Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System

Conference Paper · February 2020

DOI: 10.1145/3385209.3385231

6 authors, including:

Saurabh Pal
Bengal Institute of Technology
22 PUBLICATIONS **97** CITATIONS

Anand Nayyar
Duy Tan University
156 PUBLICATIONS **940** CITATIONS

Pijush Kanti Dutta Pramanik
National Institute of Technology, Durgapur, India
38 PUBLICATIONS **145** CITATIONS

Prasenjit Choudhury
National Institute of Technology, Durgapur
68 PUBLICATIONS **235** CITATIONS

Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System

Moutan Mukhopadhyay
Bengal Institute of Technology
Kolkata, West Bengal, India
moutanbanerjee@gmail.com

Saurabh Pal
Bengal Institute of Technology
Kolkata, West Bengal, India
saurabhpal2007@gmail.com

Anand Nayyar
Duy Tan University
Da Nang, Vietnam
anandnayyar@duytan.edu.vn

Pijush Kanti Dutta Pramanik
National Institute of Technology
Durgapur, West Bengal, India
pijushjld@yahoo.co.in

Niloy Dasgupta
Bengal Institute of Technology
Kolkata, West Bengal, India
niloydasgupta0987@gmail.com

Prasenjit Choudhury
National Institute of Technology
Durgapur, West Bengal, India
prasenjit0007@yahoo.co.in

ABSTRACT

Despite the success and the popularity of the online learning system, it still lacks in dynamically adapting suitable pedagogical methods according to the changing emotions and behaviour of the learner, as can be done in the face-to-face mode of learning. This makes the learning process mechanized, which significantly affects the learning outcome. To resolve this, the first and necessary step is to assess the emotion of a learner and identify the change of emotions during a learning session. Usually, images of facial expressions are analysed to assess one's state of mind. However, human emotions are far more complex, and these psychological states may not be reflected only through the basic emotion of a learner (i.e. analysing a single image), but a combination of two or more emotions which may be reflected on the face over a period of time. From a real survey, we derived four complex emotions that are a combination of basic human emotions often experienced by a learner, in concert, during a learning session. To capture these combined emotions correctly, we considered a fixed set of continuous image frames, instead of discrete images. We built a CNN model to classify the basic emotions and then identify the states of mind of the learners. The outcome is verified mathematically as well as surveying the learners. The results show a 65% and 62% accuracy respectively, for emotion classification and state of mind identification.

CCS Concepts

• Applied computing→Interactive learning environments • Applied computing→E-learning • Computing methodologies→Computer vision • Human-centered computing→→Gestural input

Keywords

Online learning systems; Emotion detection; Facial expression; Machine learning; CNN; Image processing; Combined emotion; State of mind

ICIIT 2020, February 19–22, 2020, Hanoi, Viet Nam

ACM ISBN 978-1-4503-7659-4/19/07…$15.00

DOI: https://doi.org/10.1145/3385209.3385231

1. INTRODUCTION

The continuous advancement in digital technologies and multimedia have proliferated the use of online learning systems (OLSs) [1] [2], settling the constraints of the face-to-face mode of learning like cost, time restriction, space requirement, unavailability, etc. The OLS advantages like ubiquity, flexibility, availability (in terms of pedagogy and resources), multimedia support, user-centric (more user control over learning), etc. have made it quite popular among all categories of learners. However, OLSs still face challenges of being not so learner-friendly. It supposes all learners same throughout a learning session, not considering their varied emotional and psychological fitment to learning. But in practice, learners are different, and so is their capability to fathom and process the learning instructions and contents. Furthermore, during a particular learning session, a learner may go through various mental and emotional states, which directly affects their learning process.

Failing to grasp and process information while learning eventually may lead the learner to get confused, feel strenuous (not comprehending), and feel bore and weary (lose attention and interest). This may trigger changes in learner behaviours, making him/her disorientated, frustrated, depressed, grim, or angry, which ultimately results in skipping and leaving out from learning sessions. Thus, an unstable and downcast emotion of a learner leads to a low or negative feeling which results in poor learning performance.

In face-to-face or human tutoring, the emotional and psychological changes in the learner are easily identified by the teacher through human's natural cognizance and experience. And based on the changed emotion and behaviour of the learner, the teacher may follow appropriate pedagogical method and approach. This brings attention to the conclusive fact that teaching and learning process needs to adapt as per the emotional and psychological state of the learner.

A learner's emotional involvement reflects the degree of engagement in learning, also referred as affective engagement [3]. As an OLS has no intelligence to recognize human emotion and the system is more mechanized, it is inherently unable to understand the affective engagement of a learner while learning. Therefore, it has to be explicitly armoured to do that. For automatic affective engagement detection, generally, two approaches are adopted [3], as shown in Table 1. For an OLS, computer vision based approach is more suitable and has a wide application scope compared to the sensor data analysis approach.

Among the other perceptual assessments (eye movement, facial expression, and gesture and posture), human face reflects one's internal emotion and psychological state most prominently. As a reason, among the three visual cue assessment approaches, the facial expression is the most potent and effective way to assess the affective engagement of a learner. Facial expression (or emotion) detection uses robust geometric and pattern matching algorithms to assess muscle tone and stretching on the face to find the instantaneous emotion of a person.

Generally, human emotions change due to internal and external stimuli and events. Figure 1 shows the basic and universally accepted human facial emotions [4]. Likewise, in a learning scenario, a learner's emotion changes as per the learning actions he/she encounters. These emotions are temporary, discrete and change frequently before one could realize it. The different emotions that develop during a learning process reflect the different states of mind of the learner. The state of mind is a long-term outlook, an overall experience of a learner, whereas emotions are the components of the state of mind. Frequent changing emotions cannot capture the actual feeling of a learner. Thus, finding a long-lasting state of mind is more useful than a temporary emotion. Analysing this would help in identifying the psychological state of the learner and its probable effect in the learning activity, which can be used as the metric for making learning more efficient and instructional pedagogy more anthropomorphized.

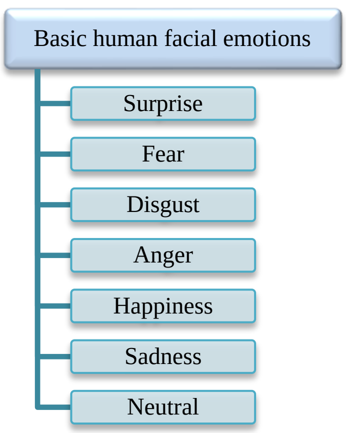

Figure 1. Basic human facial emotions

1.1 Problem Description

Most of the facial emotion detection algorithms identify one of the basic seven emotions [4], listed in Figure 1. However, human emotions are far more complex and may contain shades of more than one emotion. Therefore, the psychological states may not be revealed only through a basic emotion, but a combination of two or more emotions which may be reflected on the face over a period of time. Expression of these combined emotions is generally continuous and are not very discrete; i.e., they do not change very abruptly. So, explicitly choosing one emotion in an instant of time may fail to assess the exact emotion.

Similarly, a learner also experiences a combination of different states of mind, as listed in Figure 2, during the learning process. These states of mind of a learner may last for a short duration and transform into other emotion over time. Thus, identifying the actual emotion of a learner during the learning session is challenging.

1.2 Problem Statement

In view of the above discussions, in this paper, we precisely address the following three problems:

a) Identifying the change of emotions reflecting the state of mind of the learner during a learning session.
b) Aggregation of the emotions of the learner over time to assess the combinatorial emotions during the learning session.
c) Analyse these combinatorial emotions to consider as feedback for the learning session on learning content, pedagogical method, etc.

1.3 Contribution of this Paper

The paper has a two-fold contribution as follows:

a) We find different patterns of single or combined emotions of a learner during the learning session.
b) We propose a definite model to assess the state of mind of a learner from his/her different emotional expressions at any point of time in a learning session.

1.4 Implication of the Work

The proposed work has the following important implications in a personalized OLS and beyond:

a) Depending on the change of emotion of a learner, an intelligent OLS may adapt to a suitable pedagogical method (e.g., changing the learning contents, changing the pace of tutoring, etc.).
b) Analysing the found pattern of emotions and changes in emotions, suitable learning contents may be designed and recommended to the learner.
c) The proposed model and solution is not only limited to OLS but can be used in many other applications such as human resource analysis (e.g., during an interview), assessing driver's condition (e.g., sleepy, tired, etc.), online product feedback, real-time TV show recommendation, etc.

1.5 Organization of the Paper

The rest of the paper is organized as follows. Section 2 mentions the related work. Section 3 presents the details of detecting the emotion of a learner. In this section, the model for combinatorial emotions, the method and algorithm of detecting emotions, the classification technique, and the practical experiment are presented. The result of the experiment and its analysis are discussed in Section 4. Section 5 concludes the paper, mentioning the scope of future work, based on this paper.

Table 1. Automatic affective engagement detection approaches

	Uses	Measures	Method	Advantage	Disadvantage
Sensor data analysis based approach	Bio-sensors	• Heartbeat • EEG • Blood pressure • Skin response	Sensor data analysis	• Accurate finding of learner's physiological state. • Alertness and arousal are quite well detected.	• Requires specialized sensors which may not be convenient for a real-life learning environment. • The emotions which are detected are very less to be useful for wide online learning applications.
Computer vision based approach	Camera	• Eye movement • Facial expression • Gesture and posture	Feature extraction and analysis from captured images	• Low cost • Easy to set up and use • The wide availability of camera in different devices allows detecting learner's engagement widely supportable. • Computer vision can be trained to assess the learner, the same way as a teacher does in face-to-face teaching.	• Lacking in accuracy. • Result's accuracy depends on the dataset and algorithm used for detection.

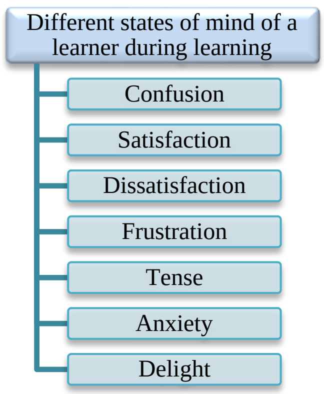

Figure 2. Different states of mind which may arise in the learner during the learning process

2. RELATED WORK

The ability of computers in detecting human emotions has encouraged its application widely in education and learning system. Emotion detection has been successful in detecting learner's alertness and attention [5], getting opinion and feedback [6], and assessing learner feeling and affective state [7]. In regular classroom learning settings, emotion detection by head posture and facial expression detection can find the student's attentiveness and synchronization rate [8]. In comparison to regular classroom learning, emotion detection has especially benefited to the e-learning system. Due to immense application scope, the research on emotion detection is quite popular. The advancement in technology has accelerated the research on emotion detection. One of the key applications of emotion detection in e-learning is personalized learning support. Whereby, learning is being adapted to be personalized based on the learner's emotion to fit as per learner suitability of learning. In [7], the facial expression is detected for emotion recognizing (sad, happy, etc.) to set the difficulty level of the task assigned to the learner, based on the learner's emotion the system tries to make the learning happy and joyful. Other applications like teaching process improvement, emotion detection, etc, are used in e-learning. While teaching over the internet, a teacher by using facial emotion detection technique can know the learner's feedback remotely and thus adapting the teaching methodology likewise [9].

Many advanced computer vision algorithms have been used to detect emotions from facial expression over the past few years. The facial expression detection involves process like facial landmark detection, facial feature extraction and classification. In [10], automatic facial expression detection was performed by extracting features by using wavelets transformation and emotion classification by K-nearest neighbours (KNN) algorithm. Typically, several number of features exist on one's face; therefore, the authors considered principal component analysis (PCA) for selecting facial features. In another approach, Krithika et al. [11] used Voila Jones algorithm and local binary patterns (LBP) for detecting a face, facial expression, eye and head movement to detect the attention, boredom etc. of the learner while learning. Machine learning and deep learning algorithms have been proved to be efficient in terms of implementation and results as compared to other contemporary systems [12]. Liyuan et al., in [13] shown that machine learning technique like linear SVM has shown a better result for identifying relevant and irrelevant facial expression using Gabor wavelet and shape feature. In another approach, HAAR cascades are used for detecting eyes and mouth feature combined with a neural network to provide much better emotion detection [14]. Guojon Yang et al. [15], proposed a deep neural network (DNN) model by using vectorized facial features. The human facial expressions are represented in vectors, allowing to train DNN with high accuracy. It is observed that among the many advanced machine learning techniques, convolution neural network (CNN) is proved to be much efficient in terms of automated feature extraction, lesser input and classification accuracy [16]. In [16], two facial expression detection methods, namely, autoencoder and CNN, have been compared. It is found, CNN statistically able to predict emotions with higher accuracy. Similarly, in [17], CNN has achieved better accuracy even by using the small-sized dataset (EMotiW). CNN not only used for detecting the face and facial expression, but it could be fine-tuned to detect important parts of the face instead of full face. Moreover, the CNN model works better over the existing models on multiple datasets, including FER-2013, CK+, FERG and JAFFE [18] [9] [19]. Conventionally emotion detection is performed on a static image, but detecting emotion detection from facial expression in a video is a challenge. Zhang et al. [20] proposed that two CNN models and a deep belief network (DBN) model, combined as a hybrid model, can extract facial expression very well from a running video.

Even though a lot of work has been done for emotion detection, they all focused on finding the basic emotion, while the complex emotions, natural to human face during the learning process, is ignored.

Table 2. Learner's state of mind, its implication in learning, and the Plutchik's combined emotion composition

Learner's state of mind	Implication in learning	Combination of emotions (Plutchik's theorem)
Confused	Learner is not sure about a concept, or feeling difficulty in comprehending the topic	Surprised + Anticipation
Satisfied	Successfully completed learning task (reading, understanding, solving. etc.)	Not Available
Dissatisfied	Not understanding the topic or learner is not happy with the content	Surprise + Sadness
Frustrated	Perform repeatedly poor for a given learning task	Not Available

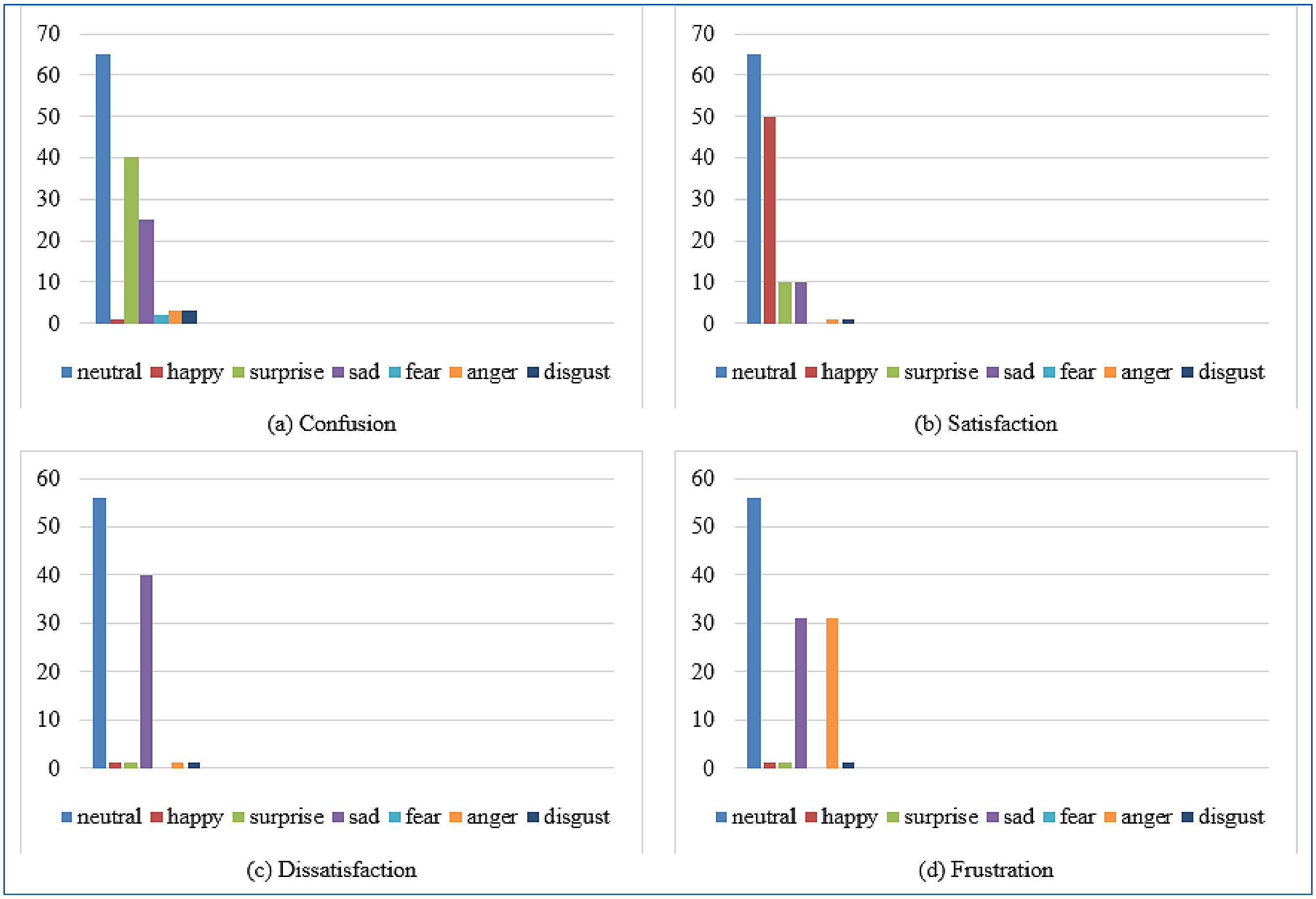

Figure 3. Basic emotion patterns for the state of minds of the surveyed learners

3. EMOTION DETECTION OF A LEARNER

3.1 Modelling the Combinatorial Emotions and State of Mind

3.1.1 Existing set of combined emotions

Assessing the learner state of mind during a learning session helps to know how does a learner feel about the learning material. Typically, humans do not explicitly express their feeling by a single emotion discretely rather by an orchestration of many emotions either simultaneously or in sequence. These emotions altogether lead to a complex emotion and often determined by a combination of the primary emotions, as postulated by Robert Plutchik [21]. Based on the general applicability for all kind of learners and wide application scope in learning, we have chosen four types of state of mind for assessment, as shown in Table 2.

It is found that the state of minds, like Satisfied and Frustrated, are not classified in Plutchik's theory of emotion [21]. Though Plutchik has theorized many emotions, not all of the complex emotions, as mentioned above, are defined by the combination of basic emotions. Furthermore, "Anticipation" is not identified as the basic emotion and is not being detected computationally [21]. Therefore, there is a need for emotion pattern finding that defines these complex emotions in terms of basic emotion combination.

3.1.2 Deriving combined emotions of a learner

To find the complex emotions of a learner from the combinations of basic emotions, a survey was conducted on 150 candidates, studying at the undergraduate level. Each student was given a small learning material, composed of reading material along with puzzles and programs. The learning material was so chosen that it might evoke the considered states of mind (Confusion, Satisfaction, Dissatisfaction, Frustration) of the learner. The facial emotions of each student while going through the learning material were recorded through video. The state of mind of each student was being monitored and noted by an expert. Similarly, the learner after completing the learning session was prompted to give feedback on his/her emotional states during the learning session; i.e., they were asked to select from the four options - Confusion, Satisfaction, Dissatisfaction, and Frustration.

Both the expert and learner feedbacks were compared to rule out the discrepancies. Cases with similar feedbacks were considered for further assessment. Only 115 candidates out of 150 learners were found to have similar feedback with that of the expert. The recorded videos were analysed to find the emotion patterns of each selected candidate. The video was split into image frames, whereby each image was analysed through a standard emotion detection API (Windows Azure) [22] for finding the basic emotion levels. The predicted confidence of each emotion class for each candidate are added. The basic emotion pattern for the state of minds is shown in Figure 3(a-d), as derived from expert's and learner's feedback. For example, Figure 3(a) shows that the learner who feels confused, exhibits higher basic emotions pattern for neutral, and shows either surprise or sad emotions, whereas in Figure 3(d), a candidate being frustrated exhibits both sad and anger emotions along with neutral.

Table 3 summarises the identified combinations of basic emotion pattern for each state of mind for 115 students after analysing their emotions. The found combinatorial emotions are neutral to the occurrence order and magnitude of the basic emotions.

Table 3. Basic emotion combination pattern for each state of mind

Combinatorial emotions	State of mind of a learner
Neutral + Surprise/sad	Confusion
Happy + Neutral	Satisfaction/delighted
Neutral + Sadness	Disappointment/dissatisfaction
Sad + Angry + Neutral	Frustrated

3.2 Emotion Detection

Emotion from the facial image could be detected by facial expression pattern matching. In this regard, machine learning tools are quite sophisticated in identifying the facial expressions and further classifying it for emotion. We chose CNN for facial expression identification and emotion detection. The reason for choosing CNN, among other available machine learning tools, is that the CNN, a supervised deep-learning algorithm, automatically identifies the features and represents the most discriminative features and hence allows for better performance. In CNN-based approaches, the input image is filtered to produce a feature map. Where each feature map is then computed through connected neural networks for recognition of facial expression and identifying the emotion class. CNN gives better accuracy than the other neural network-based classifiers [12].

3.3 Identifying Learner's State of Mind

For a learning process, it is observed that the emotion of a learner does not change instantly, and the emotion transformation happens gradually. Furthermore, the emotion changes stay on learner's face for a period of time (at least for 3 to 5 seconds), for a learning process. As a reason, complex emotion or the learner's state of mind cannot be detected by judging only one facial image. A sequence of images over a period of time is required to detect the state of mind of the learner while studying. For generalization, we consider that human emotion transformation may longs within a time frame of 6 seconds, approximately. To identify the change in emotion in the learner, a group of 6 images are taken in sequence by capturing images at a rate of one image per second. This creates a window of size 6, representing learner's images for the last 6 seconds, as shown in Figure 4. Appropriately, for each image, the facial expression is identified to assess the score of each class of basic emotion (as listed in Figure 1), as shown in Figure 5. The dominant emotion of the face image is identified as having a very high confidence score value. Since a face can display multiple shades of emotion, a different score value for each class of emotion is obtained. To normalize the classifier prediction error and excluding the minor emotion detected, a threshold value of 10% is set for selecting the appropriate emotion class score. In the window, for an image, all those emotions are selected whose confidence score value is more than 10%.

To find the emotion pattern from the 6 images in a window, the mean of the confidence score is calculated for all the respective emotions, as given by Equation (1).

$$M_X = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

Where, X = A, D, F, H, N, S, R.

The mean of the confidence of the corresponding emotions angry, disgust, fear, happy, neutral, sad and surprise are represented by M_A, M_D, M_F, M_H, M_N, M_S, and M_R respectively.

The set of the mean (M_E) of all detected emotions' confidence score for the 6 images in the window is defined in Equation (2), and the emotion pattern (E_P) is defined through Equations (3) to (6).

$$M_E = \{M_A, M_D, M_F, M_H, M_N, M_S, M_R\} \tag{2}$$

$$E_1 = \max(M_E) \tag{3}$$

$$E_2 = \max(M_E - E_1) \tag{4}$$

$$E_3 = \max(M_E - \{E_1, E_2\}) \tag{5}$$

$$E_P = \{E_1, E_2, E_3\} \tag{6}$$

Equation 6 represents three prominent emotions E_1, E_2, E_3, as selected from M_E having emotion confidence score $E_1 > E_2 > E_3$. The confidence score allows to choose the best of three prominent emotions. These three emotions, not considering their occurrence order and magnitude, show an emotion pattern for the six second time frame. The selected emotion pattern is mapped to Table 3 to find the relevant state of mind for the particular time period.

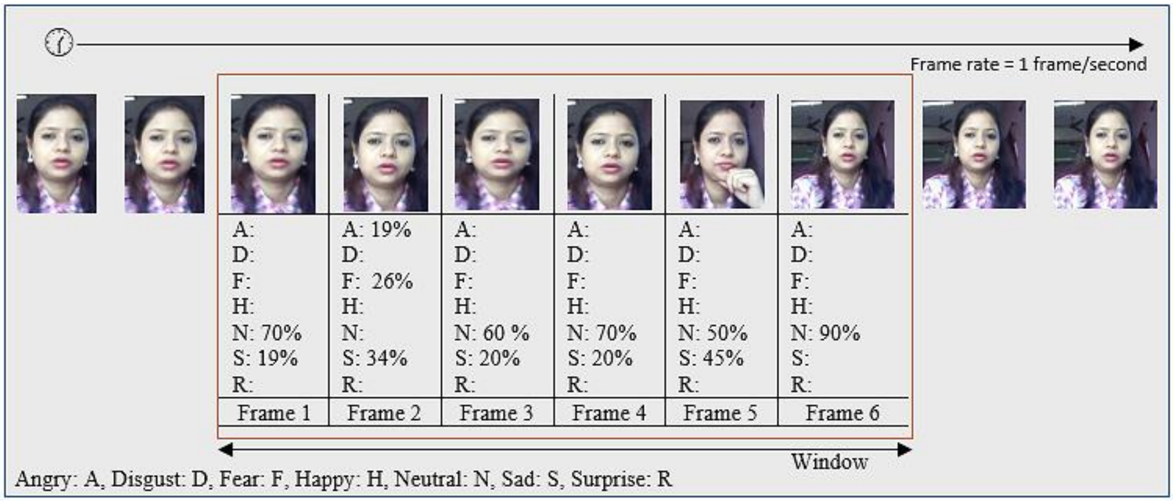

Figure 4. Basic emotion pattern recognition from a series of images

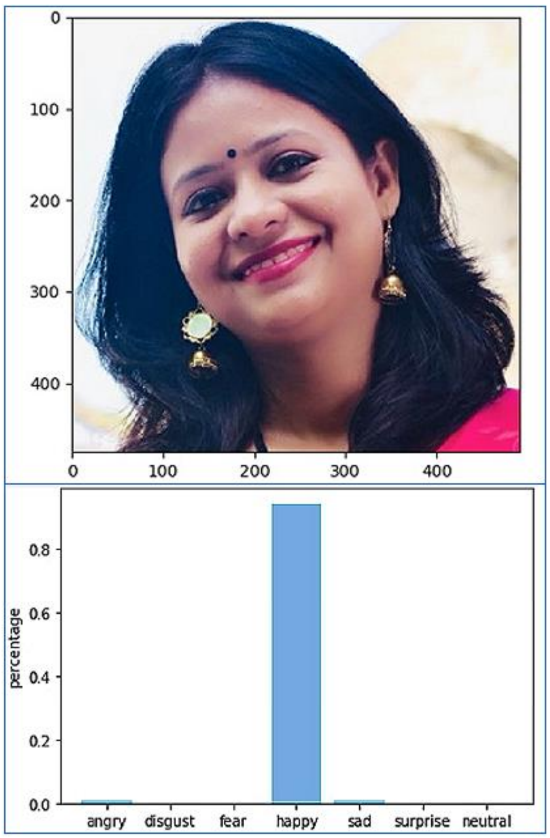

Figure 5. The confidence score of basic emotions in an image frame

3.4 Learner's State of Mind Identification Model in an E-learning System

The state of mind identification process is incorporated along with the Learning Management System (LMS) to have real-time learning adaptation, as shown in Figure 6. The web camera takes the video of the learner, from which frames (image of the learner) are grabbed at a frequency of 1 frame/sec.

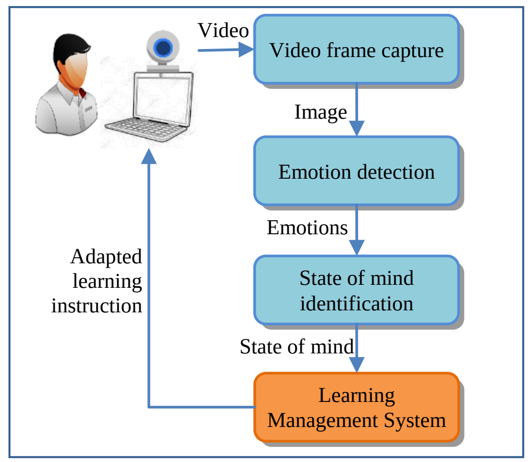

Figure 6. Emotion detection modelling in an e-learning system

The emotion detection module uses the trained CNN classifier to detect the different emotions in the image. The state of mind identification module takes a sequence of 6 images as a window frame to identify the emotion pattern of the learner for the last 6 seconds. The learner's state of mind is identified from the derived emotion pattern.

Further, the identified state of mind is sent to LMS as feedback of learner's emotion. Based on the identified state of mind of a learner, the LMS adapts the instructional or pedagogical method in real-time, making more suitable for the particular learner.

3.5 Experiment

The experiment for detecting the state of mind is conducted in the following three phases:

1. CNN, a deep learning algorithm, is being implemented for detecting emotions from images.
2. Identifying learner state of mind over a period of time.
3. Learner verification on the output for assessing the correctness of state of mind detected for learner emotion pattern.

Each step is discussed in detail in the following subsections.

3.5.1 *Working with CNN*

3.5.1.1 *Modelling and implementation*

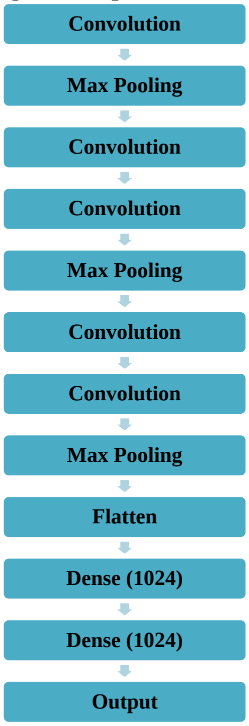

Figure 7. CNN model and its layered structure

The structure for our CNN model is given as in Figure 7. The structure consists of five convolution layers with ReLu activation function, three pooling layers, two fully connected layers, and the output layer. The functionality and the parameters set for each layer is given as:

- The convolution layer produces a feature map from the images. The convolution kernels are set at a size of 3 x 3.
- The max pooling allows reducing the dimension without losing important features and patterns. Each of the max pooling layers is set with a stride value of 2 and a pooling window of 2 x 2.
- The flatten layer converts the 2-dimensional data into 1-dimensional data that can be fed to a fully connected layer.
- The dense layer (or fully connected layer) is a combination of two or more neural network. The 1-dimensional data from flattening layer is fed as input to input nodes of each dense layer.
- The output layer has 7 nodes with SoftMax activation function, where each node stands for a different set of emotion.

The CNN model is being implemented using python. The built-up model is being further trained and tested over facial expression dataset for accuracy.

3.5.1.2 *Training step*

A CNN model requires a lot of labelled images for training. Here, the CNN model is trained with the Facial Expression Recognition (FER2013) dataset [23] [24], where each human face images in the dataset are labelled by the emotions that are reflected by the respective facial expression. The data consists of 48x48 pixel grayscale images of faces. The faces are more or less centred and occupy about the same amount of space in each image. The FER dataset consists of 28,709 samples, where 7 types of emotions are depicted, namely, Anger, Disgust, Fear, Happy, Sad, Surprised, Neutral. The FER dataset is split into two sets, training set (consisting of 80% sample data) and test set (consisting of 20% sample data). The training dataset is feed to CNN for training. Two techniques have been used to speed up the training process as:

- Image batches as input to CNN instead of using single images.
- Dropout regularization technique to get better performance on the model.

3.5.1.3 *Testing step*

We run our CNN model for around 50 epochs (considered to be as dataset run forward and backwards through CNN) and learn about the performance and accuracy of the model. After which, we use testing images to test the model. The model is also tested for real-time analysis on several input video sequences and webcam sequences for specific emotions depicted in each frame. The result is duly noted for each frame, and also errors are recorded for any misclassification that occurs; accordingly, suitable measures are taken.

3.5.2 *Learner verification*

Emotion and state of mind are not quantifiable and often are not formally expressible. Therefore, identifying the implicit learner state of mind need appropriate verification of the concerned person. For this reason, to measure the accuracy of our emotion model and emotion pattern recognizing approach, the classified emotion pattern needs learner's verification.

The approach is being assessed and verified by 40 candidates at the graduate-level course. A short online tutorial followed by a test session on machine learning is being carried over students.

While the learner is in the learning session, video recording of each candidate is being carried over to identify the learner's state of mind at different time. After the learning session is finished, the recorded video is analysed frame by frame to find the emotion pattern and thus state of mind over time. The emotion pattern detection mechanism is followed to derive the state of mind for every 6 seconds.

The learner's state of mind identified are aggregated across the entire learning session. The candidates are prompted to give their feedback for the correctness of the assessed aggregated learner's state of mind. It is being hypothesized that learner is truly able to respond to his state of mind correctly. The detected state of mind of a learner while a learning session is shown in Figure 8.

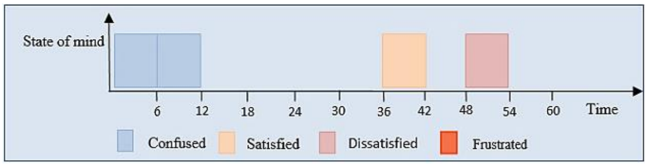

Figure 8. Learner's state of mind detected over time

4. RESULT ANALYSIS

Based on the two experiments – a) detecting and classifying emotion, b) detecting the state of mind (emotion pattern), the obtained results are shown and analysed separately for accuracy. The result is evaluated by four metrics, as described in Table 4 and using three measures, as described in Table 5 and defined in Equations (7) to (9).

Table 4. Performance evaluation parameters

True Positive	The model correctly predicts observation.
False Positive	The model incorrectly predicts observatio~~n~~.
True Negative	The model does not predict the observation which does not exist.
False Negative	The model missed in predicting the observation, or cannot recognize the observation which exists.

Table 5. Performance evaluation metrics

Accuracy	An intuitive performance measure, a ratio of correct observation made to total observation done.
Recall	The ratio of correct positive observation made to all correct and wrong observation made.
Precision	The ratio of current positive observation made to all predicted as well as non-predicted correct positive observation.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

4.1 Accuracy of the CNN Model

The CNN model is tested for accuracy, precision and recall over the test data and other real data. For the FER 2013 dataset, the model gives test accuracy of around 65%, as shown in Table 6. In FER 2013 dataset, ‘happy’ is the most dominant emotion. Thus, the CNN model trained over it gives better accuracy in predicting ‘happy’ and ‘neutral’ facial expression recognition than other emotions. Detecting ‘disgust’ and ‘fear’ have a lower accuracy rate. This inaccuracy is due to misclassifications of emotion such as deducing ‘sad’ instead of ‘fear’ and ‘angry’ instead ‘disgust’.

Table 6. The performance measure of CNN for emotion detection

Metric	Value
Accuracy	0.65
Precision	0.69
Recall	0.43

4.2 Performance Measure of the State of Mind Identification Model

The assessment for the accuracy of the proposed emotion model and emotion pattern recognizing approach is carried by manual verification over the predicted outputs. The learner verification with the predicted output for correctly identified, wrongly identified, and not identified the state of mind is shown in Table 7.

Table 7. The performance measure of the proposed approach

Metric	Value
True Positive	15
False Positive	8
True Negative	10
False Negative	7
Accuracy	0.62
Precision	0.65
Recall	0.68

4.3 Critical Analysis

Accuracy depends on the ratio of correctly identified state of mind and a total number of observations. In our approach, CNN shows 65% accuracy in classifying the basic emotions (listed in Figure 1). This led to the conclusion that 4 out of 10 emotions analysed are wrong. In our proposed approach, we used a combinatorial emotion pattern for detecting the state of mind. An incorrect emotion prediction leads to lesser accuracy in the state of mind detection.

The experimental result exhibits quite high false positive. This denotes that the model is quite susceptible to predict the state of minds that differ from learning the actual state of minds. Misclassifying the emotion ‘fear’ as ‘sad’ may lead the learner to have ‘confusion’ as a state of mind. Similarly, misclassifying ‘disgust’ as ‘angry’ wrongly detects learner to have ‘confusion’ which actually should be ‘frustrated’. The high false-negative depicts that the model often missed revealing the actual feelings of the learner. Facts like not able to detect ‘sad’ and ‘angry’ could result in ‘neutral’ feelings and thus not detecting the ‘confusion’ or ‘frustration’ state of mind. Some factors which lead to non-identification of emotions are:

- Captured images often do not contain the whole face or frontal view.
- Learner’s frequent movement, lead to instability in face recognition.
- Illumination is an important factor in capturing a clear image, poor lighting or very sharp lighting on face affects emotion detection.
- Uses of spectacles or glass lead to poor performance in emotion detection.
- Contempt learner or inexpressive learner shows no change in emotion.
- A matured learner does not show much of facial expression in a learning environment.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a method to assess the state of mind of the learners during an online learning session. We have identified four complex emotions (confusion, dissatisfaction, satisfaction, and frustration) that are combinations of the basic emotions. We also have established the fact that considering a single image capture for assessing the emotion is not sufficient. That’s why we have considered a window of six image frames, captured by the webcam, to assess the state of mind of the learner. Taking six image frames is also led by the consideration that the human emotions are continuous and are not very discrete; that is, they do not change abruptly; rather it takes a while (though very little) to change the state of mind.

In classifying learner’s emotions, our proposed CNN model performs fairly with a 65% accuracy. While the proposed model for identifying learner’s state of mind gives an accuracy of 62%.

The results show the state of mind identification for a learner is not very expressive. There is a lot of scopes to improve. Actually, considering only facial expression images are not sufficient to assess the human emotions correctly. Identification of emotions and the state of mind can be more accurate by considering other facial features like eye movement, eye blinks, eye gaze change, eyebrow movement, and other subtle facial expressions such as curves and micro-expressions.

6. REFERENCES

[1] S. Pal, P. K. D. Pramanik and P. Choudhury, “A Step Towards Smart Learning: Designing an Interactive Video-Based M-Learning System for Educational Institutes,” International Journal of Web-Based Learning and Teaching Technologies, vol. 14, no. 4, pp. 26-48, 2019.

[2] S. Pal, P. K. D. Pramanik, T. Majumdar and P. Choudhury, “A semi-automatic metadata extraction model and method for video-based e-learning contents,” *Education and Information Technologies,* vol. 24, no. 6, pp. 3243-3268, 2019.

[3] M. A. A. Dewan, M. Murshed and F. Lin, “Engagement detection in online learning: a review,” *Smart Learning Environments,* vol. VI, no. 1, pp. 1-20, 2019.

[4] M. Dubey and L. Singh, “Automatic Emotion Recognition Using Facial Expression: A Review,” *International Research Journal of Engineering and Technology,* vol. III, no. 02, pp. 488-492, 2016.

[5] S. L. Happy, A. Dasgupta, P. Patnaik and A. Routray, “Automated Alertness and Emotion Detection for Empathic Feedback during e-Learning,” in *IEEE Fifth International Conference on Technology for Education*, Kharagpur, 2013.

[6] H. H. Binali, C. Wu and V. Potdar, “A new significant area: Emotion detection in E-learning using opinion mining techniques,” in *IEEE International Conference on Digital Ecosystems and Technologies*, Istanbul, 2009.

[7] M. Saneiro, O. C. Santos, S. Salmeron-Majadas and J. G. Boticario, “Towards Emotion Detection in Educational Scenarios from Facial Expressions and Body Movements through Multimodal Approaches,” *The Scientific World Journal,* vol. 2014 (Article ID 484873), 2014.

[8] K. Fujii, P. Marian, D. Clark, Y. Okamoto and J. Rekimoto, “Sync Class: Visualization System for In-Class Student Synchronization,” in *9th Augmented Human International Conference*, Seoul, 2018.

[9] A. Sun, Y.-J. Li, Y.-M. Huang and Q. Li, “Using facial expression to detect emotion in e-learning system: A deep learning method,” in *International Symposium on Emerging Technologies for Education (SETE 2017)*, Cape Town, South Africa, 2017.

[10] J. Ou, "Classification Algorithms Research on Facial Expression Recognition," in *International Conference on Solid State Devices and Materials Science*, Hainan, China, 2012.

[11] L. B. Krithika and G. G. Lakshmi Priya, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” *Procedia Computer Science,* vol. 85, pp. 767-776, 2016.

[12] I. M. Revina and W. R. SamEmmanuel, “A Survey on Human Face Expression Recognition Techniques,” *Journal of King Saud University - Computer and Information Sciences,* 2018.

[13] L. Chenak, C. Zhoua and L. Shenb, “Facial Expression Recognition Based on SVM in E-learning,” *IERI Procedia,* vol. 2, pp. 781-787, 2012.

[14] D. Yanga, A. Alsadoona, P. Prasad, A. K. Singh and A. Elchouemi, “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment,” in *6th International Conference on Smart Computing and Communications, ICSCC 2017*, Kurukshetra, 2017.

[15] G. Yang, J. S. Y. Ortoneda and J. Saniie, “Emotion Recognition Using Deep Neural Network with Vectorized Facial Features,” in *IEEE International Conference on Electro/Information Technology (EIT)*, Rochester, Michigan, USA, 2018.

[16] P. R. Dachapally, “Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units,” *arXiv,* no. 1706.01509, 2015.

[17] H.-W. Ng, D. V. Nguyen, V. Vonikakis and S. Winkler, “Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning,” in *ACM International Conference on Multimodal Interaction,* Seattle, 2015.

[18] S. Minaee and A. Abdolrashidi, “Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network,” *arXiv,* no. 1902.01019, 2019.

[19] M. M. Taghi Zadeh , M. Imani and B. Majidi , “Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters,” in *5th Conference on Knowledge-Based Engineering and Innovation, Iran University of Science and Technology*, Tehran, Iran, 2019.

[20] S. Zhang, X. Pan, Y. Cui, X. Zhao and L. Liu, “Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning,” *IEEE Access,* vol. 7, pp. 32297-32304, 2019.

[21] M. M. Abbasi and A. Beltiukov, “Summarizing Emotions from Text Using Plutchik’s Wheel of Emotions,” in *Scientific Conference on Information Technologies for Intelligent Decision Making Support*, Ufa, 2019.

[22] Microsoft Azure, “Face,” Microsoft, 2019. [Online]. Available: https://azure.microsoft.com/en-in/services/cognitive-services/face/. [Accessed 1st December 2019].

[23] S. Li and W. Deng, “Deep Facial Expression Recognition: A Survey,” *Computer Vision and Pattern Recognition,* pp. 1-25, 2018.

[24] R. Verma, “fer2013,” Kaggle Inc, 26 May 2018. [Online]. Available: https://www.kaggle.com/deadskull7/fer2013. [Accessed 22 November 2019].