ZKG INTERNATIONAL ISSN: 2366-1313

Deep learning based Automated Student Appraisal Scheme with Computer Vision and CNN model

[1] Nidadavolu Sri Lakshmi Durga Prakash, [2] A. Chennakesava Reddy

[1]PG Scholar, Dept. of MCA, Newton's Institute of Engineering, Macherla Guntur, (A.P)

[2]Assistant Professor, Dept. of CSE, Newton's Institute of Engineering, Macherla Guntur, (A.P)

Abstract: *Researchers in the field of computer vision are always toiling away at the difficult problem of emotion detection in facial expressions. Predicting emotions from faces, detecting smiles, and other uses of facial expression are commonplace in modern technology. Facial traits differ from person to person, making this a difficult undertaking. The concept of deep learning has been used to this issue since convolutional neural networks are able to recognise such intricate elements. We've experimented with using facial-expression prediction to write reviews. Given the extensive body of literature on the topic, we decided to concentrate on practical applications of emotion detection. Our primary objective has been to develop software that can draw human sentiment into the generation of automated evaluations. We have assigned numbers to a variety of face expressions and utilised those numbers to foretell how students would evaluate a lesson.*

Keywords: *computer vision, convolutional neural network, deep learning, data normalization, features* extraction.

I. INTRODUCTION

We learn more about the mental and emotional condition of others and how they want to interact with us through the facial expressions they use. Facial expressions are extensively utilised and accepted, even when other characteristics, such as speech, text, or video, may be used to categorise emotions.

The vast potential of fully automated facial expression detection has attracted a lot of interest. It aids in the expansion of possible uses in many fields, including medicine, business, HC/HCI, psychology, robotics, AI, and many more. Humans convey a wealth of information via their facial expressions. Seven distinct human emotions—anger, contempt, fear, surprise, pleasure, sorrow, and apathy—will be categorised by the system we're developing.

In the past, FER (facial emotion recognition) has been used for a wide variety of purposes, including human-computer interaction, the diagnosis of pain

ISSN: 2366-1313

and mental diseases, the study of human behaviour, etc. This document [1] provides a history of FER analysis, including a summary of the first efforts in the field. We've developed a novel approach that has no analogues in existing literature. Our primary objective is to create a student evaluation system that can automatically produce a review based on the moods of the class as a whole. While there are a number of manual software solutions [2] in use for creating reviews, we use a computer vision and deep learning-based approach.

The document is divided as follows, as we see fit:

Our related efforts are presented in Section II. In Section III of this study, we outline our suggested technique.

The performance of our system is evaluated in Section IV, and we wrap up the study with a brief overview of our findings in Section V. Finally, Section VI demonstrates some potential avenues for developing our study further and future directions for investigation.

II. RELATED WORKS

In this part, we have reviewed various relevant research publications. In order to categorise images, the author of this article [3] first explored a method that used two distinct CNNs.

Both the universal image-based CNNs and the individual facial emotion identification system relied on photographs of faces. To demonstrate the different types of emotion, they calculated an average CNN score across all faces and pictures. For both overfitting and discriminative learning, they resorted to a variant of the edge SoftMax classifier. They achieved an accuracy of 83.9% on the validation set and 80.9% on the testing set, respectively, winning the Wild Challenge 2017 award for Emotion Recognition.

Using HOG and CNN, the authors of another research [4] disseminated a novel deep neural network architecture for the classification of human face expressions. Using deep learning and a convolutional neural network model, they taught their system to identify emotions in people's faces. The FER2013 dataset was utilised for analysis. In this work, we looked at several ways that the FER system's image processing may be improved. The method presented in this research may also extract important information from pictures by combining the models developed for the Local Binary Pattern and HOG operators.

In addition, Liyanage C. DE SILVA et al. [5] aimed to provide a visual and aural face expression recognition result in their study. They selected a pair of bilingual individuals, one speaking Spanish and the other Sinhala. They were filmed making

ISSN: 2366-1313

36 various types of emotional statements while their expressions were captured on camera. The duration of captured visual sequences was consistently constant. They determined which kind of media were more effective at eliciting a certain feeling. They evaluated people and then sought to guess how they were feeling based on their outward demeanour and tone of voice.

The goal of the research study [6] by Mao Xu et al. was to demonstrate how to effectively use transfer learning from convolution networks for the classification of facial emotions. They collected 2062 examples from four datasets associated with facial expressions (CK+, JAFFE, KDEF, and Pain expressions from PICS). With the SVM model based on their chosen dataset, they were able to attain an accuracy of 80.49 percent.

To analyse the photos in the dataset, they employed the Viola-Jones face identification method in conjunction with artificial selection. The validation set was built with ratios ranging from 0 to 0.5 so that researchers could test out a wide variety of transfer properties produced by a deep convolutional neural network.

In order to recognise face expression for picture classification, the authors of [7] calculated several kernel sizes and many filters, and then presented two distinct Convolutional Neural Network (CNN) architectures using the FER2013 dataset. Both models are almost identical with the exception of the number of filters since the second model was developed from the first. Both models employed dropout layers to reduce the number of parameters and enhance accuracy. Therefore, they achieved a 65% accuracy, which may be enhanced by additional exploration of hyper-parameters for their developed models.

Finally, H. Jung et al.'s study [8] aimed to demonstrate how deep learning methods may be used to recognise facial emotions. They started by employing Har-like characteristics to identify faces in the input photos. They sought to provide a comparison between two distinct kind of deep learning networks. At the end of the day, they concluded that CNN outperforms a deep neural network. The CK+, FER 2013 database was utilised for this study. In addition, 71774 photos were utilised for training purposes, each measuring 48 by 48 pixels. With the exception of the distaste expression, which the model consistently mis predicted, the recognition performance was excellent.

III. METHOD DESIGN

We experimented with a number of deep learning methods to see whether we could teach pictures of faces to convey feelings. The computer vision method was first employed to extract faces for use in the review generation process.

ISSN: 2366-1313

Face detection may be accomplished in a variety of ways [9]. Some examples are the Haar Cascade, MTCNN, DNN face detector of the Caffe model, and others. The DNN Face detector in OpenCV has been put to use in our project. It's a Caffe model built on SSD, and it's based on ResNet-10. The retrieved picture was then put to use in a face expression recognition and review generation procedure. We've divided the work on this project into five distinct phases.

Our suggested model's flowchart is shown in Fig. 1. Our suggested procedure calls for the development of a system that takes pictures either from the webcam on a laptop during an online lecture or from closed-circuit television cameras in the actual classroom. The next step is for the system to analyse the gathered data in order to identify the pupils' emotional states and provide a real-time summary of their moods. A facial emotion recognition (FER) model is the foundation of the system's design. Then, we've programmed an algorithm to provide reviews based on how people feel about a product. The system as a whole will run an emotion detection procedure in real time, allowing us to more accurately assess student feedback on instructors.

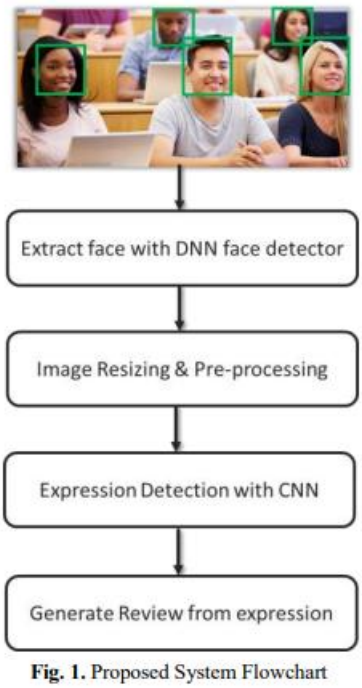

Fig. 1. Proposed System Flowchart

A. Set of Data

In our experiments, we utilised data from the Facial Expression Recognition 2013 (FER-2013) dataset [10]. This dataset is available for anybody to use, and it was first presented at the International Conference on Machine Learning (ICML). The information was compiled by Pierre-Luc Carrier and Aaron Courville, who searched Google for terms related to human emotions.

Following ICML 2013, the dataset was made publicly available for a Kaggle competition. There are a total of 35,887 grayscale photos in the dataset, with 28,709 of them having been labelled for use in training, 3,589 for validation, and 3,589 for testing. Each face picture in the 48x48 pixel dataset has been annotated

ISSN: 2366-1313

with a label indicating which of seven emotions it depicts.

Figure 2 displays several examples from our dataset together with every labelled phrase we've used. Labels range from 0 (Angry; 4593 pictures), 1 (Disgust; 547), 2 (Fear; 5121), 3 (Happy; 8989), 4 (Sad; 6077), 5 (Surprise; 4002), and 6 (Neutral; 6198) with 0 representing Angry and 6 representing Neutral. Face photos with a resolution of 48 by 48 grayscale pixels are included in the collection.

B. Priming the Data Pump

Fig. 2. Image samples of facial expressions from our dataset with labels

The dataset comes in CSV format with three columns; emotion, pixels, and usage. The pixel column contains the pixel values of the images in 1D. So, we converted the 1D list of pixels into 2D to get the images' width and height, which is 48 by 48 according to the dataset.

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

Fig. 3. Five random data according to pixels

Five random data points from our dataset, which stands in for the values of each pixel in our photographs, are shown in Fig. 3.Face photos in our collection are 48 pixels wide and tall. We scaled the photographs to match the specifications of the VGG architecture, which requires input images to be 224 by 224 pixels in size. For this improvement, we flipped the picture horizontally at random and rotated it by 30 degrees. We normalised the colour channels to rescale the pixel values from 0 to 1 since images often have pixel values between 0 and 255.Since the dataset included three distinct sets—the training set, the test set, and the validation set—we segmented it accordingly. So, we took them out of the CSV file and put them where they belonged.

C. Theoretical Constructing

Two methods of model training have been explored thus far. As a first step, we experimented with transfer learning. Our CNN model is based on the VGG-16 architecture. There are over 138,000,000 [11] different training settings within. The architecture is beautiful and straightforward, yet the network is extensive. Therefore, it would have taken a long time and a huge dataset to train the model from start. This is why we have settled on transfer learning as our method of instruction. The VGG architecture has already been trained, which means the CNN layers perform admirably as a feature extractor. As a result, we employed the CNN layers without further training as the

ISSN: 2366-1313

picture feature extractor and frozen learning parameters. The architecture's completely interconnected levels were then updated, and a new, seven-node output layer was introduced. However, the outcome was poor with this strategy. It seems that the validation loss was entrenched. This strategy has to be abandoned.

To train the model, we then construct a brand-new CNN architecture. Input greyscale photos are expected to be 48x48 in size, as required by the design. Since VGG requires input pictures to be 224x224 pixels in size with three colour channels, the fact that our dataset included images of the same size was an advantage.

Our CNN model's layering and filtering is shown in Fig. 4.All convolutional layers have been built using 3-size filters. We've utilised a stride size of 2, and we've padded all of our layers the same amount. Extra padding may be added using the same method, effectively hiding the whole picture. We utilised a dropout of 0.5 across most of our layers and implemented batch normalisation and pooling.

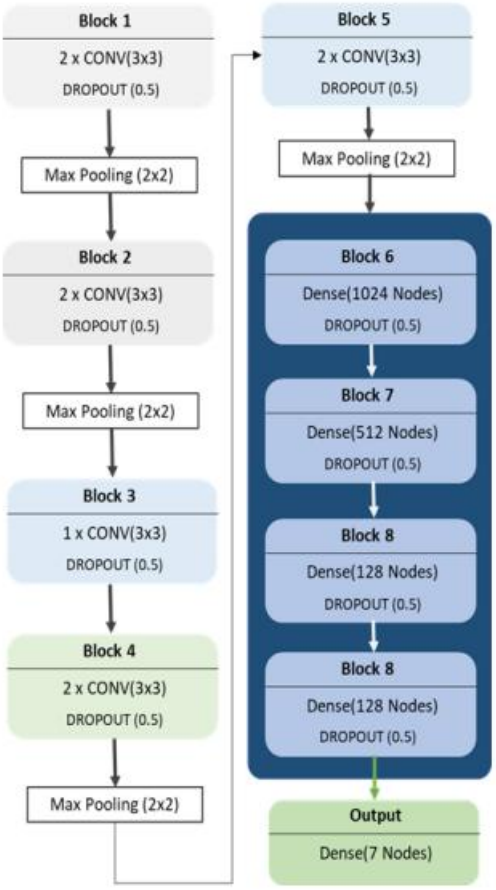

Fig. 4. Different CNN Layers and parameters

The total number of learning parameters in our model is over 13 million. To train our model, we used batch size of 32 and trained the model for about 40 epochs. We have used the following function for calculating errors.

$$CCE = -\frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j \cdot \log(\hat{y}) + (1 - y_j) \cdot \log(1 - \hat{y}_j)$$

Equation 1

The error was calculated using equation (1), the categorical cross-entropy loss function. With an initial learning rate of 0.003 D, we have additionally updated weights using the Adam optimizer. Review Production

ISSN: 2366-1313

From a single face picture, our model can infer the subject's emotional state. However, we have employed computer vision to generate ratings from a multi-faced picture. One of the most widely used techniques for accurate face detection is the DNN Face Detector built on top of OpenCV's Caffe model [12]. Thus, we have implemented this DNN classifier for face detection in image processing. Then, we applied our model to each face in order to create an emotion. We have seven distinct phrases, and we've given each one its own importance rating. Table 1 displays the results of the weighting. The scores for the various types of emotions are listed in Table 1. Here, the model's predictions are utilised along with the weights to provide an overall score. Students are given these grades based on a broad assessment of their behaviour. We contrasted survey responses from students with their participation in class. Top performers in class often show a lot of enthusiasm throughout class because they value the information being presented to them. Additionally, it has been shown that the majority of attentive pupils maintain expressionless faces throughout class. High marks are given for phrases that are neither cheerful nor neutral.

However, there are always some that zone out or become distracted by their phones or other things throughout class. As a result, they are often taken aback or frozen in dread when questions are posed, and many students who provide poor evaluations are also characterised by a lack of engagement throughout class.

We based the grades on these overarching student behaviours.

TABLE 1: EMOTIONSCORE

Emotion	Score
Happy	1
Neutral	0.7
Surprise	0.6
Fear	0.5
Sad	0.4
Angry	0.25
Disgust	0

We used an average formula to produce the review's results. The aggregate score based on all faces has been determined. Then, we have calculated an overall mean. We have categorised the results into three different categories. A favourable review has a score higher than 0.6, while a bad review has a value lower than 0.4. We have determined that a review is neutral if it falls between a score of 0.4 and 0.6. However, a student's expression in class may shift every minute. Therefore, relying on a single picture evaluation for each person might provide inaccurate results. Therefore, we retrieved the same faces from many images of the same scenario taken at the same time period. The following methodology is what we've employed.

ISSN: 2366-1313

Algorithm 1. Generate Average Review Pseudocode

```
review_count ← 0
repeatfor every 5 munities up to 1 hour:
        temp_review ← 0
        Capture images and extract faces
        temp_review ← Score(extracted_face)
        review_count ← review_count + temp_review
end loop
final_review = Average(review_count)
```

We have developed an algorithm in Algorithm1 that captures many photos at regular intervals during the course. Next, it determines which faces exist inside the photos and assigns a review score to each one. The DNN face detector included in OpenCV's Caffe model has been utilised for face extraction. The algorithm then iteratively performs this step to get a mean score. The aggregate of all computed scores is then used to determine the final score.

IV. PERFORMANCE EVALUATION

Using the FER2013 data set, we have trained our model on two distinct architectures. Our models have been evaluated using 270 photos taken from the dataset. For the sake of the model's impartial development, we kept the dataset isolated. To begin, we have trained for seven iterations using the VGG-16 architecture. In our tests, it resulted in a 54% success rate. After another 40 epochs of training with the second model, we observed that 186 of 270 test pictures could be properly predicted. Our testing revealed an accuracy of about 69%.

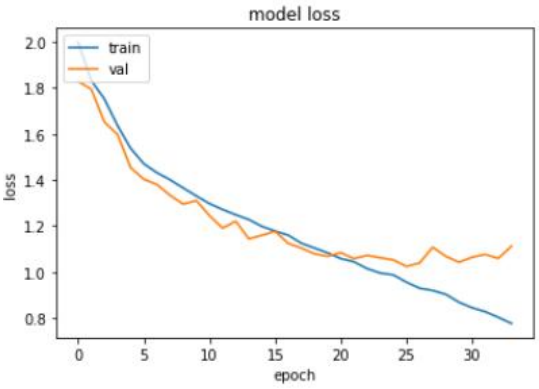

Fig. 5. Training loss Graph

Fig. 5shows the training loss and the validation loss graph. For preventing overfitting, we used early stopping by monitoring the validation loss.

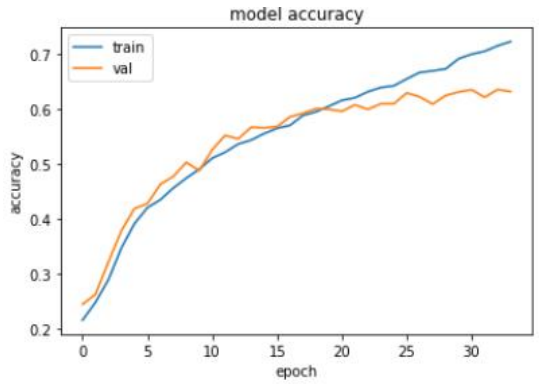

Fig. 6. Model Accuracy track

The accuracy graph of the trained model is shown in Fig. 6. The accuracy throughout both training and validation has been shown graphically.

Finally, we've taken a new way to testing the efficacy of our review system. First, we gathered 60 text evaluations from North South University students across three departments. We then sorted through the reviews, assigning good, negative, and neutral ratings to each, before calculating an overall rating for each of the involved

ISSN: 2366-1313

professors. Then, we utilised our algorithm to compile data on how many students participated in each activity throughout the online session with the same instructors. Based on our content analysis, we gave a positive rating to three different departments. Only one review was favourable, while two were indifferent, for our model. Therefore, it seems that the model is functioning well. We anticipate that, with improvements to our algorithm's emotion detector model, it will compete well with other approaches often used in teacher rating systems.

V Conclusion

In our study, we have tested two different models for determining an individual's emotional state based just on a photograph of their face. We have created an algorithm to produce reviews based on the ratings we provided to the phrase. Our primary objective was to create a system that can automatically provide reviews. Since there are existing pre-trained architectures with high accuracy for predicting facial expressions, we spent less time fine-tuning the models' accuracy. The method will provide teachers insight on their students' participation in class and give administrators a sense of the level of engagement in the classroom.

VI. LIMITATIONS& FUTURE RESEARCH

In this study, we developed a system that employs a pre-trained model to analyse face images and provide ratings and comments. Better accuracy may be achieved by training the data on pre-trained architectures such as inception, ResNet, VGG, etc., and by using a bigger dataset with images of at least 224 × 224 pixels in size. It is feasible to try out other techniques, such as Histogram of Oriented Gradients (HOG), for extracting features. Possible benefits include enhanced review prediction accuracy. IoT devices may be integrated into this system to offer instant feedback on student progress in class. Working with the picture as a whole is another method that may be tried. A Convolutional Neural Network may be trained to immediately predict a review from a picture of a classroom full of pupils, saving time and effort over dealing with individual face images. This will pave the way for the exploration of more complex structures and the use of alternative techniques, such as single-shot detectors, for the construction of more accurate models. There wasn't enough information for us to test these possibilities. However, given CNN's capabilities, we think there's a lot of room for future study to refine and

ISSN: 2366-1313

develop better models to boost accuracy and produce reviews.

REFERENCES

[1] Y. l. Tian, T. Kanade, and J. F. Cohn, "Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity," in Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 229-234, 2002.

[2] "How does a faculty evaluation system keep things fair? - Interfile", Interfile, [Online]. Available: https://www.interfolio.com/resources/blog/how-does-a faculty-evaluation-system-keep-things-fair/. [Accessed: 19- Dec- 2020].

[3] Tan, Lianzhi & Zhang, Kaipeng & Wang, Kai & Zeng, Xiaoxing & Peng, Xiaojiang & Qiao, Yu. (2017). Group emotion recognition with individual facial emotion CNNs and global image based CNNs. 549-552. 10.1145/3136755.3143008.

[4] Zafar, Sahar & Ali, Fayyaz & Guriro, Subhash & Ali, Irfan & Khan, Asif & Zaidi, Adnan. (2019). Facial Expression Recognition with Histogram of Oriented Gradients using CNN. Indian Journal of Science and Technology. 12. 10.17485/ijst/2019/v12i24/145093.

[5] L. C. De Silva, T. Miyasato and R. Nakatsu, "Facial emotion recognition using multi-modal information," Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat., Singapore, 1997, pp. 397-401 vol.1, doi: 10.1109/ICICS.1997.647126.

[6] Mao Xu, Wei Cheng, Qian Zhao, Li Ma and Fang Xu, "Facial expression recognition based on transfer learning from deep convolutional networks," 2015 11th International Conference on Natural Computation (ICNC), Zhangjiajie, 2015, pp. 702- 708, doi: 10.1109/ICNC.2015.7378076

[7] Agrawal, Abhinav & Mittal, Namita. (2019). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. The Visual Computer. 36. 10.1007/s00371-019-01630-9.

[8] H. Jung et al., "Development of deep learning-based facial expression recognition system," 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), Mokpo, 2015, pp. 1-4, doi: 10.1109/FCV.2015.7103729.

ISSN: 2366-1313

[9] V. Agarwal, "Face Detection Models: Which to Use and Why?", Medium, [Online]. Available: https://towardsdatascience.com/face-detection-models-which to-use-and-why-d263e82c302c?gi=1d1c9add7285. [Accessed: 02- Dec- 2020].

[10] "fer2013", Kaggle.com, [Online]. Available: https://www.kaggle.com/deadskull7/fer2013. [Accessed: 06- Jan- 2021].

[11] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image RecognitionarXiv e-prints, arXiv:1409.1556.

[12] "OpenCV: Deep Neural Network module", Docs.opencv.org. [Online]. Available: https://docs.opencv.org/4.0.0/d6/d0f/group__dnn.html. [Accessed: 11- Jan- 2021].

[13] Kumar, Alok & Jain, Renu. (2018). Faculty Evaluation System. Procedia Computer Science. 125. 533-541. 10.1016/j.procs.2017.12.069.

[14] Prasadu Peddi (2023), Using a Wide Range of Residuals Densely, a Deep Learning Approach to the Detection of Abnormal Driving Behaviour in Videos, ADVANCED INFORMATION TECHNOLOGY JOURNAL, ISSN 1879-8136, volume XV, issue II, pp 11-18.

[15] Naga Lakshmi Somu, Prasadu Peddi (2021), An Analysis Of Edge-Cloud Computing Networks For Computation Offloading, Webology (ISSN: 1735-188X), Volume 18, Number 6, pp 7983-7994.