*Research Article*

# Evaluating Students' Reaction to Lectures Using Facial Expression Recognition

**Arshaad Mohiadeen[1], Mohd Azam Osman[2], Abdullah Zawawi Talib[3]**

[1,2,3]School of Computer Sciences, Universiti Sains Malaysia,11800 USM, Pulau Pinang, Malaysia
[1]arshaad@student.usm.my, [2]azam@usm.my, [3]azht@usm.my

**Article History**: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

**Abstract:** Interaction between lecturers and students plays a critical role in defining one's view of a lecture. However, with an increasing number of students enrolled in universities every year and limited classroom space available, classrooms are often overcrowded. As a result, it is fairly difficult for lecturers to immediately observe the learning feedback from all students on the lecture being delivered. In this paper, we propose a tool named Facial Expression Analysis Tool (FEAT) to help lecturers in universities in evaluating the effectiveness of their lecture by evaluating their students' facial expression based on three facial expressions: bored, satisfied or confused. The tool utilizes dual CNN for detection and classification. FEATreceives the video feed via an IP camera from the classroom, and analyzes and stores the information on a cloud database. The aggregated information from the database is further filtered, and the statistical details are displayed on a visual dashboard on the web. The tool was evaluated in a real classroom environment and found to have achieved a good accuracy. The tool provides useful insights for the lecturers to better observe their students'perception on their lectures and improve their teaching approach if required.

**Keywords:** CNN, Facial Expression, Classrooms

## 1. Introduction

Education is a vital aspect of our life because it builds the necessary foundation on how we can progress as a society. The world is getting increasingly complex, seamless and dynamic, and education is the vehicle for ensuring that we can navigate this complexity with understanding, collaborationand problem-solving across cultures. Furthermore, language competenciesand educators play a crucial role in ensuring thatstudents are properly educated. Ramberg *et al.*(2018) stated that a caring teacher has previously been identified as an important factor in increasing student motivation and learning. According to Velasquez *et al.* (2013), numerous studies haveindicated that a caring teacher can positively impact learning outcomes, motivation, and social and moral development. However, Chen *et al.* (2019) and Islam et al,. 2016 mentioned that especially in universities, classrooms are often overcrowded with a large number of students which makes it strenuous for lecturers to monitor students' reaction on the lecture being delivered, and obtain immediate feedback from the students in the classroom on whether they areable to follow thelecture being delivered.

Chong (2018) discussed how written feedback conducted commonly in classrooms works. The author concluded that not much attention has been paid to the significance of socio-emotional factors in the feedback process and much less to that of learners' characteristics, and highlighted the main issue which is student's reaction during lectures that are not properly taken into account.

Therefore, in view of this issue, a tool called Face Expression Analysis Tool (FEAT) is proposed and presented in this paper. The tool can detect and keep trackthree facial expressions of students during a lecture namely feeling bored, satisfied and confused in order to observe how students perceive the lecture being delivered. The technique involves comparing a student's facial expression with images in a trained facial expression dataset (Harsh Shukla, et al, 2020). The processed information is made available in a visual format for the lecturer and accessible via a web dashboard. It can be a useful tool for a lecturer as it is a departure from usually hearing a monotone response, i.e. "Yes sir/ Yes ma'am" from the students when asked whether they are able to follow the lecture being delivered. To demonstrate the effectiveness of FEAT, a preliminary case study of using the tool wasconducted in the School of Computer Sciences, Universiti Sains Malaysia.

## 2. Materials and Methodology

### 2.1 Tools Used

Google Colab was used for building and training the model along with Amazon AWS for cloud storage. FEATwas programmed using Python 3.6 and Tensor flow 1.5.

### 2.2 Convolution Neural Network

The algorithm chosen for the tool is the Convolutional Neural Network **(CNN).** A major advantage of CNN is that it requires less memory and parameters compared to other machine learning algorithms which also additionally require preprocessing and feature extraction. As for CNN, it can handle feature extraction and preprocessing by itself. However, it requires a lot of data samples to build a model which may be arestraint for a small dataset. Nonetheless, CNN has been acclaimed for attaining good results in the field of face recognition. In this work, we have also attempted to use other algorithms such as k-Nearest Neighbors **(KNN)** and Support Vector Machine **(SVM)**. However, they have resulted inpoorer results in detecting and classifying facial expressions compared toCNN.

### 2.3 CNN Model Development

#### 2.3.1 Dataset Used

The dataset used is selected from the Google Facial Expression Comparison (FEC) dataset (See Figure 1). Only a handful of the images from this dataset have been utilized as not all images are required. The images selected for training are those onlyon the following three facial expressions: bored, satisfied and confused.

**Figure 1.** Google Facial Expression Comparison (FEC) data set

#### 2.3.2 CNN Model

The CNN model designed for this project consists of the following layers: the input layer, convolutional layer, activation layer, pooling, dense, dropout, connected layer andoutput layer. Figure 2 depicts the overall CNN structure used. It consists of three convolutional layers, two hidden layers and an output layer that uses Softmax regression model as it is a multi-face classification.

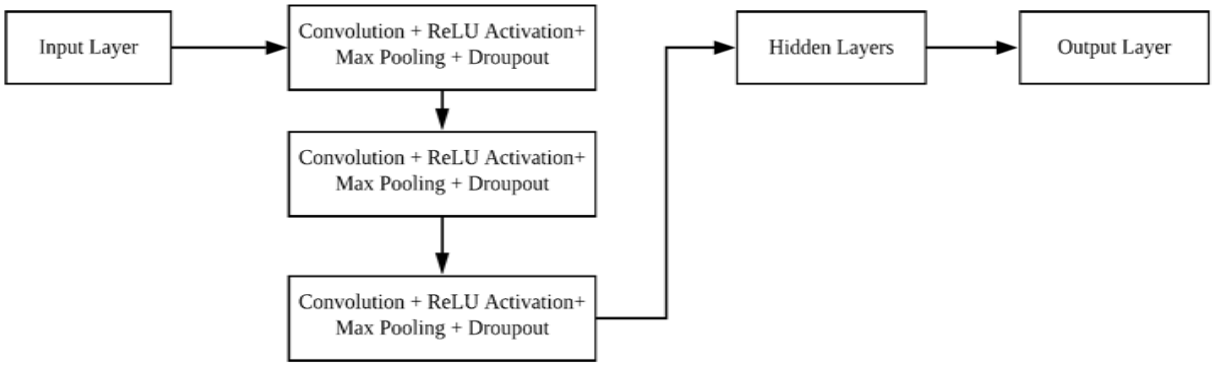


**Figure 2.** Structure of CNN

The initial input layer consists of a total sample set of 7368 images where 0.1% of the dataset were used for training validation, and thereby, 6631 images were used for training and 737 for validation purposes. The size of every image is 224*224.

After collecting the facial dataset from the images, they are then processed as the input for CNN. The next first layer right after the input layer is the first convolutional layer with 64 filters with the convolution kernel size set at 3*3, followed by activation of ReLU layer, 2*2 max pooling to reduce the spatial dimensions and a dropout of 0.5.

This is repeated two more times before entering the hidden layer where it is flattened, the dense layer output array of 512, 2*2 max pooling and a dropout of 0.5. This process is repeated once more before it goes to the output layer.The output layer includes adense layer of three neurons along with Softmax regression as it compiles.

Furthermore, testing with different combinations were experimented such as with two convolutional layers and two hidden layers, five convolutional layers and three hidden layers and many more, but most of these experiments caused the model to suffer from underfitting and overfitting, respectively.Furthermore, since we only have three classification categories, we have found a perfect balance when usingthree convolutional layers and two hidden layers.

### 2.3 Breakdown of FEAT

Figure 3 highlights the complete inner workings of FEAT i.e., after obtaining the video feed, it is split into multiple frames and sent to FEAT for analysis. FEAT computes the number of frames to be processed before entering the loop where it tries to firstly detect faces using Multi-task Cascaded Convolutional Networks **(MTCNN)**. It works in three stages whereby it initially produces candidate windows quickly through a shallow CNN. Then, it refines the windows by rejecting a large number of non-face windows through a more complex CNN. Finally, it uses a more powerful CNN to refine the result again and output five facial landmarks positions (*Zhang et al.,* 2016).

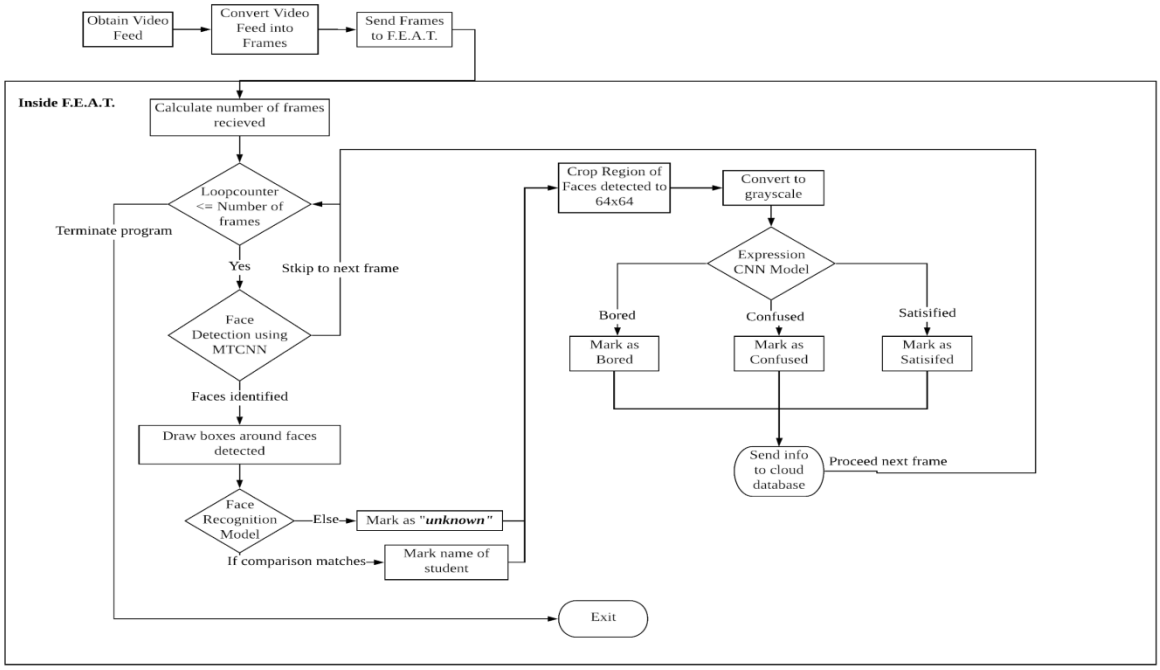


**Figure 3.** FEAT Process Flow Chart

If no faces are detected, it will proceed to the next frame, else it will try to identify the faces detected with the help of a face recognition library and a compiled image set of the students enrolled in the class. Next, their facial expressions are evaluated before compiling the results to a cloud storage database. This entire process is repeated until all frames have been processed.

FEAT can also be executed in real time, provided that it runs on a system with a decent graphics card such as

a GTX 1060Ti or higher, to handle the computational load with no bottleneck in the process.

## 3. Results and Discussion

Figure 4 shows that the CNN model built are able to attain an accuracy of 92% with 1% validation loss. The model tends to reach its peak accuracy by its 40th epoch with negligible to no changes in its accuracy and loss values.

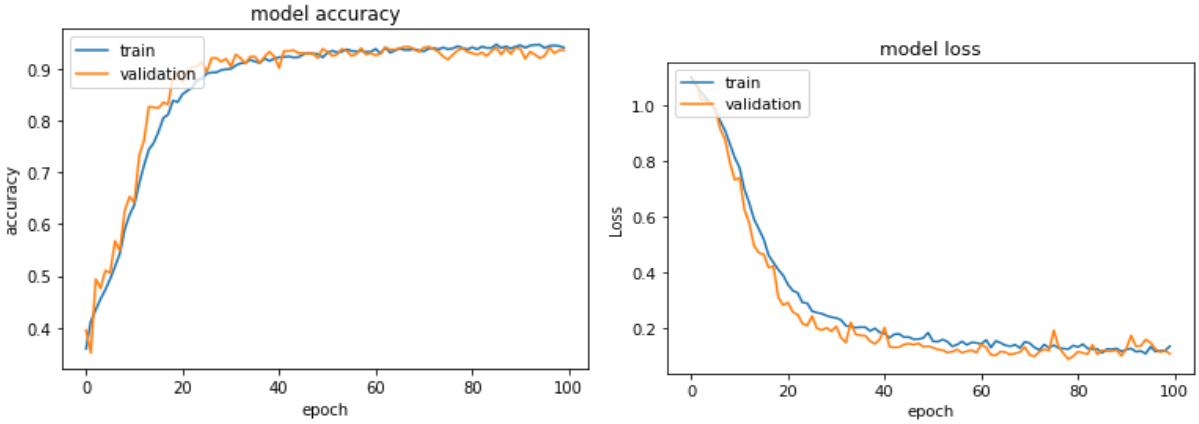


**Figure 4.** CNN model accuracy and loss rates

Even though the model and tool have been successfully deployed, there are instances where some faces are clearly misclassified. This is clearly due to the limited size of the dataset used, as CNN improves with more data to learn from. As seen in Figure 5, a bounding box is placed on the faces that have been detected using MTCNN along with a label indicating their facial expression.

Overall, FEAT was able to detect and classify about 221 faces out of a total 291,and achievean accuracy of 76% in a real-world deployment testing which was conducted within a university environment.

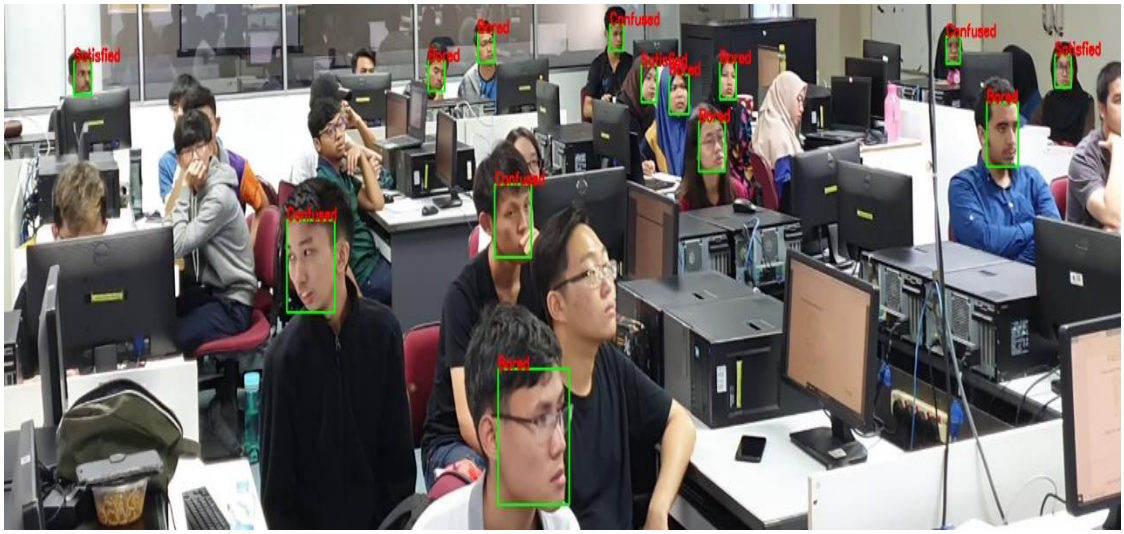

**Figure 5.** Frames from the Testing Video conducted in a Computer Lab

## 4. Challenges

First and foremost, availability of the dataset related to facial expression recognition is almost non-existent as the majority of the available datasets focus on facial emotion recognition such as CK+, JAFFE and AffectNet. The only dataset that could be used as a foundational basis for this work is the Google FEC dataset which requires manual annotation of images and classes. Subsequently, the size of the dataset used plays an important factor, since the size of the dataset is proportional to the amount of computational resources and time required to build a neural network model. However, this obstacle has been tackled with the use of Google Colab which

provides powerful resources that has help in the development of the tool.

Additionally, during the implementation, it has been found that Haar cascade classifier which is based on Viola Jones detection algorithm performs poorly in detecting multiple faces in the test cases. Different methods have been tried and MTCNN face detection has been found to be a suitable substitute that does not compromise on speed and is able to improve the number of faces detected.

In terms of limitations, there are a few aspects to be noted that can restrict the application from being fully functional as planned. The quality of the video feed, along with the angle and perspective of the student's faces seen in the feed's coverage play a huge role. The students seated at the end of the classroom are often not detected and evaluated as they appear as a tiny spec orblur, and thus contributing to its limitation. Moreover,the performance of this system is highly dependent on the hardware used as it is one of the most crucial points in running this tool. Finally, there are some partialocclusion issues faced by the system, i.e. it is quiet challenging to accurately detect those faces wearing full head covering such as hijab. Nonetheless, it is able to do so.

## 5. Conclusion

The proposed tool which uses CNN and FEC-based dataset FEAT is able to cover and analyze a wider range of audience effectively in a fraction of second without the need for any additional external inputs. The breakdown of facial detections on the dashboard into the three most common facial expressions (bored, satisfied, confused) allows the lecturer to have a better observation of thestudents' actual reaction to a lecture. This tool provides a platform for a full paradigm shift in evaluating students' reaction to a lecture instead of just responding "Yes sir/ma'am" from the students which is normally a way to mask their confusion on the concepts being taught in the class. The usage of cloud services also makes it easier for the lecturers to access their dashboard at anytime and anywhere. In a nutshell, the proposed tool is introduced to help lecturers to obtain daily feedback from a set of accumulated facial detections and see whether their teaching needs to be reviewed or improved.

## References


1. Chen, S., Dai, J., & Yan, Y. (2019). Classroom Teaching Feedback System Based on Emotion Detection. *9th International Conference on Education and Social Science (ICESS 2019)*, 940-946.
2. Chong, I. (2018). Interplay among technical, socio-emotional and personal factors in written feedback research. *Assessment & Evaluation in Higher Education*, 43(2), 185-196.
3. Harsh Shukla & Meenu Pandey. 2020. Human Suspicious Activity Recognition. *IIRJET*, V-5, I-4, CS-14 - CS-17.
4. Islam, R., Ghani, A.B.A., Kusuma, B., Theseira, B.B. (2016). Education and human capital effect on Malaysian economic growth. International Journal of Economics and Financial Issues, 6 (4), pp. 1722-1728.
5. Ramberg, J., Låftman, S. B., Almquist, Y. B., & Modin, B. (2018).School effectiveness and students' perceptions of teacher caring: A multilevel study.*Improving Schools*, 22(1), 55-71.
6. Velasquez, A., West, R., Graham, C., & Osguthorpe, R. (2013).Developing caring relationships in schools: A review of the research on caring and nurturing pedagogies.*Review of education*, 1(2), 162-190.
7. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, *23*(10), 1499-1503.