Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 10, October 2020, pg.01 – 09

Human Emotions Detection Using Hybrid CNN Approach

Nehmat Sandhu[1]; Aksh Malhotra[2]; Mandeep Kaur Bedi[3]

[1]Student, Strawberry Field High School, Sector 26, Chandigarh

[2,3]Mentor, Astirmind Solutions, Ludhiana

[1] Nehmatsandhu2412@gmail.com; [2] aksh.astirmind@gmail.com; [3] bedimandy89@gmail.com

DOI: 10.47760/IJCSMC.2020.v09i10.001

Abstract- Automated Facial emotion detection is a challenging task during human-computer interaction. In this paper, we have used the hybrid CNN approach to recognize human emotions and based upon its features, categorized them into sub-categories. This research uses a FER13 dataset for emotion recognition and trained our model accordingly to get optimal results in terms of accuracy and loss. The system performance gains average accuracy rate of 88.10%. The system has been successfully recognized seven basic emotion classes. Thus, the proposed method is proven to be effective in terms of more accuracy and minimum loss for face emotion detection.

Keywords- Image processing, facial emotion recognition, CNN, Haar Cascade, Image processing

I. INTRODUCTION

In this digital era, the latest and advanced technologies are emerging day by day. Over the past two decades, emotional facial detection has attained enormous attention. With the passage of the time and demand of the technology, many more things have been added up in the technology to make the behavioral system more secure and safe. The methods based on user behavior are usually assumed as behavioral systems.

Facial emotions examine the features of images of people's faces and give input through a digital video camera. It processes the complete facial structure, comprising distances between eyes, mouth, nose, and jaw edges. It plays a vital role in emotion recognition and detection to identify people and non-verbal communication, as well [1]. These dimensions are kept in a database and used as a contrast when a user positions before the camera. This biometric has been extended, and possibly passionately, touted as an eccentric system for recognizing potential, but so far has not seen wide recognition in high-level usage.Facial expressions make the platform for human-computer interaction. Emotions keep on changing continuously based upon the happenings that are induced as a result of impelling force. Therefore, in a real-life scenario, the detection of emotion is a challenging task [2].

It is predictable that biometric facial recognition technology will soon overtake impression biometrics as the most general form of user verification. Each face has several different breakthroughs, the dissimilar peaks and valleys that make up facial types. Each human face has around 80nodal points. The human face is a multi-dimensional that carries a lot of information about individuals, their feelings, and expressions [4]. Some of these dignified by the Facial Recognition Technologies are:

1. Space between the eyes
2. Nose Width
3. Eye sockets depth
4. Cheekbones Shapes
5. Jawlines

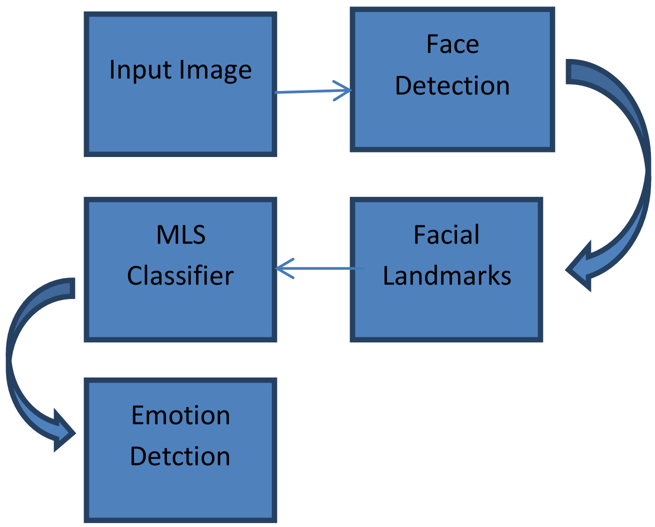

Fig. 1: Process of Emotion Detection

These nodal points are calculated based on a numerical code, called a faceprint, on behalf of the database's image. Face detection is a computer-based technology that identifies the location and the size of the human face in an image [4].

1.1 Human Emotion Detection

People can immediately notice impressions, whether he is happy or sad. The main question is, can a computer do the same? What if it could do better than the human, looking absurd, but it is possible.

Human emotions can be categorized as happy, sad, anger, surprise, disgust, neutral, and contempt. These emotions are not constant. To find out the difference between these emotions can be very challenging; even a minor difference results in different expressions [5]. Furthermore, we will focus on those portions of the face, reflecting the maximum expressions like around the mouth and eyes.

Machine learning and a neural network have been used for various tasks and to get the optimal results. Neural network approaches are highly useful to get in pattern recognition and classification. Features are the most critical aspects of any machine learning algorithm. The following four-stage process demonstrates how emotion-detection works:
1. Data Preprocessing
2. Face Detection
3. Feature Extraction
4. Classification based on the features

One of the main components of behavioral biometrics is the facial emotion recognition and its intensity.

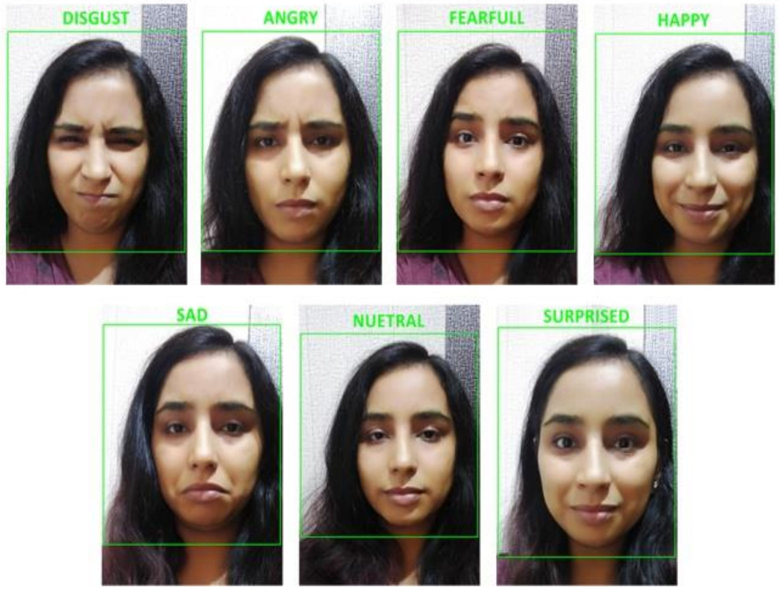

Fig. 2. Face Emotion Classifications

In Fig. 2 face emotions are classified into seven categories/ classes. These are Angry, happy, sad, disgusted, fearful, neutral, and surprised.

How it works: The following four-stage process demonstrates the method of biometric systems function:

1. Capture- The system takes a physical or an interactive sample during registration.
2. Extraction- Unique data is taken out from the section, and a template is produced.
3. Comparison- The template is then associated with a new sample.
4. Matching- The system, then chooses if the types removed from the new sample are similar or not.

In the upcoming sections,we will discuss about the related work, various methods, results, and analysis for emotion detection using CNN.

II. RELATED WORKS

Facial emotion recognition has been used in various applications such as criminal detection.

Ming et. al. [6] discussed three stages of facial expression recognition. These are facial image registration, feature extraction, and facial expression classification.This paper explained how these methods can be used for finding out the emotions of a person successfully.

Samira et al. [7] have used recurrent neural networks along with Convoluted Neural network in the form of CNN-RNN architecture to identify emotion in the video. Some other noble methods are also used in this approach to get the best and optimal results in terms of accuracy, performance, stability, and realistic. Some of the methods are time consuming and need more computing power to generate the precise solutions, but in some ways need to compromise accuracy over performance.

A. Yao et. al [8] represented a well-designed CNN architecture to recognize emotion. They proposed HOLONET methods that have three critical considerations in the network design. First is to reduce redundant filters and enhanced the non-linearity in the lower layers. They used to modify (CReLU) as an alternative of ReLU. Second, enhance accuracy and maintain efficiency in the middle layer.Third, it broadens network width and introduce the multi-scale feature extraction property.The top layers work as variants of the inception-residual structure.This method **is** more accurate than other methods.

Yelinet.al [9] discussed three layers Deep Belief Networks that gives better performance than two layered DBN's with the help of audio-video emotion recognition process. It gives better output and accuracy as compared to the existing approach.

D Y Liliana [10] used Deep CNN approach to get better emotion detection results. The author has mainly recognized 8 facial classes with the help of CK+ database that trained using different training data size. As a result, they got a 92.81% accuracy rate and improved performance.

III. METHODS

In our proposed work, we have used CNN-based using open-source dataset FER-2013.
We have proposed a hybrid CNN model and compared its accuracy and model loss over 7 different data sets to ensure its robustness.

3.1. Convolutional Neural Network

Convolutional Neural Networks (CNN) is a neural network architecture having multilayer. CNN input and output array vectors are known as a feature map. For instance, audio input has a 1D array and text input only. On the other hand, the image has a 2D array. The output map features depict the feature extraction from the given input. This CNN model has mainly three layers:

1. Convolutional filter layer
2. Subsampling
3. Classification layer

CNN considers as one of the best networks to proceed static images, attain higher accuracy, low computational cost and has a capacity to handle variable of images. No need to change the network when your dataset changes, and process data in batches. CNN can handle every kind of dataset with the least preprocessing rather than other networks such as multi-layer perceptron and recurrent neural network.

3.2 Haar Cascade

Haar Cascade classifier depends on the Haar Wavelet technique to examine the image's pixels into squares by function.It uses "integral image" concepts to detect the features. Haar Cascade uses the Ada-boost learning algorithm to choose the small numbers of important features from a large set to give an efficient result of classifiers with the help of cascading methods to detect the face in an image [12].

Haar cascade classifier works on the Viola-Jones detection algorithm. It is trained by some input faces and non-faces and trained a classifier that identifies a face.

IV. EXPERIMENTS

We have used the FER13 database and trained our model with 36000 images to get the optimal results. Fig. 3, represents the architecture of the proposed model.

4.1 Dataset used

In this work, we used the FER13 dataset to recognize all the facial expressions. It is an open-source data set that contains 35.887 grayscale, 48x48 sized face images with various emotions.Here, we have trained approximately 36000 images to detect the emotions from the faces.

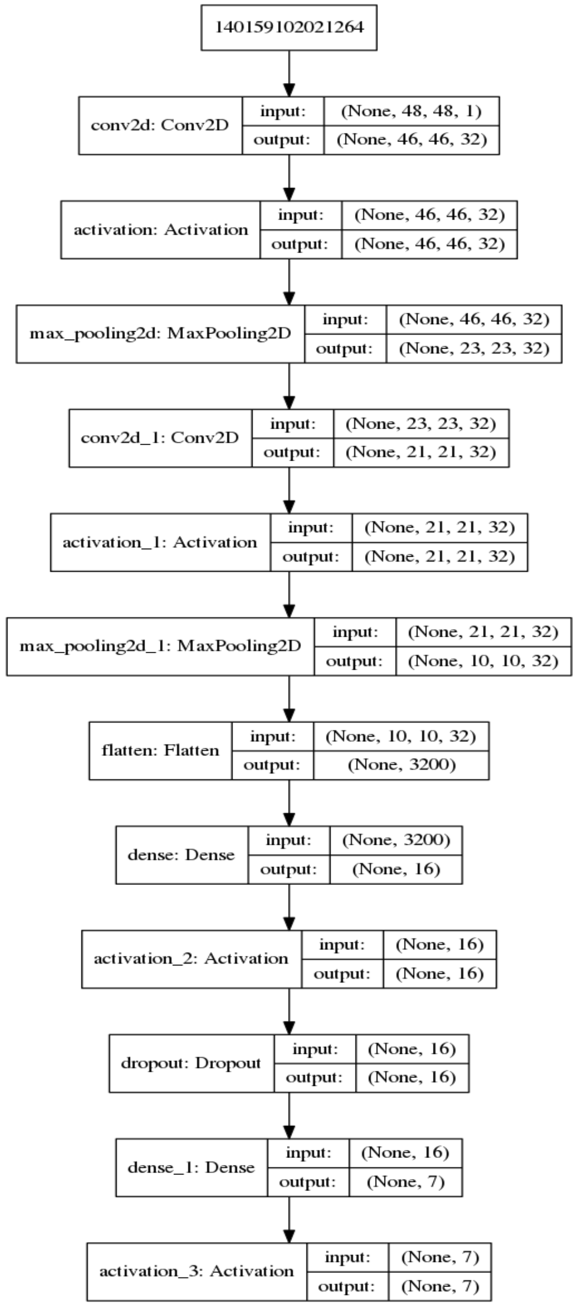

Fig 3. Architecture of Hybrid CNN

Our hybrid CNN approach is working as follow:

- First, the haar cascade method is used to detect faces in each frame of the webcam feed.
- The region of the image containing the face is resized to 48x48 and is passed as input to the CNN.
- The network outputs a list of softmax scores for the seven classes of emotions.
- The emotion with maximum score is displayed on the screen.

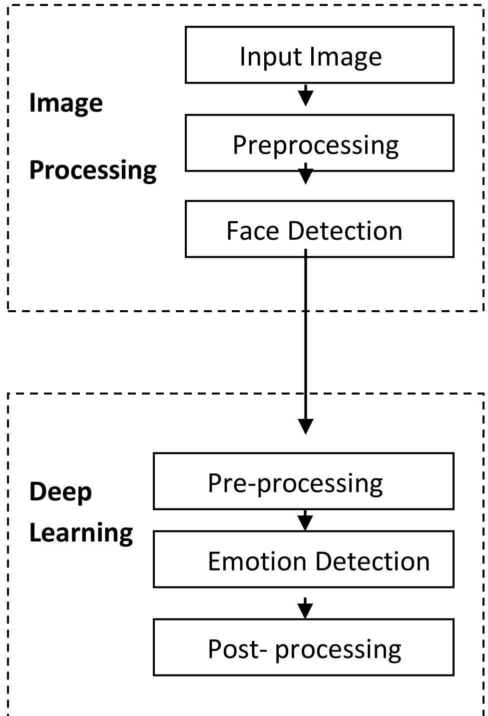

Fig 4. Flowchart of proposed model

1. Input Image: Input image will be of any size. The image will be in the form of a video frame or image. Input will feed using OpenCV, a python library.

2. Image preprocessing: In this step input image or the video frame will be preprocessed using some image processing techniques in which image compression, image cleaning, and image resizing will be included.

3*. *Face Detection: Face Detection will be performed using a cascade classifier. In this work we used harcasede.xml for face detection in the image and further on the detected face in this step, emotion detection will be performed.

4. Preprocessing: In this preprocessing face some data preprocessing steps are followed in which image grayscaling batch normalization is performed by using image generators, In our model batch size, is 64. Iimage size is 48*48

5. Emotion detection: In this phase emotion detection using deep learning techniques is performed. As our approach is to apply a multi-layer convolution neural network. In this phase grayscale image with a size of 48*48 will be feed to the trained model and face emotion is classified into 7 classes

6. Post-processing. After applying image processing and we used deep learning technique for image finally and detect the emotion accordingly.

V. RESULTS AND DISCUSSIONS

Our proposed method shows better accuracy and minimum model loss during emotion detection from an image as compared to the existing methods.

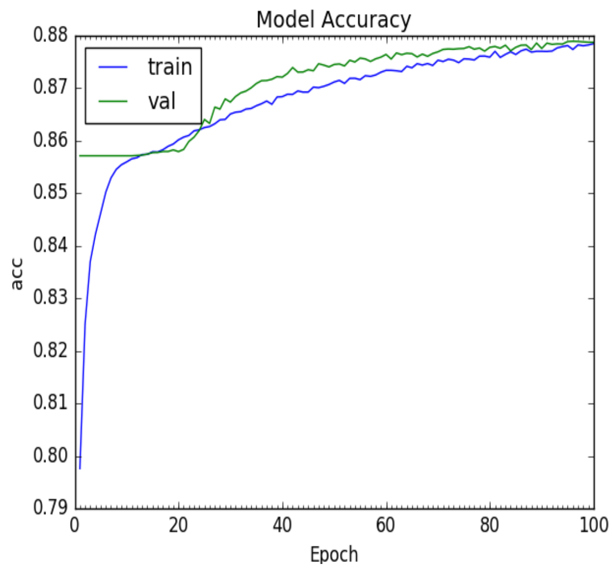

Fig. 5 Accuracy

We have trained our model to 100 epochs and accuracy is continuously increasing, using a deep multilayer convolution network. In fig 5, we attained above 88% training and validation accuracy after 100 epochs

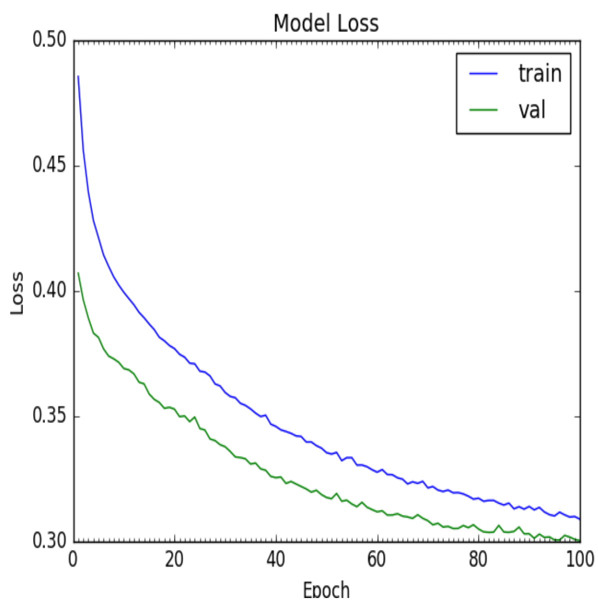

Fig. 6 Model Loss

Fig. 6 depicts, similarly for model loss with 100 epochs, by using binary_crossentropy loss function model loss decrease with epochs and we attained loss less than 0.30 with 100 epochs.

By calculating the element of each data with the help of data measurement tools, we have got different results. The accuracy rate for the angry class is 94.70%; disgusted 72.03%; fear 88%; happy 90.06%; neutral 93%; sad 87.01%, and surprised 77.4%.

TABLE I: EXPERIMENTAL RESULTS

Emotion Class	Accuracy Rate
Angry	94.70%
Disgusted	72.03%
Fear	88%
Happy	90.06%
Neutral	93%
Sad	87.01%
Surprised	77.4%

The average accuracy rate for the entire model is 86.10%

VI. Conclusion

Humans are capable to show thousands of expressions during their communication that depend upon mood, person, behavior, and many more factors.In this paper, we have discussed human emotion detection process using a neural network approach such as CNN. The features have been calculated for the three-dimensional face model. The classification of features was performed using Haar Cascading. Our experimental results show that our proposed hybrid model can detect facial emotion from an image with average accuracy of 86.10% with less than .30% model loss which is quite good as compared to the existing methods.

In the future, we can use this method on multiple platforms such social media platform to detect emotions during chats or video calls. We can use this method on various social media applications such as Twitter, Whatsapp, Messenger etc. By using this technology, we can detect mental health illness of a particular person by detecting his/her emotions while doing chats or video calls and save him from the trauma. Thus, our proposed method can be used in health sector as well.

REFERENCES

[1]. Tarnowski, P., Kolodziej, M., Majkowski, A., &Rak, R. J. (2017, June). Emotion recognition using facial expressions. In *ICCS* (pp. 1175-1184).

[2]. Dagar, D., Hudait, A., Tripathy, H. K., & Das, M. N. (2016, May). Automatic emotion detection model from facial expression. In *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)* (pp. 77-85). IEEE.

[3]. Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: a review. *Artificial Intelligence Review*, *52*(2), 927-948.

[4]. Bah, S. M., & Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. *Array*, *5*, 100014.

[5]. Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1867-1874).

[6]. Ming, Z., Bugeau, A., Rouas, J. L., &Shochi, T. (2015, May). Facial action units intensity estimation by the fusion of features with multi-kernel support vector machine. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 6, pp. 1-6). IEEE.

[7]. Wöllmer, M., Metallinou, A., Eyben, F., Schuller, B., & Narayanan, S. (2010). Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional lstm modeling. In *Proc. INTERSPEECH 2010, Makuhari, Japan* (pp. 2362-2365).

[8]. Yao, A., Cai, D., Hu, P., Wang, S., Sha, L., & Chen, Y. (2016, October). HoloNet: towards robust emotion recognition in the wild. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 472-478).

[9]. Kim, Y., Lee, H., & Provost, E. M. (2013, May). Deep learning for robust feature generation in audiovisual emotion recognition. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 3687-3691). IEEE.

[10]. Liliana, D. Y. (2019, April). Emotion recognition from facial expression using deep convolutional neural network. In *Journal of Physics: Conference Series* (Vol. 1193, No. 1, p. 012004). IOP Publishing.

[11]. Williams RJ, Zipser D (1989) A learning algorithm for continually running fully recurrent neural networks. *Neural Comput*1(2):270–280.

[12]. Phuc, L. T. H., Jeon, H., Truong, N. T. N., &Hak, J. J. (2019). Applying the Haar-cascade Algorithm for Detecting Safety Equipment in Safety Management Systems for Multiple Working Environments. *Electronics*, *8*(10), 10.

