Facial Expression Recognition using Convolution Neural Network Enhancing with Pre-processing Stages

Faten Khemakhem[1] and Hela Ltifi[12]
Email: faten.khemakhem@ieee.org, hela.ltifi@ieee.org

***Abstract*—Recognizing human expression is one of the most popular problems in the Human-Computer Interaction field. Facial Expression Recognition present a great challenge in a wide variety of areas due to varying conditions of the image, which influences expression recognition and makes this task a complex problem. The main difficulties depend on the irregular nature of the human face and the different conditions such as orientation, light and shadows. Lately, Deep learning obtained more attention as an intelligent technology to achieve robustness and offer best performance of expression recognition. Further investigations are still needed in this field in order to make the recognition process very efficient. For that, we present in this paper a new Convolutional Neural Networks model enhancing with pre-processing stages to recognize seven classes (six basic expressions and one neutral). Our approach contains two phases: normalization, and expression recognition. The result can achieve high accuracy compared to recent works with the popular facial expression databases such as CK+, JAFFE, and FER-2013.**

***Index Terms*—Expression Classification, Facial Expression Recognition, Convolutional Neural Networks, Pre-processing Stages**

I. Introduction

Current Human-Computer Interaction (HCI) applications have even more to gain the full social skills needful for robust and rich emotional interaction with human beings [1]. Modeling such interaction becomes a more challenging research topic. Facial expression is considered as a significant gesture of social interaction. Is one of the most significant non-verbal behaviors through which HCI systems may recognize an internal or affective state of human emotions. Ekman and Friesen [2] showed six basic emotional expressions, viz. anger, disgust, fear, happy, sadness, and surprise, and one neutral that are universal among human beings. These classes are the most well-known expressions since they are all perceived in similar manner, regardless of their culture.

Due to the particular importance of human expression recognition in the design of Human-Robot Interaction (HRI) and HCI systems [3], various annotated face databases have been constructed with either human actor represent basic expressions such as CK+ [4] and JAFFE [5] or faces taken spontaneously in an uncontrolled environment as FER-2013 [6]. Also, using those databases, numerous Machine-Learning (ML) algorithms have been claimed state-of-the-art performances in automated Facial Expression Recognition (FER) and efforts are constantly being taken to increase accuracy of the choosing classifier. Methods based on 2-D videos have been widely studied [7], [8], [9]. However, current FER approaches aim to classify a single face in a given image as one expression. The special terminology playing an important role in FER research. Some studies considered Action Units (AUs) [10], facial landmarks [11] or the combination in the feature extraction.

Although, traditional ML applied approaches, based on handcrafted[1] features present interesting results in classifying images captured in a supervised environment. Study of different research approaches can be found in [12]. Current studies have exhibited that these methods are not flexible to classify images captured in an uncontrolled environment and spontaneous manner.

The weakness of these algorithms is primarily caused by the fact that these systems are only capable to recognize limited or exaggerated expressions corresponding to exist in the training data. Some FER datasets have captured these frames in frontal pose and illumination well adjusted. In addition, capturing some training data is a little difficult, particularly for fear and sadness emotions, which are extremely difficult to precisely replicate and do not frequently occur in real life. Also, the form of expression varies from person to another and a person can feel more than one emotion at a time.

Handcrafted algorithms are computationally intensive due to high dimensions and are generally not robust. The selective power is usually low [13]. Unlike those approaches that use the handcrafted features, Deep Learning (DL) becomes a mainstream technique in computer visions [14]. Convolutional Neural Network (CNN) is a well-known DL architecture inspired by the natural visual perception mechanism of the living creatures [15].

The main commitment of this work is to enhance the CNN architecture with a pre-processing stage for designing FER system. Our approach is evaluated on three real-world facial expression databases (e.g. CK+, JAFEE and FER-2013). We confirm the pre-processing's potential by performing this phase with other deep models. We conduct comprehensive experiments to obtain significant results relating to traditional CNNs or recent state-of-the-art released works in terms of accuracy.

The proposed model proves the moderate improvement to predict basic expressions from images using CNN. The remainder of this paper is structured as follows: section 2 delivers the most recent related works in what we follow, while the next section describes the proposed network. In section 4, we

[1] REsearch Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia.

[2] Department of Computer Science, Faculty of Sciences and Techniques of Sidi Bouzid, University of Kairouan, Tunisia.

[1] Handcrafted allows to extract features manually based on human expertise.

978-1-7281-5052-9/19/$31.00 ©2019 IEEE

have presented experimental results and evaluated our model comparing to recent FER methods. Finally, we have concluded our observation in Section 5.

II. Related work

Current research studies in this field have focused on recognizing human expression in a single image or on video footage. Recent research works are localized on detecting, understanding, describing, and identifying features that can be extracted from images. A big number of specific features have been manually described or handcrafted with an eye for overcoming specific problems like variations in scale, occlusions and illumination [13].

Nevertheless, most FER works did not employ deep learning to recognize expression from images. Generally, those systems include three primary phases, viz. pre-processing, features extraction, and classification. In the first phase, there are four basic steps: face detection, cropping, resize and normalization. Face detection aims to remove non-face areas and background, and then crop the face area. Next pre-processing is down-sampling the resolution of all images to a single dimension [16]. The fourth is normalization in which mechanisms or processes are applied to return the image to normal condition. In the next phase, features are extracted from images data to mark changes in facial appearance caused by a specific expression. Most of the current works employ several human-crafted features including Haar features [16], Gabor wavelet coefficients [17], [18], histograms of Local Binary Patterns (LBP) [19], [20], [21], scale-invariant feature transform (SIFT) descriptors [22], Histograms of Oriented Gradients (HOG) [22], [23], [24] and 3D shape parameters [25].

In the final phase, the extracted feature sets and labels are fed into a specified classifier to train of the test data given to recognize a facial expression. In summary, most of the aforementioned approaches carry out the three phases individually and sequentially, except for a few combining the two last steps. Unlike traditional approaches based on handcrafted features, Deep Learning (DL) has been widely used in many computer vision studies, yielding state-of-the-art performance in various areas. DL is one of the reliable approaches that can be employed in facial expression analysis and emotion recognition. Various deep neural network architectures are employed to support the expression recognition problem in image analysis. For instance, Liu et al. [26] developed a Boosted Deep Belief Network (BDBN) framework for facial expression recognition. In order to characterize facial appearance changes, a training process is performed in three stages in a loopy process: (1) feature learning, (2) feature selection, and (3) classifier construction, in a unique framework. Each classifier is in charge of detecting one expression. The accuracy attained by the authors for seven expressions in the JAFFE dataset was 68% and in the CK+ dataset was 91.8%.

Mengyi et al. [27] present a 3D Convolutional Neural Networks (3D CNN) with seven successive layers and deformable action parts constraints. Specifically, their method detects different facial action parts and obtains the facial maps. This method evaluated on three datasets, CK+, MMI [28] and FERA [29], and achieved an accuracy of 92.4%, 63.4% and 56.1%, respectively.

Another FER-based approach, proposed by Andre et al. [30], used a combination of CNN and image pre-processing stages. The proposed architecture CNN is comprised of five layers: 2 convolutional layers, 2 sub-sampling layers and one fully connected layer. One classifier is considered to classify 6 and 7 learned expressions. Their experiments have been implemented using three public databases in the FER research field: CK+, JAFFE and the BU-3DFE [31] database. In this work, the authors obtained an accuracy of 95.79%, 53.57% and 71.62% respectively for seven expressions.

Recently, Ali et al. [32] implemented a new deep neural network architecture for FER problem. Specifically, their network consists of two convolutional layers each followed by max pooling and then four Inception layers. In this work, the network is a single component architecture to classify seven expressions. Their architecture produced an accuracy of 66.4% on the FER 2013 database and achieved 93,20% on CK+ for subject-independent. To evaluate the generalization ability, the authors performed a cross-database validation that reaches 64.2% on the CK+ and 34% on the FER-2013.

However, most of the previous studies do not take into account the appropriate network parameters. Adding Convolution layers leads to an important increase of network parameters. Therefore, it increases the cost of training time and space. For this reason, we aim by our proposal to achieve optimal performance improving the performance of simple CNN with pre-processing steps, parameter initialization, activation function and training settings of our model.

III. Proposed method

This research aims to recognize facial expression using DL and to show the influences of the pre-processing step in the performance of expression recognition system. Data pre-processing techniques include face detection, cropping, resizing and normalizations. In this research, we propose a new approach based on DL to classify seven basic emotions: angry, disgust, fear, happy, sad, and surprise. Our approach contains two phases: pre-processing, and expression recognition using convolution neural network model.

Figure 1 presents the general architecture of our proposed FER system. The goal is to provide a new CNN model for predicting facial expression across multiple well-known public face datasets like CK+, JAFEE and FER-2013.

A. Pre-processing

The pre-processing phase detects the face and reduces the lighting effects to some extent. Certain conditions influence facial expression recognition and make this process a complex problem [35]. These conditions include size, illumination, and contrast of images. The stages of FER pre-processing include

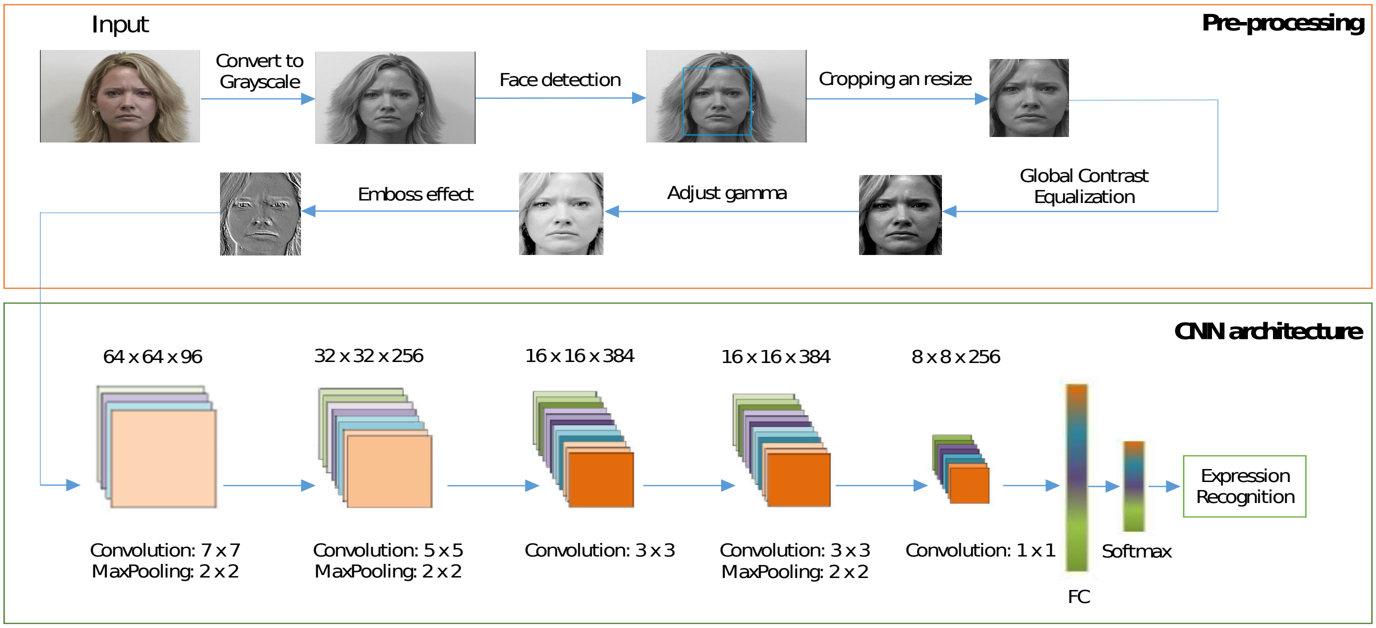

Fig. 1. The proposed CNN architecture.

face detection and face normalization[2], which are summarized in figure 1.

a) Convert RGB to Grayscale image: First, we convert the true-color image RGB to the grayscale intensity image by removing hue and saturation information while maintaining luminance.

b) Face detection: We apply a face detection using the Haar-cascade classifiers to obtain the pixel width and height of a frontal face in an image.

c) Cropping and resize: Since we have detected the geometric structure of faces in an image crop the face and we remove the background area. Finally, we perform a downsampling of the resolution of each image to be fitted into a standard size for the training stage.

d) Global Contrast Equalization: After eliminating and removing insignificant regions in the image, we will obtain higher saturation and better detail in an image. For that, We applicate the Histogram Equalization technique on the output image of the previous step. Histogram Equalization is a better technique for adjusting image intensities. The contrast enhancement of colors gives a good observation of the facial feature. This technique consists of flattening the histogram of the image by stretching the dynamic range of the grayscale. It uses the cumulative image density function.

e) Adjust gamma: Generally, some uncorrected images may appear too bright or too dark. Thus, to correct the image luminance, gamma correction is applied to correct the luminance and to control the brightness and color ratios of the input image. Adjust gamma is a nonlinear operation applied to encode and decode luminance values in an image. Gamma can be described as the relationship between an input and the resulting output. The equation to adjust the gamma of an image is presented by Equation (1):

$$I' = 256 * (\frac{I}{256})^{\gamma} \tag{1}$$

For the best picture quality, dynamic range should not reach 0 or 1. For these limit, we applied the formula with γ = 0,3.

f) Emboss effect: After the steps previously implemented, it is important that the system finds a difference in texture to extract a region of interest (ROI). This is why we employ the ”Embossing Effect” technique to create embossing effects in order to give an image a 3D look. This filter carves and stamps the image giving it relief with bumps and hollows to find all contours. Bright areas are raised and dark ones are carved that results in texture. This effect requires grayscale images.

B. Convolution Neural Networks architecture

The proposed CNN model is implemented with Keras API[3], which comprises five convolutional layers for feature extraction and fully connected layers for 7 classes. The configuration of our CNN architecture shown in Table 1.
The general expression of a convolution kernel is defined by Equation (2):

$$g(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} \omega(s,t) f(x-s, y-t) \tag{2}$$

[2] Normalization phase is used to change the range of pixel intensity value in the image which consist of global contrast normalization.

[3] Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow.

TABLE I
CONFIGURATION OF PROPOSED CNN MODEL FOR FACIAL EXPRESSION RECOGNITION

Layer Type	**Parameters**
Sofmax	7 classes
ReLU	Stride :1, Pad :0
Linear	Units : 14
Fully connected	2048 neurones
ReLU	Stride :1, Pad :0
Conv2D	Filtres 256 kernel :1 x 1 , stride :1, Pad :1
Maxpooling	P :2 x 2 , stride :2, Pad :0
ReLU	Stride :1, Pad :0
Conv2D	Filtres 384 kernel :3 x 3 , stride :1, Pad :1
ReLU	Stride :1, Pad :0
Conv2D	Filtres 384 kernel :3 x 3 , stride :1, Pad :1
Maxpooling	P :2 x 2 , stride :2, Pad :0
ReLU	Stride :1, Pad :0
Conv2D	Filtres 256 kernel :5 x 5 , stride :1, Pad :1
Maxpooling	P :2 x 2 , stride :2, Pad :0
ReLU	Stride :1, Pad :0
Conv2D	Filtres 96 kernel :7 x 7 , stride :1, Pad : 1
Input	Gray-scale image, size : 64 x 64

Where g is the filtered image, f is the original image and ω is the filter kernel. Every element of the filter kernel is considered by -a $\leq$ s $\leq$ a $\leq$ and -b $\leq$ t $\leq$ b.

First of all, we resize all input images to 64 x 64 after applying the pre-processing stage. We utilize a 7 x 7 kernel in the first convolutional layer with 96 filters only. In the second layer, we modify the kernel by 5 x 5 with 256 filters to get more details of the output of the previous layer.

To the third and fourth layers of convolution, there are 384 filters applied on the output of the second convolution layer using a 3 x 3 filter kernel . The last convolution layer is calculated by sampling using a 1 x 1 filter kernel.

The subsampling is done by max-pooling using a 2 x 2 kernel and stride of 2, except the third and fifth convolution layer, which regards the same output size of the previous convolution layers simultaneously but we change the number of filters map.

After each convolution layer, we use rectified linear units (ReLU) which is a simple activation function written as equation (3):

$$f(x) = max(0; x) \tag{3}$$

Where x is the input from a convolved image to the neuron. The ReLU is the most commonly applied activation function in DL models. It allows for avoiding the vanishing gradient problem produced by other activation functions. This problem gets worse for large networks, which have more layers (for more details see [33]).

In order to prevent our network from being oversized, we apply the dropout at the end of fully connected layers. Furthermore, we add a fully connected layer to assure that these nodes interact well and to extract the global relationship between the features. Finally, we apply a softmax function to predict each class probability for the input image.

IV. Experiment results AND EVALUATION

A. *Configurations*

The proposed architecture was built on an Intel Core i7 PC, an NVIDIA GeForce G920MX 2GB GPU, 8GB RAM and the Ubuntu operating system version LTS 18. Our system is developed with Python 3, OpenCV 3.1, TensorFlow framework and Keras API. We install Keras-GPU package on NVIDIA GPU to perform our model CNN. All experimentations were done on GPUs only. GPU-based DL algorithm has the potential to accelerate image processing [34]. Thus learning times can be reduced with GPU cores compared to a CPU. Although GPU cores are slower than CPU cores, they largely offset this disadvantage by their large number and faster memory.

B. *Datasets*

Various databases have been employed for extensive and comparative experiments in this field. Therefore, this sub-section introduces the content the used databases, which are the Extended Cohn-Kanade (CK+), JAFFE and FER-2013. Statistics describing these databases are given in table 3.

- CK+: The extended Cohn-Kanade [4] consists of 593 video sequences recorded from 123 participants (subject directories) were aged between 18 and 30 years. Most of them are female (69%). It is one of a few facial emotion recognitions. This data is available in 640 x 490 pixels on PNG format. Furthermore, it presents 6 basic emotions. We selected only the last frame from each sequence from basic emotion. To form the 7th emotion, we extracted the first frame of each subject, where it was instructed to display a neutral emotion, which results in 432 images in our experiment.

- JAFFE: Japanese Female Facial Expression [5] includes 213 images that present 7 facial emotions of ten Japanese female models. The standard size of each frame is 256 x 256 pixels. All images are exclusively in grayscale values.

- FER-2013: The Facial Expression Recognition 2013 dataset [6] is created using Google's Image Search API to search for face images. It contains 35,887 images in gray-scale. Each image is labeled as any of the seven emotions and it is resized to 48x48 pixels.

To make the model more robust to noise and slight transformations, data augmentation is employed.

C. *Experimental Analysis and Comparison*

In order to compare the obtained accuracy using different known databases, we evaluated the proposed CNN model using independent evaluation of the subject.

TABLE II
NUMBER OF IMAGES PER EACH EXPRESSION IN DATASETS

Dataset	**Labels**							**Total**	**Participants**	**Format**
	AN*	***DI****	***FE****	***HA****	***NE****	***SA****	***SU****			
CK+	45	59	25	69	123	28	83	432	123 (male and female)	.png
FER-2013	4 953	547	5 121	8 989	6 198	6 077	4 002	35 888		
JAFFE	30	29	32	31	30	31	30	213	10 (Japanese female)	.tiff

*AN, DI, FE, HA, NE, SA, SU stand for Anger, Disgust, Fear, Happiness, Neutral, Sadness, Surprised respectively.

TABLE III
AVERAGE ACCURACY (%) ON CK+, FER-2013 AND JAFFE DATABASES.

Method	**Accuracy**		
	CK+	***FER-2013***	***JAFFE***
BDBN [26]	91.80	-	68.00
3D CNN [27]	92.40	-	-
CNN with pre-processing steps [30]	**95.79**	-	53.57
Inception CNN [32]	93.20	66.40	-
Proposed method* (a)	91.34	64.30	74.89
Proposed method* (b)	93.04	68.50	76.46
Proposed method* (c)	92.12	79.63	76.20
Proposed method* (d)	93.59	**70.59**	**79.22**

* Proposed model CNN with (a) Convert RGB to Grayscale, face detection, cropping and resize (b) a + Global Contrast Equalization (c) b + Adjust gamma (d) c + Emboss effect

1) Subject independent evaluation: In our experiments, we divide each database into training, validation, and test sets in a random and strictly subject-independent manner. On average, we assign 60% for the training, 10% for the validation, and 30% for the test. Moreover, each fold on the training, validation and test sets have feebly different sample sizes in some databases. In each experiment, we trained the proposed CNN model for 400 epochs.

Compared to other related works according to protocols adopted in the major of deep learning approaches, the average recognition rate and overall accuracy of the classification are measured. As shown in Table 3, the accuracy of the proposed model is better than that of recent work, especially in FER-2013 and JAFFE. More accurately, to evaluate the pertinent work, the performance of the proposed CNN model, with face detection and cropping step only, gave significant results; viz. 91.34 in CK+, 64.30 in FER-2013 and 65.89 in JAFFE. However, this result is reported on seven facial expressions. The significant improvement observed with pre-processing steps shows that this phase has a great advantage on the CNN performance.

As previously reported, the dataset FER-2013 is more challenging than other FER databases we used. Since it contains images captured spontaneously and different age range. With FER-2013 and JAFFE, we achieved a high accuracy rate of around 70.59% and 79.22% respectively on the test set. Furthermore, Andre et al. [30], that attain the best accuracy and surpass our results, have achieved 95.79 on CK+ dataset. Our method outperforms several previous research works with a major gain and attains the second highest accuracy on this dataset with 93.59%. The confusion matrix on the test set of each database is shown in Figure 2. As we can see, our model CNN is making more mistakes for classes such as anger and surprise in JAFFE dataset.

2) Effect of pre-processing steps on state-of-the-art: To verify the advantage of using the pre-processing steps with recent DL architectures, we evaluate the performance of three models which are 3D CNN [27], Inception CNN [32] and BDBN [26] on CK+, FER-2013 and JAFFE respectively. We report results with and without pre-processing steps. Our CNN model is compared to the state-of-the-art approaches using our pre-processing phase.

In these experiments, we divide each database into training, validation, and test sets similar to the subject independent evaluation. In each experiment, we trained the model for 400 epochs. We have implemented all models in Keras according to the configuration used in the source papers.

Table 4 includes the results of 3D CNN on CK+ without and with after each pre-processing step described as follow:

- (a) Convert RGB to Grayscale, face detection, cropping and resize
- (b) a + Global Contrast Equalization
- (c) b + Adjust gamma
- (d) c + Emboss effect

Using our pre-processing steps combination, the 3D CNN performance decreases to 91.82% after the first step. But, it achieves a second-high accuracy rate of 94.40% and outperforms our model. Despite the improvement in the result, it remains slight compared to the improvement made on our model.

Table 5 provides the comparison of the Inception CNN model [32] without and with pre-processing phase on FER-2013. We notice a considerable improvement in the results obtained after each step. Note that the Inception CNN uses convolutions kernels of multiple sizes as well as pooling within one layer. Plus the fact that kernel sizes are variable, it is always good to have a general rule like this instead of a hyper parameter.

As seen in Table 6, BDBN [26] that is using the pre-processing phase significantly outperforms the original accuracy (68.00

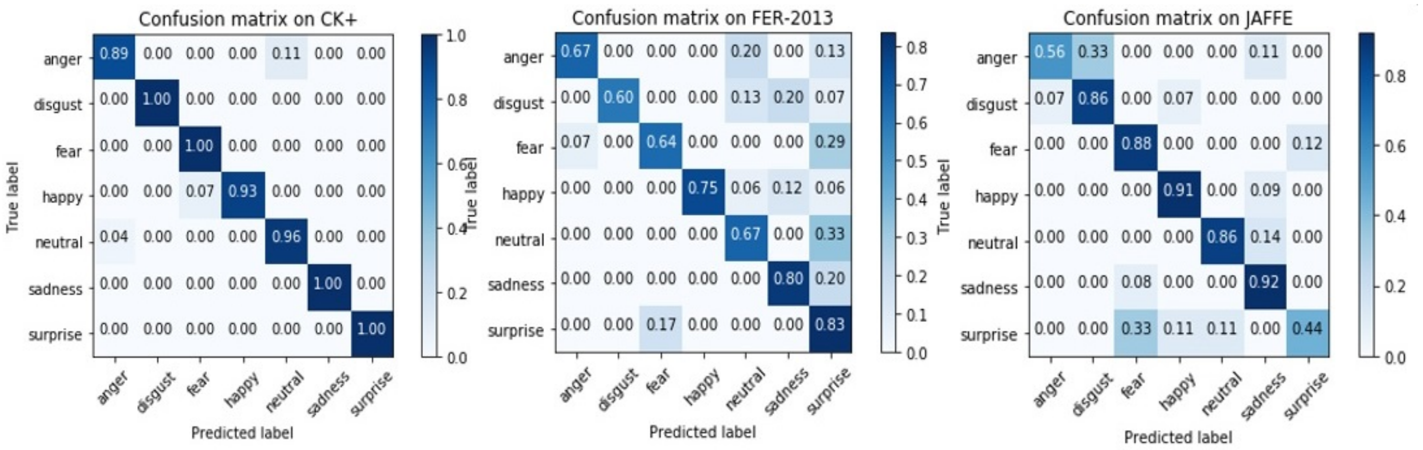

Fig. 2. The confusion matrix on CK+, FER-2013 and JAFFE.

%) on JAFFE with 71.34%.

Overall, the results obtained with our pre-processing phase show that our combination of specific algorithms provides better results with different architectures than the individual model. This exhibits the influence of our algorithms combination over other deep models.

TABLE IV
STATE-OF-THE-ART RESULT WITH DIFFERENT PRE-PROCESSING STEPS ON CK+

Steps	**Accuracy (%)**
3D CNN [27] (1)	92.40
(1) + (a)	91.82
(1) + (b)	93.00
(1) + (c)	92.89
(1) + (d)	94.40

TABLE V
STATE-OF-THE-ART RESULT WITH DIFFERENT PRE-PROCESSING STEPS ON FER-2013

Steps	**Accuracy (%)**
Inception CNN [32] (2)	66.40
(2) + (a)	67.26
(2) + (b)	68.14
(2) + (c)	69.32
(2) + (d)	72.07

TABLE VI
STATE-OF-THE-ART RESULT WITH DIFFERENT PRE-PROCESSING STEPS ON JAFFE

Steps	**Accuracy (%)**
BDBN [26] (3)	68.00
(3) + (a)	67.73
(3) + (b)	79.00
(3) + (c)	78.67
(3) + (d)	71.34

V. Conclusion and Future Works

This research aims to recognize facial expression using CNNs. The proposed method had shown the influences of the pre-processing steps over the robustness of expression recognition system. Our model achieves remarkable speed, precision compared to the state of the art systems. In the processing phase, we proposed a novel CNN architecture enhanced by pre-processing steps to classify seven facial expressions. This model included five convolution layers, each followed by the ReLU and dropout regularization.

To improve the expression recognition system, future work will focus on improving the accuracy rate in the long term through transfer learning algorithms. In addition, we plan to design and develop our real-time system based on facial and vocal expressions.

References

[1] H. Ltifi, M. B. Ayed, C. Kolski, A. M. Alimi, "HCI-enriched approach for DSS development: the UP/U approach", IEEE Symposium on Computers and Communications, Sousse, Tunisia, pp. 895-900, 2009.

[2] P. Ekman, W. V. Friesen, "Constants across cultures in the face and emotion", Journal of personality and social psychology, vol. 17, no. 2, pp. 124-129, 1971.

[3] H. Ltifi, C. Kolski, M. B. Ayed, "Combination of cognitive and HCI modeling for the design of KDD-based DSS used in dynamic situations", Decision Support Systems, Vol. 78, pp. 51-64, 2015.

[4] L. Patrick , F. C. Jeffrey , K. Takeo , S. Jason , A. Zara and M. Iain, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression", IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp. 1-8, 2010.

[5] J. L. Michael , S. Akemastu, M. Kamachi, J. Gyoba, "Coding Facial Expressions with Gabor Wavelets",' 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205, 1998.

[6] I. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, ”Challenges in Representation Learning: A report on three machine learning contests”, Neural Networks, vol. 64, pp. 59-63, April 2015.

[7] Z. Liu, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, M. Zhou, and J. Mao, ”A facial expression emotion recognition based human-robot interaction system,” IEEEICAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 668-676, 2017.

[8] L. Tao and B. J. Matuszewski, ”Is 2d unlabeled data adequate for recognizing facial expressions?” IEEE Intelligent Systems, vol. 31, no. 3, pp. 19-29, May 2016.

[9] T. Baltruaitis, P. Robinson, and L. P. Morency, ”Openface: An open source facial behavior analysis toolkit,” IEEE Winter Conference on Applications of Computer Vision (WACV), March 2016, pp. 1-10.

[10] S. Du, Y. Tao, and A. M. Martinez, ”Compound facial expressions of emotion”, PNAS, Vol. 15, pp. 1454-1462, March 2014.

[11] M. Jeong, S.Y. Kwak, B.C. Ko and J.Y. Nam, ”Driver facial landmark detection in real driving situation”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 28, pp. 1-15, 2017.

[12] E. Sariyanidi, H. Gunes, and A. Cavallaro, ”Automatic analysis of facial affect: A survey of registration, representation, and recognition”, IEEE transactions on pattern analysis and machine intelligence, Vol.37, pp. 1113-1133, 2015.

[13] L. Nanni, S. Ghidoni, S. l Brahnam, ”Handcrafted vs Non-Handcrafted Features for computer vision classification”, Pattern Recognition, vol. 71, pp. 158-172, June 2017.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Imagenet classification with deep convolutional neural networks”, Advances in neural information processing systems, vol. 60, pp. 84-90, June 2017.

[15] G. Jiuxiang, W. Zhenhua, K. Jason, M. Lianyang, S. Amir, B. Shuai, L.Ting , W. Xingxing, W. Gang, C. Jianfei and C. Tsuhan, ”Recent advances in convolutional neural networks”, Pattern Recognition, Vol 77, pp. 354-377, May 2018.

[16] D. A. Pitalokaa, A. Wulandaria, T. Basaruddina, D. Y. Liliana, ”Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition”, Procedia Computer Science, vol. 116, pp. 523-529, 2017.

[17] J. Whitehill, M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, ”Towards practical smile detection”. IEEE T-PAMI, vol. 31, no. 11, pp. 2106-2111, 2009.

[18] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. Fasel, and J. R. Movellan, ”Recognizing facial expression: Machine learning and application to spontaneous behavior”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), pp. 568-573, 2005.

[19] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer, ”Meta-Analysis of the First Facial Expression Recognition Challenge”, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 966-979, 2012.

[20] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, and L. Prevost, ”Combining AAM coefficients with LGBP histograms in the multi-kernel SVM framework to detect facial action units”, Face and Gesture Workshops, pp. 860-865, 2011.

[21] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, M. Zareapoor, ”Hybrid Deep Neural Networks for Face Emotion Recognition”. Pattern Recognition Letters, Vol. 115, no. 1, pp. 101-106, 2018.

[22] Y. Hu, Z. Zeng, L. Yin, X. Wei, X. Zhou, T. S. Huang, ”Multiview facial expression recognition”, 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1-6, 2008.

[23] M. Dahmane, J. Meunier, ”Emotion recognition using dynamic grid-based HoG features”, Face and Gesture Workshop, 2011.

[24] U. Mlakar, B. Potocnik, ”Automated facial expression recognition based on histograms of oriented gradient feature vector differences”, Signal, Image and Video Processing, vol. 9, no. 1, pp. 245-253, 2015.

[25] A. Lorincz, L. A. Jeni, Z. Szabo, J. F. Cohn, and T. Kanade, ”Emotional expression classification using time-series kernels”, IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 889-895, 2013.

[26] P. Liu, S. Han, Z. Meng, Y. Tong, ”Facial expression recognition via a boosted deep belief network”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1805-1812, 2014.

[27] M. Liu, S. Li, S. Shan, R. Wang, X. Chen, ”Deeply learning deformable facial action parts model for dynamic expression analysis”. Asian Conference on Computer Vision, pp. 143-157, 2014.

[28] M. Pantic, M. Valstar, R. Rademaker, L. Maat, ”Webbased database for facial expression analysis”, IEEE International Conference on Multimedia and Expo, pp. 1-4, 2005.

[29] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, K. Scherer, ”The First Facial Expression Recognition and Analysis Challenge”, IEEE International Conference on Automatic Face and Gesture Recognition, 2011.

[30] A. T. Lopes, E. d. Aguiar, A. F. D. Souza, T. Oliveira-Santos, ”Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order”, Pattern Recognition, Vol. 61, pp. 610-628, 2017.

[31] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, ”A 3d facial expression database for facial behavior research”, 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 211- 216, 2006.

[32] Ali M., David Chan, and Mohammad H. Mahoor, ”Going Deeper in Facial Expression Recognition using Deep Neural Networks”, IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-10, 2016.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Imagenet classification with deep convolutional neural networks”, NIPS’12 Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 1097-1105, 2012.

[34] B. N. M. Reddy, S. Shanthala, B. R. VijayaKumar, ”Performance Analysis of GPU V/S CPU for Image Processing Applications”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 5, pp. 437-443, 2017.

[35] F. Gianfelici , C. Turchetti, P. Crippa, ”A non-probabilistic recognizer of stochastic signals based on KLT”, Signal Processing, Vol. 89, pp. 422-437, 2009.