Module

Figure 2. FER system based on transfer learning in deep CNN

The suggested model is trained with new images by doing rotation, shift, scale, or flip of the original images using image cropping and data augmentation. A procedure is included in the data loader during training. A little modification performed to each image loads data from memory and modifies the memory each time. The model is less susceptible to overfitting since it is not constantly fed the same data. In the case of FER, where the dataset is small, this is favourable. The revised cost function for checking every individual image in the FER model is now computed using the loss function:

$$\mathrm{L}() = \sum_{img=1}^{N} \sum_{t=1}^{T} \log P(y_n \mid n_n^t) \qquad (8)$$

where, N is the number of images in the dataset and T is the number of alterations to be applied to each image.

4. EXPERIMENTAL RESULTS

The suggested FER system using TL on DCNN using the proposed Image Feature Stabilization Algorithm (IFSA) is tested in an experimental study. The suggested model's performance on the benchmark datasets is compared to several current approaches to ensure that the proposed model is effective.

4.1 Benchmark dataset

For the emotion recognition problem, there are a few datasets available; among them, the Extended Cohn-Kanade CK+ dataset is used in this investigation. Afraid (AF), Angry (AN), Disgusted (DI), Sad (SA), Happy (HA), Surprised (SU), and Neutral (NE) are the seven emotion classes represented in the datasets. The datasets are described briefly below, along with the reasons behind their decision. Figure 3 shows some sample images of the CK+ dataset.

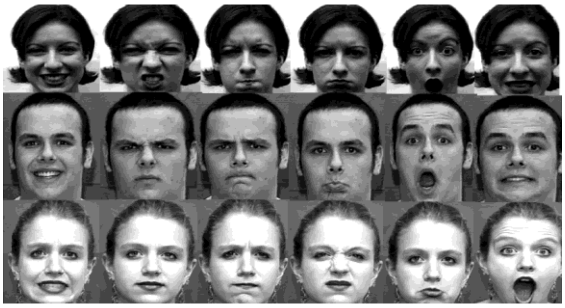

Figure 3. CK+ dataset sample images

4.2 Research setup and result analysis

In this study, OpenCV [48] is utilised to crop the face. The photos were shrunk to 224x224 pixels, which is the default input size for DCNN models that have been pre-trained. The following are the Adam optimizer's settings with the learning rate of 0.0005, beta 1 and 2 as 0.9 and 0.009 respectively. In addition, we just used a small amount of data augmentation using the following settings: Horizontal Flip, Scaling Factor: 1.1, and Rotation: 90 degrees (10 to 10). Research studies were conducted in two different modes 90 percent of available images from CK+ were randomised used as the training set, remaining 10% of images were reserved as the test set.

A 10-fold Cross-Validation (CV) was used to separate the training and test sets. The available images are separated into ten equal (or almost equal) sets in a 10-Fold CV, and the result is an average of ten independent runs, with each set serving as a test set and the remaining nine sets serving as training sets.

The suggested model's performance is compared to that of other pre-trained DCNN models using DenseNet-161. On the CK+ dataset, Table 3 shows the test set accuracies for the conventional CNN model with 2 layers, of stride 1 for a 3 x 3 size kernel, and 2 x 2 MaxPooling for varied input sizes from 48 x 48 to 360 x 360. The size of the test was chosen at random from 10% of the available data. The 50 iterations are reported as the best test accuracies.

Table 3. Test Accuracy using CNN with two layers on CK+

Input Image Size	CK+(in percentage)
48 × 48	79%
64 × 64	83%
128 × 128	92%
224 × 224	88%
360 × 360	98%

Table 4. Comparison of test set accuracies with VGG-16 for different training modes in fine-tuning

Training Model	CK+ Dataset
Only Dense Layers	63.16%
Dense Layers + VGG-16 (Block 5)	76.32%
Whole Model (Dense Layers + Full VGG-16 Base)	94.64%
Whole Model from Scratch	35.74%

The suggested method trains deeper CNN models with TL on a pre-trained model to reduce overfitting while training with a limited dataset. On the CK+ dataset, Table 4 shows the test set (with randomly selected 10% of the data) accuracies of the proposed model with VGG-16 for the different fine-tuning modes.

Table 5. Comparison of test set accuracies with different pre-trained Deep CNN models on CK+ dataset

Pre-Trained Deep CNN Model	CK+ in 10-Fold CV
VGG-16	85.9 ± 2.5%
VGG-19	94.65 ± 1.8%
ResNet-18	91.0 ± 3.3%
ResNet-34	91.0 ± 2.7
ResNet-50	91.1 ± 3.7
ResNet-152	95.1 ± 2.1
Inception-v3	97.1 ± 1.0
DenseNet-161	98.7 ± 1.3

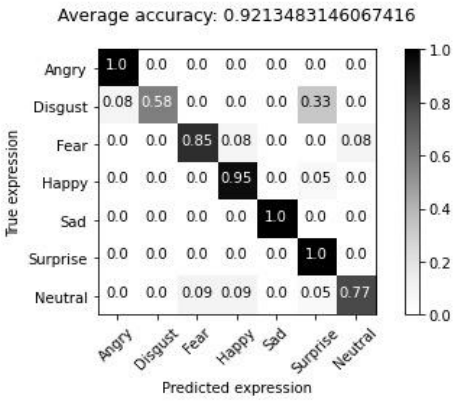

Figure 4. Confusion matrix of CK+ dataset

Table 5 shows the CK+ dataset consists of 321 labelled videos, among which 327 are annotated with eight expression labels. The total number of images in the CK+ dataset is 931 facial images, which is split into 10 folds. The proposed method is evaluated on eight distinct pre-trained DCNN models.

The proposed technique shows the highest accuracy for DenseNet-161 using CK+. The various emotions can be identified with more accuracy. In Figure 4, confusion matrix of CK+ using the DenseNet-161 model is shown and in Figure 5, a graphical representation of training and validation accuracy/loss is given.

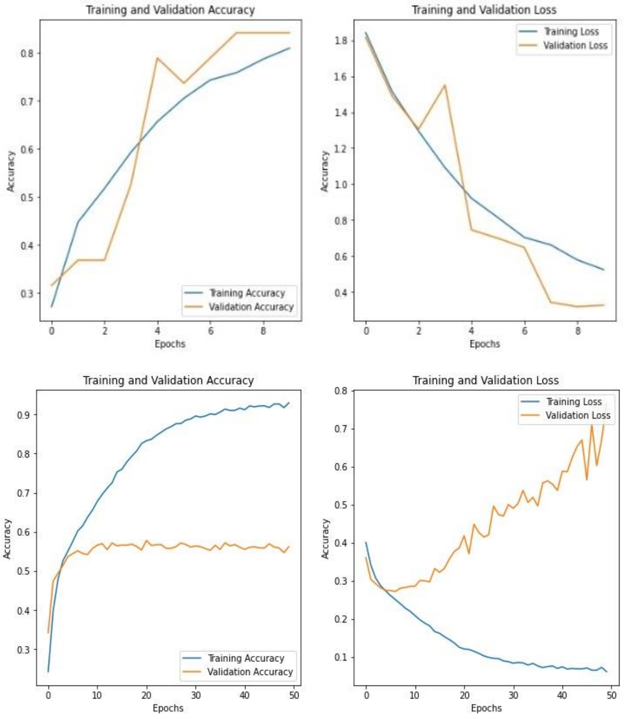

Figure 5. Epoch in X-axis and Accuracy/Loss in Y-axis using CK+ Dataset

Table 6. Comparison of the facial emotion accuracy of proposed method with existing works on KDEF, JAFFE and CK+ datasets with number of sample images

Author, Reference and Year	Samples taken for training and test division	Test set Accuracy in percentage			Methods used for selection and classification
		KDEF	JAFEE	CK+	
Joseph and Geetha [33], 2020	478:10-Fold CV	31.21			Facial geometry-based feature image extraction with different classification methods of SVM, KNN were used
Ruiz-Garcia et al. [28], 2017	980 frontal images	92.52			CNN weights are initialized using Stacked Convolutional Auto-Encoder (SCAE)
Lee et al. [31], 2012	213: 30-Fold CV		95.91		Derived feature vector from 2D discriminant locality preserving projections
Joseph and Geetha [33], 2010	213: 7-Fold CV		82.60		Appearance-based technique and classification with various SVM used for feature extraction
Lee et al. [31], 2012	210: 30-Fold CV		96.43		For facial image classification Boosting algorithm is used and Contourlet Transform for feature extraction
Proposed Method with DenseNet-161	KDEF# 4900: 10-Fold CV JAFFE# 213: 10-Fold CV CK+# 327: 10-Fold CV	98.78	97.52	98.70	Transfer leaning on pre-trained Deep CNN model employing a pipeline fine tuning technique. Feature representation using DWT with 2D-LDA and classification using SVM

4.3 Results comparison with existing methods

Using the KDEF, JAFFE, and CK+ datasets, this section compares the performance of the proposed FER technique to that of well-known emotion detection systems. Table 6 provides information on the identification accuracy of test sets, the separation of training and test data, and the differentiating characteristics of the various techniques. As demonstrated in the table, existing studies explored a range of techniques for separating training and test samples.

The various techniques used for facial image enhancement are:

·Filtering using morphological operators
·Equalization of histograms
·Noise elimination with a Wiener filter
·Contrast adjustment linear
·Median filtration
·Gaussian blur mask filtering
·Image Feature Stabilization Algorithm (IFSA)
·Decorrelation stretches

Among the various techniques using face image enhancement using the Image Feature Stabilization Algorithm (IFSA), the suggested technique which beats any other deep learning-based method in terms of performance, demonstrating the efficacy of the TL-based strategy for FER.

5. DISCUSSION

Emotion identification from facial images in uncontrolled situations when frontal view images are not always attainable is becoming more crucial for a secure and safe existence, intelligent living, and an intelligent society in today's society. To accomplish this objective, a robust FER is required since it enables emotion identification from a variety of facial images viewed from different perspectives. Face expression characteristics cannot be recovered from profile images with conventional feature extraction techniques. Consequently, utilising the DCNN model to extract FER from a high-resolution face image is the sole method for completing such a challenging assignment. The proposed FER method makes a pre-trained DCNN compatible with FER by replacing its top layers with dense layers and fine-tuning the model using facial expression data using a TL-based technique.

Every machine learning system must deal with the difficulty of picking attribute values. In the proposed FER model, only the more densely packed upper layers of the pre-trained Deep CNN are substituted with appropriate layers. After several efforts highlighting the pipeline training issue, hyperparameters such as the number of dense layers, neurons in each layer, and fine-tuning learning parameters were selected. Each parameter of a given DCNN model might be fine-tuned for each dataset (such as CK+, JAFEE, and KDEF) to improve the performance of the proposed method.

6. CONCLUSIONS

An Image Feature Stabilization Algorithm (IFSA) is suggested as an effective DCNN utilising TL and a pipeline tuning strategy for emotion detection from facial images. Experimental results demonstrate that using eight distinct pre-trained DCNN models on well-known KDEF, JAFFE, and CK+ emotion datasets with various profile views, the proposed technique obtains a recognition accuracy of 98.7±1.3 (or 100 percent). Classification accuracy might be improved by fine-tuning the hyperparameters of each pre-trained model and focusing on profile views. The most recent study, notably the performance with profile views, will be suitable for a wider range of real-world commercial applications, such as hospital patient monitoring and security surveillance. In addition, the concept of facial expression recognition might be expanded to include emotion detection via speech or body motions, allowing the creation of new industrial applications.

REFERENCES

[1] Kumar, Y., Verma, S.K., Sharma, S. (2022). Multi-pose facial expression recognition using hybrid deep learning model with improved variant of gravitational search algorithm. Int. Arab J. Inf. Technol., 19(2): 281-287. https://doi.org/10.34028/iajit/19/2/15

[2] Ekman, P. (1973). Cross-Cultural Studies of Facial Expression. In P. Ekman (Ed.), Darwin and Facial Expression: A Century of Research in Review (pp. 169-222). New York: Academic Press.

[3] Avila, A.R., Akhtar, Z., Santos, J.F., O'Shaughnessy, D., Falk, T.H. (2018). Feature pooling of modulation spectrum features for improved speech emotion recognition in the wild. IEEE Transactions on Affective Computing, 12(1): 177-188. https://doi.org/10.1109/TAFFC.2018.2858255

[4] Noroozi, F., Marjanovic, M., Njegus, A., Escalera, S., Anbarjafari, G. (2017). Audio-visual emotion recognition in video clips. IEEE Transactions on Affective Computing, 10(1): 60-75. https://doi.org/10.1109/TAFFC.2017.2713783

[5] Suja, P., Tripathi, S. (2016). Real-time emotion recognition from facial images using Raspberry Pi II. In 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 666-670. https://doi.org/10.1109/SPIN.2016.7566780

[6] Yaddaden, Y., Bouzouane, A., Adda, M., Bouchard, B. (2016). A new approach of facial expression recognition for ambient assisted living. In Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments, pp. 1-8. https://doi.org/10.1145/2910674.2910703

[7] Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J.M., Castillo, J.C., Lozano-Monasor, E., López, M.T., Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. Journal of Biomedical Informatics, 64: 55-73. https://doi.org/10.1016/j.jbi.2016.09.015

[8] Baio, J., Wiggins, L., Christensen, D.L., Maenner, M.J., Daniels, J., Warren, Z., Dowling, N.F. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveillance Summaries, 67(6): 1.

[9] Thonse, U., Behere, R.V., Praharaj, S.K., Sharma, P.S.V.N. (2018). Facial emotion recognition, socio-occupational functioning and expressed emotions in schizophrenia versus bipolar disorder. Psychiatry Research, 264: 354-360. https://doi.org/10.1016/j.psychres.2018.03.027

[10] Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.

(2010). Multi-PIE. Image Vis. Comput., 28: 807-813.

[11] Liew, C.F., Yairi, T. (2015). Facial expression recognition and analysis: a comparison study of feature descriptors. IPSJ Transactions on Computer Vision and Applications, 7: 104-120.

[12] Ko, B.C. (2018). A brief review of facial emotion recognition based on visual information. sensors, 18(2): 401. https://doi.org/10.3390/s18020401

[13] Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Asari, V.K. (2019). A state-of-the-art survey on deep learning theory and architectures. Electronics, 8(3): 292. https://doi.org/10.3390/electronics8030292

[14] Sahu, M., Dash, R. (2021). A survey on deep learning: convolution neural network (CNN). In Intelligent and Cloud Computing, pp. 317-325. https://doi.org/10.1007/978-981-15-6202-0_32

[15] Mollahosseini, A., Chan, D., Mahoor, M.H. (2016). Going deeper in facial expression recognition using deep neural networks. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-10. https://doi.org/10.1109/WACV.2016.7477450

[16] Zhao, X.M., Shi, X.G., Zhang, S.Q. (2015). Facial expression recognition via deep learning. IETE Technical Review, 32(5): 347-355.

[17] Li, J., Huang, S., Zhang, X., Fu, X., Chang, C.C., Tang, Z., Luo, Z. (2018). Facial expression recognition by transfer learning for small datasets. In International Conference on Security with Intelligent Computing and Big-data Services, pp. 756-770. https://doi.org/10.1007/978-3-030-16946-6_62

[18] Bendjillali, R.I., Beladgham, M., Merit, K., Taleb-Ahmed, A. (2019). Improved facial expression recognition based on DWT feature for deep CNN. Electronics, 8(3): 324. https://doi.org/10.3390/electronics8030324

[19] Ngoc, Q.T., Lee, S., Song, B.C. (2020). Facial landmark-based emotion recognition via directed graph neural network. Electronics, 9(5): 764. https://doi.org/10.3390/electronics9050764

[20] Pranav, E., Kamal, S., Chandran, C.S., Supriya, M.H. (2020). Facial emotion recognition using deep convolutional neural network. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 317-320. https://doi.org/10.1109/ICACCS48705.2020.9074302

[21] Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8): 5455-5516. https://doi.org/10.1007/s10462-020-09825-6

[22] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818-2826.

[23] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[24] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251-1258. https://doi.org/10.1109/CVPR.2017.195

[25] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[26] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708. https://doi.org/10.1109/CVPR.2017.243

[27] Alshamsi, H., Kepuska, V., Meng, H. (2017). Real time automated facial expression recognition app development on smart phones. In 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 384-392. https://doi.org/10.1109/IEMCON.2017.8117150

[28] Ruiz-Garcia, A., Elshaw, M., Altahhan, A., Palade, V. (2017). Stacked deep convolutional auto-encoders for emotion recognition from facial expressions. In 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1586-1593. https://doi.org/10.1109/IJCNN.2017.7966040

[29] Rawat, W., Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput, 29: 2352-2449.

[30] Kumari, J., Rajesh, R., Pooja, K.M. (2015). Facial expression recognition: A survey. Procedia Computer Science, 58: 486-491.

[31] Lee, C.C., Shih, C.Y., Lai, W.P., Lin, P.C. (2012). An improved boosting algorithm and its application to facial emotion recognition. Journal of Ambient Intelligence and Humanized Computing, 3(1): 11-17. https://doi.org/10.1007/s12652-011-0085-8

[32] Chang, C.Y., Huang, Y.C. (2010). Personalized facial expression recognition in indoor environments. In The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. https://doi.org/10.1109/IJCNN.2010.5596316

[33] Joseph, A., Geetha, P. (2020). Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. The Visual Computer, 36(3): 529-539. https://doi.org/10.1007/s00371-019-01628-3

[34] Jabid, T., Kabir, M. H., Chae, O. (2010). Robust facial expression recognition based on local directional pattern. ETRI Journal, 32(5): 784-794. https://doi.org/10.4218/etrij.10.1510.0132

[35] Zhi, R., Ruan, Q. (2008). Facial expression recognition based on two-dimensional discriminant locality preserving projections. Neurocomputing, 71(7-9): 1730-1734. https://doi.org/10.1016/j.neucom.2007.12.002

[36] Ding, H., Zhou, S.K., Chellappa, R. (2017). FaceNet2ExpNet: Regularizing a deep face recognition net for expression recognition. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017, pp. 118-126.

[37] Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 115: 101-106. https://doi.org/10.1016/j.patrec.2018.04.010

[38] Shaees, S., Naeem, H., Arslan, M., Naeem, M.R., Ali, S.H., Aldabbas, H. (2020). Facial emotion recognition using transfer learning. In Proceedings of the 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, pp. 1-5.

[39] Liliana, D.Y. (2019). Emotion recognition from facial expression using deep convolutional neural network. In Journal of Physics: Conference Series, 1193(1): 012004. https://doi.org/10.1088/1742-6596/1193/1/012004

[40] Shi, M., Xu, L., Chen, X. (2020). A novel facial expression intelligent recognition method using improved convolutional neural network. IEEE Access, 8: 57606-57614. https://doi.org/10.1109/ACCESS.2020.2982286

[41] Jin, X., Sun, W., Jin, Z. (2020). A discriminative deep association learning for facial expression recognition. International Journal of Machine Learning and Cybernetics, 11(4): 779-793. https://doi.org/10.1007/s13042-019-01024-2

[42] Porcu, S., Floris, A., Atzori, L. (2020). Evaluation of data augmentation techniques for facial expression recognition systems. Electronics, 9(11): 1892. https://doi.org/10.3390/electronics9111892

[43] Pons, G., Masip, D. (2017). Supervised committee of convolutional neural networks in automated facial expression analysis. IEEE Transactions on Affective Computing, 9(3): 343-350. https://doi.org/10.1109/TAFFC.2017.2753235

[44] Wen, G., Hou, Z., Li, H., Li, D., Jiang, L., Xun, E. (2017). Ensemble of deep neural networks with probability-based fusion for facial expression recognition. Cognitive Computation, 9(5): 597-610. https://doi.org/10.1007/s12559-017-9472-6

[45] Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, pp. 512-519. https://doi.org/10.1109/CVPRW.2014.131

[46] Bukar, A.M., Ugail, H. (2017). Automatic age estimation from facial profile view. IET Comput. Vis, 11: 650-655.

[47] Mahendran, A., Vedaldi, A. (2016). Visualizing deep convolutional neural networks using natural pre-images. Int. J. Comput. Vis, 120: 233-255.

[48] Ding, H., Zhou, S.K., Chellappa, R. (2017). FaceNet2ExpNet: Regularizing a deep face recognition net for expression recognition. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 118-126. https://doi.org/10.1109/FG.2017.23.

[49] Akhand, M.A.H., Ahmed, M., Rahman, M.M.H., Islam, M. (2018). Convolutional neural network training incorporating rotation-based generated patterns and handwritten numeral recognition of major Indian scripts. IETE J. Res., 64: 176-194. https://doi.org/10.1080/03772063.2017.1351322

[50] Zeiler, M.D., Fergus, R. (2014). Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision (ECCV 2014), Zurich, Switzerland, pp. 818-833. https://doi.org/10.1007/978-3-319-10590-1_53