Journal of Theoretical and Applied Information Technology
28th February 2021. Vol.99. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FACIAL EMOTION RECOGNITION BASED ON DEEP LEARNING TECHNIQUE

[1] ELFATIH ELMUBARAK MUSTAFA , [2]GAFAR ZEN ALABDEEN SALH

[1] Assistant professor, University of Bisha, Balgarn College of Science & Arts, Department of Computer Sciences, Saudi Arabia

[2]Assistant professor, Alneelain University, Faculty of Computer Science and Information Technology Department of Information Technology, Khartoum, Sudan

E-mail: , [1]fatih200041@yahoo.com. ,

[1]eemustafa@ub.edu.sa , [2]gafar2008_gafar2009@yahoo.com

ABSTRACT

We have created a Deep Learning network to identify the emotion of a person. It is based on seven facial expressions. (angry, disgust, sadness, happy, nature, fear, and surprise). We used extended Cohn–Kanade (CK+) database basis of (10-fold cross-validation) to identify 6 facial expressions. The Deep Learning Network scored a recognition average of 88.9%. As you can see in the confusion matrix, the expressions happy and surprised achieved the best recognition rates 98.92 and 97.23 successively. We also used, in another experiment, the JAFFE database basis of (LOOCV) and it scored a recognition average of 88%. As you can see in the confusion matrix that the expressions of fear and surprise achieved the best recognition rates 93.33 and 93.33, respectively. We compared the performance of the proposed system to similar studies that followed the same databases with the same sample and the same style. The system we used outscored other systems in the other studies. We also compared in detail the percentage of identification performance for each expression in isolation using the extended Cohn–Kanade (CK+) database. We compared our study to other studies and we found that our system did better.

Keywords: *Deep Learning ,Facial Emotion Recognition ,Leave-One-Out Cross-Validation (LOOCV), 10-Fold Cross-Validation ,recognition rates , confusion matrix*

1. INTRODUCTION

The facial expression, as one of the most significant means for human beings to show their emotions and intentions in the process of communication, plays a significant role in human interfaces. In recent years, facial expression recognition has been under especially intensive investigation, due conceivably to its vital applications in various fields including human-computer interaction, human-robot interaction, forensic, medical treatment, health-care, virtual reality, and data-driven animation, intelligent tutoring system. The main target of facial expression recognition is to identify the human emotional state (e.g., anger, contempt, disgust, fear, happiness, sadness, and surprise) based on the given facial images [1,2]. According to the Facial Action Coding System (FACS), facial expressions consist of six classes, i.e., happiness, sadness, fear, anger, disgust, and surprise. On the other hand, humans have special facial expressions relating to only themselves. Each of the facial expression groups can be represented by all kinds of facial expressions. Since the face recognition under facial expression variation and facial expression recognition from the images including different facial expressions are still challenging problems and are being done researches for a solution [2].

Many researchers have focused on facial expression recognition and recently researchers attempted to incorporate this task with an estimation of facial expression intensities. Facial expressions are naturally dynamic and they can be segmented into four temporal segments: neutral, onset, apex, and offset [5]. Neutral means no expression is shown, the onset is the instance when the muscular contraction occurs and increases in intensity, the apex is the peak of the expression, and offset is the instance when the expression starts fading away. The dynamics of facial expressions are the crucial information required for interpreting facial behavior [6]. Examples of behavioral research related to facial expressions include the study of emotion, personality, social interaction, communication, anthropology, and child development [7]. Differences in terms of physical facial appearance such as wrinkles and skin texture of different

individuals make facial expression intensity analysis a very challenging problem [3]. In general, facial expression analysis comprises recognition of facial expression, emotions and facial action units [3]. The study is based on a hypothesis that the use of deep learning techniques for the task of recognizing human emotions can improve and increase the performance of the system, and that evaluation using the cross validation method ensures that the tests are optimized.

Building a deep learning system using CNN in order to recognize the human emotions, and evaluate performance based on cross validation, what distinguishes the deep learning network is its ability to automatically extract features. That makes it strong and accurate, and this can contribute to and improve the raising and improvement of the accuracy of the level of recognition as well as providing and creating a powerful and effective mechanism for the mission of recognizing human emotions .Our work focuses on classifying a sequence of facial features in terms of emotions such as anger, disgust, fear, happiness, sadness, and surprise, as well as estimating the intensity of the expression. Research in estimating facial expression intensities is still in the early stage with few publications. Current approaches which perform facial expression recognition.

Deep Learning (DL) is an extended field of Machine Learning (ML) deal with different types of data (such as Images, speech, Text, videos …. etc.) and has a collection of algorithms useful to learn supervised and unsupervised [10] There are several types of deep learning designs, namely deep neural network (DNN), deep belief network (DBN), recurrent neural network (RNN), convolutional neural network (CNN) and convolutional deep belief networks (CDBN) [11], CNN is one of the popular deep learning methods that has been successfully applied in the classification of high dimensional data especially for image [12]. This paper will focus on CNN deep learning algorithms, which are CNN is organized in more convolutional layers with fully connected layers, deep learning network (CNN) that was built and trained, and the Softmax workbook is generally identified as an output layer for the classification problem of this network. This network has the advantage of being able to extract useful features through the network itself automatically. It should be noted that distinguishing between different feelings (the emotional state) is a difficult task regardless of the identity of the face. This is because of the individual differences of the same expression and the difficulty to identify the most accurate and precise among the available expressions. Also, the external factors, such as lighting, environment, and cameras, make it more difficult to identify.

To overcome the above challenges, we have created a deep learning framework that is based on recognizing facial expressions. The purpose of this framework is to classify expressions with high accuracy by learning useful features from the data set. These features are presented through a number of facets to identify the facial expression that contains accurate and comprehensive information about emotions. The network also creates a deep learning framework to recognize facial expressions with high accuracy by learning powerful and discriminatory features from the data set. The samples of facial expressions are selected to distinguish different facial expressions in the extended Cohn – Kanade (CK +) database. This database has been selected because it contains a large number of faces, variations in skin color, as well as variations in the facial features of different people. Our work focused on identifying facial expressions (6 expressions). The 10 -fold cross-validation was used to evaluate the performance of the system. We also discussed another database (i.e. JAFFE database). In this database, we focused on identifying 7 facial expressions and we used Leave-one-out cross-validation to evaluate the performance.

The organization of the rest of this paper is as follows: Section 2 provides a review of related works, while Section 3 presents the Research Method. The proposed approach is described in details in Section 4, also section 5 presents the validation set approach, and we present experimental results and their analysis using extended Cohn–Kanade (CK+) database & JAFFE database in section 6. Finally, the paper presents discussions and conclusion

2. RELATED WORK

Although the effect model based on basic emotions is limited in the ability to represent the complexity and subtlety of our daily affective displays [8, 9], and other emotion description models, such as the Facial Action Coding System (FACS) and the continuous model using affect dimensions, are considered to represent a wider range of emotions, the categorical model that describes emotions in terms of discrete basic emotions is still the most popular perspective for FER, due to its pioneering investigations along with the direct and intuitive definition of facial

expressions. And in this survey, we will limit our discussion on FER based on the categorical model. FER systems can be divided into two main categories according to the feature representations: static image FER and dynamic sequence FER. In static-based methods , the feature representation is encoded with only spatial information from the current single image, whereas dynamics-based methods , consider the temporal relation among contiguous frames in the input facial expression sequence. Based on these two vision-based methods, other modalities, such as audio and physiological channels, have also been used in multimodal systems to assist the recognition of expression.

The majority of the traditional methods have used handcrafted features or shallow learning (e.g., local binary patterns (LBP), LBP on three orthogonal planes (LBP-TOP), non-negative matrix factorization (NMF) and sparse learning) for FER. However, since 2013, emotion recognition competitions such as FER2013 and Emotion Recognition in the Wild (EmotiW) have collected relatively sufficient training data from challenging real-world scenarios, which implicitly promote the transition of FER from lab-controlled to in-the-wild settings. In the meanwhile, due to the dramatically increased chip processing abilities (e.g., GPU units) and well-designed network architecture, studies in various fields have begun to transfer to deep learning methods, which have achieved the state-of-the-art recognition accuracy and exceeded previous results by a large margin. Likewise, given with more effective training data of facial expression, deep learning techniques have increasingly been implemented to handle the challenging factors for emotion recognition in the wild. Figure (1) illustrates this evolution on FER in the aspect of algorithms and datasets. [8]. The early works on facial expression recognition are mostly based on handcrafted features. After the success of the AlexNet deep neural network in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), deep learning has been widely adopted in the computer vision community. Perhaps some of the works to propose deep learning approaches for facial expression recognition were presented at the 2013 Facial Expression Recognition (FER) Challenge. Interestingly, the top-scoring system in the 2013 FER Challenge is a deep convolutional neural network, while the best-handcrafted model ranked only in the fourth place. With only a few exceptions, most of the recent works on facial expression recognition are based on deep learning [. Some of these recent works proposed to train an ensemble of convolutional neural networks for improved performance, while others combined deep features with handcrafted features such as SIFT or Histograms of Oriented Gradients (HOG) [4]. The AlexNet is simply a scaled version of the LeNet with a deeper structure but is trained on a much larger dataset (ImageNet with 14 million images) with a much more powerful computational resource (GPUs). Since then, many novel architectures and efficient learning techniques have been introduced to make CNN's deeper and more powerful, achieving revolutionary performance in a wide range of computer vision applications. The annual ILSVRC event has become an important venue to recognize the performance of new CNN architectures, especially with the participation of technology giants like Google, Microsoft, and Facebook. The depth of the ``winning" CNNs has progressively increased from 8 layers in 2012 to 152 layers in 2015, while the recognition error rate has significantly dropped from 16.4% in 2012 to 3.57% in 2015[13]. This phenomenal progress is illustrated in Figure (1).

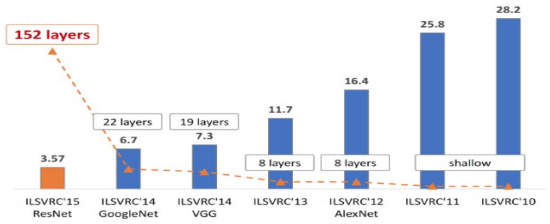

Figure 1. The evolution of the winning entries on the ImageNet Large Scale Visual Recognition Challenge from 2010 to 2015. Since 2012, CNNs have outperformed hand-crafted descriptors and shallow networks by a large margin. Image reprinted with permission [13].

3. RESEARCH METHOD

3.1 Proposed Scheme

The proposed work uses the deep learning convolutional neural network (CNN) for human facial emotion recognition, which to extract features from the input image and identify the human emotional state (e.g., anger, contempt, disgust, fear, happiness, sadness, and surprise) Figure (2).

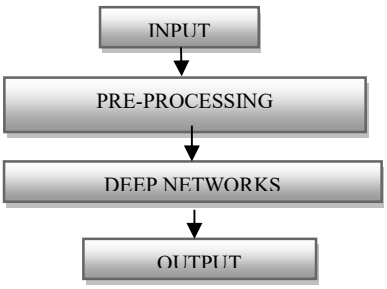

Figure 2. Human Facial Emotion Recognition System Diagram

Considering the dominance of CNNs in the computer vision field and inspired by recent research [13, 14] which has shown that Off-the-Shelf CNN Features work very well for multiple classifications and recognition tasks, we investigate the performance of state-of-the-art CNNs pre-trained on the JAFFE database &, *Cohn-Kanade database* for the facial emotion recognition task. We first review some popular CNN architectures and then present our framework for facial emotion recognition using these CNN Features. We will now analyze each architecture in detail and highlight its notable properties.

A. ALEXNET: ILSVRC 2012 winner: In 2012, Krizhevsky et al. [15] achieved a breakthrough in the large-scale ILSVRC challenge, by utilizing a deep CNN that significantly outperformed other hand-crafted features resulting in a top-5 error rate of 16.4%. AlexNet is a scaled version of the conventional LeNet and takes advantage of a large-scale training dataset (ImageNet) and more computational power (GPUs that allow for 10x speed-up in training). Tuning the hyperparameters of AlexNet was observed to result in better performance, subsequently winning the ILSVCR 2013 challenge [16, 13].

B. VGG: ILSVRC 2014 runner-up: In 2014, Simonyan and Zisserman from Oxford showed that using smaller letters (3 × 3) in the convolutional layer leads to improved performance. The intuition is that multiple small letters in sequence can emulate the effects of larger ones. The simplicity of using small-sized letters throughout the network leads to very good generalization performance. Based on these observations, they introduced a network called VGG which is still widely used today due to its simplicity and good generalization performance. Multiple versions of VGG have been introduced, but the two most popular ones are VGG-16 and VGG-19 that contain 16 and 19 layers, respectively. The detailed architecture of VGG is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (16) and all fully connected layers (2) to generate the CNN Features for the iris recognition task. [13]

C. GoogLeNet AND INCEPTION: ILSVRC 2014 winner: In 2014, Szegedy et al. from Google introduced the Inception v1 architecture that was implemented in the winning ILSVRC 2014 submission called GoogLeNet with a top-5 error rate of 6.7%. The main innovation is the introduction of an inception module, which functions as a small network inside a bigger network. The new insight was the use of (1×1)convolutional blocks to aggregate and reduce the number of features before invoking the expensive parallel blocks. This helps in combining convolutional features in a better way that is not possible by simply stacking more convolutional layers. Later, the authors introduced some improvements in terms of batch normalization and re-designed the later arrangement in the inception module to create Inception v2 and v3. Most recently, they added residual connections to improve the gradient _ows in Inception v4 [49]. The detailed architecture of Inception v3 is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (5) and all inception layers (12) to generate the CNN Features for the iris recognition task. [17,13]

D. RESNET: ILSVRC 2015 winner: In 2015, He et al,. from Microsoft, introduced the notion of residual connection or skip connection which feeds the output of two successive convolutional layers and bypasses the input to the next layer [18]. This residual connection improves the gradient flow in the network, allowing the network to become very deep with 152 layers. This network won the ILSVRC 2015 challenge with a top-5 error rate of 3.57%. The detailed architecture of ResNet-152 is presented in the Appendix. In this paper, we extract the outputs of all convolutional layers (1) and all bottleneck layers (17) to generate the CNN Features for the iris recognition task.

E. DENSENET: In 2016, Huang et al. [19] from Facebook proposed DenseNet, which connects each layer of a CNN to every other layer in a feed-forward fashion. Using densely connected architectures leads to several advantages as pointed out by the authors: “alleviating the vanishing-gradient problem, strengthening feature propagation, encouraging feature reuse,

and substantially reducing the number of parameters."

3.2 CNN Architecture

We used the CNN architecture. The input image will be of size 64*64 and this needs to be flattened to be passed through the convolution layer, so we are reshaping it again to -1*64*64*1. 1 represents the color code as grayscale. We are using 3 convolution layers with The first layer of the CNN is a convolutional layer with filter 32, kernel size 5, stride size of 1, zero paddings and relu activation. Followed by a max-pooling layer of size 5x5 with stride size. The third layer is also a convolutional layer, with filter size 50 and the stride size of 1, zero padding and relu activation. Followed by a max-pooling layer of size 5x5 with stride size, the first six layers are arranged alternately in this pattern, except that the fifth layer is with filter of size 80, stride size of 1, zero paddings and relu activation, also followed by a max-pooling layer of size 5x5 with stride size, if you notice for each layer the filter count is increasing, because the initial layer represents the high-level features of the image and the deeper layers will represent more detailed features and so they usually have number of filters. After three convolution layers, we have one dropout layer and this is to avoid overfitting problem. And once the image passes through the convolution layers, it has to be flattened again to be fed into fully connected layers (it's called a dense layer). We have 2 dense layers and the first one is having 512 neurons and relu activation, this is also arbitrary and we can have the neuron count as per our choice. the second dense layer will have only 7 neurons as we have only seven classes to classify, usually, the number of neurons in the output layer will be equal to the number of classes in our problem. This layer will use softmax activation. softmax activation will calculate the probabilities of each target class over all possible target classes and the sum of all the probabilities will always be 1. The input will be classified into any of the target class based on the higher probability value in softmax. We are using Adam optimizer with "categorical-crossentropy" as loss function and learning rate of 0.001. We train our model for 200 epochs (for every epoch the model will adjust its parameter value to minimize the loss) and the accuracy we got here is around 99%.

Figure 3. CNN architecture from the input image of size 64x64x1 to the final output. The output sizes of the intermediate layers are indicated. There are three convolutional layers, three max-pooling layers, one activation layer, and one softmax layer in our CNN.

Note that the CNN architecture is not unique. However, the parameters of the filters in the convolutional layers and the size of max pooling operators must be consistent to allow meaningful computations. For our datasets, each input image of size 64x64x1 leads to an output of size 1x1x2 after the forwarding propagation of the 9 layers (see Figure 2). The classification error is defined using the 1x1x2 tensor with each component corresponding to the score for the category of cancerous or non-cancerous nodules. According to [16], the batch normalization technique [15] allows much fewer epochs to converge than the dropout technique. Therefore, we applied batch normalization in all our simulation tests.

4. THE VALIDATION SET APPROACH

Suppose that we would like to estimate the test error associated with fitting a particular statistical learning method on a set of observations. The validation set approach, displayed in Figure (4), is a very simple strategy validation for this task. It involves randomly dividing the available set of observer-set approach-into two parts, a training set and a validation set or hold-out set. The validation set, hold-out set model, is fit on the training set, and the fitted model is used to predict the responses for the observations in the validation set. The resulting validation set error rate typically assessed using MSE in the case of a quantitative response provides an estimate of the test error rate. [20]

The example of The validation set approach was used on the Auto data set to estimate the test error that results from predicting mpg using polynomial functions of horsepower. Left: Validation error estimates for a single split into training and validation data sets. Right: The validation method was repeated ten times, each time using a different random split of the observations into a training set and a validation set. This illustrates the variability in the estimated test MSE that results from this approach. [20]

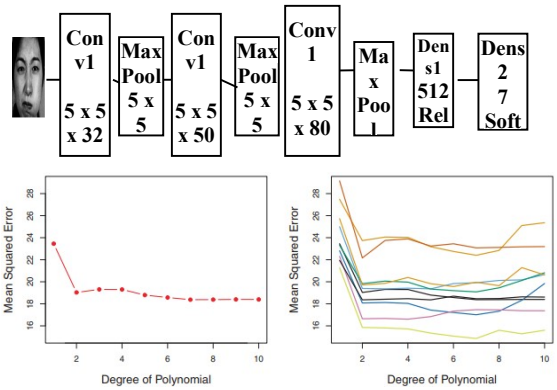

Figure 4.: Left: Validation error estimates for a single split into training and validation data sets. Right: The validation method was repeated ten times [20]

F. Leave-One-Out Cross-Validation:

Like the validation set approach, LOOCV involves splitting the set of observations into two parts. However, instead of creating two subsets of comparable size, a single observation (x_1, y_1) is used for the validation set, and the remaining observations $\{(x_2, y_2),..., (x_n, y_n)\}$ make up the training set. The statistical learning method is fit on the n − 1 training observations, and a prediction $\hat{y}_1$ is made for the excluded observation, using its value x1. Since (x_1, y_1) was not used in the fitting process, $MSE_1 = (y_1 - \hat{y}_1)\ 2$ provides an approximately unbiased estimate for the test error. But even though MSE_1 is unbiased for the test error, it is a poor estimate because it is highly variable, since it is based upon a single observation (x_1, y_1). We can repeat the procedure by selecting (x_2, y_2) for the validation data, training the statistical learning procedure on the n − 1 observations $\{(x_1, y_1), (x_3, y_3),..., (x_n, y_n)\}$, and computing $MSE_2 = (y_2 - \hat{y}_2)2$. Repeating this approach n times produces n squared errors, MSE1,..., MSEn. The LOOCV estimate for the test MSE is the average of these n test error estimates:

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n} MSE_i.$$

A schematic of the LOOCV approach is illustrated in Figure 5.

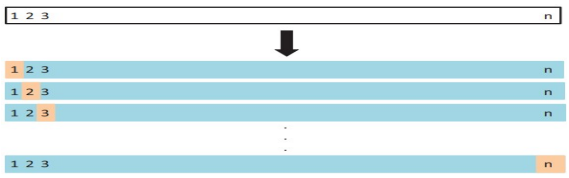

Figure 5: A Schematic Of The LOOCV Approach

G. K-Fold Cross-Validation

An alternative to LOOCV is a k-fold CV. This approach involves randomly k-fold CV dividing the set of observations into k groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining k−1 folds. The mean squared error, MSE_1, is then computed on the observations in the held-out fold. This procedure is repeated k times; each time, a different group of observations is treated as a validation set. This process results in k estimates of the test error, MSE_1, MSE_2,..., MSE_k. The k-fold CV estimate is computed by averaging these values,

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k} MSE_i$$

Figure (6) illustrates the k-fold CV approach. It is not hard to see that LOOCV is a special case of k-fold CV in which k is set to equal n. In practice, one typically performs k-fold CV using k = 5 or k = 10. What is the advantage of using k = 5 or k = 10 rather than k = n? The most obvious advantage is computational. LOOCV requires fitting the statistical learning method n times. This has the potential to be computationally expensive (except for linear models fit by least squares, in which case formula (5.2) can be used). But cross-validation is a very general approach that can be applied to almost any statistical learning method. Some statistical learning methods have computationally intensive fitting procedures, and so performing LOOCV may pose computational problems, especially if n is extremely large. In contrast, performing 10-fold [20]. A schematic of the *k-fold* approach is illustrated in Figure (6):

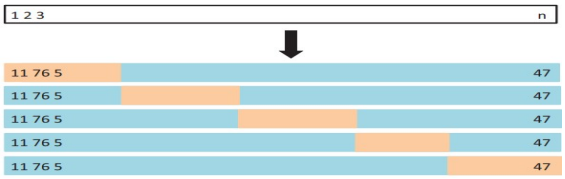

Figure 6 Illustrates The K-Fold CV Approach =5

5. RESULTS AND ANALYSIS

We utilized the extended Cohn–Kanade (CK+) database to evaluate the proposed framework of facial expression recognition in this paper. The CK+ database released in 2010 is the extension of the Cohn–Kanade (CK), which has become one of the most widely used benchmark databases for evaluating the recognition performance of algorithms [21]. Especially, the type of emotion states in CK+ is increased to eight categories and all labels of emotion are amended and validated to improve the performance of the database, meanwhile, the difficulty of expression recognition is increased greatly. The CK+ database consists of

593 image sequences from 123 subjects, including eight basic emotion categories, which are anger, contempt, disgust, fear, happy, sadness, surprise and neutral, as shown in Figure (7). It should be mentioned that the duration of image sequences varies from 10 to 60 frames, with the same criterion beginning at the neutral frame and ending at the peak expression frame. Considering not all of the sequences were labeled as one of seven basic emotions, we only chose those with labels, including 327 sequences [21] [1]. Besides, the neutral frame and four peak frames of each sequence were selected for prototypic expression recognition. Hence, a total of 1218 images were utilized for the experiment, including 178 anger, 233 disgust, 97 fear, 279 happy, 111 sadness, and 320 surprises. For the sake of evaluating the performance of the proposed approach for expression recognition, we employed the (k-fold CV) leave-10-fold-out cross-validation method. More specifically, the data set was separated into 10 subjects with roughly the same number of images, where nine subjects were used as the training dataset and the remaining one was seem as the testing dataset. One thing to note here is that the images of each subject are not duplicated, therefore, the testing data is not included in the training data and even from different individuals. Finally, we reported the mean accuracy of 10 recognition results obtained by treating each subject as the testing set. Specifications of the computer on which the proposed system was implemented. Intel® Core ™ i5-7200U (2.5 GHz base frequency up to 3.1 GHz with Intel® Turbo Boost technology, 3 MB cache and 2 cores).

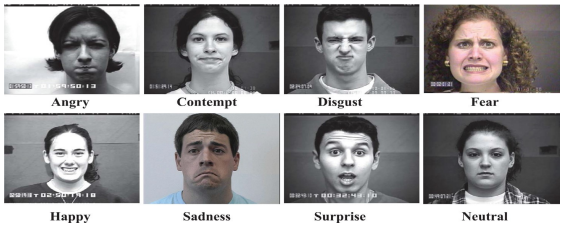

Figure 7. Samples Of Eight Facial Expressions Of The Extended Cohn–Kanade Database. [1]

In this paper, 1,218 images were chosen including one of the six expressions, four images for each expression of 123 subjects. The images are selected from the last frame of each sequence. In the facial expression algorithm, after being detected the face locations in all images by the Viola-Jones algorithm, they were cropped and scaled to 64 x 64 pixels resolution. To evaluate the performance of the proposed facial expression algorithm in this paper, leave-one-subject-out cross-validation method was used [2, 22, 23]. In the leave-10-fold-out evaluation method, the test set was composed of images of one subject, out of the 10 total and the remaining images were used to generate the training set. The training set does not contain any image of the test subject at each trial. Since the test image is not included in the training set, all images in the Cohn-Kanade database (ck+) were separated into 10 sets, and all images of one subject were included in the same set.

Through Table (1) shows that, the average number of training images is 1095, the average number of the testing images is 112.2 and the average of the training accuracy is 0.998, the average of the training loss is 0.0054, the average of the training time is895.11, the average of the validation accuracy is 0.889, the average of the validation loss is 1.174, the average of the validation time is 1.0.42, through the Table (1) shows the recognition ratios for each set. Each set is trained separately, then the test sets of evaluation samples (validation set), and then calculate confusion matrix , which is shown through the Table (2) and through this matrix note the following , Angry expression is confused with disgust 3.93, fear 2.24, sadness 1.68, surprise 1.68 whereas disgust expression is confused with angry 2.72, Happy 1.36, sadness 0.90, surprise 2.27, and fear expression is confused with angry 4.30, disgust 4.30, happy 2.15, sadness 6.45, surprise 8.6 and happy expression is confused with fear 1.07, and sadness expression is confused with angry 3.63, disgust 4.54, fear 4.54, happy 0.90, surprise 6.36. And surprise expression confused with fear 2.46, sadness 0.30 . It is clear that the expressions happy and surprise achieved the best recognition rates 98.92 and 97.23, respectively, compared to other expressions when they were recognized, and they were less confused, and that the expression fear achieved the lowest percentage of recognition 74.19, that is, the algorithm encountered difficulty distinguishing expression Fear compared to getting to know other expressions.

Table 1: Recognition Rates For Different Classifiers Of The Proposed Algorithm On The Extended Cohn–Kanade (Ck+) Database

	Tr-img	Val-img	Tr-acc	Tr-loss	Tr-time	Val-acc	Val-loss	Val-time
Set1	1082	136	1.00	0.0013	929.31	0.91	0.72	0.75
Set2	1137	81	0.998	0.008	888.40	0.89	1.782	0.22
Set3	1150	68	0.999	0.0053	898.05	0.947	0.624	0.192
Set4	1086	132	0.999	0.0024	871.06	0.875	1.060	0.372
Set5	1098	120	0.994	0.0131	875.29	0.838	2.083	0.347
Set6	1115	103	0.995	0.0077	879.66	0.901	1.254	0.285
Set7	1099	119	0.997	0.0069	871.79	0.937	0.974	0.336
Set8	1097	121	0.999	0.0039	874.37	0.923	0.778	0.342
Set9	1066	152	1.00	0.0017	862.11	0.869	1.005	0.430
Set10	1028	190	0.999	0.0041	1001.1	0.799	1.465	1.019
Mean	**1095**	**122**	**0.998**	**0.0054**	**895.11**	**0.889**	**1.174**	**0.429**

Table 2: Recognition Rates Of The Proposed Algorithm For The Extended Cohn–Kanade (CK+) Database

	An	Di	Fe	Ha	Sa	Su
An	**90.44**	3.93	2.24	0	1.68	1.68
Di	2.72	**92.72**	0	1.36	0.90	2.27
Fe	4.30	4.30	**74.19**	2.15	6.45	8.60
Ha	0	0	1.07	**98.92**	0	0
Sa	3.63	4.54	4.54	0.90	**80**	6.36
Su	0	0	2.46	0	0.30	**97.23**

In the second experiment, the proposed algorithm is evaluated on the JAFFE database. The JAFFE database includes 213 grayscale facial expression images relating to 10 Japanese females. Seven different facial expressions are consisting of happy, sadness, fear, surprise, angry, disgust and neutral. Each subject has two to four different images with a resolution of 256 x 256 pixels for each expression. Figure (8) shows some images comprising seven basic facial expressions from the JAFFE database. In this paper, the images of JAFFE are posed as a seven-class classification problem [2] . We took 3 images from each expression of 7 expressions of its own, so the number of images taken from one person is 21 images. The work performance of the identification algorithm was evaluated according to the leave - one - out (LOOCV) method, validation sets were created and established by transferring part of the training samples. This validation set is used for testing. after the proposed deep learning network has completed the training of samples for this training set. We have relied on the idea of excluding the same facial expression for each of 10 people out of 21 facial expressions to represent a validation set witch is for testing, the number of images for each validation set is 10 images, and the number of all validation sets is 21. The number of training sets becomes 21 training, corresponding to 21 validation sets. Table (3) shows data for 21 sets, as well as training and test results for each set.

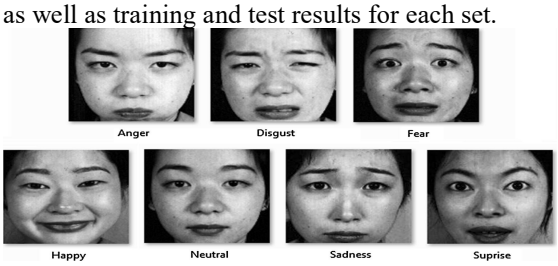

Figure 8 Facial expression samples of the JAFFE database [2]

The average number of training images is 200, the average number of the testing images is 10 and the average of the training accuracy is 0.99, the average of the training loss is 0.0086, the average of the training time is 335.78, the average of the validation accuracy is 0.88 , the average of the validation loss is 0.327 , the average of the validation time is 0.0318, through the Table (3)

Table 3: Recognition Rates For Different Classifiers Of The Proposed Algorithm On The JAFFE Database

	Tr-accuracy	Tr-loss	Tr-time	Val-accuracy	Val-loss	Val-time
Set1	0.995	0.0082	319.36	0.80	0.024	0.018
Set2	0.994	0.0132	307.93	1.0	0.0177	0.015
Set3	1.0	0.0052	313.47	0.90	0.020	0.015
Set4	1.0	0.0098	323.99	0.70	1.116	0.031
Set5	1.0	0.0102	451.74	1.0	0.0711	0.027
Set6	0.990	0.0204	432	0.80	0.895	0.064
Set7	1.0	0.0070	319	0.9	0.57	0.048
Set8	1.0	0.0062	315	1.0	0.072	0.031
Set9	0.995	0.0139	313	0.90	0.189	0.031
Set10	1.0	0.0015	311	1.0	0.0020	0.031
Set11	0.995	0.0076	330	0.60	1.303	0.031
Set12	1.0	0.0066	313	1.0	0.114	0.015
Set13	1.0	0.0104	358	0.70	1.64	0.028
Set14	1.0	0.0037	340	1.0	0.0070	0.034
Set15	0.995	0.0197	336	1.0	0.052	0.044
Set16	1.0	0.0078	336	0.70	0.0163	0.031
Set17	1.0	0.0073	327	0.9	0.0445	0.032
Set18	1.0	0.0069	349	0.80	0.0164	0.050
Set19	1.0	0.0034	328	0.90	0.50	0.031
Set20	1.0	0.0047	316	1.0	0.0207	0.031
Set21	1.0	0.0084	312	0.90	0.183	0.031
Mean	**0.99**	**0.0086**	**335.78**	**0.88**	**0.327**	**0.0318**

shows the recognition ratios for each set. Each set is trained separately, then the test sets of evaluation samples (validation set), and then calculate confusion matrix, which is shown through the Table (4) and through this matrix note the following , Angry expression is confused with disgust 6.66, happy 3.33, Neutral 3.33, whereas disgust expression is confused with angry 6.66, fear, 6.66 Happy 3.33, and fear expression is confused with disgust 3.33, sadness 3.33, happy expression is confused with neutral 13.33, sadness expression is confused with angry 3.33, disgust 3.33, happy 6.66, neutral 6.66. And surprise expression confused with fear 1.0, sadness 1.0.

Table 4: Recognition Rates For Different Classifiers Of The Proposed Algorithm On The JAFFE Database

	An	*Di*	*Fe*	*Ha*	*Ne*	*Sa*	*Su*
An	***90.0***	*6.66*	*0*	*3.33*	*3.33*	*0*	*0*
Di	*6.66*	***83.33***	*6.66*	*3.33*	*0*	*0*	*0*
Fe	*0*	*3.33*	***93.33***	*0*	*0*	*3.33*	*0*
Ha	*0*	*0*	*0*	***86.66***	*13.33*	*0*	*0*
Ne	*0*	*0*	*3.33*	*0*	***90.0***	*6.33*	*0*
Sa	*3.33*	*3.33*	*0*	*6.66*	*6.66*	***80.0***	*0*
Su	*0*	*0*	*3.33*	*0*	*0*	*3.33*	***93.33***

It is clear that the expressions fear and surprise achieved the best recognition rates 93.33 and 93.33, respectively, compared to other expressions when they were recognized, and they were less confused, and that the expression sadness achieved the lowest percentage of recognition 80.0, that is, the algorithm encountered difficulty distinguishing expression sadness compared to getting to know other expressions, and it also getting highest confused with the other expressions.

In the third experiment, we made a performance comparison on a group of previous studies in the same field according to JAFFE database shown in Table (5), Figure 9. This shows that the algorithm performance achieved 88%, which is better than the concerned previous studies' ratios. This indicates the optimum of the present system with noticeable quality and efficiency on the ideal training sets samples, which had a great impact on reaching this good percentage. We also compared the performance of our proposed methods, with the previous studies in the same field according to the extended Cohn–Kanade (CK+) database shown in Table (6), Figure 10, Figure 11. This shows that the algorithm performance achieved 88.9%.

Table 5: Comparing The Proposed Methods With Different State-Of-The-Art Methods On JAFFE Database/ 7 Class

COMPARISON METHODS	RECOGNITION RATE (%)
KYPEROUNTAS [28]	85.92
ZHENGDONG [29]	65.77
HORIKAWA [30]	67.00
BIN J, [31]	79.30
WONG J, [32]	83.84
PROPOSED	88.0

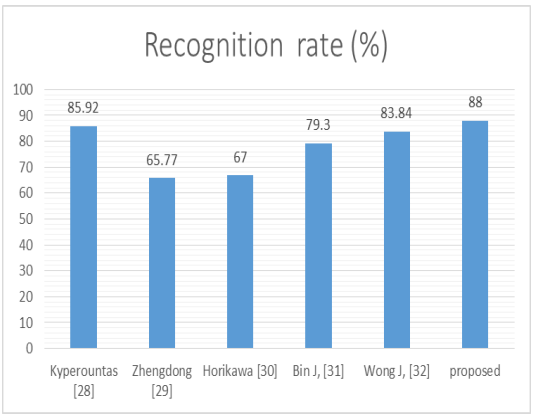

Figure 9: Comparing The Proposed Methods With Different State-Of-The-Art Methods On JAFFE Database / 7 Class

TABLE 6: Comparing The Proposed Methods With Different State-Of-The-Art Methods On Extended Cohn–Kanade (Ck+) Database/ 6 Class

	An	*Di*	*Fe*	*Ha*	*Sa*	*Su*	*Mean*
Suk.M. [24]	77.8	83.1	72	92.8	67.9	96.4	81.7
HCRF [25]	51.7	0.00	35.3	90.2	25.0	95.6	49.6
HCORF[26]	69.0	61.3	41.2	95.1	37.5	91.2	65.9
VSLSRF[27]	69.0	41.9	35.3	55.7	43.8	82.4	54.7
Siti /method1[3]	82.8	67.7	64.7	98.4	12.5	95.6	70.3
Siti/method2 [3]	82.8	58.1	41.2	98.4	18.8	89.7	64.8
proposed	90.4	92.7	74.1	98.9	80	97.2	**88.9**

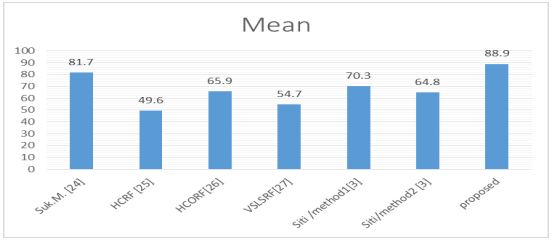

Figure 10: Comparing The Proposed Methods With Different State-Of-The-Art Methods On Extended Cohn–Kanade (Ck+) Database/ 6 Class

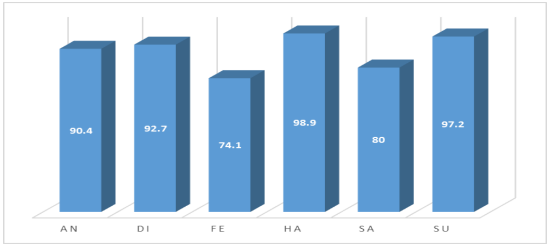

Figure 11: Comparing The Stats (Anger, Disgust, Fear, Happiness, Sadness, Surprise)For Proposed Methods On Extended Cohn–Kanade (Ck+) Database/ 6 Class

Based on the previous results, we notice that our system has tested its performance on two standard databases, and achieved a preference of 88.9% over the extended cohn–kanade (ck +) database/6 class compared to similar studies [24] [25] [26] [27][3].

It and also a preference of 88.0% over the JAFFE database / 7 class compared to [28] [29] [30] [31] [32].

6. DISCUSSIONS AND CONCLUSIONS

Classifying the emotional condition of a person is very important in many fields in our daily life. Emotion affects our lives so it is an aspect of our feeling. Identifying the emotion can be used in many fields in our life, such as education (identifying learner's feelings), industry (employee's feeling), commerce (neuro-marketing area), can manufacturing (controlling the aggressiveness of drivers) and so on. In this paper, we have discussed the importance of identifying feelings through facial expressions by using any one of the methods of deep learning. Our experiments showed that the features of CNN can be specified to identify the emotional feeling through the face. Using the comparative performance, we created CNN networks and apply them to identify the emotional feeling. We found that the precise of identification in our system was superior to the systems in the other studies using two databases (i.e. extended Cohn–Kanade (CK+) database and JAFFE database). The results indicated that CNN features can be used to identify the feeling case. Then elicits the visible features effectively in faces images and this will help in doing the difficult duties. Using CNN networks in eliciting the automatic features is an important matter and it can be used in many huge applications. We notice that many challenges while dealing with deep learning networks. We should be aware of when using complex statistical. We need strong GPUs to achieve the training. Deep learning enables the network to modify itself and create better cases for the problem under investigation. We recommend using the style of cross-validation in the evaluation process of the system, as we did in this paper. It includes many factors such as using all the samples in the training and identification processes. Most of the studies in the field of identification follow this method.

REFERENCES:

[1] Nianyin Zeng a, Hong Zhang a, Baoye Song b, ∗, Weibo Liu c, Yurong Li d, e, Abdullah M. Dobaie f, Facial expression recognition via learning deep sparse autoencoders, http://dx.doi.org/10.1016/j.neucom.2017.08.043 0925-2312/© 2017 Elsevier B.V. All rights reserved.

[2] Ays¸egu¨l Uc¸ar • Yakup Demir • Cu¨neyt Gu¨zelis¸ A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering, _ Springer-Verlag London, Neural Comput & Applic (2016) 27:131–142 DOI 10.1007/s00521-014-1569-1.

[3] Siti Khairuni Amalina Kamarol, Mohamed Hisham Jawad, Heikki K¬ alvi ¬ainen,Jussi Parkkinen, Rajendran Parthiban, Joint Facial Expression Recognition and Intensity Estimation Based on Weighted Votes of Image Sequences, Pattern Recognition Letters (2017), doi: 10.1016/j.patrec.2017.04.003.

[4] MARIANA-IULIANA GEORGESCU1,2, RADU TUDOR IONESCU 1,3, (Member, IEEE), AND MARIUS POPESCU1,3, Local Learning With Deep and Handcrafted Features for Facial Expression Recognition, 2169-3536 2019 IEEE.

[5] T. Fang, X. Zhao, O. Ocegueda, S. K. Shah, I. A. Kakadiaris, 3D facial expression recognition: A perspective on promises and challenges, in IEEE Int. Conf. Automat. Face and Gesture Recognition and Workshops,2011, pp. 603–610.

[6] O. Rudovic, V. Pavlovic, M. Pantic, Kernel conditional ordinal random fields for temporal segmentation of facial action units, in: Proc. 12th Eur. Conf. Comput. Vision, 2012, pp. 260–269.

[7] M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski, Measuring facial expressions by computer image analysis, Psychophysiology 36 (2) (1999) 253–263.

[8] Shan Li and Weihong Deng, Member, IEEE, Deep Facial Expression Recognition: A Survey arXiv:1804.08348v2 [cs.CV] 22 Oct 2018.

[9] B. Martinez and M. F. Valstar, "Advances, challenges, and opportunities in automatic facial expression recognition," in Advances in Face Detection and Facial Image Analysis. Springer, 2016, pp. 63–100.

[10] Mohamed Galal, Magda M. Madbouly, Adel El-Zoghby "Classifying Arabic Text Using Deep Learning " Journal Of Theoretical And Applied Information Technology 15th December 2019. Vol.97. No 23 (JATIT) , Issn: 1992-8645 , E-Issn: 1817-3195 pp. 3412- 3422 .

[11] Amri, A'inur A'fifah and Ismail, Amelia Ritahani and Zarir, Abdullah Ahmad (2017) Convolutional neural networks and deep belief networks for analyzing the imbalanced class issue in handwritten dataset. International Journal on Advanced Science, Engineering and Information Technology, 7 (6). pp. 2302-2307. ISSN 2088-5334 E-ISSN 2460-6952.

[12] Jingwei Too, A. R. Abdullah, N Mohd Saad, N Mohd Ali, T. N. S. Tengku Zawawi, Featureless EMG pattern recognition based on convolutional neural network, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 14, No. 3, June 2019, pp. 1291~1297, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v14.i3.pp1291-1297

[13] Kien Nguyen 1, (Member, IEEE), Clinton Fookes1, (Senior Member, IEEE), Arun Ross2, (Senior Member, IEEE), And Sridha Sridharan1, (Senior Member, IEEE), Iris Recognition With Off-The-Shelf Cnn Features A Deep Learning Perspective, Digital Object Identifier 10.1109/Access.2017.2784352, 2169-3536 2017 IEEE.

[14] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, ``CNN features off-the-shelf: An astounding baseline for recognition," in Proc. IEEE CVPR, May 2014, pp. 512_519.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ``ImageNet classi_cation with deep convolutional neural networks," in Proc. Adv. Neural Inf. Pro-cess. Syst., 2012, pp. 1097_1105.

[16] M. D. Zeiler and R. Fergus, ``Visualizing and understanding convolutional networks," in Proc. Eur. Conf. Comput. Vis. (ECCV), Aug. 2014, pp. 818_833.

[17] C. Szegedy et al., ``Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1_9.

[18] K. He, X. Zhang, S. Ren, and J. SUN, ``Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770_778.

[19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, ``Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2261_2269.

[20] Gareth James • Daniela Witten • Trevor Hastie Robert Tibshirani , statistical learning with applications in r an introduction to , with applications in r, ISSN 1431-875x ,ISBN 978-1-4614-7137-0 ISBN 978-1-4614-7138-7 (ebook), doi 10.1007/978-1-4614-7138-7 ,springer new york Heidelberg Dordrecht London, library of congress control number: 2013936251,© springer science+business media new york 2013 (corrected at 6th printing 2015)

[21] P. Lucey, J. Cohn, T. Kanade, J. Saragih, 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, 2010, pp. 94–101.

[22] zhang l, tjondronegoro d (2011) facial expression recognition using facial movement features. ieee trans affect compt 2(4):219–229

[23] kyperountas m, tefas a, pitas i (2010) salient feature and reliable classifier selection for facial expression classification. pattern recognition 43(3):972–986 .

[24] suk m.h., prabhakaran b. real-time mobile facial expression recognition system—a case study; proceedings of the ieee conference of computer vision and pattern recognition workshops (cvprw); columbus, oh, usa. 28 June 2014; pp. 132–137.

[25] s. b. wang, a. quattoni, l.-p. morency, d. Demirjian, t. Darrell, hidden conditional random fields for gesture recognition, in: ieee conf. comput. vision and pattern recognition, vol. 2, 2006, pp. 1521–1527.

[26] m. kim, v. pavlovic, hidden conditional ordinal random fields for sequence classification, in: machine learning and knowledge discovery in databases, Springer, 2010, pp. 51–65.

[27] r.walecki, o. rudovic, v. pavlovic, m. pantic, variable-state latent conditional random fields for facial expression recognition and action unit detection, in: ieee int. conf. automat. face and gesture recognition and workshops, 2015, pp. 1–8.

[28] kyperountas m, tefas a, pitas i (2010) salient feature and reliable classifier selection for facial expression classification. pattern recognit 43(3):972–986

[29] zhengdong c, bin s, xiang f, yu-jin z (2008) automatic coefficient selection in weighted maximum margin criterion. in: 19th international conference on pattern recognition. tampa, fl, pp 1–4

[30] horikawa y (2007) facial expression recognition using kcca with combining correlation kernels and kansei information.,international conference on computational science and its applications. kuala Lumpur, malaysian, pp 489–498

[31] bin j, guo-sheng y, huan-long z (2008) comparative study of dimension reduction and recognition algorithms of dct and 2dpca. in: international conference on machine learning and cybernetics. kunming, china, pp 407–410

[32] wong j, cho sa (2010) face emotion tree structure representation with probabilistic recursive neural network modeling. neural comput appl 19(1):33–54