Multimedia Tools and Applications (2019) 78:25321–25348
https://doi.org/10.1007/s11042-019-7651-z

Students' affective content analysis in smart classroom environment using deep learning techniques

Sujit Kumar Gupta[1] · T. S. Ashwin[1] · Ram Mohana Reddy Guddeti[1]

Received: 17 June 2018 / Revised: 12 February 2019 / Accepted: 15 April 2019 /
Published online: 24 May 2019

Abstract

In the era of the smart classroom environment, students' affective content analysis plays a vital role as it helps to foster the affective states that are beneficial to learning. Some techniques target to improve the learning rate using the students' affective content analysis in the classroom. In this paper, a novel max margin face detection based method for students' affective content analysis using their facial expressions is proposed. The affective content analysis includes analyzing four different moods of students', namely: High Positive Affect, Low Positive Affect, High Negative Affect, and Low Negative Affect. Engagement scores have been calculated based upon the four moods of students as predicted by the proposed method. Further, the classroom engagement analysis is performed by considering the entire classroom as one group and the corresponding group engagement score. Expert feedback and analyzed affect content videos are used as feedback to the faculty member to improve the teaching strategy and hence improving the students' learning rate. The proposed smart classroom system was tested for more than 100 students of four different Information Technology courses and the corresponding faculty members at National Institute of Technology Karnataka Surathkal, Mangalore, India. The experimental results demonstrate the train and test accuracy of 90.67% and 87.65%, respectively for mood classification. Furthermore, an analysis was performed over incidence, distribution and temporal dynamics of students' affective states and promising results were obtained.

Keywords Facial expression · Structure of moods · Mood classification · Affective content analysis · Students' engagement · Inception-V3

Sujit Kumar Gupta and T. S. Ashwin contributed equally to this work.

✉ T. S. Ashwin
ashwindixit9@gmail.com

Sujit Kumar Gupta
sujitnik16@gmail.com

Ram Mohana Reddy Guddeti
profgrmreddy@nitk.edu.in

[1] Department of Information Technology, National Institute of Technology Karnataka Surathkal, Mangalore, India

1 Introduction

Creating a good classroom climate to improve the learning rate is a challenging task. The students' engagement in the classroom is an important factor as it improves the learning rate and the academic performance of an educational institute. Bomia [4] defined the students' engagement as a student's willingness, needs, desire, motivation and success in the learning process. The teaching style and strategy are also other factors to create the students' attention in the classroom. Students' engagement is analyzed using various ways such as self-reports, manual observation, computer vision based techniques for emotion & behavior analytics, pre-test & post-test performance-based analysis, conversation-based text analysis. The existing literature contains the student's emotion recognition from computer vision and physiological signals for a single student. But for large classrooms, it is very difficult to recognize the students' emotions. Using tools such as clickers, subtle stones, physiological signals from sensors are obtrusive [5, 8, 17, 24, 31, 37, 42]. Computer vision based techniques are unobtrusive. According to students' obtrusive and unobtrusive feedback in the classroom, the faculty member must adapt to the needs of students for a better teaching-learning process. Expectations from the faculty member with their affect states and how they interact with the students in the classroom will play a vital role in the students' learning process. It is important for a faculty member to be focused, setting immoderate expectations for students' achievement, and working hard to develop good relationships with students. To address these issues many modifications, adaptations and changes are observed in several significant areas of classroom environments [34].

Students' affect states play a vital role in understanding the positive classroom climate for any type of teaching-learning process. If the teaching strategy is not effective, then the students will not be interested to learn in the classroom. Even if the topic is interesting to the students, the students can feel bored and the learning rate of the classroom decreases [32]. In [14, 28, 35] authors indicated that the students' boredom mood in the classroom and the educational environment is one sign of a lack of engagement. For example, the students who are bored in the classroom may take less interest and will not be paying attention to the lecture, subsequently becoming even more bored over time. Such students ultimately end up in getting poor academic results, involving in many in-disciplinary actions and even facing the risk of dropping out.

There are many students' engagement analysis methods based on various factors considered [11, 13, 17, 23, 24, 27, 29, 32, 35, 36, 42, 43]. Some works also considered the students' feedback, problem-solving techniques, student's posture analysis using sensors but these are either inefficient or obtrusive [14, 40]. There are several methods which perform the engagement analysis of students' in the e-learning environment using facial expressions [2, 9, 18, 30, 33]. These techniques are neither tested nor experimented on large classrooms with multiple students in a single image frame. Also, most of the works use either basic emotions such as happiness, sadness, fear, disgust, anger, and surprise [4, 19] or affect emotions such as bored, frustrated, eureka, engaged, etc. [20]. There are some works on positive and negative human affect states [5], but there are fewer works on analyzing the students' positive and negative affect states [11]. Hence, this motivates us to explore the area of positive and negative affect states of students using their facial expressions in a classroom environment for group engagement analysis (multiple students in a single image frame). To address the above-mentioned issues, an architecture is proposed with a core idea of analyzing the students' affect states using their facial expressions for group engagement analysis in a classroom environment. This analyzed students' affect state can be used by the faculty member as feedback to improve their teaching style and strategy so that the learning rate of all the students present in the classroom can be enhanced.

Most of the existing methods on students' engagement analysis are addressed only for the e-learning environment with a single person in a single image frame. Further, these works did not involve the feedback mechanism for the faculty member to improve the teaching process further. There exist several works which provide the computer vision-based students' engagement analyzed data with proper visualizations for the faculty member as feedback to understand the analyzed data easily [12, 35].

Hence, in this paper, a novel method is proposed to overcome the drawbacks of the existing techniques where the students' engagement is classified into high positive, low positive, high negative and low negative affect states. The classified data is analyzed using statistical methods, and also the engagement scores are calculated. This analyzed data is used as feedback to the faculty member in the form of various visualizations such as transition diagrams, affect flow diagram, pie chart, etc. to further improve the positive classroom climate. Further, the proposed system provides the students' affect feedback of each classroom to the expert faculty member to improve the teaching style and strategy of the newly joined instructor or the teaching assistant.

The key contributions of this paper are as follows:

- An optimized and efficient unobtrusive students' affect state recognition in the classroom environment using students' facial expressions and classification of these affect states into high positive, low positive, high negative and low negative affect states.
- A novel group engagement analysis of students and providing the analyzed data for the faculty members as the feedback to further improve the teaching-learning process.
- Creation of a students' affect database for the classification of positive and negative affect states of students in the teaching learning process.

Rest of the paper is organized as follows. Section 2 discusses the related work on existing students' engagement analysis using facial expressions. Section 3 describes the proposed methodology. Section 4 discusses the experimental results, and finally, Section 5 concludes the paper with future directions.

2 Related work

Students' engagement analysis using facial expressions in a classroom environment is a challenging task. The students' engagement has a significant impact on the classroom climate. Works in [31] observed that the students who are more emotionally engaged in the classroom are generally interacting with the faculty member. However, emotionally disengaged students are not interacting with the faculty member. There are various other methods to calculate the students' engagement such as examination, assessments, quizzes, scores, test and other academic measurements [12, 20]. But these methods are not sufficient to calculate the impact of teaching in the classroom environment as the student can have prior knowledge about the subject or can self-learn the concept and then take the examination.

Managing large classrooms with more number of students is difficult. These issues are observed in countries with a high population density. Several research works are carried out to address this issue, and in some works [10, 29], it is found that the Asian students are less engaged when all the students present in the classroom are considered. Here the participation level of students in learning the subject in a classroom is less. Further, it is concluded that there are approximately 20% of students who asked questions during the lecture period and also shown interest in the classroom teaching-learning process. This is because of the disinterest in ongoing lessons which are difficult to manage by one faculty member and/or less effective teaching strategy.

The students' engagement score is important to improve the learning rate as it significantly affects the academic performance of an educational institution. The researchers also observed that many students feel bored, less interaction with the faculty member, unmotivated and bad classroom climate from the academics is also one of the reasons for poor engagement score of students [13, 17, 24]. But these works did not mention how to collect each students' emotion or mood during the teaching-learning process.

Some researchers tried to improve the students' engagement score in the classroom using traditional methods such as collecting feedback from each student, questionnaire survey, etc. [7, 8, 42]. According to their results, it is observed that the traditional methods have poor efficiency for immeasurable studies of the problem [36].

In [4, 26, 34], the authors recognized the students' engagement using the facial expressions of students using machine learning techniques. But their system works for a single student in a single image frame, and also, they did not give any feedback review system to the faculty member for further improving the teaching process. In [16], authors detected the students' face and carried out eye gaze analysis to find the engagement level of the students but eye gaze analysis is not sufficient to calculate the engagement score properly.

Further, there are other works on students' engagement and affective content analysis [1, 4, 6, 14, 28, 32, 34, 35, 40] but, they are either obtrusive (usage of the physiological sensors on the body or the tools used by the students for providing affective content as a feedback makes it easily noticeable and draws attention) or fail to give better accuracy. There are many state-of-the-art techniques to identify the students' mood which are summarized in Table 1, but these works were not enough to improve the teaching-learning process and also do not provide any affect/expert feedback for the further improvement in the teaching-learning process.

3 Proposed methodology

In this paper, students' affective content analysis method is proposed to calculate the students' engagement score and thus helps in improving the students' learning rate with the faculty member development process by providing the students' engagement score as feedback to the faculty member for a better positive classroom climate. The moods are predicted using students' facial expressions in the classroom environment. Figure 1 shows the system architecture for one-hour classroom video lecture which is divided into smaller video clips of 5-6 minutes. The faces of all students are recognized using the proposed Max-Margin Face Detection technique (MMFD). From these recognized and cropped face images, individual mood classification is performed using the convolutional neural network (CNN) architecture with the help of proper training data. The predicted mood of each student is combined for each frame to calculate the group or classroom engagement score.

3.1 Steps involved in proposed methodology

The proposed system introduces a novel mood analysis method in the classroom environment to analyze the students' engagement where engagement scores are calculated and these scores are provided to faculty members as feedback for further improvement in the teaching-learning process and thus, improving the students' learning rate. The proposed system uses

Table 1 Summary of existing works

Works	Methodology	Merits	Limitations
Whitehill et al. [39]	Engagement analysis from students' facial expressions using machine learning.	Classified the students' engagement into high and low engagement.	Context factors and time is not considered.
Ahlfeldt et al. [1]	Analyze the students' engagement using PBL.	Defined the various factors which affects the engagement levels.	Relation between engagement score and understating concept is not considered.
Burnik et al. [6]	Analyze the students' engagement using Kinect device.	Monitor the students' attention during lecture using gaze and behaviour cues.	No feedback is given to either teacher or student to improve further.
Thomas et al. [34]	Analyze engagement based on facial behaviour cues.	Classified students' engagement into distract and engaged.	Context factors are not considered.
D'Mello et al. [11]	The affect states analysis based on problem solving.	Classified students' engagement into Curiosity, Boredom, Frustration and Happiness.	Posture analysis using sensor on every desk make it unobtrusive.
Subramainan et al. [32]	The students' engagement model based on emotion is animated by an agent-based social simulator.	Try to identifying the students' misbehavior, motivation and participation.	Video affective content, is not used.
Sagayadevan et al. [28]	Analyze the engagement based on academic outcomes.	Identifies the relationship between students and faculty interaction, engagement and academic outcomes.	Only questionnaires based analysis is performed .
Ventura et al. [35]	Analyze the engagement based on video summaries assessment	Classified students' engagement into engage and non-engage.	No feedback to faculty for further improvement of the teaching-learning process.
Guo et al. [14]	Affective video production based on students' engagement	Build the public datasets based on MOOC video usage.	They cannot identify whether the student is actively paying attention to the video or not.
Wong et al. [40]	Impact of both mechanics and dynamics of a serious game.	Both the user engagement and concentration are considered for the learner engagement.	User's enjoyment and concentration may vary with situation and it is considered for evaluation.

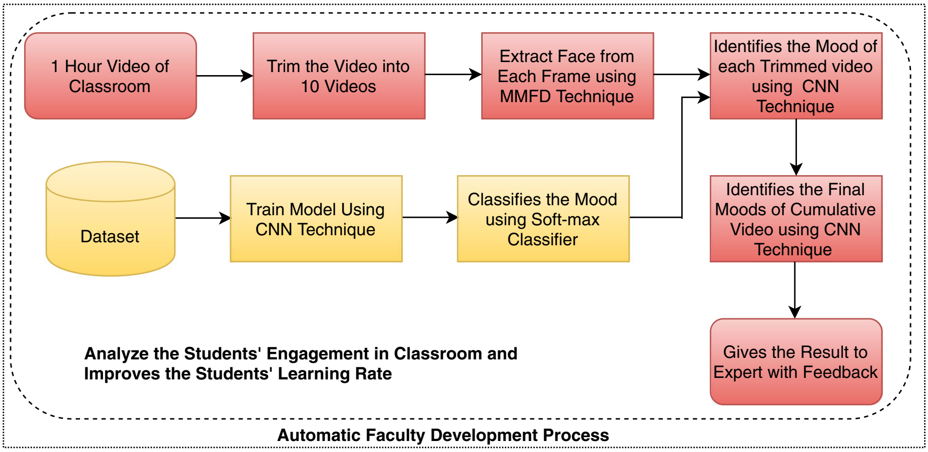
1 Hour Video of Classroom
Trim the Video into 10 Videos
Extract Face from Each Frame using MMFD Technique
Identifies the Mood of each Trimmed video using CNN Technique
Dataset
Train Model Using CNN Technique
Classifies the Mood using Soft-max Classifier
Identifies the Final Moods of Cumulative Video using CNN Technique
Gives the Result to Expert with Feedback
Analyze the Students' Engagement in Classroom and Improves the Students' Learning Rate
Automatic Faculty Development Process

Fig. 1 The flow diagram of the proposed system architecture

a rotating high definition camera, equipped in the classroom, which streams the video of approximately one hour class period. For a class period of one hour video, the proposed architecture shown in Fig. 4 follows the below-mentioned steps.

1. Every one-hour classroom video is segmented into ten video segments where each video segment consists of 5 to 6 minutes duration.
2. Pre-process the data (extract frames, resize the frames as fixed size (uses 800px X 600px)).
3. Detect the students' faces from each frame using the proposed Max-Margin Face Detection (MMFD) technique.
4. Predict the mood of each student from the detected face using the proposed Modified-Inception V-3 Model.
5. Get the sub-final mood of each video segments as Max(Avg(all four mood scores from all frames)) as shown in the transition state of one-hour classroom video (Fig. 16).
6. Finally, calculate the final mood of one-hour video lecture of the classroom using Max(Avg(all four sub-final mood score from 10 video segments))
7. If any affect (High Negative or High Positive) is present for more than 70% time duration for the entire class then intimate the teacher about the same and ask the teacher to change the teaching strategy or style.
8. If *special case* occurs then provide the entire video segment to an expert for further analysis and feedback for improvement in teaching strategy.
9. Give feedback after each class based on the analyzed result to the subject expert as well as the faculty member.

Special Case- Each one hour class consists of 10 video segments with a mean video segment length of 5-6 minutes. If the predicted group moods of 4-5 minutes of entire 5-6 minutes video length are consistently present in only one of the major moods such as high positive or high negative affect states then considered this as the special case. Here, more than 70% of the entire video segment is in only one state for the entire class. By providing this entire video segment to the expert helps in analyzing the group engagement score with the context and other external parameters and thus helps in providing better feedback for the faculty members to improve the learning rate of students.

The detailed explanation of the proposed architecture, workflow and the steps involved in the proposed architecture are given in the following subsections of Section 3. The complete work-flow of the proposed affective content analysis is shown in Fig. 2.

3.2 Optimal video length for student engagement

Affect feedback for small segments of one-hour video data is better than providing one affect feedback for one-hour video data. It is important to generalize the video length of each video segment for one-hour classroom data. Guo et al. [14] defined how video production decisions affect students' engagement in online educational videos. Authors measured engagement by analyzing how long students can watch each video. Similarly, continuously how long a student can be engage in the classroom for a fixed period. Authors observed that the median engagement time for video length is 6 minutes long and was close to 100%. The average engagement time of any video is at most 6 minutes, regardless of its length. And engagement time decreases as video duration is more. For instance, the average students spent around 3 minutes on videos that are longer than 12 minutes, which means that they are

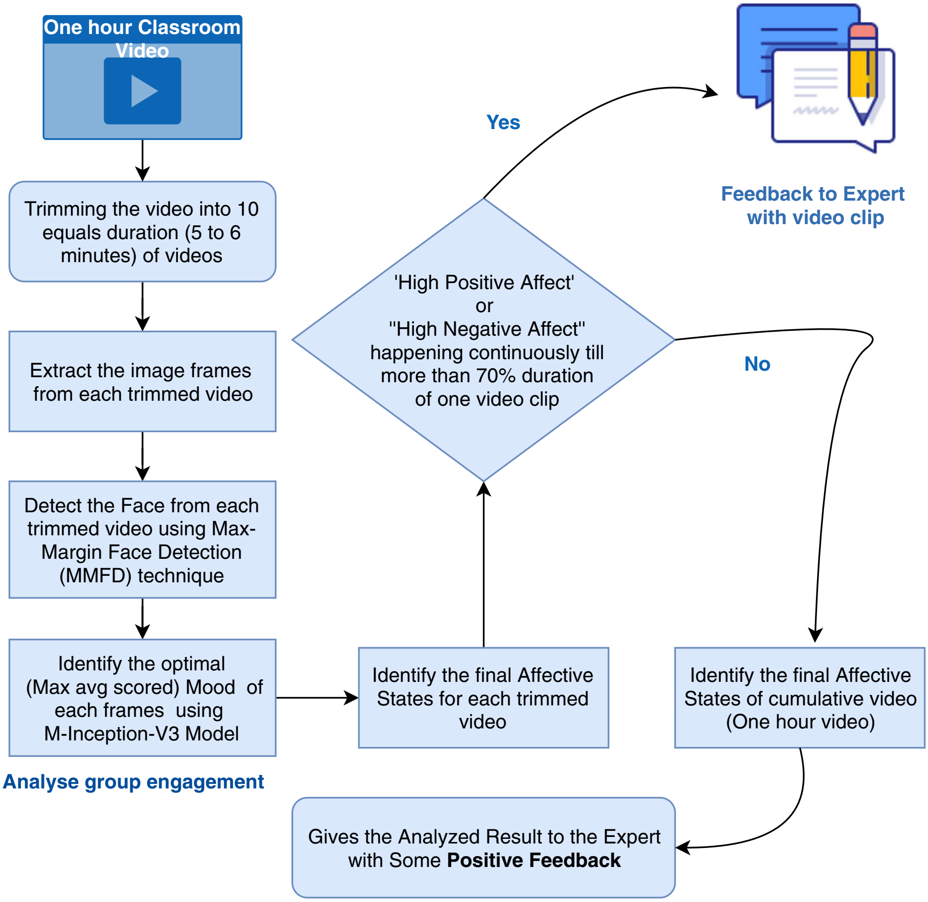
One hour Classroom Video
Trimming the video into 10 equals duration (5 to 6 minutes) of videos
Extract the image frames from each trimmed video
Detect the Face from each trimmed video using Max-Margin Face Detection (MMFD) technique
Identify the optimal (Max avg scored) Mood of each frames using M-Inception-V3 Model
Analyse group engagement
Identify the final Affective States for each trimmed video
'High Positive Affect' or "High Negative Affect" happening continuously till more than 70% duration of one video clip
Yes
No
Feedback to Expert with video clip
Identify the final Affective States of cumulative video (One hour video)
Gives the Analyzed Result to the Expert with Some Positive Feedback

Fig. 2 Flow chart of proposed students' affective content analysis

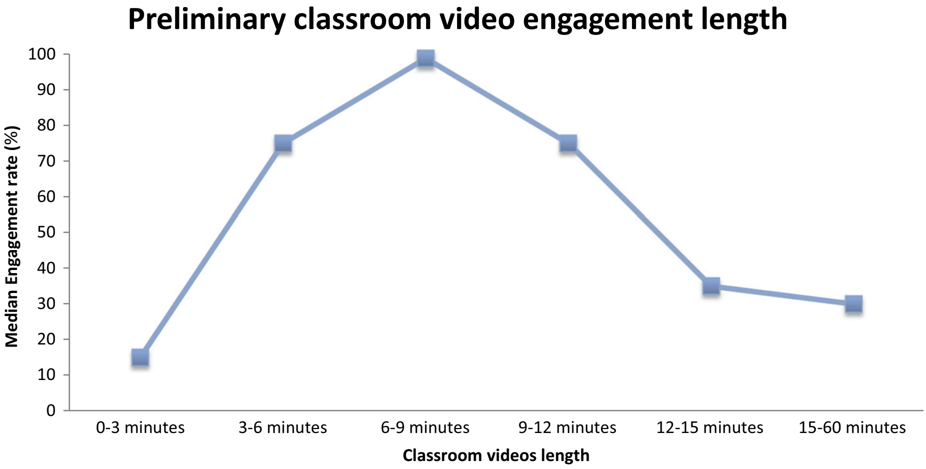

Fig. 3 Sample engagement rate of the classroom video segment

engaged with less than a quarter of the content. Figure 3 shows that the median engagement time is at most 6 minutes, regardless of total video length.

Guo et al. [14] observed that for a large classroom video lecture, the length of each segment of video lecture is in between 5 and 6 minutes. Figure 3 shows that the student engagement rate dropped, such that the median engagement rate with 9-12 minute video is $\approx 75\%$ and the median engagement rate with 12-15 minute videos is $\approx 35\%$. The maximum median engagement time for a video of any length is 6 minutes. Therefore, videos longer than 6-9 minutes is likely to be futile.

3.3 Face detection

In computer vision, detecting and tracking the face is a challenging task. There are many techniques such as Viola-Jones [36], HAAR Cascade classifier [30], Histogram of the oriented gradient (HOG) [9], Local Binary Pattern (LBP) [2], Linear Discriminant Analysis [15], Scale Invariant Feature Transform (SIFT) [21, 22], Independent Component Analysis (ICA) [3], Eigen signatures to detect the face in real time [41]. But the main limitation of these methods is that tilted and occluded faces will not be detected with good accuracy.

To analyze the mood of students in the classroom using their facial expression, detection of their faces for all its variants is a vital part. But this system requires a robust method to recognize the face even if it is tilted, occluded or with a very difficult recognizable pose.

Dalal and Triggs [9] introduced the Histogram of Oriented Gradients (HOG) feature for detecting the face within a sliding window framework. They performed detection of the frontal face using HOG based features extraction and classification, but the accuracy is less for the detection of tilted faces.

In this paper, the HOG with face detection technique is used by incorporating Max-Margin Object Detection (MMOD) [18] technique after generating the sampling window as an optimizer. This optimizer is designed to convolve over the entire window and thus optimizes the performance of a face detection system in terms of the number of missed detections and false alarms of the final output obtained from HOGs. It is observed that a single rigid HOG filter can outperform a state-of-the-art model if the HOG filter is learned via MMOD [18]. This technique does not play out any sub-sampling, but it optimizes overall sub-windows. Hence, the proposed MMFD method is a hybrid approach of combining both HOG and MMOD to improve the face detection method which is linear for the learned parameters such as HOG or sack of-visual-word models [18]. It is also evident that a solitary inflexible HOG channel can beat a cutting-edge deformable part display on the Face Detection Dataset and Benchmark (FDDB) when the HOG filter is learned via MMOD. This method is more accurate to detect all varied faces positioned frontal, tilted up/right/left/down and occluded faces.

Hence, the proposed MMFD model is trained to detect the following cases:

- No need to pay the attention while capturing the image in the classroom.
- The technique works even in pose variation, occlusion, and illumination.
- Pre-aligned face is not required.

Thus the proposed MMFD method can detect frontal, every pose of tilted faces with 99.56% accuracy. Figure 4 shows sample of the detected face using the proposed method.

3.4 Mood and affective analysis

A novel convolutional neural network architecture is proposed for classifying the human moods from dynamic facial expressions. The transfer learning is used on the fully connected layers of an existing convolutional neural network which is pre-trained for human emotion classification. A variety of datasets, as well as own created image dataset, are used to train the model. Affect, emotions and moods are the three different terminologies used in proposed work and are closely related. A broad range of feelings that a person generally exhibits

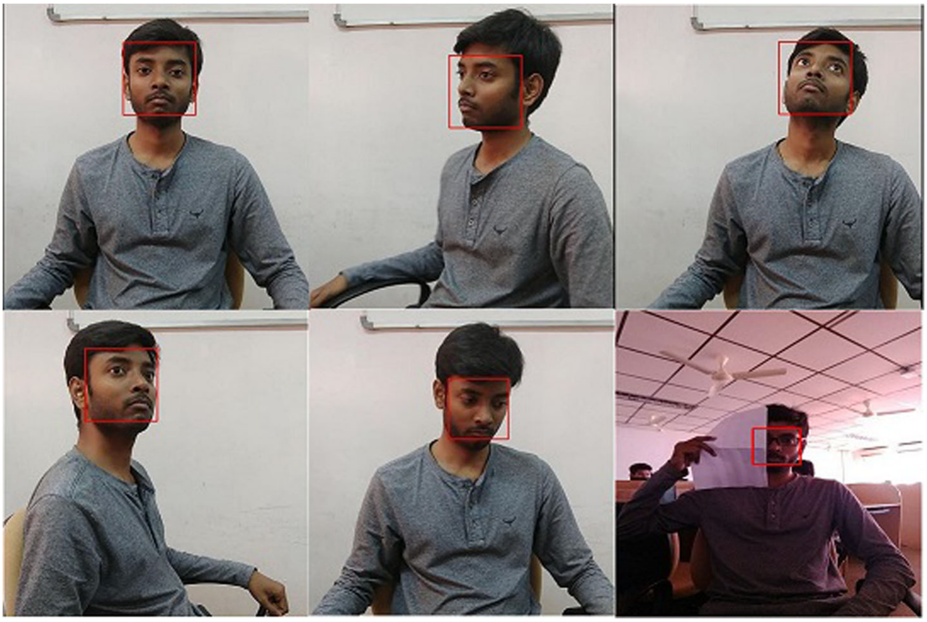

Fig. 4 Some sample of detected face in different orientation

is covered in affect. It is an umbrella concept that encloses both emotions and moods. Emotions are the enormous feelings that are directed at something. Moods are feelings that are less enormous and often.

3.4.1 Mood structure

The mood of students should be either positive or negative based on activities considered in the classroom. Langton [26] defined the positive and negative mood based on basic emotions. On the basis of ”the structure of mood” [19] and ”Two-factor structure of affect” [38], a cluster of emotions is proposed and classified into positive and negative categories, considering all the mood states of both positive and negative affect of students instead of isolating one particular emotion. For e.g., in Fig. 5 the excited emotion is marked with the high positive affect and bored is marked with the low positive affect. Similarly, nervous is marked with the high negative affect while relaxed is marked with the low negative affect. Finally, clustered emotions are categorized into four affect states as shown in Table 2. Further, the neutral is considered under low negative affect for students' engagement analysis as it is observed in most of the students during the teaching-learning process.

Mood analysis of students in the classroom is mainly focused on:

- faculty (newly recruited faculty member or teaching assistant) development program for improving the teaching-learning process in the classroom.
- Analyze the students' affect state in the classroom environment.
- Improve the students' learning rate based on students' engagement score by analyzing the students' affect state.

3.5 Measuring the engagement in classroom

It is possible to measure the students' engagement from classroom videos, but the true engagement is impossible to measure without direct observation and questioning, which is

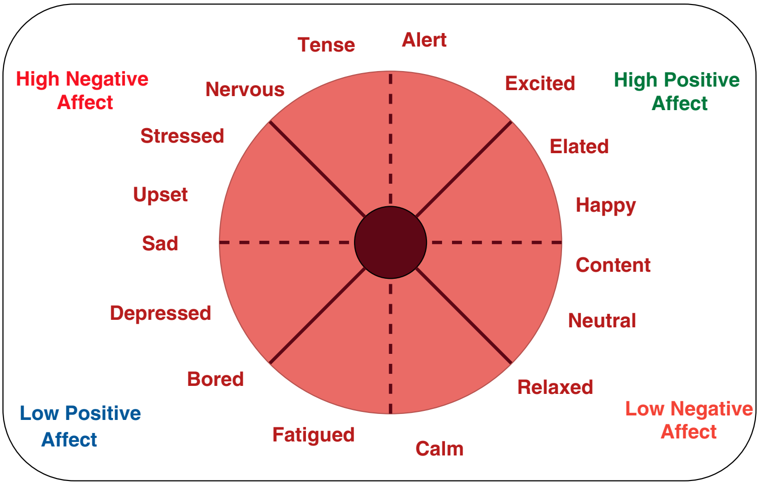

Fig. 5 Proposed mood structure

Table 2 Mood classification based on emotions

Emotion	Corresponding mood
Alert	High Positive Affect
Excited	
Elated	
Happy	
Content	Low Negative Affect
Relaxed	
Calm	
Neutral	
Fatigued	Low Positive Affect
Bored	
Depressed	
Sad	
Upset	High Negative Affect
Stressed	
Nervous	
Tense	

in-feasible at scale. Hence, the mood analysis is performed using the facial expression of students in the classroom to measure the engagement and learning rate.

3.6 Proposed classification method

This section discusses the proposed methods to predict the mood of multiple students present in the classroom.

Method to Identify the Mood The following steps are involved to identify the students' mood in the classroom of every one-hour video lecture.

- One hour classroom video is trimmed into 10 equal slots with duration of (5-6 minutes) video segments $V_j \subseteq \{V_1, V_2, V_3, ..., V_{10}\}$
- Extract the frames from each video segment.
- Detect the students' faces from each frames using the proposed MMFD.
- Identify the mood (class-name and score) of each detected face using the M-Inception-V3 model, Fig. 6 shows the M-Inception v3 architecture.
- Sub-final mood of each video segments=Max(Avg$_i$(all four mood scores from all frames))

$$Avg\_Mood\_score\_of\_video\_segment\_V_i = \frac{\sum_{Mood_i Score}}{Number\_of\_detected\_Moods}$$

- Final mood of one hour video= Max((Avg$_i$(all four sub-final mood score from 10 videos segments)) {i=Affective states, j=Video segments}

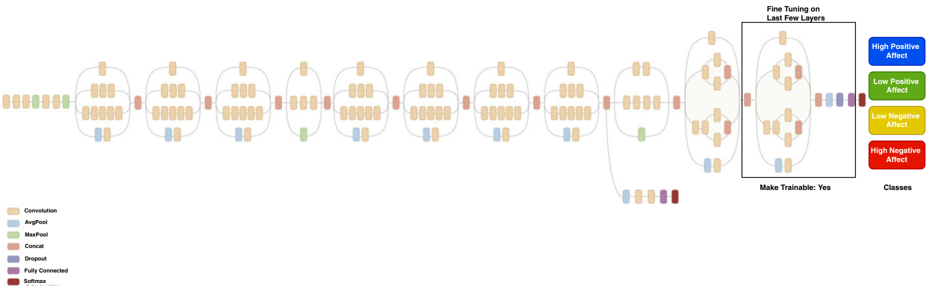

Fig. 6 M-Inception-V3 architecture

The students' affect states are classified using the modified version of inception-v3 (M-Inception-V3) model. In M-Inception-V3 model, the last 3 layers of 48 deep layer Inception-V3 model is modified and also the hyper-parameters such as loss function is optimized. The hyper-parameter is optimized to reduce the loss and thus obtained the optimal architecture. Fine tuning is performed for last 3 layers, the learning rate was initialized with the value of 0.046 and optimized the loss function using gradient descent to reduce the loss and to improve the performance of the model. In the last layer, softmax classifier is used to classify the four different affect states. The proposed M-Inception-V3 architecture is shown in Fig. 6.

3.7 Automatic faculty development process

After obtaining the analyzed affect content data and based on the feedback obtained from the expert (expert prepares a detailed report on the improvements to be incorporated in the teaching strategy and provides this report as feedback). This report is used by the newly recruited faculty members so that they can improve the teaching-learning process and also can create a positive classroom climate in order to improve the learning rate. Figure 7 shows the feedback based teaching-strategy improvement process.

Fig. 7 Feedback based teaching-strategy improvement process

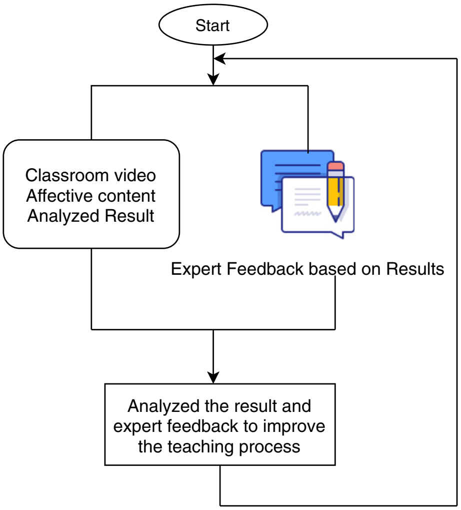
Start
Classroom video Affective content Analyzed Result
Expert Feedback based on Results
Analyzed the result and expert feedback to improve the teaching process

Table 3 Participant students details

Courses	Male students	Female students
Undergraduate	75	35
Graduate	62	35

4 Dataset creation

The proposed model is trained with the created dataset, and these videos consist of 95–110 Indian students as shown in Table 3. Four videos from each student with four different affect states are collected. Each affect video contains approximately 1 million frames.

- First video contains Alert, Excited, Elated, Happy facial expressions of students.

Table 4 Affect States and their corresponding definitions

Affect state	Affect state definition
Alert	Happy expression with raising hand
Exited	The expression shows to be ready to learn new things and gives a happy expression.
Elated	The students' expression after getting the answer.
Happy	Satisfaction with performance, feelings of pleasure about the material.
Content	The expression which shows the satisfied with what one is or has; not wanting more or anything else.
Neutral	Pay attention the lecture without any facial expressions.
Relaxed	The expression shows is less active and more calm and happy.
Calm	Be quiet and without worry pay attention in the classroom.
Fatigued	Confused expression from a long duration.
Bored	The expression which not taking interest in the classroom.
Depressed	From long duration not getting anything in the classroom.
Sad	Feelings of melancholy, beyond negative self-efficacy.
Upset	Because of not getting anything in the classroom getting angry.
Stressed	A state of mental or emotional strain.
Nervous	Slightly afraid in the classroom because of some cause.
Tense	Characterized by a strain upon the nerves

- Second video contains Content, Neutral, Relaxed, Calm facial expressions of students.
- Third video contains Fatigued, Bored, Depressed, Sad facial expression of students and
- Fourth video contains Upset, Stressed, Nervous, Tense facial expression of students.

These four videos are categorized into four different mood classes such as high positive affect, low positive affect, low negative affect, and high negative affect, respectively. Table 2 shows the mood classification based on emotions. At the time of data creation, the students are provided with a list of affect states (Alert, Excited, Elated, Happy, Content, Neutral,

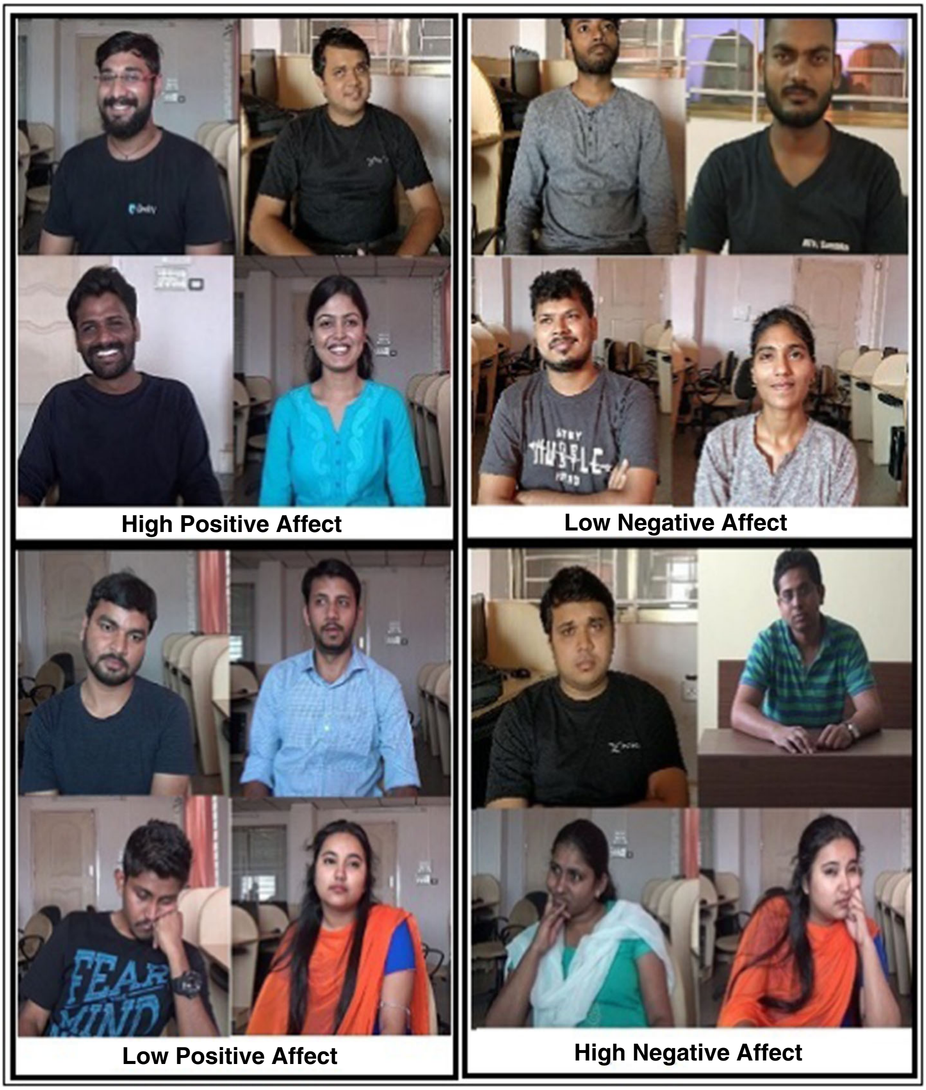

Fig. 8 Sample snapshot of database with students' face in different affect states

Relaxed, Calm, Fatigued, Boredom, Depressed, Sad, Upset, Stressed, Nervous and Tense) with definitions as shown in Table 4. The expressed affect states by the students are manually verified by a panel of three members consisting of two faculty members and an expert annotator along with the students' self-annotation.

Figure 8 shows a sample image snapshot of students during the creation of dataset with posed expressions.

5 Results and discussion

The proposed architecture is trained and tested for more than 100 graduate and undergraduate students of National Institute of Technology Karnataka Surathkal, Mangalore, India. One-hour classroom videos are divided into video segments of 5-6 minutes each. Then pre-processing is performed by involving the frame generation of each segment. Using the proposed MMFD technique, the faces of all the students present in every frame are detected, and the mood of each student in each frame using their facial expressions in the classroom is analyzed. After analyzing the students' mood, only some video segments are used as a training set, and the proposed model is trained using the deep learning technique. For the remaining video segments with the same students, the deep learning technique is used to analyze the students' affect states. The proposed M-Inception-V3 architecture is used to train the model.

5.1 Experimental setup:

Experimental setup includes a high definition rotating camera (16MP) in the classroom for data capturing. Tesla M40 GPU server with 48GB RAM and 300 GB scrap space are used for computations.

5.2 Student identification and mood analysis:

The proposed model is tested with state-of-the-art face detection and mood recognition techniques. The proposed Max-Margin Face Detection (MMFD) technique is used for face detection. This technique detects every possible pose of faces (i.e., tilted, occluded face) with better accuracy. Figure 4 shows the sample detected faces using MMFD technique. The proposed M-Inception-V3 model is used to recognize the students' mood.

Table 5 shows a comparison of the proposed face detection technique with other state-of-the-art techniques. In order to evaluate the robustness of the proposed architecture, a standard database which contains all the face image variants are required. Thus, two standard datasets, namely: YALE [25] and FDDB [16] are selected which contain different pose, different illuminations, rotation invariant, multi-view and other image variants such as tilted, occluded, etc. The most popular face detection techniques such as HAAR Cascade Classifiers, Viola Jones, and HOG methods are tested on these datasets. These techniques gave good accuracy for the frontal face but failed to perform better for tilted and occlude faces as shown in Table 5. Proposed MMFD gave an average accuracy of 99.56% on these datasets for all image variants of the face data but computation time is very high on the CPU when compared to other techniques.

Figure 9 showed the comparison of the proposed face detection technique with other state-of-the-art techniques which are tested on the FDDB and YALE datasets for accuracy

Table 5 Comparison of face detection techniques using FDDB and YALE datasets

Techniques	Dataset	Number of images	Number of face detected	Accuracy rate (%)
Haar Cascade Classifiers	YALE	166	153	92.16
	FDDB	3500	3315	94.71
Viola Jones	YALE	166	157	94.57
	FDDB	3500	3317	94.77
HOG Method	YALE	166	161	96.98
	FDDB	3500	3374	96.40
Proposed MMFD Method	YALE	166	165	99.39
	FDDB	3500	3491	99.74

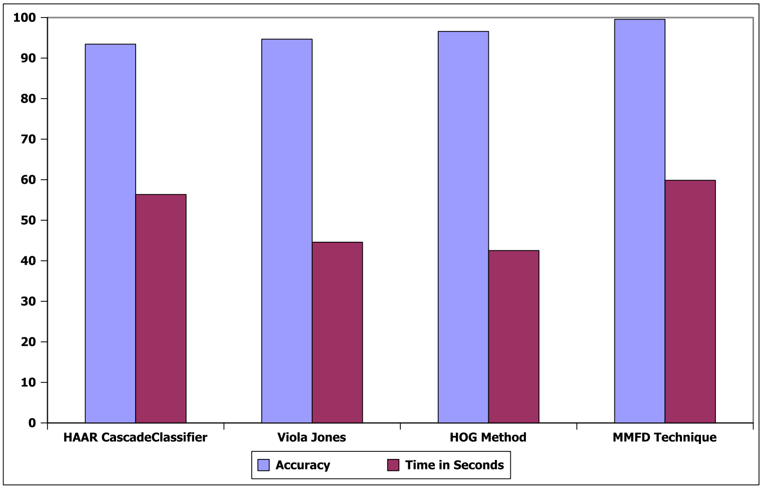

Fig. 9 Comparison with various Face Detection Techniques

Table 6 CPUs vs GPUs for face detection technique

For 100 frames	CPU	GPU
Time in Seconds	5.54	3.16

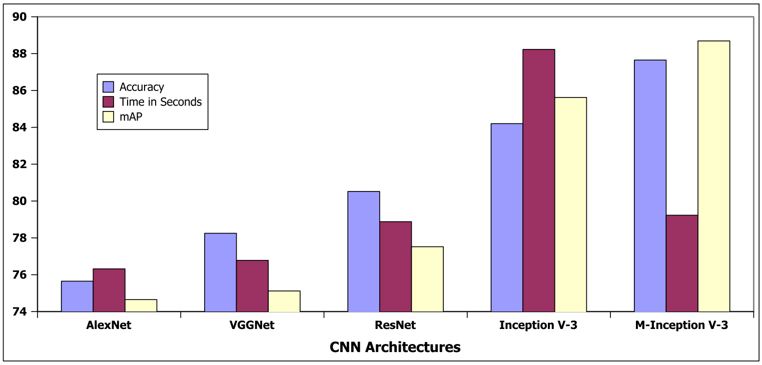

Fig. 10 Comparison with various Mood classification CNN Architectures

and time in seconds with HAAR Cascade Classifier, Viola Jones, HOG and proposed MMFD technique. The HAAR Cascade Classifier, Viola Jones, HOG methods are less accurate than the proposed method as it uses only the hand-crafted features for face detection whereas the proposed method is trained over a set of training samples. The proposed model takes more computation time to detect a single face and have to minimize the execution time we used Tesla M40 GPUs. The comparison of MMFD with CPU and GPU for a sample 100 frames is shown in Table 6. It is observed from Table 6 that the time taken for long classroom videos will have a significant reduction in their processing time using GPUs.

Figure 10 shows the comparison of the proposed model to affect state classification with other state-of-the-art architectures. The collected data is trained and tested with AlexNet, VGGNet, ResNet, Inception-V3, and M-Inception-V3. From Fig. 10, it is observed that some architectures gave better accuracy but took more time for computation and vice versa, whereas the proposed model gives better accuracy of 87.65% with minimal computation time. The confusion matrix for the four different affect states is shown in Table 7.

Figures 11 and 12 show the sample image frames taken during the posed expressions (the students are told a prior about the mood which they are supposed to express) of affect state recognition dataset creation. These images are considered only for the training of the dataset. The 10-fold cross validation is performed, and this validation includes student-independent (the students present in training are not present in the testing) 10-fold cross-validation [11].

Table 7 Confusion matrix for affect state classification

	High positive affect	Low positive affect	Low negative affect	High negative affect
High Positive Affect	0.845	0.131	0.017	0.007
Low Positive Affect	0.024	0.928	0.035	0.013
Low Negative Affect	0.017	0.074	0.895	0.014
High Negative Affect	0.014	0.038	0.11	0.838

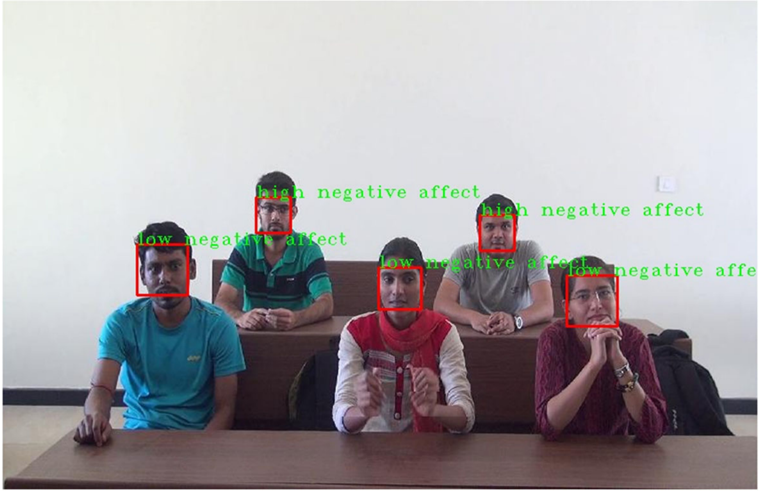

Fig. 11 Screenshot of affect state classification for the posed expression classroom data

There are not only captured images from rotation cameras which are placed at the proper height to collect the students' faces, but also the collected data from various possible angles so that the robustness of system can be improved as shown in Fig. 13. The sample image

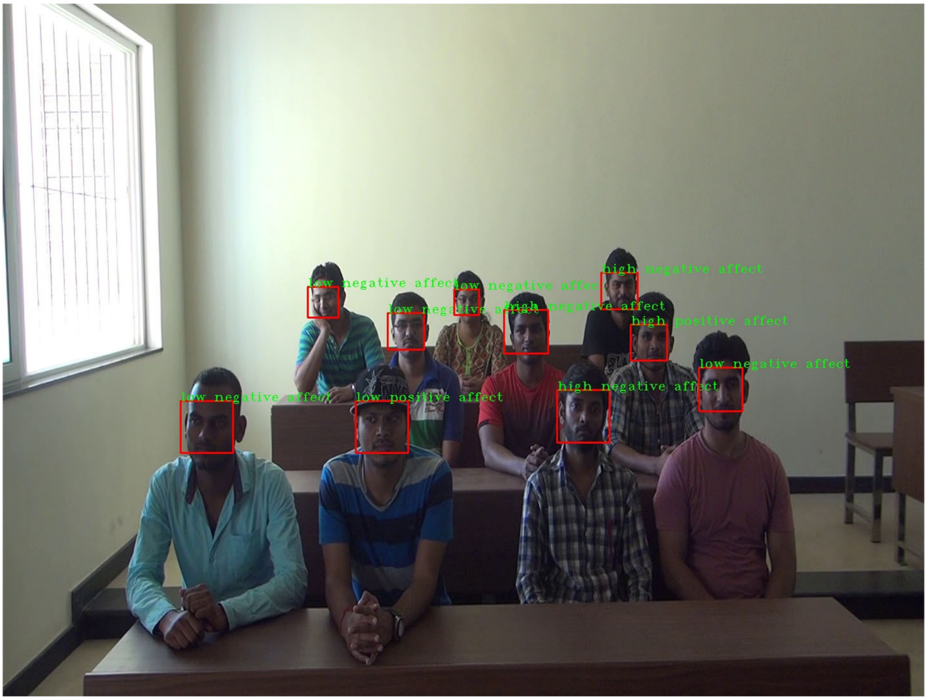

Fig. 12 Screenshot of affect state classification for the posed expression classroom data

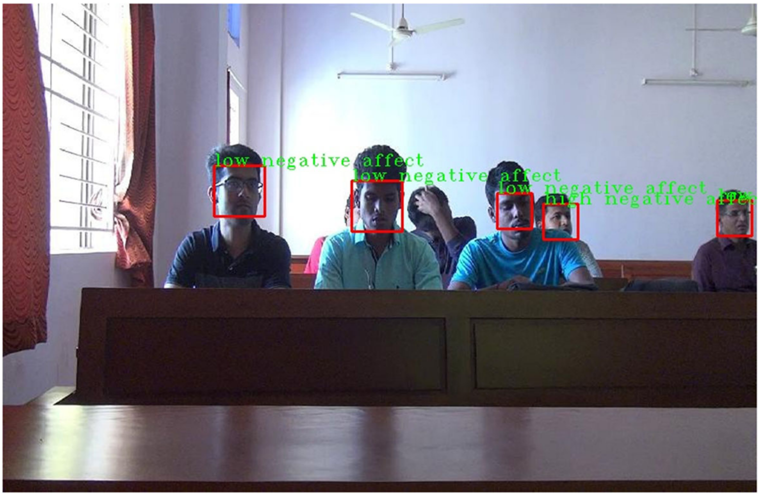

Fig. 13 Screenshot of affect state classification for the classroom for spontaneous or natural facial expressions

frame of Fig. 13 is taken when the lecture is in progress and hence the expressions are spontaneous. So this sample image frame shows the working of the proposed model for detected affect state of students in a classroom environment which includes the occluded faces, and the students are observed from a different angle when compared to the training dataset.

5.3 Further analysis

The standard analysis on student's affect states as mentioned in [11] is performed on the created dataset. The analyzed results are mentioned in the following subsections. Further, the impact of using the analyzed affective content as feedback to the faculty members are also discussed.

5.3.1 The frequency of affect states

Table 8 shows the analysis for a sample of 20 minutes of classroom video with 16 post-graduate students of the National Institute of Technology Karnataka Surathkal, Mangalore, India. The affect state frequency analysis is performed on these students. The predicted affect states are used for the statistical analysis using repeated measure ANOVA test and observed that there is a significant difference in the proportion of affect states experienced by the students $F(4, 73100) = 521.83$, $MSe = 0.011, n^2 = 0.411$. The Bonferroni posthoc test revealed the following pattern $((HighPositive = LowPositive) > HighNegative)$ with $(p < 0.05)$ and tried to isolate these affect states using base as neutral (which is present in low negative affect) using the chance $(Chance = (1 - M_{low\text{-}positive})/N_{affect} = (1 - .311)/4)$ and performed t-test analysis on the data and observed that there are only routine and sporadic affect states for the proposed four affect state classification of students. Similar results were observed when the same test was conducted for the entire dataset collected from students present in the classroom.

Table 8 Distribution of affect states

Affect states	Frequencies		Proportions		One-sample t-test		
	N	P	M	SD	t(15)	p	d
Routine							
High Positive	12	0.750	0.113	0.089	3.890	<0.010	0.330
Low Positive	14	0.875	0.102	0.106	3.110	<0.001	0.410
Sporadic							
High Negative	9	0.562	0.038	0.045	0.201	0.017	0.036
Low Nagative	16	1.000	0.439	0.459			

N = number of students that experienced the affect state at least once

P = proportion of students that experienced the affect state at least once

M = median and SD = standard deviation

5.3.2 Temporal dynamic affective states

Results are also analyzed for the persistence of the affect states. Persistence refers to a property in which it is observed that the affect state (S_t) at time t is also observed at time $t + 1$ (S_{t+1}). An affect state (S_{t+1}) can be considered to be persistent if it is experienced at one time interval increases the likelihood of experiencing the state at the subsequent time interval i.e. ($S_t \rightarrow S_{t+1}$). Similarly, an affect state is transient if it is experienced at one time interval decreases the likelihood that will be observed at $t + 1$. Finally, for a random affect state, if it is observed at time t and it is not related to the probability of its occurrence at $t + 1$.

The likelihood metric (1) is used in an attempt to characterize the affect states along this tripartite classification scheme. The metric quantifies the likelihood that the current affect state (S_t) influences the next affect state (X) after correcting for the base rate of X. According to this metric, if $L(S_t \rightarrow X) \approx 1$, then the state X reliably follows state (S_t) above and beyond the prior probability of state X. If $L(S_t \rightarrow X) \approx 0$, then X follows ($S_t$) at the chance level. Furthermore, if $L(S_t \rightarrow X) < 0$, then the likelihood of state X following state (S_t) is much lower than the base rate of X.

$$L(S_t \rightarrow X) = \frac{P(X|S_t) - P(X)}{1 - P(X)} \tag{1}$$

The main goal is to assess the likelihood that affect state (S_t) observed at time t is also observed at time $t + 1$ (S_{t+1}). This can be easily accomplished by modifying the metric such that the current state (S_t) and the next state (X) are the same (2).

$$L(S_t \rightarrow S_{t+1}) = \frac{P(S_{t+1}|S_t) - P(S_{t+1})}{1 - P(S_{t+1})} \tag{2}$$

In order to detect the significant affect state persistence, the likelihood of each state repeating itself to a hypothesized mean of 0 (normalized base rate) is compared using a one-sample t-test. The results of the tests are presented in Table 9 where it appears that the data supports a one-way classification scheme (persistent) instead of a three-way classification scheme, as there are no instances of random and ephemeral states.

It is observed from Table 9 that there are no random and ephemeral affect states in the proposed affect state classification. This infers that the four different affect states have a

Table 9 Persistence of affect states

Affective states	Descriptive measurement (Likelihood)		One-sample t-test			
	M	SD	t	df	p	d
Persistent						
High Positive Affect->High Positive Affect	0.133	0.236	3.330	12	0.009	0.460
Low Positive Affect->Low Positive Affect	0.890	0.222	3.960	16	0.003	0.580
Low Negative Affect->Low Negative Affect	0.630	0.131	1.890	08	0.063	0.310
High Negative Affect->High Negative Affect	0.102	0.188	2.420	12	0.034	0.380

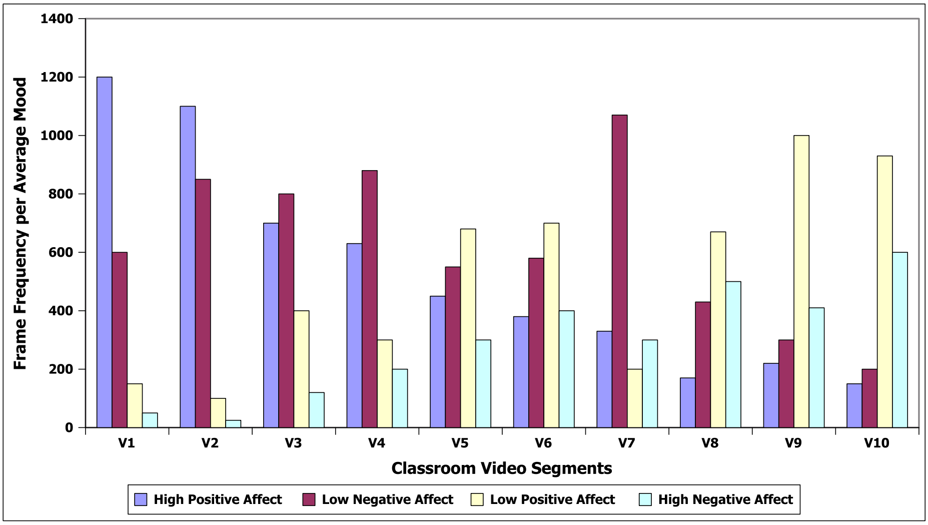

Fig. 14 Avg mood percentage of complete one hour classroom

significant impact on the students' moods and emotions. Its prediction is sufficient when compared to other learning-centered and basic emotions like eureka, fear, angry, etc.

Figure 14 contains a sample classroom video clip of 20 minutes where the duration of every segmented video is 2 minutes, 300 frames from each video segment is extracted at the rate of 5 frames/second. It is observed that the first segment video affect state has 2732 judgments, the next subsequent video affect states has judgments of 2880, 2882, 2901, 2800, 2880, 2830, 2820, 2810, 2753. The distribution of affect states for a particular student may be different but when the entire class is considered, there exist enough instances of affect states for possible likelihood in the temporal dynamics of affect states. It is also observed that similar results are obtained for the entire collected data.

Figure 15 shows the affect states of one sample classroom video segment. The segmented video length duration is 6 minutes, low positive affect state is observed in the first minute of the video, and low negative affect state is observed in the next minute of the video. There is a change from one state to another state with respect to the low negative state (neutral state). So the final average affect state of this segment is low positive affect. The proposed

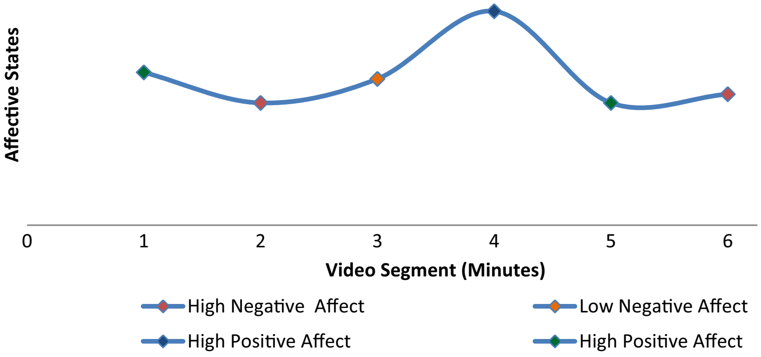

Fig. 15 Sample of One Segmented classroom video' Affective States

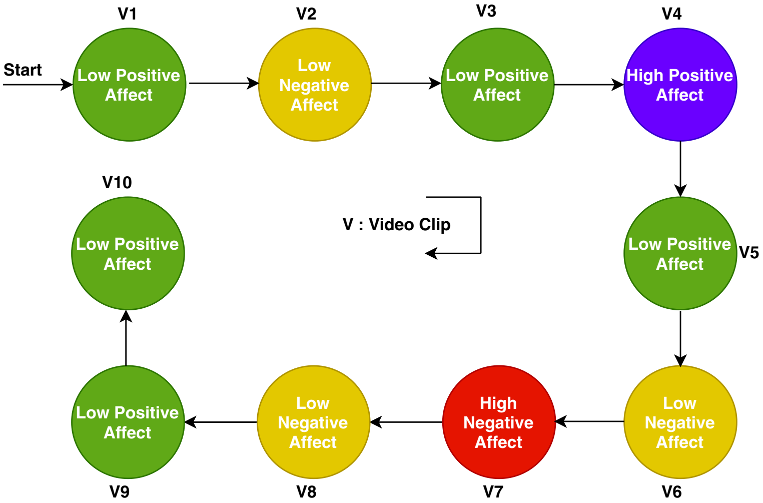

Fig. 16 Transition state of the one hour classroom video w.r.t. affect state

methodology not only provides the average engagement score of each video to the faculty member but also provides the graphical representation of the analyzed information as shown in Fig. 15. Thus the faculty member can visualize the video sections where most of the students are at high negative state and address the issue of negative affect state. Further, the faculty member is provided with details such as the transition of group or classroom affect states for the entire class using affect state transition diagram as shown in Fig. 16.

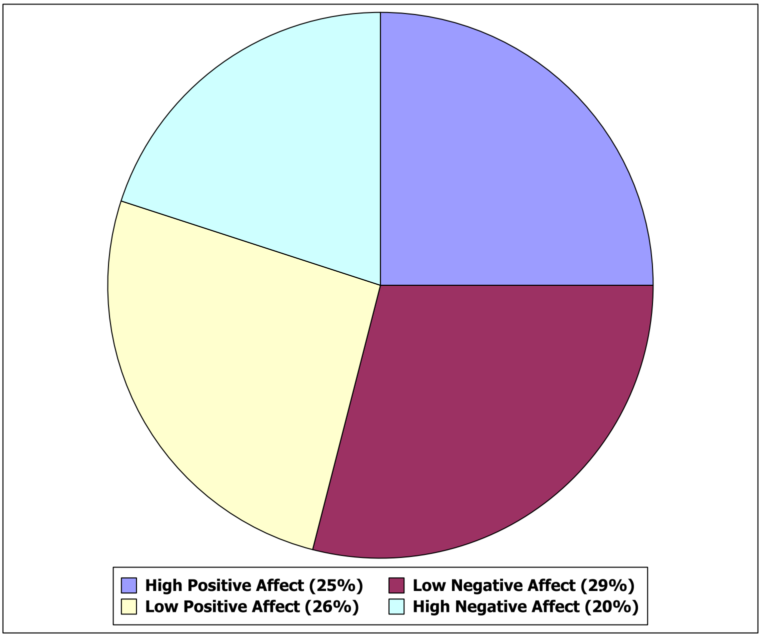

Fig. 17 Avg mood score of a sample of one hour classroom video

Table 10 Likert analysis for affect feedback to the faculty members

Likert analysis (0-5 scale where, 5 being the highest)

Q No.	Question	Likert score
1	Overall, the affective feedback system is useful and has a scope	4.1
2	Expert feedback is useful	3.9
3	Automated Affect feedback is useful	
	a) Engagement score	3.7
	b) Affect video content raw data	1.9
	c) Minute to minute transition diagram	3.4
	d) overall transition diagram	2.3
	e) Affect bar chart	3.1

After calculating the average mood for each segment of classroom video, affect states are combined for the entire one-hour class to obtain the overall results as shown in Fig. 17. Here, the number of instances for the observed affect states are presented in a bar chart with the data in the form of a percentage. The high positive score is 25%, the low positive score is 26%, the low negative score is 29%, and the high negative score is 20% of one-hour classroom video. With the help of these results, faculty members and experts can analyze the learning rate of the classroom and make the teaching process more effective. The expert also provides feedback to faculty members to improve the teaching process according to these results and transition states.

5.3.3 Feedback to the faculty members

If ’high positive affect’ is observed continuously for more than 70% duration of one video segment, then the following could be the reasons.

- Very good teaching style and strategy.
- Students’ interest is high.

If ’high negative affect’ is observed continuously for more than 70% duration of one video segment, then the following could be the causes for it.

- Very poor teaching style or strategy.
- Students’ interest is low.

By incorporating the suggestions given by the experts and thus improving the content/style/strategy one can create a positive classroom climate which improves the students’ engagement within the academic learning time.

The proposed affective content analysis as feedback to the faculty members is tested for 7 different courses, and the impact of the proposed system is analyzed using 5 Point Likert Scale analysis and the results are shown in Table 10. It is observed from Table 10 that the feedback mechanism is useful and promising, still a lot of improvement in the automatic feedback mechanism is required.

6 Conclusion

The students’ affective content based on their facial expressions in the classroom is analyzed using the proposed MMFD and M-Inception-V3. The corresponding students’ engagement score based on the different predicted students' affect states using the proposed method for the entire classroom is considered as one group. This analyzed classroom engagement group score is provided as feedback to both the faculty member (instructor) and the expert for further improvement of the teaching-learning process. The proposed affective content analysis using the classroom data performed better for both the face detection and the affect state classification with an accuracy of 99.56% and 87.65% respectively. Also, it is observed that the proposed affect state classification has a significant impact on the likelihood of affect state when compared to the basic and learning-centered emotions, since all the proposed affect states are only persistent in temporal dynamics of affect states.

In the future, the proposed methodology can be optimized to perform better with fewer GPU computations. The entire proposed method can be performed in real-time so that this can be used in any learning environment.

Compliance with Ethical Standards

Ethics statement The experimental procedure and the video content shown to the participants were approved by the Institutional Ethics Committee (IEC) of the National Institute of Technology Karnataka Surathkal, Mangalore, India. The participants were also informed that they had the right to quit the experiment at any time. The video recordings of the subjects were included in the database after their written consent for the use of videos for research purpose. Some subjects also agreed to use their face images in the research articles.

References

1. Ahlfeldt S, Mehta S, Sellnow T (2005) Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use. Higher Educ Res Dev 24(1): 5–20
2. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Trans Pattern Anal Mach Intell 28(12):2037–2041
3. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13(6):1450–1464
4. Bomia L, Beluzo L, Demeester D, Elander K, Johnson M, Sheldon B (1997) The Impact of Teaching Strategies on Intrinsic Motivation
5. Broeckelman-Post MA (2008) Faculty and student classroom influences on academic dishonesty. IEEE Trans Educ 51(2):206–211
6. Burnik U, Zaletelj J, Košir A (2017) Kinect based system for student engagement monitoring. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE, pp 1229–1232
7. Camelia F, Ferris TLJ (2016) Validation Studies of a questionnaire developed to measure students' engagement with systems thinking. IEEE Transactions on Systems, Man, and Cybernetics, Systems
8. Camelia F, Ferris TLJ (2017) Undergraduate students' engagement with systems thinking: results of a survey study. IEEE Trans Syst Man Cybern: Syst 47(12):3165–3176
9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005. CVPR 2005. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, vol 1, pp 886–893
10. Dasari B (2009) Hong Kong students' approaches to learning: cross-cultural comparisons. US-China Educ Rev 6(12):46–58
11. D'Mello SK, Lehman B, Person N (2010) Monitoring affect states during effortful problem solving activities. Int J Artif Intell Educ 20(4):361–389
12. Farhan M, Aslam M, Jabbar S, Khalid S (2018) Multimedia based qualitative assessment methodology in eLearning: student teacher engagement analysis. Multimed Tools Appl 77(4):4909–4923
13. Fredricks JA, Blumenfeld PC, Paris AH (2004) School engagement: Potential of the concept, state of the evidence. Rev Educ Res 74(1):59–109

14. Guo PJ, Kim J, Rubin R (2014) How video production affects student engagement: An empirical study of mooc videos. In: Proceedings of the first ACM conference on Learning@ scale conference. ACM, pp 41–50
15. Izenman AJ (2013) Linear discriminant analysis. In: Modern multivariate statistical techniques. Springer, New York, pp 237–280
16. Jain V, Learned-Miller E (2010) Fddb: A benchmark for face de-tection in unconstrained settings. University of Massachusetts, Amherst, Technical Report UM-CS-2010-009 2(7):8
17. Kazmi A (2010) Sleepwalking through undergrad: Using student engagement as an institutional alarm clock. Coll Q 13(1):n1
18. King DE (2015) Max-margin object detection. arXiv:1502.00046
19. Langton N, Robbins SP, Judge TA (2013) Fundamentals of organizational behaviour. Pearson Education, Canada
20. Lehman BA, Zapata-Rivera D (2018) Student emotions in Conversation-Based assessments. IEEE Trans Learn Technol 11(1):41–53
21. Lowe DG (1999) Object recognition from local scale-invariant features. In: 1999. The proceedings of the seventh IEEE international conference on Computer vision. IEEE, vol 2, pp 1150–1157
22. Luo J, Ma Y, Takikawa E, Lao S, Kawade M, Lu B-L (2007) Person-specific SIFT features for face recognition. In: 2007. ICASSP 2007. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, vol 2, pp II–593
23. Macal CM, North MJ (2005) Tutorial on agent-based modeling and simulation. In: 2005 proceedings of the winter Simulation conference. IEEE, pp 14–pp
24. Mann S, Robinson A (2009) Boredom in the lecture theatre: an investigation into the contributors, moderators and outcomes of boredom amongst university students. Br Educ Res J 35(2):243–258
25. Minear M, Park DC (2004) A lifespan database of adult facial stimuli. Behav Res Methods, Instrum Comput 36(4):630–633
26. Monkaresi H, Bosch N, Calvo RA, D'Mello SK (2017) Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans Affect Comput 8(1):15–28
27. Qin J, Zhou Y, Lu H, Ya H (2015) Teaching Video Analytics Based on Student Spatial and Temporal Behavior Mining. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. ACM, pp 635–642
28. Sagayadevan V, Jeyaraj S (2012) The role of emotional engagement in lecturer-student interaction and the impact on academic outcomes of student achievement and learning. J Scholarsh Teach Learn 12(3):1–30
29. Sayadi ZA (2007) An investigation into first year engineering students' oral classroom participation: a case study. PhD diss., Universiti Teknologi Malaysia
30. Schmidt A, Kasiński A (2007) The performance of the haar cascade classifiers applied to the face and eyes detection. In: Computer Recognition Systems 2. Springer, Berlin, pp 816–823
31. Skinner EA, Belmont MJ (1993) Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. J Educ Psychol 85(4):571
32. Subramainan L, Mahmoud MA, Ahmad MS, Yusoff MZM (2016) Evaluating students engagement in classrooms using agent-based social simulation. In: 2016 2nd International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR). IEEE, pp 34–39
33. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2818–2826
34. Thomas C, Jayagopi DB (2017) Predicting student engagement in classrooms using facial behavioral cues. In: Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education. ACM, pp 33–40
35. Ventura J, Cruz S, Boult TE (2014) Improving Teaching and Learning through Video Summaries of Student Engagement
36. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154
37. Walberg HJ, Anderson GJ (1968) Classroom climate and individual learning. J Educ Psychol 59(6p1):414
38. Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Personal Soc Psychol 54(6):1063
39. Whitehill J, Serpell Z, Lin Y-C, Foster A, Movellan JR (2014) The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Trans Affect Comput 5(1):86–98

40. Wong A (2016) Student perception on a student response system formed by combining mobile phone and a polling website. Int J Educ Dev Inf Commun Technol 12(1):144
41. Yang M-H, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58
42. Yazzie-Mintz E (2007) Voices Of students on engagement: a report on the 2006 high school survey of student engagement. Center for Evaluation and Education Policy, Indiana University
43. Young MS, Robinson S, Alberts P (2009) Students pay attention! Combating the vigilance decrement to improve learning during lectures. Act Learn High Educ 10(1):41–55

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Sujit Kumar Gupta received his B.Tech degree in 2015 from Feroze Gandhi Institute of Engineering and Technology, Rae Bareli (U.P.), India. He completed his M.Tech degree from National Institute of Technology Karnataka Surathkal, Mangalore, India. His research interests include Affective Computing, Internet of Things, Deep Learning and Cloud Computing. He is the student member of IEEE.

T. S. Ashwin received his BE degree in 2011 from Visveswaraya Technological University, Belgaum, Karnataka, India and M.Tech degree in 2013 from the Manipal University, Manipal, Karnataka, India. Currently, he is pursuing his full-time Ph.D. at National Institute of Technology Karnataka Surathkal, Mangalore, India. His research interests include Affective Computing, E-Learning, Deep Learning and IoT. He is the student member of IEEE and has more than 23 International Conference publications so far which includes an IEEE Region 10 conference's best paper award.

Ram Mohana Reddy Guddeti received his B.Tech from S.V. University, Tirupati, Andhra Pradesh, India in 1987; M.Tech from Indian Institute of Technology, Khargpur, India in 1993 and Ph.D. from The University of Edinburgh, U.K in 2005. Currently, he is the Professor and Head, Department of Information Technology, National Institute of Technology Karnataka Surathkal, Mangalore, India. His research interests include Affective Computing, Big Data and Cognitive Analytics, Bio-Inspired Cloud and Green Computing, Internet of Things and Smart Sensor Networks, Social Multimedia and Social Network Analysis. He is a Senior Member of both IEEE and ACM; Life Fellow of IETE (India); Life Member of ISTE (India) and Life Member of Computer Society of India. He has more than 200 research publications in reputed and peer reviewed International Journals, Conference Proceedings and Book Chapters.