ISSN 2278-3091

Volume 9, No.1.4, 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse0691.42020.pdf

https://doi.org/10.30534/ijatcse/2020/0691.42020

# Facial Expression Recognition: A Convolutional Neural Network Approach

**N. A. Ali[1], A. R. Syafeeza[2], A.S Jaafar[3], M.K Fitri Alif[4], P. Marzuki[5]**

[1]Centre for Telecommunication Research & Innovation (CeTRI), UTeM, Malaysia, alisa@utem.edu.my

[2]Centre for Telecommunication Research & Innovation (CeTRI), UTeM, Malaysia, syafeeza@utem.edu.my

[3]Centre for Telecommunication Research & Innovation (CeTRI), UTeM, Malaysia, shukur@utem.edu.my

[4]Control and Mechatronics Eng Department, UTM, Malaysia, mohdfitrialif@gmail.com

[5]Department for Machine Learning and Signal Processing (MLSP) , UTeM, Malaysia, marzuki1992@gmail.com

**ABSTRACT**

Facial expression recognition systems have been researched for years remains a challenging problem in computer vision. Facial expressions manifest the emotions of the person and tend to be different between individuals. In this paper, a robust Convolutional Neural Network(CNN) architecture is proposed for the facial expression recognition problem. Two datasets employed in the experiments which were Japanese Female Facial Expression(JAFFE) and Cohn-Kanade(CK+). A comparative study of the impact of each image pre-processing operation in accuracy rate is presented. By determining these ideal elements, a greater accuracy can possibly be attained. CNN was exploited to a set of seven emotions from numerous facial expressions namely, happy, sad, angry, surprise, disgust, fear and neutral. Experimental result shows that the proposed CNN solution achieved accuracy of 91.76% in the CK+ database.

**Key words :** Convolutional neural network, deep learning, facial expression recognition.

## 1. INTRODUCTION

One of the most essential information on understanding human emotions are facial expression. For many year facial emotions have been studied. Emotional instability affected the transformation of the expression at human face automatically [1]. Emotional expression has several actions associated with it, such as face movements and body gestures, changes in voice tone, physiological changes in the skin resistance and facial flushing [2]. Shih [3] stated in his book entitle “Image Processing and Pattern Recognition” that facial expressions play a major role in connecting with people since it expresses the emotions of the human in interaction between human surrounding, Shan et.al [4] express that facial expression can be divided into seven categories which were happy, sad, angry, fear, surprise, disgust and neutral.

Facial expression recognition system has been developing for years. Face and object recognition systems are applicable to real life applications, such as human-computer interaction, surveillance, pervasive computing, and access control (non-intrusive security) [5]. Hamester et.al [6] stated that surrounding environment, position of the subject and skin colours are among of restriction that can lead to lack of capability of facial expression recognition system. Despite the continuous research efforts, accurate facial expression recognition under un-controlled environment still remains a significant challenge.

A promising approach that can be utilized to address facial expression recognition issues is a Convolutional Neural Network (CNN). It has been applied to several applications, such as recognition of human face [7–9], classification of human gender [10], identification of finger-vein [11–13], recognition of license plate [14] and other applications. Lecun et al. [15] proposed the classical Convolutional Neural Network in 1998; the LeNet-5 which was utilized in handwritten digit character recognition. LeNet-5 architecture is shown in Figure 1.

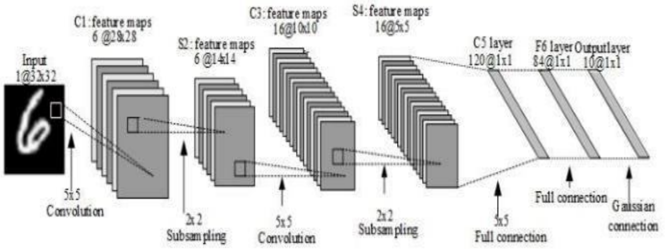

**Figure 1:** The LeNet-5 CNN Architecture [15]

CNN provides a great promise of solution in order to solve complex pattern recognition problems. A CNN is a hierarchical multi-layered neural network that combines segmentation, feature extraction, and the classification process in one trainable module. Hence, it accepts raw data images with minimal image processing. CNN completes feature extraction and classification at once, and needs not to design feature filters with the ability of learning features [16,17]. This paper focuses on designing the CNN model and architecture, particularly for facial expression recognition that would save computational time and cost. The cross - validation method has been used to determine the best parameters and model of CNN architecture. It has been proven that by applying this method, the accuracy can be measured.

This paper is organized as follows: The next section describes the database used and the proposed approach. The results and analysis are shown in Section 4. Finally, Section 5 concludes this paper.

## 2. RESEARCH METHOD

The methodology for this research is divided into three sections. The first section discusses the two standard databases used, the second section discusses applied pre-processing method and the final section discusses the CNN design. This system ran on 2.3 GHz Intel Core i5 2.5GHz with an NVIDIA GeForce processor that has 610Mb memory. The Python 3.5.2 language, Open Source Computer Vision Library (OpenCV) and Tensorflow were used for pre-processing and to design the CNN model.

### 2.1 Database

There were two databases used in this paper to classify the images of human faces into seven classifications of expression which were Japanese Female Facial Expression (JAFFE) and Cohn-Kanade (CK+) database.

The Japanese Female Facial Expression (JAFFE) database [18] that has seven total facial expressions, namely happy, sad, surprise, angry, disgust, fear and neutral. JAFFE database consists of 213 grayscale frontal faces pose images from 10 Japanese females. In this database, there are 4 images taken from each subjects in the six basic expressions and one image of the neutral expression from each subjects. All images in the dataset are stored in grayscale of 256×256 pixel arrays with 8-bit precision. Figure 2 shows the sample image of JAFFE database.

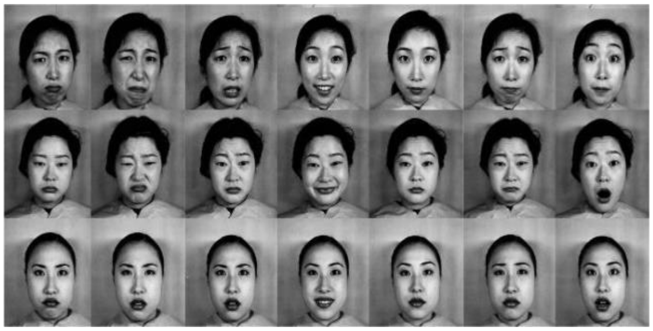

**Figure 2:** Sample images of JAFFE Database [18]

The Cohn-Kanade (CK+) database comprises of the images of 100 university students with the age between 18 and 30 years old. The subjects involving of 65% female, 15% African America and 3% Asian or South American [19]. The students were instructed to perform the six basic expressions that begin and end with neutral expression. All the images in the database are stored grayscale of 640×480 pixel arrays with 8-bit precision. The database contains images for the following expression; neutral, angry, disgust, fear, happy, sad and surprise. Some examples of the CK+ database images are shown in Figure 3.

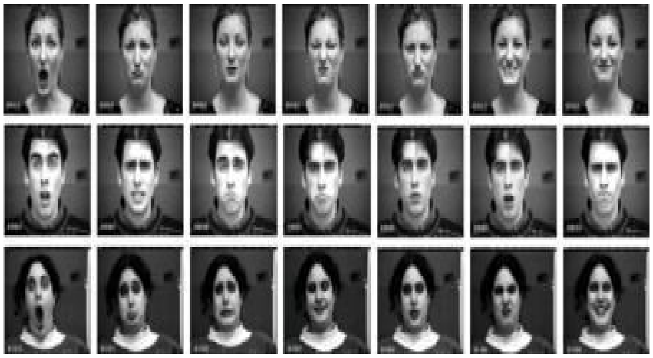

**Figure 3:** Sample images of CK+ Database [18]

### 2.2 Data Pre-processing

The input to the network is expected to be in the term of facial image. However, it can be difficult for the deep network to handle high variations in the facial pose and lighting conditions. Thus, it becomes necessary to pre-process the input to make the faces more uniform. The pre-processing step can be divided to six parts as shown in Figure 4.

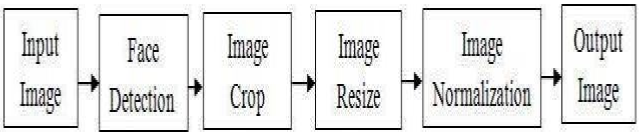

**Figure 4:** Pre-processing procedure

The first part is input image from the image of JAFFE and CK+ databases. The second part is face detection; extracted from two databases used. After the face was detected, the face was cropped and resized to 56×46 pixels. The image was then normalized using min-max normalization algorithm which produces pixel value within the range of -1.0 to 1.0. Output images were stored in numeric data and divided to two parts; 100 training normalized images and 40 testing normalized images. The cross validation was used to find the best parameters and the best accuracy between those two databases. The training was carried out 100 times where during the validation, the best period during the training would configure the best network weights. The best network weights are selected and used to compute the accuracy of the test phase.

### 2.3 Design of CNN

The CNN design for facial expression recognition consists of four layers. In designing a neural network architecture, the parameters (convolution and subsampling of layers) are to be reduced to the greatest extent possible, also known as reducing the model complexity. This approach results in four layers 4-14-60 model topology as shown in Figure 5. While Figure 6 shows the convolution and subsampling process of the proposed CNN architecture.

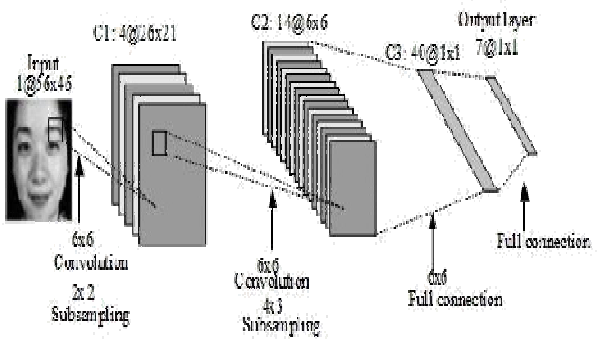

**Figure 5:** Proposed CNN Design

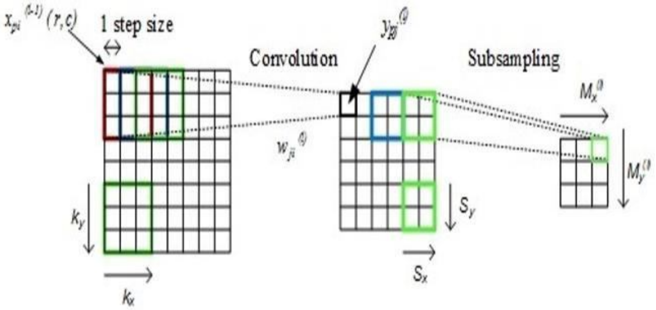

**Figure 6:** Convolution and subsampling process

In the convolution or subsampling process, the convolution kernel $w_{ji}^{(l)}$ is convolved with the input feature map $M_i$, with the subsampling incorporated as a skipping operation in the convolution process. The average pooling operation in a subsampling layer is not performed. Instead, only the sizes of the kernels $(S_x,\ S_y)$ are represented as skipping factors between subsequent convolutions in $x$ and $y$ directions respectively. Equation (1) gives the sizes of the output map $(M_x,\ M_y)$, in $x$ and $y$ direction.

$$M_x^{(l)} = \frac{M_x^{(l-1)} - \left(K_x^{(l-1)} - S_x^{(l)}\right)}{S_x^{(l)}};\ M_y^{(l)} = \frac{M_y^{(l-1)} - \left(K_y^{(l-1)} - S_y^{(l)}\right)}{S_y^{(l)}} \quad (1)$$

where ($K_x$, $K_y$) denotes the convolution kernels, index ($l$) indicates the layer.

The proposed system operates in two main phases which are training and test. During the training, the system receives a training data comprising grayscale images of faces with their respective expression and learns a set of weights for the network. To ensure that the training performance is not effected by the order of presentation the examples, a few images are separated as validation and are used to choose the best set of weights out of a set of trainings performed with samples presented in different orders. During the test, the system receives a grayscale image of face and output the predicted expression by using the final network weights learned during the classification of training. The output produced determines the classification of the expressions represent integer numbers (0 – angry, 1 – disgust, 2 – fear, 3 – happy, 4 – sad, 5 – surprise and 6 – neutral). The overview of the proposed facial recognition system is shown in Figure 7.

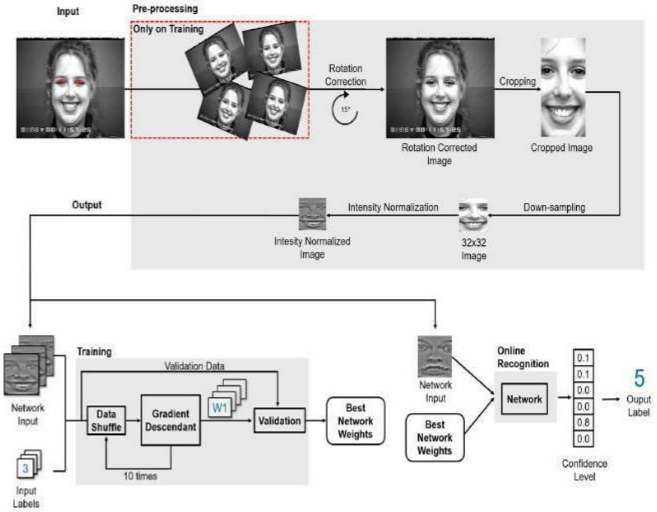

**Figure 7:** Overview of proposed facial expression recognition systems [20]

## 3. RESULT AND DISCUSSION

This section presents the result of performance for the proposed system which separated into training and testing phase.

### 3.1 Cohn-Kanade(CK+) Database

The CK+ database was separated into 7 groups of expression where in each of the groups there was none subject overlapping between them. Each subject in the groups would be go through the pre-processing step which consist of rotation correction, image cropping, image flatten or down-sampling and intensity normalization. The training was carried out 100 times where during the validation, the best epoch during the training would configure the best network weights. The best network weights are selected and used to compute the accuracy of the test phase. Figure 8 show the illustration of the intermediate layer where the best set of weights were choosing to determine the expressions.

**Figure 8:** Illustration of the intermediate layer for CK+ Database

The training of the network takes about 45 minutes where for each training set only takes up to 80 seconds. The accuracy of the proposed system that classify into the seven expressions using the CK+ is shows in Table 1.

**Table 1:** The accuracy of seven expression for CK+

| Expression | Neutral | Angry | Surprise | Disgust | Fear | Happy | Sad |
|---|---|---|---|---|---|---|---|
| Precision (%) | 100 | 70 | 100 | 100 | 62 | 80 | 75 |

Based on the Table 1, disgust and surprise expression achieved accuracy rate of 100% whereas fear and angry expression achieved the smallest recognition rate. This shows that the pixel space of the two expressions were not well separated as the two expressions are very similar to each other. The pixel space cannot determine the best set of weight between the pixels in the two expressions. The average accuracy of the seven expressions for the CK+ recorded to be up to only 85%. Figure 9 shows the accuracy of CK+ database during training where it achieved 91.76% during training. The accuracy increased as the epoch value increased.

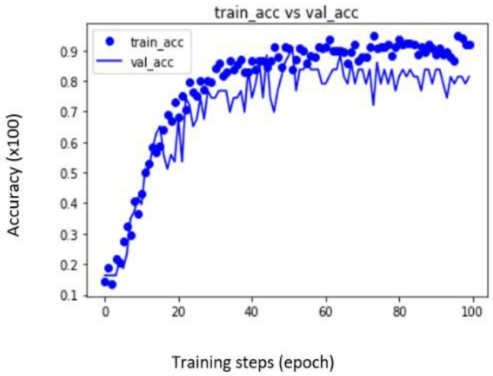

**Figure 9:** The accuracy of CK+ database during training

The normalized of the confusion matrix for the seven expressions of CK+ database shown in Table 2.

**Table 2:** The normalized confusion matrix on CK+ database

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Neutral | 67 | 11 | 0 | 22 | 0 | 0 | 0 |
| Angry | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| Disgust | 0 | 9 | 82 | 0 | 0 | 9 | 0 |
| Fear | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| Happy | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Sad | 0 | 17 | 0 | 17 | 17 | 50 | 0 |
| Surprise | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

Table 2 shows that the sad expression has the lowest configuration of accuracy since it mostly confused with the angry, fear and happy expressions. Disgust expression also cannot determine between angry and sad expressions. Many cells outside of the main diagonal expressions that slightly above value of zero showed that the samples were misclassified between the seven expressions.

### 3.2 Japanese Female Facial Expression (JAFFE) dataset

The JAFFE database contains images from only 10 subjects where same as the CK+ database, the images were separated into seven categories of expression. The training phase was carried out the 170 samples of image and validate on the 43 samples of image. Same as CK+ database, JAFFE database undergoes pre-processing step before training. The training phase was carried out 100 times where it shows that the accuracy increases up to 84.54% shown in Figure10.

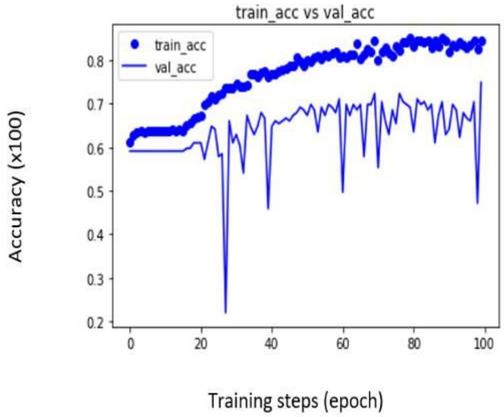

**Figure 10:** The accuracy of JAFFE database during training

The best set of weights were used to determine the expression of the images in train. The training of the network takes about 20 minutes where each training set takes about 20 seconds. The best network weights were selected and used to compute the accuracy of the test set. The accuracy of the seven expressions after the training and testing phase were recorded in Table 3.

**Table 3:** The accuracy of seven expression for JAFFE

| Expression | Neutral | Angry | Surprise | Disgust | Fear | Happy | Sad |
|---|---|---|---|---|---|---|---|
| Precision (%) | 82 | 14 | 100 | 91 | 100 | 63 | 40 |

Based on Table 3, angry and surprise were the expression that shows the lowest recognition among the seven expressions. This was due to incapability of the network weights to determine the expressions since the pixel space of these two expressions were not separated respectively. The average accuracy of the seven expressions for the JAFFE database only reached up to 67%. The intermediate layer as in Fig. 11 through the three layer that learned the kernel and generate maps for each convolutional layer. The image selected to be testing would configure the feature maps of each pixel in image by interpolation of the nearest pixel. The normalized confusion matrix for the seven expressions of CK+ database was shown in Table 4 where disgust expression has the lowest accuracy compared to the other expressions. It was mostly tangled with the neutral expression where it shown from the sample images of disgust was misclassified as neutral expression.

**Table 4:** Normalized confusion matrix for JAFFE Database

| | Neutral | Angry | Disgust | Fear | Happy | Sad | Surprise |
|---|---|---|---|---|---|---|---|
| Neutral | 98 | 1 | 1 | 0 | 0 | 0 | 0 |
| Angry | 71 | 14 | 14 | 0 | 0 | 0 | 0 |
| Disgust | 53 | 7 | 27 | 0 | 7 | 0 | 7 |
| Fear | 50 | 0 | 0 | 0 | 0 | 0 | 50 |
| Happy | 6 | 0 | 19 | 0 | 62 | 0 | 12 |
| Sad | 40 | 30 | 10 | 0 | 0 | 0 | 20 |
| Surprise | 0 | 8 | 0 | 0 | 0 | 0 | 92 |

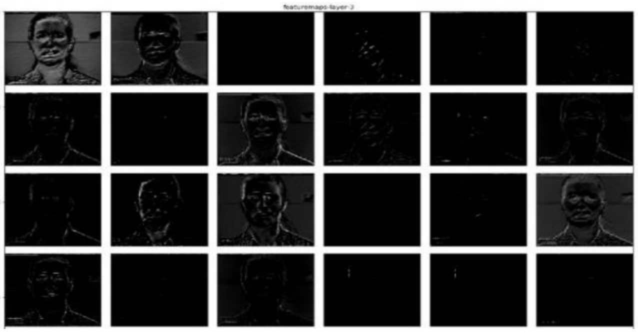

**Figure 11:** The illustration of intermediate layer for JAFFE database

Table 5 summarized that comparison between the accuracy of CK+ database and JAFFE database. As it can be seen, CK+ shows the greatest performance of accuracy with 91.76% rather than JAFFE with 84.54% accuracy. JAFFE show lower accuracy due to small number of database. Convolutional Neural Network mostly emphasized a big amount of data in order to adjust the parameters. The required data amount that methods for any facial expression recognition needed in the training phase to achieve a good accuracy.

**Table 5:** Comparison between CK+ and JAFFE Database

| Database | CK+ | JAFFE |
|---|---|---|
| **Accuracy (%)** | 91.76 | 84.54 |

The best performance of the facial expression recognition system by using the CK+ and JAFFE database was achieved by using the both spatial and intensity normalization. Table 6 summarized the performances of each step in the pre-processing steps.

**Table 6:** Comparison of pre-processing step carried out in the experiment

| Pre-processing | Accuracy (%) |
|---|---|
| None | 41.36 |
| Rotation correction | 43.65 |
| Image cropping | 56.78 |
| Down-sampling | 48.64 |
| Intensity normalization | 65.43 |
| Down-sampling and intensity normalization | 86.67 |

## 4. CONCLUSION

The proposed CNN model for facial expression recognition system achieved accuracy of 91.76% in the CK+ database. The combination of the CNN and the pre-processing steps has been increased the performance accuracy of the system where it shows higher accuracy to classify the database into seven facial expressions.

This combination provides simpler solution where it takes less times to train and classify if the system is use in real time application. However, in order to increase the performance accuracy, the CNN need a large amount of data to be able to learn the set of features which provides the best models of desired classification in future.

## ACKNOWLEDGEMENT

Authors are grateful to Universiti Teknikal Malaysia Melaka for the financial support through short term grant PJP/2019/FKEKK(11D)/S01644.

## REFERENCES

1. E. Owusu, Y. Zhan, and Q. R. Mao (2014), **Expert Systems with Applications A neural-AdaBoost based facial expression recognition system**, *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3383–3390.
2. V. Pilla, A. Zanellato, C. Bortolini, H. R. Gamba, G. B. Borba, and H. Medeiros, **Facial Expression Classification Using Convolutional Neural Network and Support Vector Machine**. Available at: https://pdfs.semanticscholar.org/d300/50cfd16b29e43ed2024ae74787ac0bbcf2f7.pdf.
3. F. H. Shih(2010), **Image Processing and Pattern Recognition**, Image Processing and Pattern Recognition, Wiley, 2013, ch. 2, pp. 45-47.
4. K. Shan, J. Guo, W. You, D. Lu, and R. Bie(2017), **Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural Network Structure**, pp. 123–128.
5. S. Afaq,A. Shah,M. Bennamoun,F. Boussaid(2015), **Neurocomputing Iterative deep learning for image set based face object recognition**, *Neurocomputing*,pp. 1–9.
6. D. Hamester, P. Barros and S. Wermter, **Face expression recognition with a 2-channel Convolutional Neural Network**, *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, 2015, pp. 1-8, doi: 10.1109/IJCNN.2015.7280539
7. A. R. Syafeeza, M. Khalil-Hani, S. S. Liew, and R. Bakhteri. **Convolutional Neural Networks with Fused Layers Applied to Face Recognition**. *Int. J. Comput. Intell. Appl. 2015*; 14:1550014. https://doi.org/10.1142/S1469026815500145

8. A. R. Syafeeza, M. Khalil-Hani, S. S. Liew, and R. Bakhteri. **Convolutional neural network for face recognition with pose and illumination variation**. *Int. J. Eng. Technol. 2014*; 6:44–57.
9. Syazana Itqan, Khalid and Syafeeza, Ahmad Radzi and Norhashimah, Mohd Saad. **A MATLAB-Based Convolutional Neural Network Approach for Face Recognition System**. *Journal of Bioinformatics And Proteomics Review. 2016*; 2:1-5.
10. S. S. Liew, M. Khalil-Hani, S. Ahmad Radzi, And R. Bakhteri. **Gender classification: a convolutional neural network approach**. *Turkish J. Electr. Eng. Comput. Sci. 2016*; 24:1248–1264.
11. M. R. Devi, (2019) **An Efficient Technique to Classify Human Activity using Convolutional** *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1, pp. 59–67, 2019. https://doi.org/10.30534/ijatcse/2019/1381.32019
12. Nazeer, S. A., Omar, N., & Khalid, M. (2007). **Face recognition system using artificial neural networks approach**. *2007 International Conference on Signal Processing, Communications and Networking* (pp. 420-425). IEEE
13. K. S. Itqan, A. R. Syafeeza, F. G. Gong, N. Mustafa, Y. C. Wong, and M. M. Ibrahim. **User identification system based on finger-vein patterns using Convolutional Neural Network**. *ARPN J. Eng. Appl. Sci*. 2016; 11:3316–3319.
14. Bhandary, Abhir & k b, Sudeepa & Chokkadi, Sukhada &
    M S, Sannidhan. (2019**). A Study on various state of the art of the Art Face Recognition System using Deep Learning Techniques**. *International Journal of Advanced Trends in Computer Science and Engineering*. 8. 1590-1600. 10.30534/ijatcse/2019/84842019.
15. Y. LeCun et al. **Backpropagation Applied to Handwritten Zip Code Recognition**, *Neural Computation, 1989*; 1:541–551.
16. Y. Liu and Y. Chen(2017), **Recognition of facial expression based on CNN-CBP features**, pp. 2139–2145, 2017.
17. I. Song, H. J. Kim, and P. B. Jeon. **Deep learning for real-time robust facial expression recognition on a smartphone.** *IEEE International Conference on Consumer Electronics, 2014*. pp. 564–567.
18. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. **Coding facial expressions with Gabor wavelets**. *3rd IEEE International Conference on Automatic Face and Gesture Recognition, 1998*. pp. 200–205.
19. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, **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*, 2010, pp. 94–101.
20. Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, **Identity-aware convolutional neural network for facial expression recognition**, *Automatic Face & Gesture Recognition (FG 2017)*, 2017 12th IEEE International Conference on. IEEE, 2017, pp. 558–565.
21. A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. OliveiraSantos(2017), **Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order**," *Pattern Recognit.*, vol. 61, no. September, pp. 610–628 https://doi.org/10.1016/j.patcog.2016.07.026