An Intelligent Approach for Predicting Emotion Using Convolution Neural Network

Sumit Das¹, Manas Kumar Synyal², Sourav Kumar Upadhyay¹, Supriyo Chatterjee¹

¹ JIS College of Engineering, Information Technology, Kalyani, 741235, India
² University of Kalyani, Department of Business Administration, 741235, India

sumit.das@jiscollege.ac.in

Abstract. Emotions have started controlling not only the way as humans; interact with other living beings but also the way we interact with computers. Emotions have started controlling our every decision like going to a shop again, purchasing a particular product, helping a person, and many others. The developments in the field of artificial intelligence and computer vision have further attracted the attention of people to work in this area. In this work we have designed a real-time emotion recognition system that can recognize the emotions of a person from his facial expressions. The system uses machine learning techniques for performing the task and has been implemented using the python programming language. This system can have a lot of useful applications in real-life scenarios such as old age health monitoring, determining the comfort level of a patient during medical treatment, recognizing the emotions in patients suffering from neurological diseases, not dispensing the money from an ATM if the withdrawer is nervous, determining tiredness or sleepiness during driving and raising an alert, facial emotion detection in interviews, taking feedback of customers visiting a store and later using that for fine-tuning market strategies and many others. It believes this work will be very useful to critical heath care monitoring and management in this advance era of artificial intelligence.

Keyword. Convolutional Neural Network(CNN), Facial Action Coding System(FACS), Facial Expression Recognition, Graphics Processing Unit(GPU), Linear Discriminant Analysis(LDA) and Principal Component Analysis(PCA).

1. Introduction
The machine learning is the emerging field in the era of artificial intelligence and soft computing, where emotional intelligence playing a significant role to percept the emotion of people. The fact that emotions play a crucial role in the life of living beings is undisputed. It is taught from our childhood days that our emotions should not control our actions. The reason for this is that emotions can affect the outcomes of an action. The most important are emotions that play an important role in interpersonal relationships. In short, emotions play a great role in motivation, information, and communication. The importance of emotions, as discussed above in brief, has made ‘emotion recognition’ one of the most interesting areas of research. Human emotions and intentions are expressed through face expressions and deriving an efficient and effective feature is that the
The fundamental component of the facial expression system. Facial expressions convey non-verbal cues, which play a vital role in interpersonal relations. Automatic recognition of facial expressions will be a vital component of natural human-machine interfaces; it should even be employed in behavioral science and in clinical practice. An automatic face expression recognition system has to solve the subsequent problems: detection and placement of faces in a very cluttered scene, facial feature extraction, and face expression classification.

In this article, facial expression recognition, the system is implemented using a convolution neural network. Facial images are classified into seven face expression categories namely Anger, Disgust, Fear, Happy, Sad, Surprise, and 'Neutral. A dataset from Kaggle has been used to train and test the classifier.

The proposed system is a deep learning application, which has been trained using the dataset from the Kaggle Emotion Recognition Challenge. The application captures the image of the face of the user through the webcam and thereafter predicts the emotion by analyzing the facial expression. The result is displayed alongside the window in which the face is visible. The result is in terms of the percentage chance that the user is having a particular emotion. This emotion recognition is done in real-time and this system, with a slight change, can be used to recognize emotions from the images taken by any other camera.

Functional Requirements are that the application should capture images clearly and the recognition of emotion should be on a real-time basis. In this work the inputs are data set and user data; the outputs are results and accuracy will be elaborated.

2. Literature Survey
The artificial intelligence[1] and machine learning[2] are the emerging application in various fields such as intelligent healthcare system, several subcomponents of industry and research and development sectors. A paper in face recognition research[3], introduces the current advances and applications in countenance recognition from face detection, trait extraction, categorization, and ethnic expression recognition. Facial features recognition is one among the new a skin condition in recent years. It applies in the emotional analysis, pattern recognition, and societal interaction. The strategies of trait extraction were divided into many totally different characteristic classes. In line with the facial expression recognition history and achievements, the occurrence of ethnic countenance recognition and also the trend of face expression recognition are given.

In a article of infants and kid of the insane, of picture and monument, of cats and dogs and monkeys, and of the ways during which folks in numerous culture specific their feelings, Darwin's insight not been surpassed by smart science. This perfect edition of Darwin's masterwork contains a substantial new opening and afterword by Paul Ekman [4]. The optical flow emotionless face recognition article [5] presents a way that uses optical flow to educated guess striated muscle actions, which can then be recognized as facial expressions. Facial expressions were the results of muscular actions that were activated by the nerve impulse produced by emotions. The muscle actions origin the movement and twist of facial skin and facial expressions. Since facial skin has the texture of a fine-grained organ that helps in dig out the optical flow, we are going to extract muscle actions from the exterior appearance. The scientists were ready to construct a countenance recognition system supported optical flow in sequence. They explore the recognition methodology, the optical-flow fields of skin movement were assessed in muscle win sows, and each of that defines one most important route of shortening to correctly pull out muscle movement. A fifteen-dimensional characteristic vector is utilized to signify the leading active points in terms of the flow discrepancy from end to end and local unique areas. The initial research indicates associate correctness of roughly sorts of expressions such as pleasure, annoyance, hatred, and shock.
The Coding-analysis-interpretation emotion recognition paper illustrates a computer vision scheme for monitoring facial movement by using a best possible inference optical flow process together with geometric, corporeal and gesture based vibrant models describing the facial formation. This system produces a dependable parametric demonstration of the face-independent muscle stroke group, as well as an precise estimate of facial gesture[6].

In an article, scientists are extensively applied pattern recognition because of the discrimination power and robustness [7]. The hidden markov model is a extremely consistent way of recognition. In this study, the scientists have projected a way of using moment invariants as description and hidden markov model as detection process in facial expression detection. The series of four common expressions are recognized and achieve precision as far above the ground as 96.77%.

In a study scientist proposes a framework, real-time face gesture recognition in the interactive computation. The main contributions of this work, they projected a network configuration and features learning for embed hidden-markov-model and apply this optimized embedded hidden-markov-model for recognition. In this article, the embedded hidden-markov-model uses 2D discrete cosine transform coefficients as the examination vectors conflicting to previous hidden-markov-model, which use image pixel intensities to shape the inspection vectors. This anticipated system reduces the complexity of the system. It offers flexible structure and used in real-time applications. Experimental results make obvious that the projected approach is an efficient process to recognize facial appearance [8]. In another paper, Scientist proposes the way of facial features of ethnic minorities based on face detection expertise. The researchers formed a face database repository of ethnic minorities and extract facial trait using face detection technique. Then they built a deformable model of those characteristics to obtain all facial features. In addition, the multi feature classifier is accepted to carry out learning on the dataset and eventually discriminate facial characteristics of ethnic minorities [9].

3. Methodology

The facial emotion expression detection technique is implemented with convolutional artificial neural net (CNN). The building block diagram of the framework is shown in following figures:

![Figure 1 Training Phase of the model](image-url)
At the time of training, the structure take input a training dataset containing gray-scale representation images of faces by their own appearance label and learns a grouping of weights for the CNN and training step took as input a picture with a face. After that, intensity normalization is applied to the image and the normalized images are went to train the CNN. To make sure that the training show is not overwhelmed by the order of presentation of the examples. The subset of dataset that is validation dataset is in use to settle on the decisive best set of weights out of a group of trainings executed with samples obtainable in numerous orders. The output of the training step may be a set of weights that achieve the most effective result with the training data. Again at the time of testing, the structure is taken a gray-scale image of a face from test dataset, and output the expected expression by using the ultimate network weights learned throughout training. Its output could be a single number that represents one among the seven basic expressions.

The figure 3 depicts the architecture of CNN, which contains an input-layer, several convolutional-layers, various fully-connected-layers, and finally an output-layer. The CNN is meant with some modification on LeNet Architecture [10]. The architecture of the Convolution Neural Network utilized in the model.
**Input-Layer:** The input-layer has a predetermined, fixed-dimension, where the image have to be pre-processed prior to it fed into the layer. The normalized gray-scale images of size 48 by 48 pictures from the Kaggle are used for the training, validation, as well as testing. For testing propose laptop webcam images also are used, within which face is detected and cropped using OpenCV Haar Cascade Classifier and normalized.

**Convolution and Pooling Layer:** Convolution and pooling is finished supported execution. Here every group has N images and CNN filter weights are updated on those batches. Each convolution layer takes image group input of 4-dimension N x Colour-Channel x The width x The height. The feature map or filter for convolution is additionally four-dimensional such input features map number, output features map number, width of filter and height of filter. In each convolution layer, 4D convolution is computed between image groups and perform mapping. After convolution, the sole arguments that change are picture width as well as height. New image width = old image width – filter width + 1 New image height = old image height – filter height + 1. After each convolution layer, down-sampling/sub-sampling is completed for dimensionality reduction. This process is named as the pooling. The maximum-pooling and mean-pooling are two well-known pooling techniques. During this work, max pooling is completed after the convolution. The pool-size of 2 by 2 is taken, which splits the image into a grid of blocks each of size 2x2 and takes a maximum of 4 pixels. After pooling only height and width are affected. Two convolution layer and pooling layer are utilized in the architecture. At the primary convolution layer size of the input, the image batch-size is Nx1x48x48. Here, the dimensions of the image batch-size is N. The quantity of colour-channels is 1 and the both image height-width are 48-pixels. The convolution with a feature map of 1x20x5x5 results image batch is of size Nx20x44x44. After the convolution pooling is completed with a pool size of 2 by 2, which ends up in a picture group of size Nx20x22x22. this can be followed by the second the convolution layer with a feature map of 20x20x5x5, which ends up in a picture batch of size N20x18x18. this is often followed by a pooling layer with pool size 2 by 2, which ends from a picture group of size N20x9x9.

**Fully Connected Layer:** The fully connected layer is stimulated by the way neurons spread signals throughout the brain. This layer receives the number of input traits and converts features all through layers associated with trainable weight vectors. There are two hidden layers, the size of fist one is 500 units and the second one is 300 units which are applied in the fully-connected layer. The weights of CNN are trained by feed forward propagation and then feed backward propagation is used to propagate errors in back. The Back-propagation starts for calculating the difference between forecast and true value and back computation of the load fine-tuning needed to each layer, which are capable to manage the training swiftness and consequently the complexity of the design by fine tuning the excited parameter. These layers include learning rate, momentum, regularization parameter, and decay

**Output Layer:** The output from the second-hidden layer is associated to the output layer having 7-different classes. Using the Soft-max activation function, the output is acquired using the chances for each of the seven classes. The category with the best likelihood is that the prediction of expected class.

4. Implementation

The input layer has predetermined and fixed dimensions. Therefore image have to be pre-processed prior to it feed into the layer. The normalized gray-scale images of size 48by48 pixels from Kaggle dataset are utilised for training, validation and testing. For testing propose laptop webcam images are also used, in which face is detected and cropped using OpenCV Haar Cascade Classifier and normalized.

Now let us dig deeper into the training datasets, at first batch normalization is done to improve the speed, performance and stability of ANN here some simple mean and standard deviation is occurred and the same is done with the testing set also. After that re-scaling and re-centering is done. Next the
convolution layers will be introduced and two of the initial convolution layer will be consisting of 64 nodes and the third one will be having 128 numbers of nodes in each of the layer max-pooling will be introduced at the end of the layer, it is done for down sampling on input representation for reduction of dimensionality and allowing for assumptions to be made about feature contained in the sub regions. After that the global average will be calculated and after that Soft-max procedure will be done. Now what is soft-max, well it is a squashing function which limits the output into range of 0-1 means directly interpreted as probability and it helps to understand the decisions of the model for every given input and why the model is making such decisions by analyzing the probability. in the algorithm the loss reduction technique has been used to enhance the efficiency of the model. If the efficiency of the model is not satisfactory then number values in the dataset has to be increased or the quality of the data sets need to be improved. For that image pre-processing can be done here each of the image can be distorted in various ways to generate a new set of images in this way the input dataset quality can be enhanced. After generating the output the result can be stored in a file for later usage and in this way the entire model works.

The dataset from a Kaggle Facial Expression Recognition Challenge (FER2013) [11] is used for the training and testing. It contains pre-cropped, 48-by-48-pixel gray-scale images of faces each labelled with one of the seven emotion classes such as (i) anger, (ii) disgust, (iii) fear, (iv) happiness, (v) sadness, (vi) surprise, and (vii) neutral. Dataset has training set of 35887 facial images with facial expression labels. The dataset has class imbalance issue, since some classes have large number of examples while some has few. The dataset is balanced using oversampling, by increasing numbers in minority classes. The balanced dataset contains 40263 images, from which 29263 images are used for training, 6000 images are applied for testing, and 5000 images are applied for validation.

5. Result and Discussion

The emotion of the people is very important in this advance AI era, the sub component AI that is machine learning are the significant part of medical diagnosis[12]. CNN architecture for face expression recognition as mentioned above was implemented in Python. Together with Python programming language, OpenCV, NumPy, pandas, Keras, imutils, scikit-learn, tensor flow, and CUDA libraries [13] were used. The training image batch size was taken as 30, while the filter map is of size 20x5x5 for both the convolution layer. A validation set was won to validate the training process. Within the last batch of each epoch invalidation cost, validation error, training cost, training error is calculated. Input parameters for training are image set and corresponding output labels. The training process updated the weights of feature maps and hidden layers supported hyper-parameters like learning rate, momentum, regularization, and decay. During this system, the batch-wise learning rate was used as 10e-5, momentum as 0.99, and regularization as 10e-7 and decay as 0.99999.

Some results given by the application have been given below:

Figure 4 Angry Expression
According to the sample training model it signifies as the expression of "Anger" and the geometric analytical model of facial expressions justifies the results as shown in Figure 4.

![Figure 4: Scared Expression]

Figure 5 Scared Expressions
According to the sample training model it signifies as the expression of "Scared" and the geometric analytical model of facial expressions justifies the results as shown in Figure 5.

![Figure 5: Happy Expression]

Figure 6 Happy Expression
According to the sample training model it signifies as the expression of "Happy" and the geometric analytical model of facial expressions justifies the results as shown in Figure 6.

![Figure 6: Sad Expression]

Figure 7 Sad Expression
According to the sample training model it signifies as the expression of "Sad" and the geometric analytical model of facial expressions justifies the results as shown in Figure 7.
According to the sample training model it signifies as the expression of "Neutral" and the geometric analytical model of facial expressions justifies the results as shown in figure 8. This emotion recognition system can be used for monitoring and controlling patients’ condition where facial expression could be utilised to assess the patient conditions intelligently by a machine [14] to avoid risk in case of acute low-back-pain patients[15].

6. Conclusion

In this work, a LeNet architecture based six-layer convolution neural network is implemented to classify human facial expressions that is (i) anger, (ii) disgust, (iii) fear, (iv) happiness, (v) sadness, (vi) surprise, and (vii) neutral. The system has been evaluated using accuracy. The classifier achieved an accuracy of 70%. It has been noticed that the system recognizes negative emotions less accurately as compared to the positive ones. This system can be used for monitoring and management of the patient’s criticalness in the field of AI healthcare system.

Future Work: within the future, we are aiming to implement some method so the system also can display an emoji to represent the sort of emotion. We’ll also attempt to customize this application for specific use cases. We are going to also try and increase the accuracy score of the system especially for the popularity of negative emotions.

Acknowledgments

The thanks to Dr(Prof) Manas Kumar Sanyal, University of Kalyani, for his valuable expertise advised and encouragement for such kind of development. Finally, thanks to IT department, Management, JIS College of Engineering, JIS GROUP for providing all kinds of facilities along with encouragement.
Reference

[1] S. Das and M. K. Sanyal, “Application of AI and Soft Computing in Healthcare: A Review and Speculation,” vol. 8, no. 11, p. 21, 2019.

[2] S. Das, A. Dey, A. Pal, and N. Roy, “Applications of Artificial Intelligence in Machine Learning: Review and Prospect,” International Journal of Computer Applications, vol. 115, no. 9, pp. 31–41, Apr. 2015, doi: 10.5120/20182-2402.

[3] T. Wu, S. Fu, and G. Yang, “Survey of the Facial Expression Recognition Research,” in Advances in Brain Inspired Cognitive Systems, Berlin, Heidelberg, 2012, pp. 392–402, doi: 10.1007/978-3-642-31561-9_44.

[4] “Darwin Online: Darwin’s Publications,” 2020. http://darwin-online.org.uk/contents.html (accessed Aug. 03, 2020).

[5] K. Mase, “Recognition of Facial Expression from Optical Flow,” IEICE TRANSACTIONS on Information and Systems, vol. E74-D, no. 10, pp. 3474–3483, Oct. 1991.

[6] I. A. Essa and A. P. Pentland, “Coding, analysis, interpretation, and recognition of facial expressions,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757–763, Jul. 1997, doi: 10.1109/34.598232.

[7] Y. Zhu, L. C. De Silva, and C. C. Ko, “Using moment invariants and HMM in facial expression recognition,” Pattern Recogn. Lett., vol. 23, no. 1–3, pp. 83–91, Jan. 2002, doi: 10.1016/S0167-8655(01)00108-8.

[8] Xiaoxu Zhou, Xiangsheng Huang, Bin Xu, and Yangsheng Wang, “Real-time facial expression recognition based on boosted embedded hidden Markov model,” in Third International Conference on Image and Graphics (ICIG’04), Dec. 2004, pp. 290–293, doi: 10.1109/ICIG.2004.118.

[9] X. Duan, C. Wang, X. Liu, Z. Li, J. Wu, and H. Zhang, “Ethnic Features extraction and recognition of human faces,” in 2010 2nd International Conference on Advanced Computer Control, Mar. 2010, vol. 2, pp. 125–130, doi: 10.1109/ICACC.2010.5487194.

[10] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M. J. Rosato, “A 3D Facial Expression Database For Facial Behavior Research,” in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 2006, pp. 211–216, doi: 10.1109/FGR.2006.6.

[11] “Challenges in Representation Learning: Facial Expression Recognition Challenge,” 2020. https://kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge (accessed Aug. 04, 2020).

[12] S. Das, “A Comparative Study for Prediction of Heart Diseases Using Machine Learning by Sumit Das, Manas Kumar Sanyal, Sourav Kumar Upadhyay :: SSRN,” 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3526776 (accessed Sep. 13, 2020).

[13] “GPU support,” TensorFlow, 2020. https://www.tensorflow.org/install/gpu (accessed Aug. 04, 2020).

[14] S. Das and M. K. Sanyal, “Machine intelligent diagnostic system (MIDs): an instance of medical diagnosis of tuberculosis,” Neural Comput & Applic, Apr. 2020, doi: 10.1007/s00521-020-04894-8.

[15] S. Das, M. K. Sanyal, and D. Datta, “Artificial Intelligent Embedded Doctor (AIEDr.): A Prospect of Low Back Pain Diagnosis,” International Journal of Big Data and Analytics in Healthcare (IJBDAH), Jul. 01, 2019. www.igi-global.com/article/artificial-intelligent-embedded-doctor-aiedr/247457 (accessed Sep. 13, 2020).