A Novel DAAM-DCNNs Hybrid Approach to Facial Expression Recognition to Enhance Learning Experience*

Rayner Alfred[1][0000−0002−3080−3264], Rayner Henry pailus[1] and Joe Henry Obit[1][0000−0002−3900−2228], Yuto Lim[2][0000−0001−7597−5660], and Haviluddin Sukirno[3][0000−0003−0016−1413]

[1] Creative Advanced Machine Intelligence Research Centre, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia
ralfred@ums.edu.my, rayner.pailus@gmail.com, joehenry@ums.edu.my

[2] School of Information Science, Japan Advanced Institute of Science and Technology, Access 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
ylim@jaist.ac.jp

[3] Dept of Informatics, Universitas Mulawarman, Samarinda, Kalimantan Timur, Indonesia
haviluddin@unmul.ac.id

Abstract. Many machine learning models are applied on facial expression classification and there are three main issues affecting the performance of any algorithms in classifying emotions based on facial expressions, and these issues include image illumination, image quality and partial features recognition. Many approaches have been proposed to handle these issues. Unfortunately, one of the main challenges in detecting and classifying facial expression process is minimal differences of features between different types of emotions that can be used to differentiate these different types of emotions. Thus, there is a need to enrich each type of emotion with more relevant extracted features by having a more effective approach to extract features that can be used to represent each type of emotions more effectively and efficiently. This work addresses the issue of improving the emotion recognition accuracy by introducing a novel hybrid approach that combines the Depth Active Appearance Model (DAAM) and Deep Convolutional Neural Networks (DCNNs). The proposed DAAM and DCNNs model can be used to assist one in identifying emotions and classify learner involvement and interest in the topic which are plotted as feedback to the instructor to improve learner experience. The proposed method is evaluated on two publicly available datasets namely, JAFFE and CK+ and the results are compared to the state-of-the-art results. The empirical study showed that the proposed DAMM-CNNs hybrid method managed to perform the face expression recognition with 97.4% for the JAFFE dataset and 96.9% for the CK+ dataset.

* Supported by Universiti Malaysia Sabah under Grant No: SDN0057-2019.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_11

Keywords: Emotions Classification · Facial Expression Recognition · Depth Active Appearance Model · Deep Convolutional Neural Networks · Deep Learning · Feature Extractions

1 Introduction

As early as the twentieth century, Ekman and Friesen[1] defined six basic emotions based on cross-culture study [2], which indicated that humans perceive certain basic emotions in the sameway regardless of culture [3]. These prototypical facial expressions are anger, disgust, fear, happiness, sadness, and surprise. Motivations are closely related to emotions. An emotion is a mental and physiological feeling state that directs our attention and guides our behaviour. There is a need to detect and classify one's emotion in classroom with more effectively and efficiently as emotions affect learning and education. This is due to the fact that students' emotional experiences can impact on their ability to learn, their engagement in school, and their career choices. In fact, facial expression is the most frequently used nonverbal communication mode by the students in the virtual classroom. In addition to that the facial expressions of the students are significantly correlated to their emotions which helps to recognize their comprehension towards the lecture [4].

Emotion detection and classification can be performed based on neurophysiological measurements, behaviour patterns, speech [6] and facial expressions [7]. Electroencephalography (EEG) received considerable attention from researchers, since it can provide a simple, cheap, portable, and ease-to-use solution for identifying emotions [8]. However, this approach requires intervention during the process of learning and it does not provide a seamless approach to detect and recognize emotions. In a classroom, body movements and gestures can also be used to detect and classify students' emotions. Analyzing body movements and gestures also helps in emotion detection with the help of machine learning [9]. The body movements, posture, and gestures change significantly with changes in emotions. This is the reason why we can generally guess a person's basic mood with a combination of his hand/arm gestures and body movements. However, body movements and gestures are seldom used to assess emotion in classroom due to the limitation of body movements and gestures during the process of learning [10]. In speech emotion recognition, several different classifiers and different methods for features extraction can be developed to analyze extracted features that include Mel-frequency cepstrum coefficients (MFCC) and modulation spectral (MS) features [5]. A recurrent neural network (RNN), multivariate linear regression (MLR) and support vector machines (SVM) techniques are widely used in the field of emotion recognition for spoken audio signals.

Facial Recognition is a useful emotion detection technique in which through the identification of facial features using machine learning (e.g., deep learning), features of important facial regions are extracted and analyzed to classify facial expressions [11]. Major facial features used in emotion detection through machine learning and these features include eyes, nose, lips, jaw, eyebrows, mouth (open/close), and more. By having an efficient and effective feature extraction method, more enriched facial representation can be obtained for effective emotion detection through identifying facial features. One of the main challenges in detecting and classifying facial expression process is the limitation in extracting relevant features that can be used to differentiate different types of emotions. In addition to that, insufficient datasets used for training, data bias and inconsistent annotations are very common among different facial expression datasets due to different collecting conditions and the subjectiveness of annotating. In fact, it remains challenging for most researchers to design and implement algorithms to classify facial expressions under different light conditions, poses, and backgrounds and across people of different ages, genders, and ethnicities [12].

Although pure expression recognition based on only visible face images can achieve promising results, incorporating with other models into a high-level framework can provide complementary information and further enhance the robustness although this will add more complexities in designing and analyzing the results produced by multi modals [13]. Nevertheless, the fusion of other modalities, such as infrared images, depth information from 3D face models and physiological data, is becoming a promising research direction due to the large complementarity for facial expressions.

Thus, there is a need to enrich each type of emotion with more relevant extracted features by having a more effective approach to extract features that can be used to represent each type of emotions more effectively and efficiently. This work addresses the issue of improving the emotion recognition accuracy by introducing a novel hybrid approach that combines the Depth Active Appearance Model (DAAM) [15] and Deep Convolutional Neural Networks (DCNNs). DAAM is a well-known statistical model that is used to detect face model by training the image and matching with new image through shape and appearance of the image which efficiency in detecting face position. DAAM has better performance than other methods in eliminations of the influence of different facial region size, head pose and lighting condition and thus can effectively increase the recognition accuracy [16]. DCNNs is well known a deep learning algorithm that is used in computer vision task such as face recognition or object detection. As it has been shown that the method combining Active Appearance Model (AAM) and deep learning achieves significant segmentation accuracy in Prostate segmentation on 3D MR images [17], it is expected that the proposed DAAM-DCNNs hybrid approach to emotion classification is able to perform better than the stand alone CNNs methods. The proposed DAAM and DCNNs model can further be used to assist us in identifying emotions and classify learner involvement and interest in the topic which are plotted as feedback to the instructor to improve learner experience [18]. In this paper, the effects of using different values of the parameters used in the DCNNs are also investigated. These parameters include the number of epochs, the percentages of dropout and validation dataset used in this work.

The remainder of this paper is organized as follows. Section 2 discusses related works in emotion detections and classifications. Section 3 presents the formalization of the novel hybrid approach that combines the Depth Active Appearance Model (DAAM) and Deep Convolutional Neural Networks (DCNNs) in detecting and classifying emotions. Section 4 describes the experimental setup and discusses the obtained results related to emotion detections and classifications. Finally, the paper is concluded in Section 5 by drawing a conclusion and providing some future works.

2 Related Works

CNN has been extensively used in diverse computer vision applications, including Face Expression Recognition [19], increase agriculture productivity [20] and diseases detection [21]. This section presents several related works according to three categories; total solution for face expression recognition using CNN, feature extractions using CNN and finally applying CNN as the main classifiers.

2.1 Complete Solutions for Face Expression Recognition

There are several well-known CNN architectures used in diverse computer vision applications. For face expression recognition, AlexNet [22][23], VGGNet [24][25] and Inception [26] are well-known CNN architectures.

AlexNet contains eight layers with weights; the first five are convolutional and the remaining three are fully connected.[22] In facial expression recognition, AlexNet network can be improved by the aid of Batch Normalization (BN) layer to the existing network.[27] In contrast, VGG16 contains sixteen layers with weights in which the first thirteen are convolutional and the remaining three are fully connected. Meanwhile, the VGG19 contain nineteen layers with weights in which the first sixteen are convolutional and the remaining three are fully connected.[25] With VGGNet deep convolutional neural network, having a deeper network architecture and a 3×3 small convolution kernel and a 2×2 small pool kernel, the recognition rate is significantly improved, and the number of parameters is only slightly larger than that of the shallow layer.[24] Inception V3 has 22 layers with weights deep network, the first twenty one are convolutional and the remaining one is a fully connected.[28] Agrawal and Mittal investigated the effects of CNN parameters namely kernel size and number of filters on the classification accuracy using the FER-2013 dataset.[19]

A Faster R-CNN (Faster Regions with Convolutional Neural Network Features) [29] for facial expression recognition has been proposed that utilizes Region Proposal Networks (RPN) [30]. Firstly, the facial expression image is normalized and the implicit features are extracted by using the trainable convolution kernel. Then, the maximum pooling is used to reduce the dimensions of the extracted implicit features. After that, RPNs [30] is used to generate high-quality region proposals, which are used by Faster R-CNN for detection. Finally, the Softmax classifier and regression layer is used to classify the facial expressions and predict boundary box of the test sample, respectively. It was reported that the mean Average Precision (mAP) is around 0.82.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_11

Facial expression recognition using Efficient Local Binary Pattern (LBP) for feature extraction and General Regression Neural Network (GRNN) [31] for classification has also been presented. GRNN trains the network faster and does not require iterative training procedure. The proposed algorithm with the optimum window sizes of 64×64 improves the recognition rate.

Real-time convolutional neural networks for emotion and gender classification has also been studied [32]. In this work, two models were proposed in which they are evaluated in accordance to their test accuracy and number of parameters. The first model relies on the idea of eliminating completely the fully connected layers. The second architecture combines the deletion of the fully connected layer and the inclusion of the combined depth-wise separable convolutions and residual modules. They reported accuracies of 96% in the IMDB gender dataset and 66% in the FER-2013 emotion dataset.

A Video-based Emotion Recognition Using Deeply-Supervised CNN (DSN) architecture has been proposed to recognition emotion in Video [33]. The proposed DSB takes the multi-level and multi-scale features extracted from different convolutional layers to provide a more advanced representation of emotion recognition. CNN and LSTM based facial expression analysis model for a humanoid robot was also proposed and studied. First, a convolutional neural network (CNN) is used to extract visual features by learning on a large number of static images. Second, a long short-term memory (LSTM) recurrent neural network is used to determine the relationship between the transformation of facial expressions in image sequences and the six basic emotions. Third, CNN and LSTM are combined to exploit their advantages in the proposed model [34].

2.2 Applying Image Pre-Processing Prior to CNNs Classification

A simple solution for facial expression recognition that uses a combination of Convolutional Neural Network and specific image pre-processing steps has also been proposed [35]. In this work, several pre-processing techniques are applied to extract only expression specific features from a face image and explore the presentation order of the samples during training. The proposed method achieved competitive results when compared with other facial expression recognition methods with 96.76% of accuracy in the CK+ database [36].

Automatic facial expression recognition based on a deep convolutional-neural-network structure is also introduced in which a face detection based on Haar-Like Feature is applied before feeding them into the CNNs for facial expression recognition [37]. The best recognition result is obtained when the learning rate, η, is 0.5 in JAFFE [38], and 0.7 in CK+ [36].

A 3D facial expression recognition (FER) algorithm using convolutional neural networks (CNNs) and landmark features and masks was introduced by Huiyuan and Lijun [39], which is invariant to pose and illumination variations due to the solely use of 3D geometric facial models without any texture information. The results show that the CNN model benefits from the masking, and the combination of landmark and CNN features can further improve the 3D FER accuracy

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_11

2.3 Features Extraction Using CNNs Coupled with Other Machine Learning Classifiers

Several works have been conducted in which the deep neural network (particularly a CNN) is used as a feature extraction tool and then apply additional independent classifiers such as Random Forest (RF), Support Vector Machine (SVM) and k Nearest Neighbour (k-NN).

For instance, a novel method for automatically recognizing facial expressions using Deep Convolutional Neural Network(DCNN) features and SVM is proposed [40]. In this work, automatic facial expression recognition using DCNN features is investigated. Two publicly available datasets CK+ [36] and JAFFE [38] are used to carry out the experiment. Pre-processing step involves detecting and cropping the frontal faces using OpenCV2. Then facial features are extracted using the DCNN framework and the facial expression classification is performed using a SVM.

Another facial expression recognition method has also been introduced that makes full use of CNNs to detect face features globally and locally and then combines these global and local generic features for improving accuracy in facial expression recognition using the Support Vector Machine (SVM) classifier.

3 DAAM-DCNNs Hybrid Approach to Facial Expression Recognition

The methodology of the paper consists of three parts (i.e., face detection, face alignment, feature refinement, classification) which is in Fig. 1 described below.

Before applying Active Appearance Model, firstly we collect enough face images with various shapes as training set. The faces are automatically detected using viola Jones algorithm, which is one of the most used algorithms for face detection, the system is trained with face and non-face images, this technique can easily detect single and multiple faces from images and video. Next, an affine transformation (rotation, scaling and translation) is used to normalize the face geometric position [41]. This transformation allows creating perspective distortion. The affine transformation is used for scaling, skewing and rotation. At the end of this process, all faces across an entire dataset will:

1. Be centered in the image
2. Be rotated that such the eyes lie on a horizontal line (i.e., the face is rotated such that the eyes lie along the same y-coordinates)
3. Be scaled such that the size of the faces are approximately identical

Finally, these captured faces are then passed to the DAAM algorithm for feature refinement in order to obtain the mean shape of all the faces in order to construct shape model for facial expression classification. In feature refinement, the Depth Active Appearance Model (DAAM) [42] method is incorporated for both shape and texture information from facial images which has shown strong potential in a variety of facial recognition technologies. AAM allows accurate,

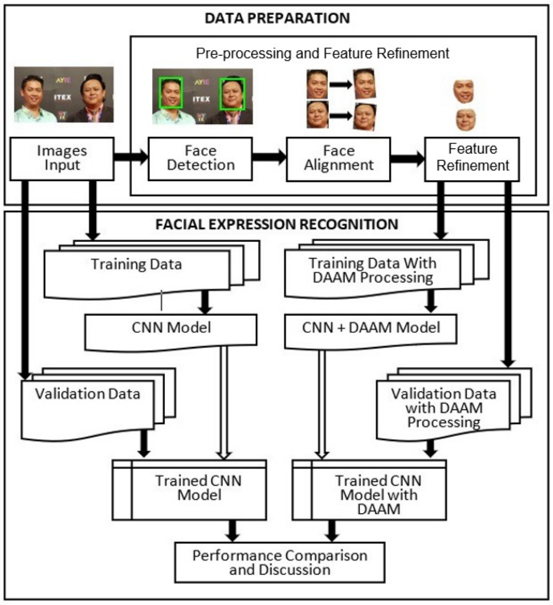
DATA PREPARATION
Pre-processing and Feature Refinement
Images Input
Face Detection
Face Alignment
Feature Refinement
FACIAL EXPRESSION RECOGNITION
Training Data
Training Data With DAAM Processing
CNN Model
CNN + DAAM Model
Validation Data
Validation Data with DAAM Processing
Trained CNN Model
Trained CNN Model with DAAM
Performance Comparison and Discussion

Fig. 1: DAAM-DCNNs Hybrid framework for Facial Expression Recognition system

real-time tracking of human faces in 2D and can be extended to track faces in 3D by constraining its fitting with a linear 3D morphable model. It then was extended to Depth Active Appearance Model (DAAM) [15], where a new constraint is introduced into AAM fitting that uses depth data from a commodity RGBD camera (Kinect). AAD consists of two combined models: Shape and Texture models. The shape model represents the shape of each model image and the texture model warps each image so that its landmark points match the mean shape. The PCA is applied to the normalized values and the model of texture.

3.1 Convolutional Neural Network

Fig. 2 displays the architecture of the CNN model. The CNN architecture comprises 3 layers: two consecutives convolutional-pooling layers and a fully-connected classification layer. The two convolutional pooling layers use the same fixed 5 x 5 convolutional kernel and 2 x 2 pooling kernel, but have 64 and 128 neurons, respectively. The last layer has 256 neurons, which are all connected to the final six neurons for all siz types of facial expression classification.

In this work we apply a deep learning algorithm (e.g., Convolutional Neural Network) in implementing the facial expression recognition. The input image was an RGB image of 224x224 pixels. So, input size = 224x224x3. The **Convolution layer** applies a 2D convolution of the input feature maps and a convolution kernel. The first hidden layer is a convolutional layer that applies 64 filters with size

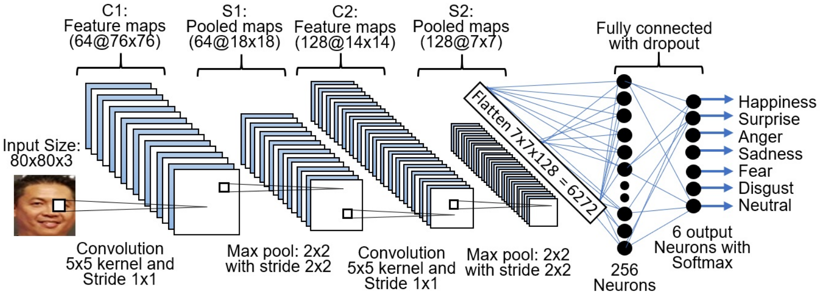

Fig. 2: CNN architecture comprises 3 layers: two consecutives convolutional-pooling layers and a fully-connected classification layer.

5x5 each and stride 1x1. Then, the second hidden layer is the **Pooling layer** that applies a Max pooling function over a spatial window without overlapping (pooling kernel) per each output feature map. This helps to reduce processing time and also helps to reduce overfitting. Each Max Pooling layer has configuration — windows size = 2x2 and stride = 2x2. Thus, we half the size of the image at every Pooling layer. The third hidden layer is a **Convolutional layer** that uses 128 filters with size 5x5 each and stride 1x1. Then the fourth hidden layer will be the **Pooling layer** that applies a Max pooling function. Each Max Pooling layer has configuration — windows size = 2x2 and stride = 2x2. Thus, we half the size of the image at every Pooling layer.

Then next is a **Flatten layer** that converts the 2D matrix data (7x7)x128 kernels to a 1D vector with 6272 parameters before building the fully connected layers. After that we will use a **Fully connected layer** with 17x17x128 = 6272 neurons and *Relu* activation function. Then, we will use a regularization layer called Dropout. It is configured to randomly exclude 20% of neurons in the layer in order to reduce overfitting. Finally, the output layer which has 7 neurons for the 7 classes and a *Softmax* activation function to output probability-like predictions for each class that includes happiness, surprise, anger, sadness, fear, disgust, and neutral.

4 Experimental Setup and Results

In this work, the effects of CNN parameters namely the number of epoch, the percentage of dropout for the fully connected network (see Fig. 2) and the percentages of validation datasets are investigated using the CK+ [43] and JAFFE [44] datasets.

4.1 Datasets

In this work, two publicly available facial expression datasets were used to evaluate the proposed DAAM-DCNN hybrid method, namely CK+ [43] and

JAFFE [44] datasets. CK+ dataset includes 327 video sequences acted out by 118 participants which is labeled with one of these motions: *anger*, *contempt*, *disgust*, *fear*, *happiness*, *sadness* and *surprise*. Each sequence of video consists of approximately 15 to 35 frames, only last frame is used to recognize facial expression. Every sequence starts with the neutral emotion and the last frame depicts the emotion and only last frame is used to recognize facial expression which is for the corresponding label. Japanese Female Facial Expression (JAFFE) [44] consists 213 facial expressions acted by ten subjects. It consists of 30 *anger*, 29 *disgust*, 32 *fear*, 31 *happiness*, 30 *neutral*, 31 *sadness* and 30 *surprise* expressions. All 327 sequences of the CK+dataset and 213 images from JAFFE dataset are used for evaluating the proposed model.

4.2 Investigated Parameters

In this experiment, we vary the number of epochs in order to observe the performance of the proposed DAAM-DCNN hybrid method. We also vary the percentage of dropout and also the percentage number of samples used to validate the performance of the proposed DAAM-DCNN hybrid method. All the values for all the parameters used in this experiment are listed in Table 2.

Table 1: Values of parameters Investigated in this work.

	Parameters		
Data Sets	**Epoch**	**% of Dropout**	**% of Validation Dataset Used**
CK+	15, 30, 50, 100	5, 10, 25, 50	10, 20, 30, 40
JAFFE	15, 30, 50, 100	5, 10, 25, 50	10, 20, 30, 40

In the DAAM-DCNNs method, the initial values for the number of epochs (**E**), the percentage of dropout (**D**) and validation dataset (**V**) are set to 15, 25 and 20 respectively. Then, the best value for **E** is determined by varying its values (e.g., 13, 30, 50 and 100) and at the same time maintaining the values for **D** and **V**. Once, the best value for **E** has been determined, this value is then used to determine the best value for **D** by varying its values (e.g., 5, 10, 25 and 50) and at the same time maintaining the values for **V**. Finally, once the best value for **D** has been determined, these **E**'s and **D**'s values are then used to determine the best value for **V** by varying its values (e.g., 10, 20, 30 and 40) and at the same time maintaining the values for **E** and **D**.

Once all the parameters' values are determined that will produce the optimise recognition rate for the two publicly available datasets, JAFFE and CK+, the same setting will be used to determine the recognition rate for CNN alone without the DAAM preprocessing steps. THe recognition rate obtained will then be compared to the one obtained from the DAAM-DCNNs framework. The results obtained on this work will also be compared to the all results that were published previously [45][46].

4.3 Results and Discussion

Table 2: Comparison of recognition rate with different values of epochs, dropouts and validation percentages.

Dataset	**V (Validation), D (Dropout), E (Epoch)**	**Values**	**Recognition Rate Acc (%)**
CK+	D = 25%, V = 20%, Epoch	15	66.8
		30	75.0
		50	76.5
		100	**77.5**
	E = 100, V = 20%, Dropout (%)	5	**86.3**
		10	85.1
		25	80.2
		50	75.6
	E = 100, D = 5%, Validation (%)	10	87.2
		20	91.3
		30	93.9
		40	**96.9**
JAFFA	D = 25%, V = 20%, Epoch	15	80.7
		30	78.9
		50	83.2
		100	**85.8**
	E = 100, V = 20%, Dropout (%)	5	**92.8**
		10	89.3
		25	83.2
		50	75.1
	E = 100, D = 5%, Validation (%)	10	93.6
		20	92.8
		30	**97.4**
		40	96.4

Table 2 tabulates all the results obtained in work. In this work, the effects of varying the values for the number of epoch, the percentage of dropout for the fully connected network (see Fig. 2) and the percentages of validation datasets are investigated. As shown in Table 2, **V** stands for the percentage of *validation* dataset, **D** stands for the percentages of *dropout* and the **E** stands for the number of *epoch* used in the experiments.

JAFFE Dataset As for the JAFFE dataset, the best recognition rate was 97.4% in which the values for epoch, dropout percentage and percentage of samples used for validation were 100, 5% and 30% respectively for dataset JAFFE. It can be observed that the number of epochs has positive relationship with the recognition rate. The higher is the epoch number, the better is the recognition rate as shown in Table 2. Based on Table 2, the best recognition rate is 85.8% having epoch's, dropout's and validation's values of 100, 25% and 20%. Since epoch of 100 produced highest recognition rate, this epoch's value is maintained and we vary other parameters' values (e.g., dropout's and validation's percentage). Based on the results obtained, a higher recognition rate can be obtained when the dropout percentage value is lowered to 5% as shown in Table 2. It can be observed that, the lower the dropout's value, the better is the recognition rate of the proposed DAAM-DCNN hybrid method. At this point the best recognition rate is 92.8%.

Finally, as the epoch's and dropout's values are maintained as 100 and 5%, the value of the percentage of validation dataset is varied and based on the results, it showed that the best recognition rate can be achieved is 97.4%, when the percentage of validation dataset is set to 30% as shown in Table 2.

CK+ Dataset As for the CK+dataset, the best recognition rate was 97.4% in which the values for epoch, dropout percentage and percentage of samples used for validation were 100, 5% and 30% respectively. When observing the patterns of recognition rate with respect to the number of epochs, percentage of dropout and also the percentage of samples used for validations, similar patterns can be observed in which the number of epoch has positive relationship with the recognition rate. The higher is the epoch number, the better is the recognition rate as shown in Table 2. Based on Table 2, the best recognition rate for CK+ dataset is 77.5% having epoch's, dropout's and validation's values of 100, 25% and 20%.

Similarly, the epoch's value of 100 is maintained in order to produce better recognition rate, other parameters' values are varied (e.g., dropout's and validation's percentage). Based on the results obtained, a higher recognition rate can be obtained when the dropout percentage value is lowered to 5% as shown in Table 2. It also can be observed that, the lower the dropout's value, the better is the recognition rate of the proposed DAAM-DCNN hybrid method. At this point the best recognition rate is 86.3%. Finally, as the epoch's and dropout's values are maintained as 100 and 5%, the value of the percentage of validation dataset is varied and based on the results, it showed that the best recognition rate can be achieved is 96.9%, when the percentage of validation dataset is set to 40% as shown in Table 2. Table 3 shows the recognition rate of the CNN alone without the DAAM preprocessing stage with predefined parameters'values for the number of epochs, the percentages of dropouts and percentages of validation dataset used. Based on Table 3 , the recognition rates obtained for the CK+ and JAFFA datasets are 87.3% and 92.1% respectively. Finally, Table 4 shows the comparison of recognition rate obtained by the proposed DAAM-DCNNS model with state-of-art literature results.

Table 3: Recognition rate using CNN alone with predefined values of epoch, dropout and validation percentages.

	Parameters			
Data Sets	**Epoch**	**Dropout (%)**	**Validation (%)**	**Recognition Rate (%)**
CK+	100	5	40	87.3
JAFFE	100	5	30	92.1

Table 4: Comparison of recognition rate obtained by the proposed DAAM DCNN model with state-of-art literature.

	Recognition Rate	
Data Sets	**Methods**	**Recognition Rate (%)**
CK+	**DCNN**	**87.3**
	DAAM-DCNN	**96.9**
	Shan *et al.*[49](2009)	89.1
	Jeni *et al.*[50](2011)	96.0
	Kahou *et al.*[51](2015)	91.3
JAFFA	**DCNN**	**92.1**
	DAAM-DCNN	**97.4**
	Lyon *et al.*[45](1999)	92.0
	Zhao *et al.*[46](2011)	81.6
	Zhang *et al.*[47](2011)	92.9
	Mlakar *et al.*[48](2015)	87.8

5 Conclusion

Facial expressions convey the emotional state of an individual to the observers. An efficient and effective method to recognize facial expressions has been proposed in this paper. The hybrid approach that combines the DAAM and DCNNs can be used to effectively extract relevant features to be used for facial expression recognition. The proposed method is evaluated on two publicly available datasets name, JAFFE and CK+ and the results are compared to the state-of-the-art results. The empirical study showed that the proposed DAMM-CNNs hybrid method managed to perform the face expression recognition with 97.4% for the JAFFE dataset and 96.9% for the CK+ dataset. The proposed model can be adopted to any generic facial expressions recognition dataset that either involves recognition in static images or video sequences. No retraining or extensive pre-processing techniques are required to adopt the proposed method for facial feature extraction. The proposed DAAM and DCNNs model can be used to assist us in identifying emotions and classify learner involvement and interest in the topic which are plotted as feedback to the instructor to improve learner experience. However, the performance of any CNNs architecture can be tuned by changing the size of the filters, the number of strides, the number of

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_11

convolutional layers, the number of layers for the fully connected network, the number of neurons for each layer and the learning and activation functions used in learning any data. Thus, future work involves exploring ways or approaches used to automatically opitmize the CNNs parameters in order to learn any data presented to the deep learning algorithms.

Acknowledgements This work was supported in part by the Universiti Malaysia Sabah internal grant no. SDN0057-2019 (Biometric Patient Authentication System Using Face Recognition Approach for Smart Hospital).

References

1. Ekman, P., and Friesen, W. V., Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124. (1971)
2. Huang, G., Cui, J., Alam, M., and Wong, K. H., Experimental Analysis of the Facial Expression Recognition of Male and Female. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering (pp. 1-5). (2019, October)
3. Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S., The influences of emotion on learning and memory. Frontiers in psychology, 1454. (2017)
4. Sathik, M., & Jonathan, S. G., Effect of facial expressions on student's comprehension recognition in virtual educational environments. SpringerPlus, 2, 1-9. (2013)
5. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K., Mahjoub, M. A., & Cleder, C., Automatic speech emotion recognition using machine learning. (2019)
6. Sun, L., Zou, B., Fu, S., Chen, J., & Wang, F., Speech emotion recognition based on DNN-decision tree SVM model. Speech Communication, 115, 29-37. (2019)
7. Marechal, C., Mikolajewski, D., Tyburek, K., Prokopowicz, P., Bougueroua, L., Ancourt, C., & Wegrzyn-Wolska, K., Survey on AI-Based Multimodal Methods for Emotion Detection. High-performance modelling and simulation for big data applications, 11400, 307-324. (2019)
8. Alarcao, S. M., & Fonseca, M. J., Emotions recognition using EEG signals: A survey. IEEE Transactions on Affective Computing, 10(3), 374-393. (2017)
9. Shen, Z., Cheng, J., Hu, X., & Dong, Q., Emotion recognition based on multi-view body gestures. In 2019 ieee international conference on image processing (icip) (pp. 3317-3321). IEEE. (2019, September)
10. Nafisi, J. S. A., Gesture and body-movement as teaching and learning tools in western classical singing (Doctoral dissertation, Monash University). (2013)
11. Ko, B.C., A brief review of facial emotion recognition based on visual information. sensors, 18(2), p.401. (2018)
12. Valstar, M.F., Mehu, M., Jiang, B., Pantic, M. and Scherer, K., Meta-analysis of the first facial expression recognition challenge. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(4), pp.966-979. (2012)
13. Ringeval, F., Schuller, B., Valstar, M., Gratch, J., Cowie, R., Scherer, S., Mozgai, S., Cummins, N., Schmitt, M. and Pantic, M., Avec 2017: Real-life depression, and affect recognition workshop and challenge. In Proceedings of the 7th annual workshop on audio/visual emotion challenge (pp. 3-9). (2017, October)

14. Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Cowie, R. and Pantic, M., Summary for AVEC 2016: Depression, mood, and emotion recognition workshop and challenge. In Proceedings of the 24th ACM international conference on Multimedia (pp. 1483-1484). (2016, October)
15. Smolyanskiy, N., Huitema, C., Liang, L. and Anderson, S.E., Real-time 3D face tracking based on active appearance model constrained by depth data. Image and Vision Computing, 32(11), pp.860-869. (2014)
16. Wang, L., Li, R. and Wang, K., A Novel Automatic Facial Expression Recognition Method Based on AAM. J. Comput., 9(3), pp.608-617. (2014)
17. Cheng, R., Roth, H.R., Lu, L., Wang, S., Turkbey, B., Gandler, W., McCreedy, E.S., Agarwal, H.K., Choyke, P., Summers, R.M. and McAuliffe, M.J., Active appearance model and deep learning for more accurate prostate segmentation on MRI. In Medical imaging 2016: Image processing (Vol. 9784, pp. 678-686). SPIE. (2016)
18. Krithika, L.B. and GG, L.P., Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. Procedia Computer Science, 85, pp.767-776. (2016)
19. Agrawal, A. and Mittal, N., Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. The Visual Computer, 36(2), pp.405-412. (2020)
20. Alfred R., Obit J.H., Chin C.P.-Y., Haviluddin H., Lim Y., Towards paddy rice smart farming: A review on big data, machine learning, and rice production tasks, (2021) IEEE Access, 9, art. no. 9389541, pp. 50358 - 50380, DOI: 10.1109/ACCESS.2021.3069449
21. Alfred R., Obit J.H., The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review (2021) Heliyon, 7 (6), art. no. e07371, DOI: 10.1016/j.heliyon.2021.e07371
22. Pedraza, A., Gallego, J., Lopez, S., Gonzalez, L., Laurinavicius, A. and Bueno, G., Glomerulus classification with convolutional neural networks. In Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21 (pp. 839-849). Springer International Publishing. (2017)
23. Krizhevsky, A., Sutskever, I. and Hinton, G.E., Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90. (2017)
24. Jun, H., Shuai, L., Jinming, S., Yue, L., Jingwei, W. and Peng, J., November. Facial expression recognition based on VGGNet convolutional neural network. In 2018 Chinese Automation Congress (CAC) (pp. 4146-4151). IEEE. (2018)
25. Gopalakrishnan, K., Khaitan, S.K., Choudhary, A. and Agrawal, A., Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and building materials, 157, pp.322-330. (2017)
26. Ning, C., Zhou, H., Song, Y. and Tang, J., Inception single shot multibox detector for object detection. In 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 549-554). IEEE. (2017)
27. Chen, X., Yang, X., Wang, M. and Zou, J., Convolution neural network for automatic facial expression recognition. In 2017 International conference on applied system innovation (ICASI) (pp. 814-817). IEEE. (2017)
28. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9). (2015)

29. Li, J., Zhang, D., Zhang, J., Zhang, J., Li, T., Xia, Y., Yan, Q. and Xun, L., Facial expression recognition with faster R-CNN. Procedia Computer Science, 107, pp.135-140. (2017)
30. Ren, S., He, K., Girshick, R. and Sun, J., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28. (2015)
31. Talele, K., Shirsat, A., Uplenchwar, T. and Tuckley, K., Facial expression recognition using general regression neural network. In 2016 IEEE Bombay Section Symposium (IBSS) (pp. 1-6). IEEE. (2016)
32. Arriaga, O., Valdenegro-Toro, M. and Plöger, P., Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557. (2017)
33. Fan, Y., Lam, J.C. and Li, V.O., Video-based emotion recognition using deeply-supervised neural networks. In Proceedings of the 20th ACM international conference on multimodal interaction (pp. 584-588). (2018)
34. Li, T.H.S., Kuo, P.H., Tsai, T.N. and Luan, P.C., CNN and LSTM based facial expression analysis model for a humanoid robot. IEEE Access, 7, pp.93998-94011. (2019)
35. Talele, K., Shirsat, A., Uplenchwar, T. and Tuckley, K., Facial expression recognition using general regression neural network. In 2016 IEEE Bombay Section Symposium (IBSS) (pp. 1-6). IEEE. (2016)
36. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I., 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). IEEE. (2010)
37. Shan, K., Guo, J., You, W., Lu, D. and Bie, R., Automatic facial expression recognition based on a deep convolutional-neural-network structure. In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 123-128). IEEE. (2017)
38. Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J., Coding facial expressions with gabor wavelets. In Proceedings Third IEEE international conference on automatic face and gesture recognition (pp. 200-205). IEEE. (1998)
39. Yang, H. and Yin, L., CNN based 3D facial expression recognition using masking and landmark features. In 2017 seventh international conference on affective computing and intelligent interaction (ACII) (pp. 556-560). IEEE. (2017)
40. Mayya, V., Pai, R.M. and Pai, M.M., Automatic facial expression recognition using DCNN. Procedia Computer Science, 93, pp.453-461. (2016)
41. Jain, A.K., Fundamentals of digital image processing. Prentice-Hall, Inc.. (1989)
42. Edwards, G.J., Cootes, T.F. and Taylor, C.J., 1998. Face recognition using active appearance models. In Computer Vision—ECCV'98: 5th European Conference on Computer Vision Freiburg, Germany, June 2–6, 1998 Proceedings, Volume II 5 (pp. 581-595). Springer Berlin Heidelberg. (1998)
43. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I., 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). IEEE. (2010)
44. Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J., Coding facial expressions with gabor wavelets. In Proceedings Third IEEE international conference on automatic face and gesture recognition (pp. 200-205). IEEE. (1998)

DOI: 10.1007/978-3-031-36027-5_11

45. Lyons, M.J., Budynek, J. and Akamatsu, S., Automatic classification of single facial images. IEEE transactions on pattern analysis and machine intelligence, 21(12), pp.1357-1362. (1999)
46. Zhao, X. and Zhang, S., Facial expression recognition based on local binary patterns and kernel discriminant isomap. Sensors, 11(10), pp.9573-9588. (2011)
47. Zhang, L. and Tjondronegoro, D., 2011. Facial expression recognition using facial movement features. IEEE transactions on affective computing, 2(4), pp.219-229. (2011)
48. Mlakar, U. and Potočnik, B., Automated facial expression recognition based on histograms of oriented gradient feature vector differences. Signal, Image and Video Processing, 9, pp.245-253. (2015)
49. Shan, C., Gong, S. and McOwan, P.W., 2009. Facial expression recognition based on local binary patterns: A comprehensive study. Image and vision Computing, 27(6), pp.803-816. (2009)
50. Jeni, L.A., Takacs, D. and Lorincz, A., High quality facial expression recognition in video streams using shape related information only. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (pp. 2168-2174). IEEE. (2011)
51. Kahou, S.E., Froumenty, P. and Pal, C.J., Facial Expression Analysis Based on High Dimensional Binary Features. In ECCV Workshops (2) (pp. 135-147). (2014)