

Extension of MDNFM For Smart Classroom Activity Monitoring

Rahul Kumar Pandey

Research Scholar, MSIT
MATS University
Raipur Chhattisgarh India
e-mail:Ravi.care18@gmail.com

Dr. Gyanesh Shrivastava

Research Supervisor, MSIT
MATS University
Raipur Chhattisgarh India
gyaneshnkshrivastava@gmail.com

Dr. Umesh Kumar Pandey✉

Co-Research Supervisor,
United Institute of Technology
Prayagraj Uttar Pradesh India
umesh6326@gmail.com

ABSTRACT

One application of the automated smart learning process is monitoring the student's activities in the classroom. However, when indulged in taking lectures, the instructors or the faculty cannot monitor the activities of the students appropriately. Therefore, traditional methods employ several face detection algorithms to monitor the activities of the students.

However, the results obtained through these methods are inaccurate, and hence an efficient algorithm is required to predict the active state of the student in the classroom. Hence, we use Multi-tasking Deep Neuro-Fuzzy Model (MDNFM) for the activity monitoring (AM) of the students in the classroom. Initially, the images of the students in the classroom are captured through web cameras and other accessories placed in the smart class. Then, the acquired image is passed to the Capture, Transform, and Flow (CTF) tool for storing and transferring for further processing of images.

Keywords: *activity prediction, multi-face detection, feature extraction, smart classroom.*

I Introduction

The traditional lecture and note-taking approach have lost its effectiveness with the modern-day growth of education. Moreover, since every student is interested in different subjects, different modes of teaching and learning techniques are needed to enrich conceptual growth and development. Therefore, the educational institutes should be responsible for providing opportunities or openings to the students to gain academic growth and interest during their childhood [1].

A modern classroom management system is a system that helps teachers to create a more appropriate classroom environment for students. For example, this software can be employed to restrict students' access to distracting websites. It also motivates and encourages more students by offering and involving several activities. Moreover, classroom management tools assist teachers in improving student behavior by giving feedback. This helps teachers give parents an impression of how their child is doing in the classroom.

In addition, teachers can monitor the student's progress, make sure they have achieved the goal, and tell their parents about the student's progress. Thus the classroom management software can help with the learning and the student management system. The parents and teachers get a broader perspective on student development and behavior [2]. Fig.1 gives the classification of the tracking management system.

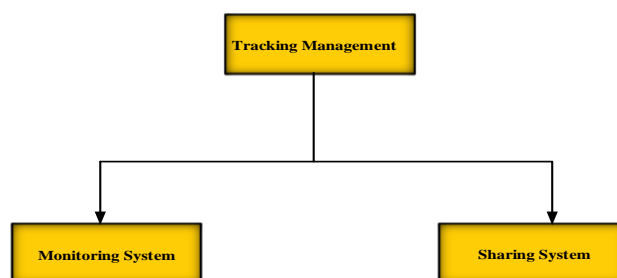


Figure 1: Classification of Tracking Management

Instead of teachers observing the classroom hours, this software analyzes the video based on the frame, and the behavior examples of expert identification match the arrangement of these measurements. Through this, teachers can analyze the students individually, and the teachers can recognize that the students learn better in the morning than in the afternoon. The automatic software can give a student behavioral feedback over weeks or a month. The use of this method can evaluate the curriculum activities in hundreds of classrooms without any researchers for observations in the classroom. Engagement in measuring students in a classroom is a difficult problem [3].

Achieving a high strength in students' engagement is essential for academic performance and teaching. Ultimately, students are keen to retain valuable information and develop a meaningful and comprehensive knowledge of a subject or topic if they find the lessons interesting, engaging, rewarding, meaningful, and important [4]. But, of course, producing this kind of student engagement in the modern classroom needs both the right student engagement plans and the right student engagement activities.

When combined with an experienced teacher, these elements can help students maintain concentration and invest expressively in their studies, eventually successful learning outcomes and making lessons much more valuable.

II Related Work

Chae YN et al.(2016). Presented color filtering-based face detection method scheme to identify the face quickly. This design uses a color filtering which passes through the areas that does not have faces utilizing a face color membership function.

Li et al. (2018) recommend a solution to the problem of light variation by combining histogram equalization (HE) with a Gaussian low-pass filter (GLPF).

Khan et al. (2018) planned a method based on the Swarm method (S.M.). The Discrete Wavelet Transform offers efficient extraction of appropriate features, reduces set-up time, and increases the recognition accuracy.

Khan S et al. (2020) presented the YOLO Algorithm for face detection. The overall picture of the development of this scheme has two stages. The first stage involves the face detection step, and the second step is training the system based on a user's image input.

Sun et al. (2019) designed an Analysis of covariance (ANCOVA) program to study the impact of instructional support on students' self-efficacy in computer-assisted learning.

Zhang et al. (2018) introduced an eye tracking program called VLEYE. It collects important areas of interest (AOI) and associates them with the records of eye activity.

Abbasi and Khosravi (2020) recommend using genetic algorithms in the particle filter method. Initially, the frame having eye images is obtained and is preprocessed.

Farhan et al. (2018) introduced video exploration software tools based on multimedia to determine students' alertness.

Wu Y *et al.* (2017) proposed a convolution neural network for the regression of facial landmarks. This method doesn't need auxiliary labels beyond the landmarks. Here they used vanilla CNN and extracted the representations from the input of every layer.

Kim H *et al.* (2018) proposed faster RCNN facial landmark extraction (FLE). Here they utilized separation and FLE.

Feng ZH *et al.* (2019) proposed regression-based CNN architecture. In this architecture, they proposed an online HAEM approach. For experimentation, they use two datasets they are AFLW and COFW.

Drouard V *et al.* (2017) proposed using a linear regression method. Here it learns with both head-present boundaries and bouncing box-to-confront arrangements, with the end goal that, at runtime, both the head-present points and bounding box shifts are anticipated.

Wang Y *et al.* (2019) offered a deep neural network with two branches of coarse arrangement and fine reversion level. The assessment with a course Coarseto- Fine measure, where the precise present boundary assessment is trailed by coarse classification. The two errands are accomplished through profound learning models, which share a similar full-picture convolutional highlight map.

Hsu HW *et al.* (2018) proposed head pose estimation based on quaternions with multiple regression loss. The organization is trained with quaternion portrayal, which maintains a strategic distance from the gimbal lock issue and prompts robust execution during testing.

Zhang et al. (2017) proposed GazeNet, an initial appearance on account of the estimation process of gaze. In light of a 16-layer VGG profound convolutional neural organization.

Bah SM and Ming F (2020) proposed a face recognition system based on a Local binary pattern. However, the capacity of the LBP face recognition calculation depends on the exactness of execution of the feature extraction and assessment stage, which also significantly relies on the idea of both the data face pictures and the arrangement/reference pictures looking into the face connection measure.

Raychaudhuri *et al.* (2019) planned a new technique to identify the moving objects from the surveillance video. The first step in their method is background modeling.

Cao *et al.* (2018) used a network to detect the Region of Interest (ROI). It consists of two stages: initialization, training, and detection. In the initialization stage, the system is trained to obtain immediate output.

Matava C et al. proposed a faster R-CNN scheme. The dataset was separated into two classes for use in preparing and testing. They utilized three convolutional neural organizations (CNNs): ResNet, Inception, and MobileNet.

Bhavana D et al. (2020) proposed a Binary algorithm in an automated and smart attendance system for speech output Computer vision. Facial recognition is composed via the automatic presence system in a classroom atmosphere.

Mahmood S et al. (2019) proposed a framework based on teacher-student interaction (TSI) to improve the delivery and efficiency of regular exercises with the social and ecological associations. Here Raspberry Pi is utilized inside the study hall setting, mounted with a camera to catch the understudy's outward Appearances.

Ashraf R et al. (2018) introduced an Image Retrieval (I.R.) approach for searching low-level visual image features from the database to retrieve. The characteristics of the low-level image are the texture, shape, and color acquired utilizing CBIR. Initially, the color image is changed into YCbCr (luma blue-difference and red-difference Chroma component) by the canny detector, and then the feature edge is taken from the Y-luminance matrix.

Pandey R.K et al. (2021) introduce Student Monitoring Model using Face Detection & Activity Classification for Smart Classroom

Zhu X et al. (2019) proposed a Distance learning strategy for individual reidentification. In this, proposed a Projection Hard and Easy Negative examples mining approach based Distance learning (PHEND) learns a projection network and a separation metric.

III MATERIALS AND METHODOLOGY

- **Fuzzy Membership Functions**

Generally, the classification method is hardly used to identify the optimal threshold value for the face activity decision-making process. This section investigates the attention assessment using the fuzzy-based face detection method. Based on its assumptions, the face regions of the eye and lips have darker areas than the skin. Therefore, this method uses skin color components as fuzzy input in the threshold range level of [0 and 1] same as the fuzzy output system.

To obtain the features, the fuzzy membership function is designed with the considered features of eyes, lips, edge distance, and feature movement between the facial features [5], [6]. The membership function obtains the possible number of detection results. This can be split into four types: [0, 1, 2, and 3]. Where zero represents that the located eyes cannot be positioned in the skin region, one and up to 3 define that the eye is situated with the wrong decision. As shown in Fig 2 (a) and (b), the low (L) and high (H) region is obtained using the membership function.

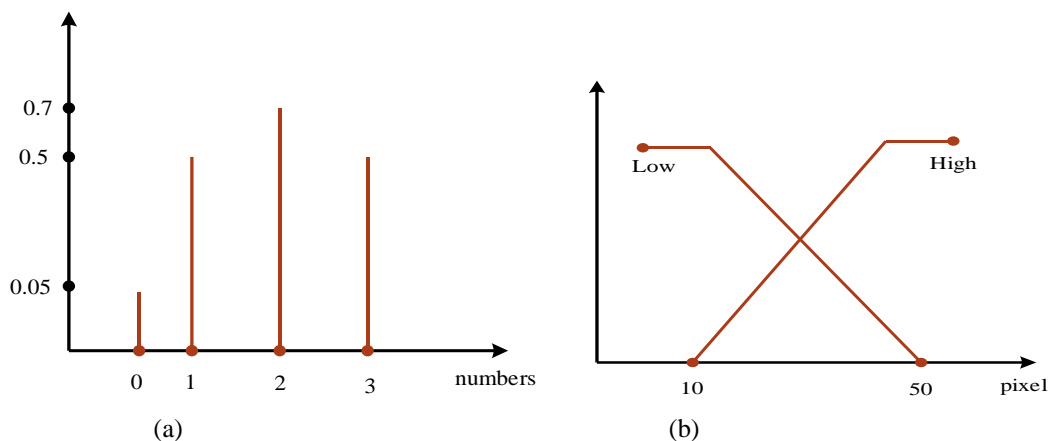


Figure 2: (a) and (b) Detecting eye and lip region degree of membership function

The student's facial features are advanced to determine the attention states at different decision levels. This method examines motion detection under the degree of attention membership function as Low (L), Middle (M), and high (H) levels. The detection of edge and central distance membership function and the attending degree of membership function output is shown in Fig 2.1 (a) and (b), and 2.2

- Leaving and Drowsiness Fuzzy Rules

The facial behavior features of the eye and lips are judged depending on the obtained detection results [7]. For this, if the facial features of eyes and lips don't have the detect position, the student cannot be in the classroom. On the other hand, if the face detection becomes normal and the eyes and lip features cannot be detected, the student indicates drowsy. According to the elements, the size and position area are applied with the fuzzy rules. The fuzzy rule table is shown in Table 1.

The combination of drowsiness and fuzzy rules is varied in the facial feature indicators of eye closing duration [8]. This indicator can overcome the limitations of various measurements. It considers three drowsy states, low state, medium state, and high state. These defined rules choose the indicators that minimize the wrong detection result. The medium state represents the two eye levels to the medium drowsy state. If the activity level is high, the drowsiness output level increases due to the undesired impacts of conversation or turning. The function output is selected based on the minimum and maximum values for classification. The eye closing system for drowsiness detection is shown in Figure 2.3

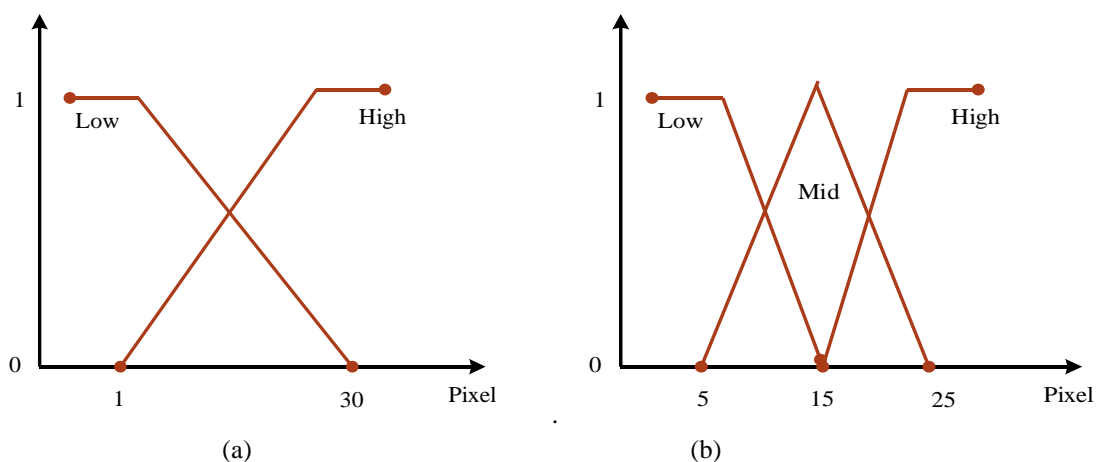


Figure 2.1: (a) and (b) Detecting central and edge distance degree of membership functions

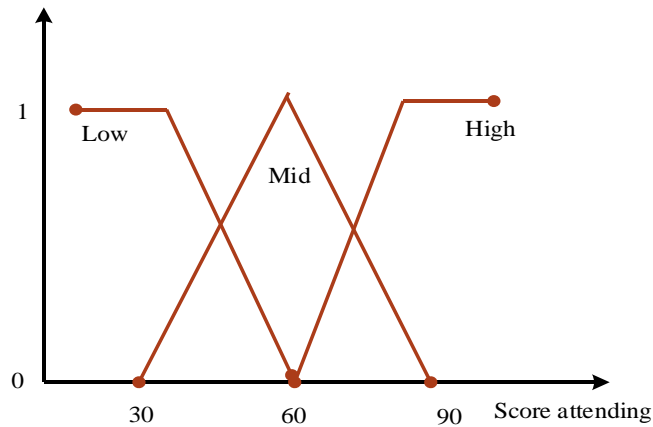


Figure 2.2: Output fuzzy membership function

Table 1: Fuzzy rule table

Input (Eye value)	Input (Lip value)	Output (Attending)
0	Low	Bad
0	High	Bad
1	Low	Bad
1	High	Mid
2	Low	Good
2	High	Good
3	Low	Good
3	High	Good

• **Head Turning and No Fuzzy Motion Rules**

In a fuzzy inference system, the head turning detection is determined based on facial edge and feature distance, and the central point of motion features. If the features are relatively small and the central point of motion feature becomes high or mid, the position of the head is turned. No moving feature is detected in each frame when the central point of motion feature becomes normal, and the facial edge distance becomes low from the detected result. Thus, fuzzy rules determine the facial edge and feature distance, and the central point of motion features [9]. The center and edge distance fuzzy rule is shown in Table 2

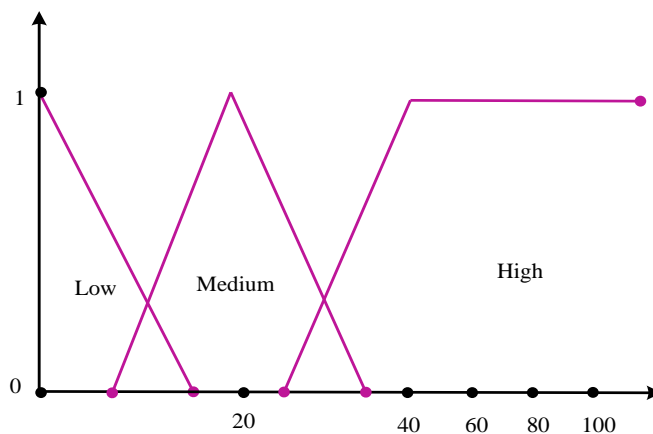


Figure 2.3: Detecting eye closing duration of fuzzy membership function

- **Data Annotation**

Data annotation is the critical factor in identification tasks where the quality of annotations leads to the success of the machine learning applications. For the emotion recognition, the pose pattern and the behavioral features have to be annotated from the collected data of the students to determine their estimated affective state from them. The video clips are mostly collected from unconstrained environments like laboratories, hostel rooms, open ground, crowded places, school campuses, and classrooms with different illumination variations [10]. The data collected are in video clips with 24fps where the blurred and the repeated sections in the frames are preprocessed. The repeated frames define the exact behavior of the student in the single image frames or the two frames that depicts the same emotion. After processing the image frames, the video clips are then manually annotated by the experts that exactly define the actual emotion of the students. The annotation of emotions is mostly classified into confusion, anger, fear, sad, happy, frustrated, boredom, neutral, and engaged. The rating score is provided for these emotions from 0 to 3 according to the level of intensity

Table 2: Fuzzy rule table.

Input (Edge value)	Input (central point value)	Output (Attending value)
Low	High	Bad
Low	Mid	Bad
Low	Low	Mid
High	High	Good
High	Mid	Good
High	Low	Bad

IV Multi-tasking Deep Neuro-Fuzzy Model (MDNFM)

In our work, automatic classification of students' attention in classrooms is performed under two categories like active and inactive clusters. The clustering is performed based on the Multi-tasking Deep Neuro-Fuzzy Model (MDNFM), which estimates the students' attentiveness based on the features learned and trained by the proposed approach.[11] Here, the classroom activities are collected by the web camera fixed inside the classroom that captures the images of the scenes are forwarded to the Capture, Transform and Flow (CTF) tool. The collected images from the classrooms are then stored in a data warehouse and analyzed further. The proposed MDNFM model is classified into three stages.

1. Pre Processing
2. Multi face detection
3. Activity classification

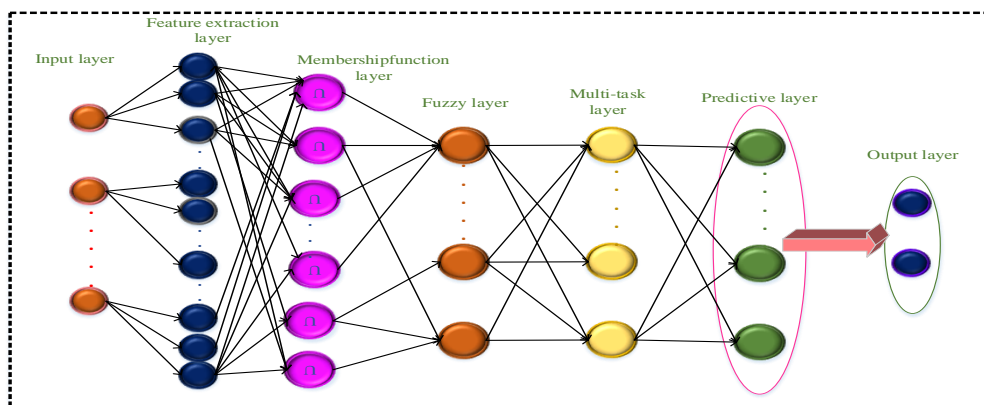


Figure 3: Illustration of various layers in MDNFM

- **Input layer:**

The image obtained after preprocessing stage reaches the input stage that consists of different color models like N-RGB (Normalized- Red, Green, Blue components), N-HSV (Normalized-Hue, Saturation, Value), N-YCbCr (Normalized-Chroma components of green, blue, red). The proposed approach is trained based on the various skin colors in different color spaces. The variation of image size across the frames is normalized and thus makes the system not sensitive to image size.

- **Feature extraction layer:**

In this layer, multiple features are extracted simultaneously to estimate multiple faces at the same instant time. The following are some of the features learned for the multiple face detection from the images;

- ✓ Nose width,
- ✓ Width of the mouth,
- ✓ Distance between the eyes,
- ✓ The horizontal and vertical distance between the mouth and nose,
- ✓ Distance between the left eye and the left side of the nose,
- ✓ Distance between the right eye and the right side of the nose,
- ✓ The vertical and horizontal distance between eye and nose,
- ✓ The horizontal and vertical distance between the left and right edge of the face to the left and right side of the nose

Variant features of the face are indicated with specific pixels placed at the lower and upper portions of the rectangular region. The pixel with the face region is denoted as P_f in which the color model ranges from 0 to 255. Two different pixel measures are considered to face and background for face detection. The output image is categorized with two other conditions;

$$R(P_f) = \begin{cases} 1 & \text{if } P_f \text{ is a face pixel} \\ 0 & \text{if } P_f \text{ is a non face pixel} \end{cases} \quad (1)$$

- **Multi-task layer**

In the final layer, the differentiation between the skin and non-skin pixels and the transition of these regions is determined. The fuzzy probabilities in terms of skin and non-skin pixel differentiation are estimated as follows;



(a)



(b)

Figure 4 : (a-b) Input image



(a)



(b)

Figure 4.1: (a-b) skin Region Extraction

$$P_s = \sum_{i=0}^{255} \Phi(x_i) \mu_s(x_i) \quad (2)$$

$$P_{NS} = \sum_{i=0}^{255} \Phi(x_i) \mu_{NS}(x_i) \quad (3)$$

A similar procedure was followed for other color models like N-HSV and N-YCbCr. Two categories define the binary image obtained as pixel and non-skin pixel region. Here four connectivity-based blob algorithm is used to estimate the skin region localization. Thus the face region is recognized based on the color of the skin. From the binary image I, the segmentation is performed from the face image with pixel size $H = m \times n$, and the extraction of the region R_h done separately.

The extracted region is estimated as the face if it satisfies the following condition as follows,

$$\frac{N_c}{H} \geq l \quad (4)$$

Where N_c is the number of kin pixels present in the region R_h , and l defines the threshold measure that possesses the face in the image.

V Experimental Results and Analysis

This section discusses the metrics, dataset details, results, and comparative analysis between the existing and the proposed approach in detail. For the performance evaluation of the proposed MDNFM model, we use the self-collected dataset consisting of the emotions and activities of the 55 students, respectively. The simulation was carried out under the MATLAB 2019a platform. The input image from the self-collected dataset is fed into the MDNFM model for the performance estimation. Figure 3 shows the input image acquired during the lecture in the classroom.

After the image acquisition process, preprocessing is performed, and then extraction of multi-face based on the skin color feature is performed. The image obtained after multi-face detection is shown in figure 4.1.

Based on the features obtained from the skin region pixel, the students' faces in the frame are estimated from the multiple features learned from the face of the students. The multiple faces detected from the input image shown in the figure. 4.2.

From the detected faces, the student activity is predicted based on the three features, including lip movement, eye gaze, and head position. Then, from the fuzzy rule concept, the attentive factor of the student is measured, and group the captured faces into the active and inactive cluster sets. The students are classified into active and inactive groups shown in figure 4.3(a-d).

Inferred from the result that the values of the highly attentive and neutral emotions achieve a higher accuracy rate while the sleepy emotion attains a lower rate. This is due to the faster recognition of highly attentive emotion than sleepy emotion. The existing models like LDA, Deep CNN, SVM, and KNN. [12] are utilized to compare the proposed MDNFM model, and the result shows a higher accuracy rate and lower error for the proposed model than the earlier techniques.

From table 4, we perform experiments in terms of different approaches in the real-time learning smart classroom environment. Here, the baseline models achieve the minimum precision rate. At the same time, the deconvolution and the bilinear interpolation show a better outcome. Compared to earlier models, the proposed method achieves an improved performance with a performance prediction of accurate bounding boxes than the other approaches.

Table 3: Comparison of overall and recognition accuracy of each emotion for different methods

Techniques	Highly attentive (%)	Sleepy (%)	Distracted (%)	Bored (%)	Neutral (%)	Non-attentive (%)	Overall Accuracy (%)
LDA	88.9	25	55.6	62.5	55.6	37.5	56.5
SVM	88.9	25	66.7	63.5	88.2	62.5	71.7
KNN	77.8	37.5	33.3	63.5	86.7	75	68.9
Deep CNN (Alexnet+FC6+LDA)	92.6	47.2	66.7	73.6	90.4	77.8	77.2
MDNFM	93.5	82.6	77.8	87.6	92.5	84.7	91.5



(a)



(b)

Figure 4.2: (a-b) Multiple face detection from the Input image

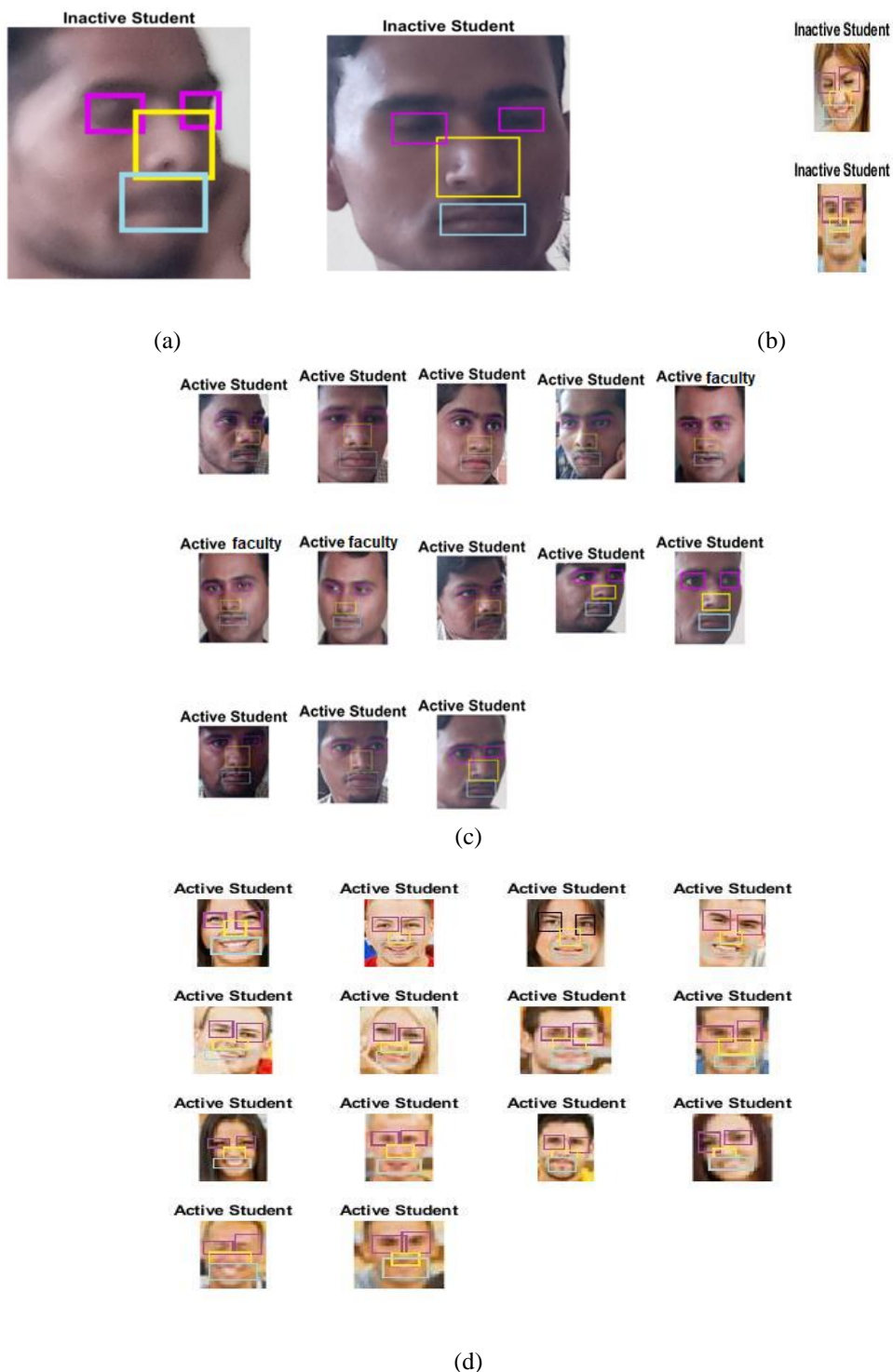


Figure 4.3 (a-d): Classification of Active and inactive cluster set

Table 4 Comparison result of Average precision (A.P.)

	AP@0.5	AP@0.7
R-FCN	0.6080	0.3904
Faster-RCNN+RPN	0.6309	0.4705
RFCN+DCN	0.6713	0.4822

RFCN+FP (DECONV)	0.7245	0.4636
RFCN+FP (BILINEAR)	0.6709	0.4432
RFCN+FP+DCN	0.7477	0.5506
MDNFM	0.7845	0.6256

VII Conclusions and Future Scope

The presented MDNFM approach is designed to predict students' attention accurately in the smart classroom environment. This approach monitors the student's activity in real-time and predicts the attentive state of the student.

Here the student's actions like head pose, eye movement, lip position, and motion features are estimated from the input facial features that classify the result into active and inactive cluster set detection process. After the image acquisition, three main processes are performed: preprocessing, multiple face detection, and activity classification.

The MDNFM model performs skin color extraction to detect the multiple faces from the group of students in the image. Then, the fuzzy rules were formulated to differentiate the skin and non-skin pixels based on the estimated facial features. In the final stage, the predictive layer learns the features from the input images and results in the emotions into inactive and active classes. Experimental results evaluated in terms of accuracy and cross-validation measure show that the proposed model achieves greater improvement than the other state-of-the-art approaches.

In the future, the proposed model may be subjected to modifications considering the following aspects as listed below.

- This model can incorporate additional security aspects using biometrics in attendance marking and performance estimation.
- More facial features were considered for detecting the activities performed by the students.
- A daily report of the deeds of the students is sent to the parent and guardian of the student. Also, periodically the assessment can be made using the technological tools available in smart learning.
- Development of more accurate models required less complexity and efficient training of the model for acquiring better scope in the prediction process

Funding: Authors did not receive any funding for this research work.

Conflict of Interest: All Authors declare that none of the authors has any conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by the authors.

References

- [1] Khan, S. A., Ishtiaq, M., & Shaheen, MM. (2018). Face recognition under varying expressions and illumination using particle swarm optimization. *Journal of computational science*, 28, 94-100.
- [2] Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., & Jeon, G. (2018). Content-based image retrieval by using color descriptor and discrete wavelet transform. *Journal of medical systems*, 42 (3), 44
- [3] Cebrián, G., Palau, R. & Mogas, J. (2020). The Smart Classroom as a means to the development of ESD methodologies. *Sustainability*, 12(7), 3010-2020
- [4] Bdiwi, R., Runz, C. de., Faiz, S., & Cherif, A. A. (2019) Smart Learning Environment: Teacher's Role in Assessing Classroom Attention. *Research in Learning Technology*, 27.
- [5] C. H. Yang. "Fuzzy fusion for attending and responding assessment system of affective teaching goals in distance learning." *Expert Systems with Applications*. Vol.39, no. 3, pp. 2501-8, 2012.
- [6] M. Ilbeygi, H. Shah-Hosseini. "A novel fuzzy facial expression recognition system based on facial feature extraction from color face images." *Engineering Applications of Artificial Intelligence*. Vol.25, no.1, pp. 130-46, 2012.
- [7] R. V. Priya. "Emotion recognition from geometric fuzzy membership functions." *Multimedia Tools and Applications*. Vol. 78, no. 13, pp.17847-78, 2019.

- [8] M. Sathik, S. G. Jonathan. "Effect of facial expressions on student's comprehension recognition in virtual educational environments." *Springer Plus*. Vol.2, no. 1, P. 455.
- [9] J. Jo, S. J. Lee, K. R. Park, I. J. Kim, J. Kim. "Detecting driver drowsiness using feature-level fusion and user-specific classification." *Expert Systems with Applications*. Vol.41, no. 4, pp. 1139-52, 2014.
- [10] X. Ochoa, K. Chiluiza, G. Méndez, Luzardo G, Guamán B, Castells J. "Expertise estimation based on simple multimodal features". In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pp. 583-590, 2013.
- [11] Pandey,R.K.,Faridi,A.A.,&Shrivastava,G.(2021). attentiveness Measure in Classroom Environment using Face Detection, 2021 6th ,(ICICT),1053-1058, <https://doi.org/10.1109/ICICT50816.2021.9358600>
- [12] Liang D, Liang H, Yu Z, Zhang Y. Deep convolutional BiLSTM fusion network for facial expression recognition. *The Visual Computer*. Vol. 36, no. 3, pp. 499-508, 2020.
- [13] Hemachandra, S., and R. V. S. Satyanarayana. "Co-active neuro-fuzzy inference system for prediction of electric load." *International Journal of Electrical and Electronics Engineering Research* 3.2 (2013): 217-222.
- [14] Rajani, B., and P. Sangameswara Raju. "POWER QUALITY IMPROVEMENT USING MULTI CONVERTER UNIFIED POWER QUALITY CONDITIONER WITH PQ THEORY." *International Journal of Electrical and ElectronicsEngineering Research (IJEER)* ISSN: 185-200.
- [15] Barhate, Vinay. "A review of distinguishing schemes for power transformer's magnetizing inrush and fault currents." *International Journal of Electrical and Electronics Engineering Research* 3.2 (2013): 277-284.
- [16] SALIH, HASAN WAHHAB, A. K. Bhardwaj, and SURYA PRAKASH. "Load frequency control of hybrid system using industrial controller and implement fuzzy controller practically using PLC." (2013).