iJRASET
International Journal For Research in
Applied Science and Engineering Technology


INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: V Month of publication: May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5439

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

# A Brief Survey on Facial Emotion Recognition

Mayur S Sharma[1], K Vishal Warrier[2]
[1, 2]Dayananda Sagar College of Engineering

***Abstract: This paper introduces a series of methodologies to accurately tackle the classification of emotion and compare the various methodologies and identify the ideal solution. The progression of the decades of scientific research has been conducted for developing methods for automated emotion recognition. Now, there is an extensive literature proposing and evaluating various methods, leveraging techniques from multiple fields, such as signal processing, computer vision, machine learning, and speech processing. Given an image of arbitrary size, the job is to identify an emotion of a human face appearing in the image. Face detection in complex environments is disputing since the faces may appear in different scales, different head poses, and orientations. External factors also play a vital role; for instance, the lighting conditions, facial expressions, and shadows are few other sources of variations that need to be taken into account. The approach is to yield a better classification performance implemented in real-world scenarios with fewer exemptions and more generalized accuracy.***

## I. INTODUCTION

Emotion recognition is a method of identifying human emotion. People vary widely in accuracy at recognizing the emotions of all people. The use of modern technology to help people with emotion identification is a relatively nascent research area. Nowadays, most of the work has been going on in the area of automating the recognition of facial expressions from video, audio, written expressions from the text, and physiology as measured by wearables technology. Humans show plenty of cognitive changes in their abilities to recognize emotion over the generations. The progression of the decades of scientific research has been conducted for developing methods for automated emotion recognition. Now, there is an extensive literature proposing and evaluating various methods, leveraging techniques from multiple fields, such as signal processing, computer vision, machine learning with speech processing. These different methodologies and techniques are employed to interpret emotions such as Bayesian networks, Gaussian Mixture Models, and Hidden Markov Models.

## II. METHODOLOGIES

Following sections have a brief review of literature about different ways of recognition and classification of emotions :

*A.* The first proposed solution we go through is a paper where it focuses on the extraction of discriminative features from salient facial patches is vital in emotion recognition. The accuracy of the detection of facial landmarks improves the maximum localization of the salient patches on face images. Here, this paper proposes a framework for expression recognition by using the appearance features of selected facial patches. Facial patches are extracted, which are active during emotion elicitation, depending on the position of facial landmarks. These active patches are curated to obtain the salient patches, which comprise essential features for classification of each pair of expressions, thereby choosing different facial patches as salient for different pairs of expression classes. The one-against-one classification method is used using these features. Moreover, an automated learning-free facial landmark detection technique is proposed, where it results in a similar performance as that of other modern landmark detection methods, yet requires significantly less execution time. The proposed method is found to perform well consistently in different environments, hence, providing a solution for expression recognition in low-resolution images. Experiments on CK+ and JAFFE facial expression databases show the effectiveness of the proposed system.

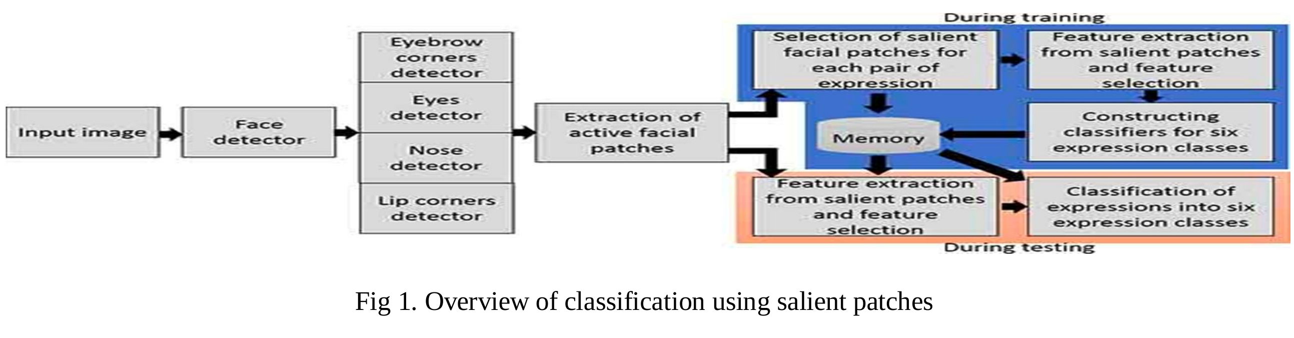



Fig 1. Overview of classification using salient patches

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

1) *Advantages*
a) Improves the localization of the salient patches on face images.
b) Predicts expression recognition in low resolution images.
2) *Disadvantages*
a) Requires huge data storage.
b) Requires huge computational power.

*B.* This paper presents a fuzzy logic based emotion recognition system. The system is comprised of an image processing stage, followed by an emotion recognition stage. In the image processing stage, the subject's facial features (eyes, mouth, etc.) are extracted. Next, the identifying points strategically placed on the face are extracted from each facial feature. In the emotion recognition stage, the identifying points are used to fuzzify and determine the strength of different facial actions. These strengths are then used to determine the subject's displayed emotion. The Japanese Female Facial Expression database was used to evaluate the system's performance resulting in an overall successful detection rate of 78.8%.

1) *Advantages*
a) Improves the accuracy of emotion prediction .
b) Improves its advantage of phycology.
2) *Disadvantages*
a) Accuracy is very less.
b) Requires huge fuzzy sets to predict emotions.

*C.* An emotion recognition method from a set of face images is proposed in this paper, which can recognize seven emotions of humans, i.e., six basic expressions in addition to neutral (anger, sad, surprised, disgust, fear, happy). The proposed method uses the GLCM approach for feature extraction and the nearest neighbor (NN) for classification. The fuzzy Euclidean distance is used with GLCM providing the texture characteristics of an input image through second-order statistical measurements. Because of the existence of vagueness and uncertainty in the discriminant features extracted from various emotional face images, the fuzzy measure is involved in the NN classifier to recognize the emotions of faces with more accuracy. The experiments show excellent efficiency compared to some other feature extraction and facial emotion recognition methods.

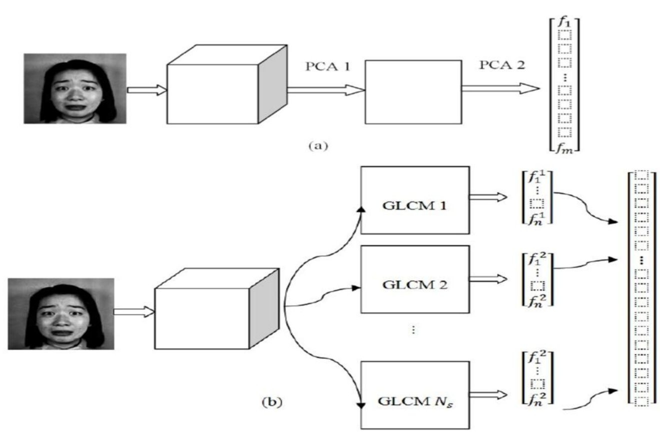



Fig 2. Two approaches for facial extraction (a) with feature reduction (b)without feature reduction

1) *Advantages*
a) Good efficiency in facial emotion classification.
b) Increase efficiency in feature extraction.
2) *Disadvantages*
a) Accuracy is very less.
b) Requires huge fuzzy sets to predict emotions.

*D.* This paper proposes a method for detecting facial regions by combining a Gabor filter and a convolutional neural network. Gabor filter is used in the first stage, which extracts the most intrinsic facial features. As a result of this Gabor-two-filters transformation, we obtain four sub-images. The second stage of the method uses the application of the convolutional neural network to these four images. This approach yields better classification performance in comparison to the results obtained by the convolutional neural network alone.

*1) Advantages*

*a)* Improve better classification performance of facial detection.

*b)* Increase efficiency of extracts intrinsic facial features.

*2) Disadvantages*

*a)* Not suitable for emotion detection.

*b)* Can't consider the all features to for face detection.

*E.* Here they present a novel and efficient Deep Fusion Convolutional Neural Network (DF-CNN) for multi-modal 2D+3D Facial Expression Recognition (FER). DF-CNN uses a feature extraction subnet, a feature fusion subnet, and a softmax layer. In particular, each textured 3D face scan is represented as six types of 2D facial attribute maps, all of which are jointly fed into DF-CNN for feature learning and fusion learning, which results in a highly concentrated facial representation (32- dimensional). Prediction is calculated in two ways: 1) learning linear SVM classifiers using the 32-dimensional fused deep features; 2) directly performing softmax prediction using the 6-dimensional expression probability vectors. Very Different from existing 3D FER methods, DF-CNN combines feature learning and fusion learning into a single end-to-end training framework. To demonstrate the effectiveness of DF-CNN, conducted comprehensive experiments to compare the performance of DFCNN with handcrafted features, pre-trained deep features, fine-tuned deep features, and on three 3D face datasets. Usually, in most of the cases, DF-CNN consistently achieved the best results. This is the first work of introducing deep CNN to 3D FER and deep learning-based feature-level fusion for multi-modal 2D+3D FER.

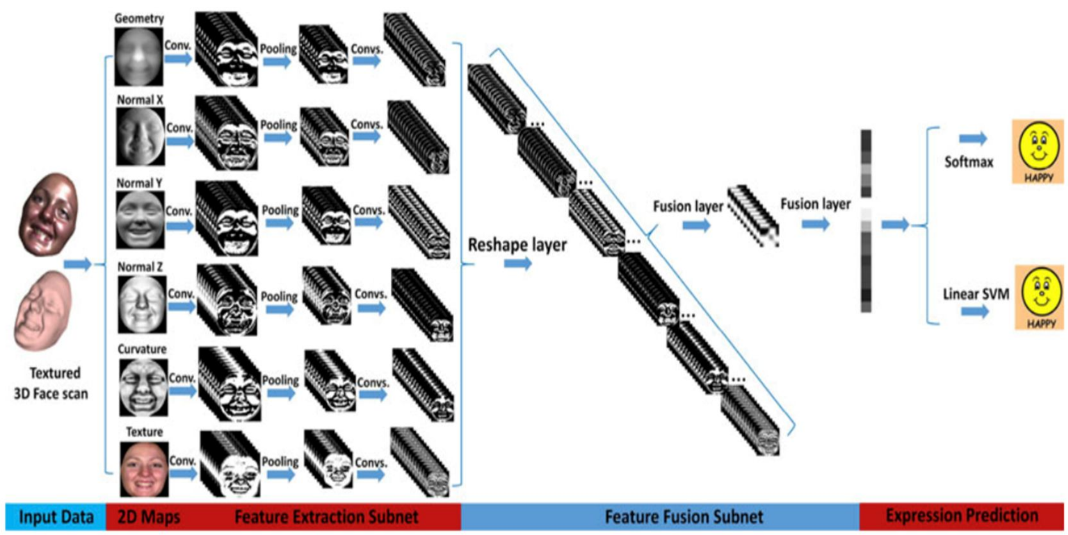



Fig 3. Pipeline of the proposed DF-CNN based multimodal 2D+3D FER approach

*1) Advantages*

*a)* Improves the localization of the salient patches on face images.

*b)* Predicts expression recognition in low resolution images.

*2) Disadvantages*

*a)* Requires huge data storage.

*b)* Requires huge computational power.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

*F.* Due to the different modalities emotions could be expressed, automatic affect recognition might be a challenging task. Applications can be found in various domains, including multimedia retrieval as well as human-computer interaction. Past few years, deep neural networks have had great success in determining emotional states. Due to this success, the paper proposes an emotion recognition system using auditory and visual modalities. To capture the emotional content for different styles of robust, speaking features have to be extracted. We use a CNN (convolutional neural network) to extract features from the speech, while for the visual modality, a deep residual network of fifty layers is employed. Additionally, to the importance of feature extraction, a machine learning algorithm needs also to be insensitive to outliers while having the ability to model the context. Long short term memory networks are utilized in order to tackle this problem. The system is then trained in an end-to-end fashion where—making use of an advantage of the correlations of each of the streams— that we manage to significantly outperform, on the lines of concordance coefficient of correlation, traditional approaches supported visually handcrafted as well as auditory features for the prediction of spontaneous and natural emotions on the RECOLA database of the AVEC 2016 research challenge on emotion recognition.

*G.* An unsolved problem in computer vision is the Automated emotion recognition from facial images. Although Recent methods achieve a near-human accuracy under controlled scenarios, the popularity of emotions within the wild remains a challenging problem. Looking at the advances in Deep learning has supposed a breakthrough in many computer vision tasks, including facial expression analysis. Notably, the Deep Convolutional Neural Networks usage has reached the most straightforward ends up in the recent public challenges. This state of the art algorithms brings out the usage of ensembles of CNNs, which can outperform an individual CNN classifier. Two key considerations that are a significant influence on the results are: (i) the ultimate classification rule that assembles the results of the committee, and (ii) the planning of CNN's involves the adjustment of parameters that allow diversity and complementarity within the partial classification results. At the time of this paper, they propose enhancing the committee assembling with the introduction of supervised learning for the ensemble computation. The validation reveals an accuracy five percent higher with regards to the previous state of the art results supported averaging classifiers and a four percent to the bulk voting rule.

*H.* One of the first critical features of a human being is facial expressions, which can be used to determine the emotional state at a particular moment. The paper uses the Deep Neural Network and CNN to develop a facial emotion recognition model that categorizes a facial expression into seven various emotions: Surprised, Sad, Neutral, Happy, Disgusted, Angry, and Afraid. This paper compares the performance of 2 existing deep neural network architectures with our proposed architecture, the Venturi Architecture on the bases of training testing loss, training-loss, testing accuracy, and accuracy. The paper makes use of the Karolinska Directed Emotional Faces dataset containing a collection of 4900 pictures of human facial expressions. 2 layers of feature maps were used to convolute the features from the images, and then it was passed on to the deep neural network with up to 6 hidden layers. The proposed Venturi architecture shows significant accuracy improvement compared to the modified triangular architecture and rectangular architecture.

*I.* Here, in this paper, they have used a method of finding out the coefficients describing elements of facial expressions. The features could also be taken from a 3D model using a Microsoft Kinect sensor. The classification of the features were performed using K-NN and MLP Neural network. Light conditions and changes of the head position are the main factors that affect the quality of the emotion recognition systems, therefore, methods in which 3D models are used are far more promising. In this experiment, Microsoft Kinect is used for 3D face modelling which has a infrared emitter and two cameras ( one records visible light and the other operates in IR which measure the depth). The Kinect projects 121 strategically located points on the face which characterises the positions of the muscles on the face which is responsible for emotions. These are called the "Action Units". A total of six action units are provided which detects the muscle movements. The distance between each AUs tells us the type of emotions expressed.

| ES | neutral | joy | surprise | anger | sadness | fear | disgust |
|---|---|---|---|---|---|---|---|

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AU0 | 0.21 | 0.77 | -0.10 | 0.30 | 0.17 | -0.11 | 0.91 |
| AU1 | -0.06 | 0.09 | 0.60 | -0.07 | -0.04 | 0.20 | 0.13 |
| AU2 | -0.25 | 1.00 | -0.49 | 0.06 | -0.37 | -0.60 | 0.88 |
| AU3 | -0.21 | 0.00 | -0.13 | 0.04 | -0.09 | -0.17 | 0.00 |
| AU4 | -0.04 | -0.47 | 0.58 | -0.19 | -0.02 | 0.28 | -0.32 |
| AU5 | -0.23 | -0.30 | 0.10 | -0.34 | -0.27 | -0.02 | -0.39 |

Fig 4. Facial Expressions and corresponding AU

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue V May 2020- Available at www.ijraset.com*

## III. CONCLUSION

We chose [13] as our reference paper as the method used here (2GF+CNN) provides us with a clear understanding methodology of the feature extraction by the Gabor Filter and the classification of emotions using the CNN. This method, along with another experiment without using the Gabor Filter allows us to gain a tangible evidence on how the accuracy and the performance is increased by the Gabor Filter. After the Gabor filter applied, the system learning became faster, and the accuracy has improved. The learning speed of the convolutional neural network has increased profoundly. This is because the Gabor filter extracts the image sub-feature and gives the neural network. The convolutional neural network receives several sub-features and takes one step further in extracting the emotions from the faces.

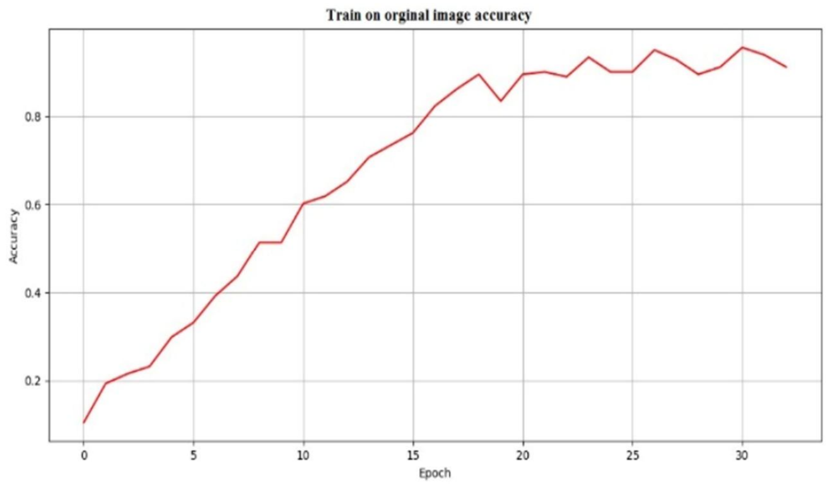



Fig 5. Accuracy with only CNN

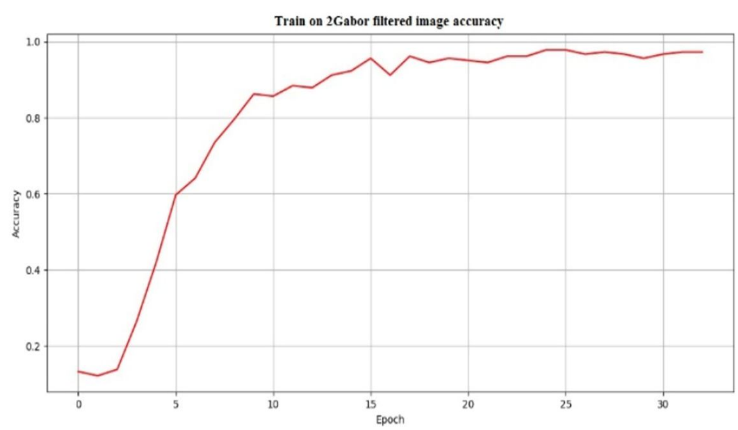


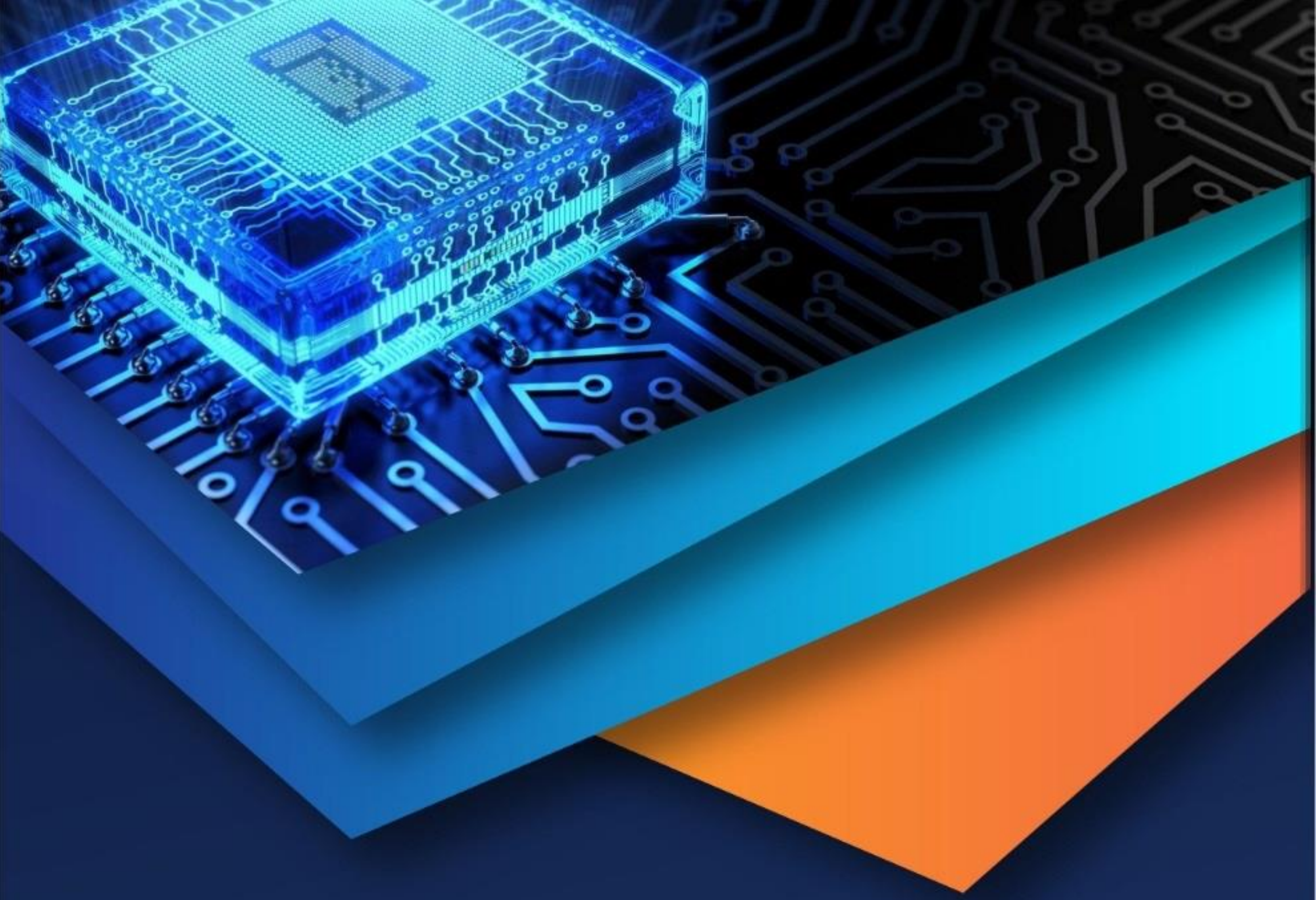

Fig 6. Accuracy with 2Gabor Filter + CNN

Indeed, the classification accuracy was influenced by the way users play specific facial expressions. In real conditions, classification accuracy can be affected by many additional factors. When you feel real emotions, facial expressions can vary greatly - may be exposed to a greater or lesser extent. The advantage of the proposed solution is that it achieves high face detection rates and real-time performance due to no exhaustive searching on the whole image.

In the above-proposed solutions, each solution is based on different environmental conditions and situations. The quality and the accuracy of the solutions working is questionable and most certainly arguable. Nevertheless, in the stepping stones of the development towards a cleaner and accurate classifiers, these solutions prove to be solid ground foundations.

## REFERENCES

[1] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," in IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1-12, 1 Jan.-March 2015, doi: 10.1109/TAFFC.2014.2386334.
[2] A. Nicolai and A. Choi, "Facial Emotion Recognition Using Fuzzy Systems," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, 2015, pp. 2216-2221, doi: 10.1109/SMC.2015.387.
[3] M. Imani and G. A. Montazer, "GLCM features and fuzzy nearest neighbor classifier for emotion recognition from face," 2017 7th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, 2017, pp. 8-13, doi: 10.1109/ICCKE.2017.8167879.
[4] Kwolek B. (2005) Face Detection Using Convolutional Neural Networks and Gabor Filters. In: Duch W., Kacprzyk J., Oja E., Zadrożny S. (eds) Artificial Neural Networks: Biological Inspirations – ICANN 2005. ICANN 2005. Lecture Notes in Computer Science, vol 3696. Springer, Berlin, Heidelberg
[5] H. Li, J. Sun, Z. Xu and L. Chen, "Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network," in IEEE Transactions on Multimedia, vol. 19, no. 12, pp. 2816-2831, Dec. 2017, doi: 10.1109/TMM.2017.2713408.
[6] Tzirakis, Panagiotis & Trigeorgis, George & Nicolaou, Mihalis & Schuller, Björn & Zafeiriou, Stefanos. (2017). End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. IEEE Journal of Selected Topics in Signal Processing. PP. 10.1109/JSTSP.2017.2764438.
[7] G. Pons and D. Masip, "Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis," in IEEE Transactions on Affective Computing, vol. 9, no. 3, pp. 343-350, 1 July-Sept. 2018, doi: 10.1109/TAFFC.2017.2753235.
[8] A. Verma, P. Singh and J. S. Rani Alex, "Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), Osijek, Croatia, 2019, pp. 169-173, doi: 10.1109/IWSSIP.2019.8787215.
[9] Tarnowski, Paweł & Kołodziej, Marcin & Majkowski, Andrzej & Rak, Remigiusz. (2017). Emotion recognition using facial expressions. Procedia Computer Science. 108. 1175-1184. 10.1016/j.procs.2017.05.025.
[10] M. M. Taghi Zadeh, M. Imani and B. Majidi, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Tehran, Iran, 2019, pp. 577-581, doi: 10.1109/KBEI.2019.8734943.




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY