Journal of Autonomous Intelligence (2024) Volume 7 Issue 2
doi: 10.32629/jai.v7i2.790

ORIGINAL RESEARCH ARTICLE

Emotion sensitive analysis of learners' cognitive state using deep learning

S. Aruna[1,2,*], Swarna Kuchibhotla[1]

[1] *Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*

[2] *IT Department, Vasavi College of Engineering, Hyderabad 500031, India*

*** Corresponding author:** S. Aruna, 173030121@kluniversity.in, s.aruna@staff.vce.ac.in

ABSTRACT

The assessment of the state of mind of a student has traditionally been a troublesome task. The advances in deep learning have given analysts new opportunities to try and do therefore. Most state of mind methods focus principally on attention, failing to account for the significance of human emotions. Emotions are significant in laptop vision and a good deal of analysis is conducted exploitation human feelings. Our objective is to propose an emotion-sensitive analysis of individuals' mental state, specifically focusing on students' attention levels. This analysis will be carried out in a non-intrusive manner by detecting both head posture and emotions. To achieve this, we employ a multi-task learning approach that utilizes convolutional neural networks (CNNs). These networks are capable of simultaneously identifying facial expressions, locating facial landmarks, and estimating head position, all in real-time. Face alignment is additional assessed by estimating the pinnacle position and face alignment. The estimation of the pinnacle cause and alignment of the face is additional employed by the trainer to live the learner's span. Experimental results show that the technique will accurately verify students' emotions with a ninety-four accuracy rate.

Keywords: landmark location; cognitive state; head pose estimation

ARTICLE INFO

Received: 26 June 2023
Accepted: 19 July 2023
Available online: 1 December 2023

1. Introduction

According to the control-value theory of achievement emotions, the impact of cognitive activation teaching strategies on students' effective utilization of Cognitive and Meta-Cognitive (CMC) learning strategies is influenced by their perceptions of control, such as self-efficacy, and their emotions related to achievement, specifically enjoyment and boredom. In other words, the theory suggests that the effects of these teaching tactics are mediated by students' appraisal of control and their emotional experiences of enjoyment and boredom. Despite assertions of cross-cultural and domain consistency, there is minimal and contradictory empirical evidence to support this notion[1]. It is necessary to do sequential research to determine if students' emotions impact the use of cognitive learning techniques, or vice versa. Both orientations are feasible, according to the control-value theory of achievement emotions. Enjoyment is a good, motivating emotion that belongs to the category of action emotions. It occurs when an individual's personal aspirations and objectives are in sync. Boredom is a deactivating and unpleasant feeling that is also classified as an action emotion. Boredom occurs when an individual's personal aspirations differ from his or her professional ambitions. When there is direct association with the learning material, cognitive

learning methods assist the learner in processing information. Learners can use metacognitive learning strategies to help them apply cognitive learning tactics more effectively. Rehearsal, organisation, and elaboration strategies are all examples of cognitive learning strategies[2]. As the importance of improving academic performance among young individuals grows, their academic identity and well-being are increasingly significant. The way they perceive and experience their academic achievements and setbacks significantly impacts their academic accomplishments and overall effectiveness. To bring forth this goal, gain knowledge or nurture new skills, 'cognitive emotions' play a critical part by indicating the state or "flow" of emotions in a student when challenged with difficult activities. Measuring the emotive components of learning in creative teaching models has primarily relied on metrics and self-generated reports, ignoring the possibility of emotion detection in the real-time, through video monitoring of facial gestures[3]. Head posture estimation, the programmed estimation of the orientation of the head relative to a camera-centred arrange framework, may be an issue of both hypothetical and viable significance. In the event that an acknowledgment framework needs to prosper under real-life situations, it should handle critical varieties of head poses, and programmed location of the head is anticipated to be a highly likely addition to human-computer interfacing in the near future[4]. Modelling human head posture may be a challenging issue in computer vision and flag handling. It is alluring since this head pose flag gives us imperative meta-information almost communicative motions, notable districts in a scene based on centre of consideration[5], bunch discovery, swarm behavioural elements and following[6], and inconsistency detection. In domains where near level iris/eye following isn't conceivable, human head posture is the foremost vital highlight in evaluating human focus-of-attention.

Figure 1 represents a typical classroom environment with a wide range of cognitive abilities. Typically, we can detect that some pupils are concentrated, while others are not, based on their head posture and the blackboard as a reference point. Another important component is recognising each student's facial expression and categorising it into one of six fundamental emotion variations: fear, surprise, happiness, sadness, anger, disgust[7]. Their aggregate impact will be equated to a Boolean quantity indicating the student's attention span.

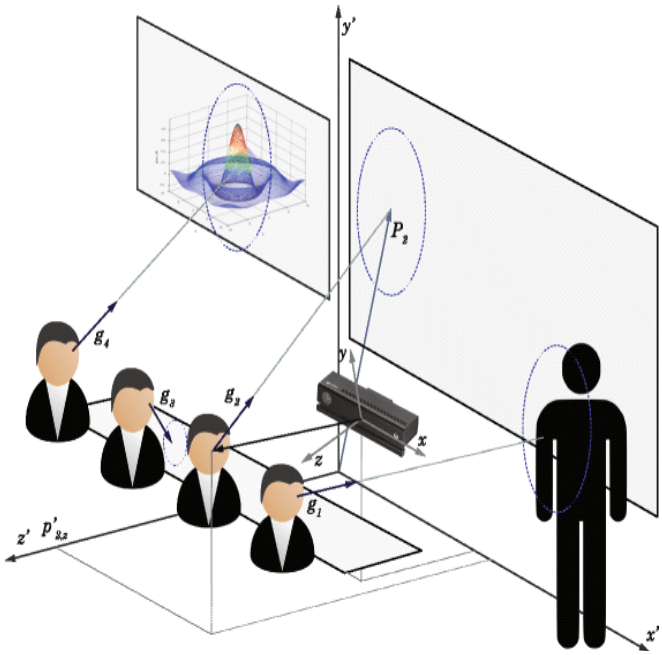

Figure 1. A case of estimation in the class.

2. Related work

Traditional facial expression recognition algorithms are ineffective in real-time. This is based on the convolutional neural network algorithm's recognition and detection mechanism. This is built on the LeNet-5 network, which has been optimised and improved in terms of network structure and internal structure. To overcome the problem of network model overfitting caused by diverse features, add batch normalisation[7]. The mental balance or incoherency between new data and current information while learning by means of a MOOC course causes epistemic feelings such as curiosity, delight, bewilderment, and fear in a MOOC environment. The proposed study used a mix technique of deep learning and Social Network Analysis (SNA) to find examples of epistemic feelings regarding communications on a MOOC stage by collecting data from 1190 Chinese learners. The findings uncovered that 4 patterns emerged from the data: core, neighbour, scattered, and peripheral learners. These patterns tended to expand relationships through votes and establish deep through remark and answer collaborations[8].

A study looks into students' profiles, taking into account their learning styles, emotional variables, and study tempo. There are 406 social science students in the sample (34.1 percent first-year, 27.6 percent second-year, 16.7 percent third-year and 21.6 percent are fourth-year students). They filled out the following questionnaires: (a) the Approaches to Learning and Studying Inventory (ALSI); (b) the Student Experience of Emotions Inventory; and (c) the Emotion Regulation Questionnaire. (d) the Sense of Coherence Scale and (e) the Need for Cognition-short form. CFA, Cluster Analysis, MANOVA, Discriminant Analysis, and the Decision Tree Model were employed in the statistical analysis[9]. The face expression recognition model network reaches the degree of convergence after 100 iterations, according to an experimental finding. The model recognition rate in the training set was 98.76 percent, while the model accuracy rate and F1 value in the test set were both 1. The model can correctly identify and discriminate seven facial expressions: angry, disgust, fear, happiness, sadness, surprise, and neutral. The suggested methodology can catch changes in students' facial expressions in online classrooms, allowing teachers to have a real-time understanding of their students' learning status[10]. AutoTutor is a computer tutoring programme that mimics human tutoring and converses with pupils in normal language. Twenty-four people took turns with the student and the AutoTutor for a total of 200 turns. To construct a record of previous and next set of emotions, two series of attributes and emotions were concatenated into one row. To perform classification for emotional state labelling, feature extraction techniques such as multilayer-perceptron and naive Bayes were used on the dataset[11]. An Eye Tracker is used to record the parameters of a student's eye gaze. At the conclusion of the experiment, an eye tiredness detection test is performed to detect eye weariness. The eye metrics were statistically analysed, and it was discovered that they have a substantial link with cognitive load[12]. Full-online, hybrid, blended, synchronous, and asynchronous online learning are all gaining in popularity. Teachers are finding it difficult to assess their pupils' engagement without having direct interaction with them. In this regard, an intelligent application for teachers is being created to recognise students' emotions and determine their levels of interest during a presentation[13].

E-learning refers to learning that takes place via the use of electronic media. The learners in most E-learning systems are always passive. Many studies are being conducted to convert passive learners into active learners. The recognition of head movements and facial emotions is the focus of this study[14]. When this is used for training players and ludology experience in professional gaming, customer satisfaction analysis for broadcasters and streamers, or monitoring the concentration of the driver, head posture and emotion shifts present major issues. Because of the growing popularity and usability of depth sensors, it is now possible to collect significant amounts of three-dimensional (3D) data for analysis[15].

One of the most important applications of machine learning in affective computing is the capacity to recognise users' emotions for the goal of emotion engineering. Electroencephalography (EEG), face image

processing, and speech inflections are some of the more prevalent ways of emotion recognition[16,17]. The method of determining whether or not a face is present in a picture is known as face detection and tracking. In social communication, the face is very significant. According to psychological research, the nonverbal element of social communication is the most informative route[18,19].

3. Methodology

Dataset:

The Amplified CohnKanade (CK+) data set is an expanded adaptation of the CohnKanade (CK) data set containing 593 video groups and still pictures consisting 6 generic emotions and 1 particularly unbiased emotion. Still, images and recordings are taken in a laboratory environment. 123 subjects between the ages of 18 and 30 were selected for these recordings and image recordings. The determination for every picture is 640 × 490 pixels and 640 × 480 pixels, and the dim esteem is eight 8bit exactness. The precision of identifying the essential emotions for distinctive strategies on the CK+ data set appears underneath the charts. The images in the dataset may not be consistent due to various factors like lighting conditions, camera lens quality, etc., we try to process the images with image processing and filtering and sharpening the images such that we can hope to get better results in the CNN model.

MTCNN Algorithm:

MTCNN (Multi-task Cascaded Convolutional Networks) is a framework designed to address the tasks of face detection and face alignment simultaneously. It employs a three-stage convolutional network approach to identify faces and accurately locate facial landmarks, including the eyes, nose, and mouth.

MTCNN has three stages.

Stage 1: the Proposal Network (P-Net):

The initial phase of the MTCNN framework employs a fully convolutional network (FCN) instead of a traditional convolutional neural network (CNN) architecture. The key distinction lies in the absence of a dense layer within an FCN. This stage, known as the Proposal Network, is utilized to acquire candidate windows and their corresponding bounding box regression vectors.

Bounding box regression is a widely used technique for precisely determining the location of boxes when detecting objects of specific pre-defined classes, such as faces in this context. Once the bounding box vectors are obtained, a refinement process is applied to merge overlapping regions. This refinement step aims to combine and consolidate overlapping regions, resulting in a final output of candidate windows. These candidate windows have undergone refinement to effectively reduce the volume of candidates, ensuring a more manageable dataset for subsequent processing as shown in **Figure 2**.

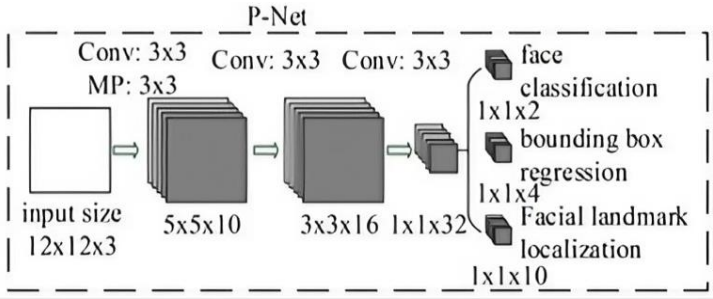

Figure 2. P-Net.

Stage 2: the Refine Network (R-Net):

The candidates generated by the P-Net are subsequently inputted into the Refine Network. It's important to note that unlike the previous stage, the Refine Network is a CNN that incorporates a dense layer in its architecture. The purpose of the Refine Network, or R-Net, is to further reduce the number of candidates, perform bounding box regression for calibration, and utilize non-maximum suppression (NMS) to merge overlapping candidates. The output of the R-Net includes a determination of whether the input is a face or not, a 4-element vector representing the bounding box coordinates of the face, and a 10-element vector for precise facial landmark localization as shown in **Figure 3**.

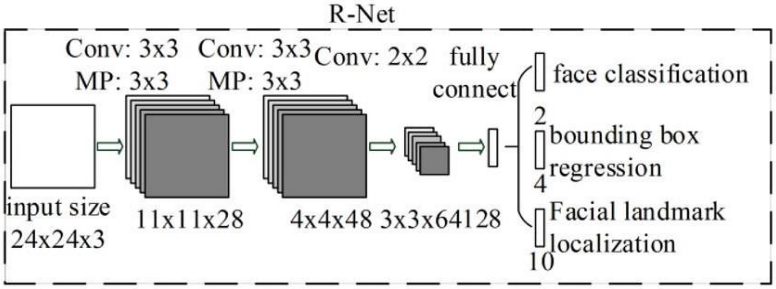

Figure 3. R-Net.

Stage 3: the Output Network (O-Net)

Similar to the R-Net, the Output Network serves a comparable purpose as shown in **Figure 4**. However, its primary objective is to provide a more comprehensive description of the face by accurately determining the positions of the five facial landmarks, namely the eyes, nose, and mouth.

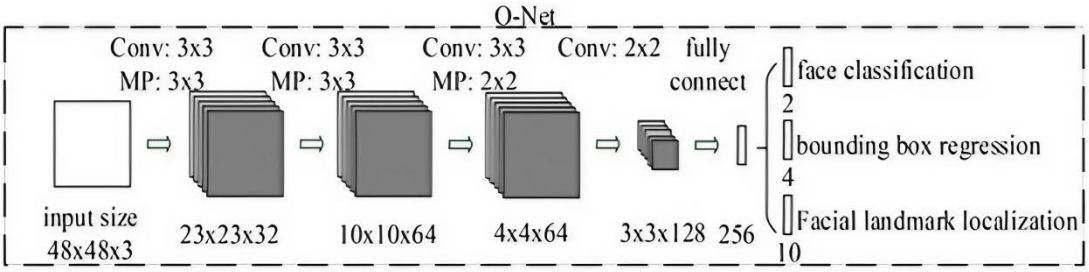

Figure 4. O-Net.

Mathematical representation of the model

Layers in CNN model:

The initial 4 convolution layers with ELU activation function can be discribed as:

1) $hl^{l}_{nj} = conv^{l}_{n+1,j} = \mathrm{Act}(u^{l}_{2n,j})$

Act

After the 4 convolution layers, there are two flatten layers:

2) $h^{l}_{Z} = \{a_1 h^{l}_{1}, a_2 h^{l}_{2}, a_3 h^{l}_{3}, a_4 h^{l}_{4}\}$

3) $out^{l}_{f1} = Flatten(F_{n}\, Act(h_{z}))$

4) $out^{l}_{f2} = Flatten(F_{n+1}\, Act(out^{l}_{f1}))$

Finally, there is a dense layer:

5) $out^{l}_{F} = Dense(Den_{c}, Act_{softmax}(out^{l}_{f2}))$

Proposed algorithm:

The projected system is associate degree execution of various assignments strategy cascaded with a CNN, i.e., Multi-Task Cascaded Convolutional Neural Networks. it's associate degree calculation comprising of three stages, that acknowledge the bounding boxes of faces in an exceeding image in conjunction with the facial points of interest. Every net is created by passing its inputs through the CNN repeatedly, which returns scored candidate bounding boxes, taken when by maximum concealment. The expanded Multi-Task Cascaded Convolutional Neural Networks could also be a three-stage cascaded net and further finetunes the process.

Initially, multiple image pyramids were created using the existing photographs. These pyramids were then fed into the P-Net to generate regression vectors for the bounding boxes. The purpose of this P-Net is to refine the uncertain images before sending to next stage. To make sure of extra possible yield, it grasps an amazingly lean convolutional arrange. The candidate windows are subsequently passed through the R-Net to further refine and eliminate non-face candidates. The R-Net employs a dynamically complex convolutional neural network architecture. In the next step, the refined output is fed into the O-Net to precisely extract detailed facial features.

Subsequently, 0.33 internet includes a similar shape to the second arrange but has no other CNN. Unlike a multi-task CNN, in this particular approach, the third stage architecture addresses three distinct tasks simultaneously. These tasks include the detection of facial expressions, regression of bounding boxes, and identification of points of interest. Additionally, the method also incorporates the estimation of head posture as an integral part of its process. The below stated strategy contains 3 distinctive nets: the P-Net, R-Net, and O-Net as regarded in **Figure 5**.

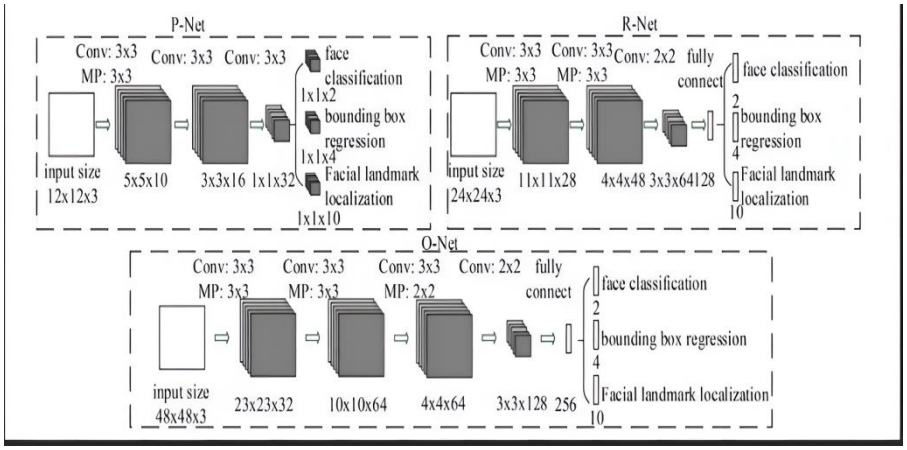

Figure 5. Three separate nets in an MTCNN model: P-Net, R-Net, and O-Net.

Facial feeling acknowledgment is an approach to figure out how a human is feeling owing to his facial expressions. These expressions are categorized into 6 vital emotions namely: astonish, nauseate, bliss, outrage, worry, and pity. A spotlight set is extricated from the making ready statistics to get the essential and unique traits of discourse signals. A trained model is created by feeding input from surrounding units and target values of emotion classes into an SVM (Support Vector Machine) to calculate the ready display. Once the expressions are determined, points of interest are identified using a combination of HOG (Histogram of Oriented Gradients) features, a sliding window approach, an image pyramid, and a linear classifier.

The proposed framework is as confirmed in the bottom **Figure 6**. This contemplates centers across the headway of a cognitive country exam framework sensitive to feelings. To recognize a student's cognitive country in a diffused way, it's very vital to set off feelings and concentration by using the pinnacle pose and

expressions. Thus, the extra intrigue to recognize the sentiments makes it vital to research the advent of confronting and escalated estimation of feelings. A profound mastering primarily based on cascading CNN is offered to tackle the overfitting problem and carry out those distinctive assignments at the same time. The match employments five focuses on equal time. The Multi-Task Cascaded Convolutional Neural Networks employments a five focuses facial factor of hobby on the facial factor of hobby discovery. We can calculate pitch, roll and yaw of the head by evaluating head posture as regarded in **Figure 7**. The values of those factors are tuned, with improvements in the metrics.

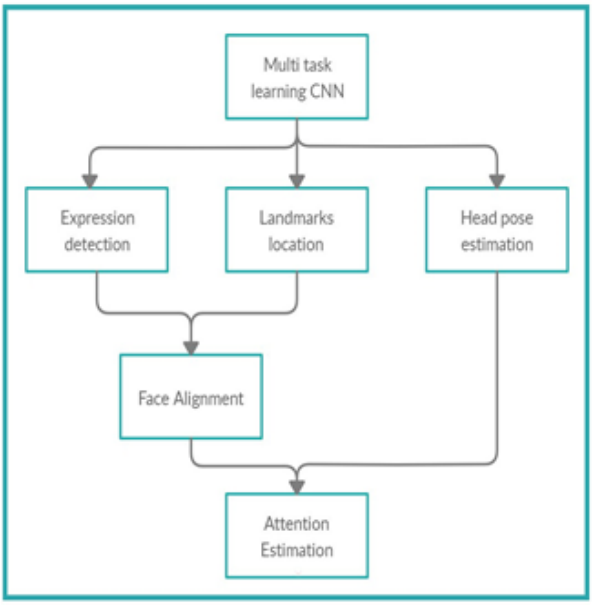

Figure 6. The suggested system's design.

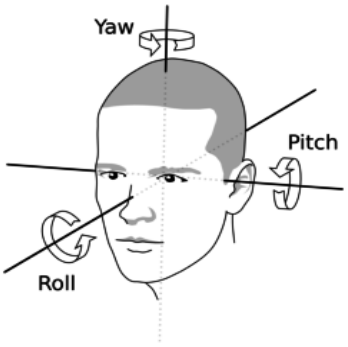

Figure 7. Angles of yaw, roll and pitch based on individual person movements.

We regularly utilize perspectives to differentiate one human's face to other face. Assuming a learner might have a large nose or big eyes, the points of interest found will help distinguish the face. Basically, to make relative changes, select the required operation, get the changed comparison network, and perform the point element on the original image. The best way to easily fill a dark column is to use the most accurate calculation. The numerical representation of the focus (the point of interest) is just that someone's conflict has turned into a grid. The next step in face recognition is to calculate some points of interest by comparing these points of interest. But they are just points of interest that are meaningful to people[20].

Though these methods help catch various parts of a face, it cannot conclude that the photos belong to the same person. Later, machine learning strategies could be used to turn conflicts into numerical representations. Assuming that the human face vector representation contains full appreciation for some of the 128 points, there is no mistake. You can use the Euclidean distance to get a numerical representation of

the conflict and then compare the vector to some of the past vectors to determine the difference between the vectors. There is no mysterious distance which can be utilized to determine if two persons are the same. You can determine the distance and the precision of the test. The confrontation stamp finder starts at the bounding box and bypasses the identified faces. Use the default prepared classifier for the end face.

After getting the viewpoint of interest, try to measure the face. The 2DPOI found in the showdown basically transforms into a head shape. Therefore, given a 3D demonstration of a non-exclusive human head, you are ready to compare 3D focus for almost any point of interest. Interesting facial focus captures hard and non-hard distortions of the face with a very conservative expression and is essential for a variety of errands, including facial examinations. As more applications move from still images to recordings, the global persistence of points of interest becomes more important. So, to speak, there are many records with points of real interest.

Transfer learning approach to save time and resources from having to train multiple machine learning models from scratch to complete similar tasks. As an efficiency saving in areas of machine learning that require high amounts of resources such as image categorisation or natural language processing.

4. Results

The framework is prepared using Cohn-Kanade+ dataset. The data utilized for examination is live nourish from the computer web cam. The understudy's feeling is derived from the expressions and portrayals facially. The description of each emotion and the corresponding facial muscles involved are derived based on Darwin's universal theory of emotions. The pictures within the frame of pixels are put away as clusters and are spoken to as 24 × 24 bit. A picture pyramid is made to recognize the face of every estimate. In conclusion, we ought to make different copies of the pictures in several sizes to discover unmistakably measured confront inside the pictures. In some cases, a picture may comprise as if it were a chunk of a confront interfering into the outline in the edge. Then, the net may conclude that it is outside the boundary using a bounding box. For each box, a cluster of a settled measure is made, and the estimates of pixel of an isolated cluster are marked. Since we have numerous 24 × 24 pixels clusters, the boxes are resized to 48 × 48 pixels. The result of the discovery of facial expressions that appeared in **Figure 8**.

Figure 8. Facial emotion identification.

The Multi-Task Cascaded Convolutional Neural Networks employments five focuses on facial points of interest for facial point of interest locations. The five focuses are either eye ends, the nose tip and either mouth ends. We ordinarily utilize facial points of interest to distinguish faces. Assume an individual has a wide set of eyes or a large nose. The points of interest found offer assistance to distinguish the faces. After obtaining the facial points of interest, we make an endeavour to urge the measurement of the face. The 2D points of interest found on the face essentially acclimate to the head shape. Hence, being able to get around comparing 3D focuses on almost all the points of interest when a 3D model of a bland head is given. In spite of the fact that the head posture may be gotten by points of interest, the learning calculation sensibly leverages the affiliations among different assignments, ordinarily inciting to the advancement of a person's execution. The yield of head position estimation.

Pitch, roll and yaw are measures of student head development. Head roll, pitch, and yaw points are pre-calculated and then evaluated by the student's observations as shown in **Figure 9**. These points are used to consider metrics that determine if the visual centre of the understudy display is on the whiteboard. Consideration is high for positive emotions and the determined position is displayed as a measure. If the emotions are negative and the determined head position is used as a measurement, the consideration at this point is straightforward. If the calculated head position does not match the metric at this point, the considerations are irrelevant to the student's feelings.

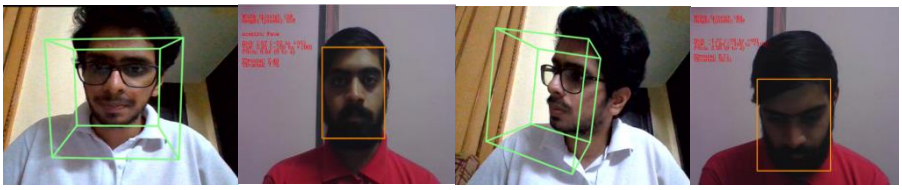

Figure 9. Estimation of head pose and estimation of attention.

The Amplified CohnKanade (CK+) data set is an expanded adaptation of the CohnKanade (CK) data set containing 593 video groups and still pictures consisting 6 generic emotions and 1 particularly unbiased emotion. Still, images and recordings are taken in a laboratory environment. 123 subjects between the ages of 18 and 30 were selected for these recordings and image recordings. The determination for every picture is 640 × 490 pixels and 640 × 480 pixels, and the dim esteem is eight 8bit exactness. The precision of identifying the essential emotions for distinctive strategies on the CK+ data set appears underneath the charts. The images in the dataset may not be consistent due to various factors like lighting conditions, camera lens quality etc. we try to process the images with image processing and filtering and sharpening the images such that we can hope to get better results in the CNN model.

Investigating prepared SVM classifiers from highlights chosen by Adaboost, prepared on the edge yields of chosen Gabor highlights. At any rate, we prepared SVM's on the nonstop yields of the chosen channels. We, for the most part, call these combined classifiers AdaSVM. As shown in **Figure 10**, the detection rate of nausea manifestations in AdaSVM is very good.

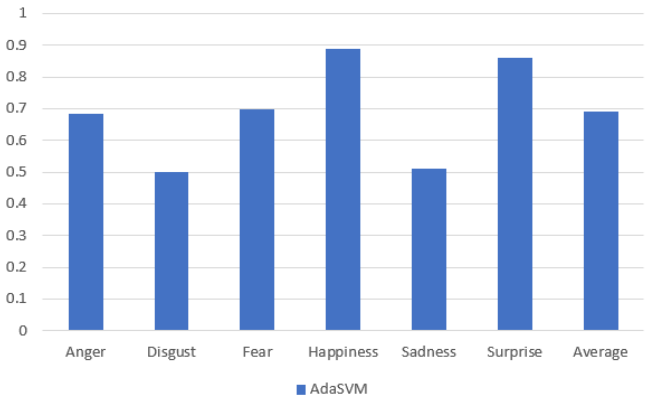

Figure 10. AdaSVM emotion detection estimate on CK+ dataset.

Adaboost is not only a decision strategy, but also a fast classifier. The benefit of this is it highlights chosen subordinate to the already chosen highlights. Adaboost beats AdaSVM in almost every expression, but unfortunately, it's about the same. **Figure 11** shows the position of Adaboost on the CK+ dataset.

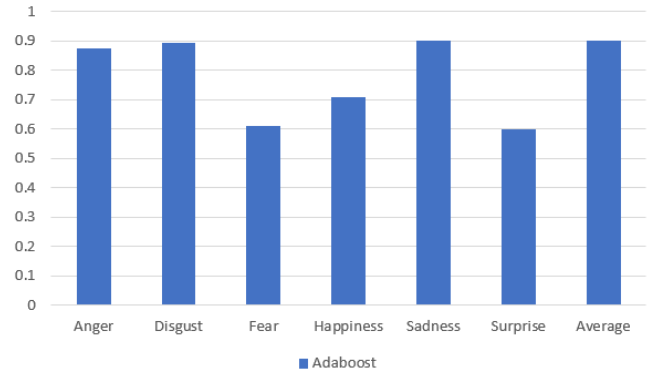

Figure 11. Adaboost emotion detection estimate on CK+ dataset.

The Rankboost strategy on the dataset was more accurate in perceiving pitiful expressions in contrast to Adaboost, AdaSVM as appeared in **Figure 12**. The acknowledgment rate of emotions cheerful and shock were nearly the same as above mentioned methods.

With a large number of input highlights, overfitting is expected in managed learning. It is worth noting that while adapting the test by preparing for error minimization using an unregulated discriminative model, the complexity of the test increases relatively with VC measurements. Therefore, the rule is retained to push the presentation of the rank boost strategy as well. As shown, the accuracy of phrase recognition is significantly improved compared to rank boost.

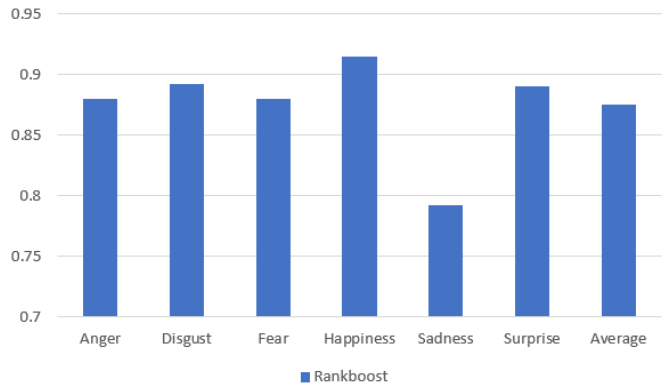

Figure 12. Rankboost's emotion detection estimate on CK+ dataset.

Figure 13 shows Regrankboost detection estimation. **Figure 14** almost shows the Multi-Task Cascaded Convolutional Neural Networks expression detection rate in the CK+ dataset, with a typical confirmation rate of 94% for the Multi-Task Cascaded Convolutional Neural Networks strategy, which is the highest compared to other strategies. Recently, some strategies have achieved 98% accuracy. In any case, when testing the run, use the so-called top [far] peak contour.

The **Figure 15** shows the confusion matrix obtained after training the MTCNN model with the CK+ dataset.

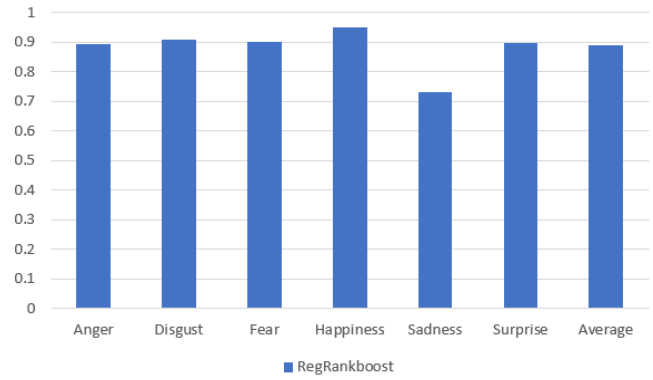

Figure 13. RegRankboost's emotion detection estimate on the CK+ dataset.

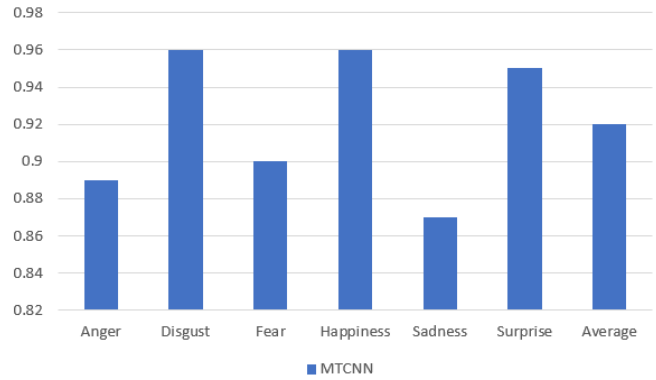

Figure 14. MTCNN emotion detection estimate on CK+ dataset.

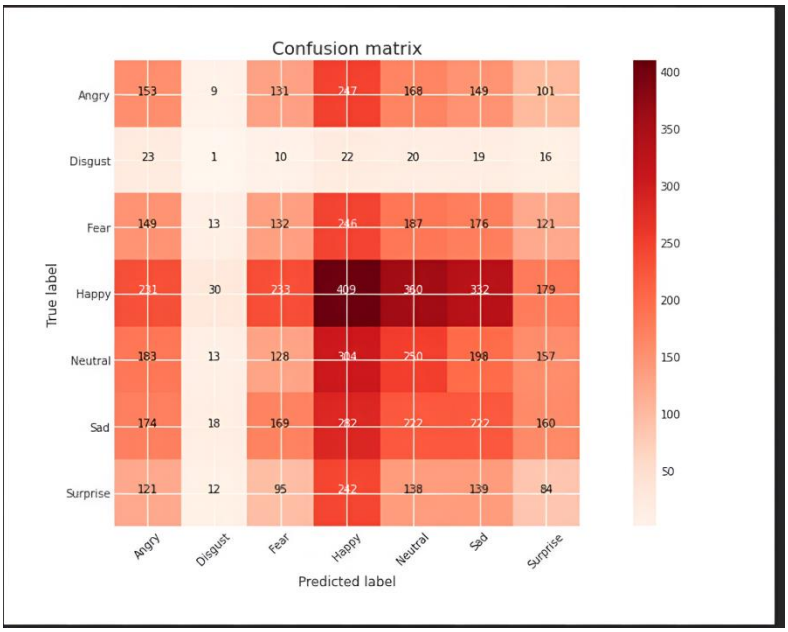

Figure 15. Confusion matrix.

5. Comparative analysis

Here, a review of the proposed framework for learning emotion-sensitive uncertain cognitive states, along with various images from public databases such as the CCNU dataset, which may be a class dataset. It will be fully considered. This dataset was acquired with the introduction of video monitoring through CCTV in the classroom and contains images of learners with poses of head going from −90° to +90° along a

momentary representation. **Figure 16** shows a comparison of expression detection rates for different strategies in the CK+ dataset.

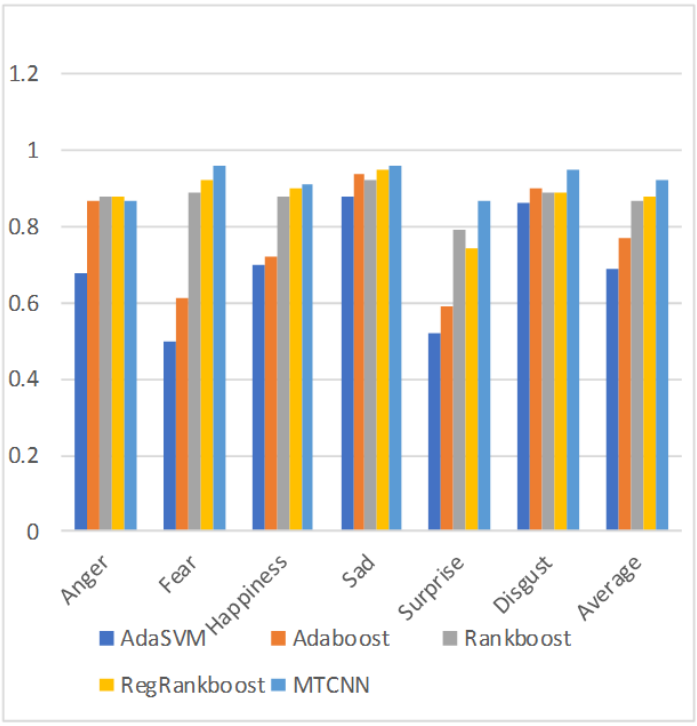

Figure 16. Different methods' emotion detection estimates on the CK+ dataset.

6. Conclusion and future scope

Examining a learner's cognitive state in an unpretentious way can be a tedious task, and the related task of assessing emotions and reasoning are two notable tasks for analysts. The eye tracking method outperforms head pose calculation in accurately measuring a person's attention. Even if an individual maintains a correct head pose, their level of concentration may not be accurately determined. However, an eye tracking system can effectively identify whether a person is genuinely focused, thereby providing a more precise assessment of their attention. To investigate this issue, we present an emotion-sensitive cognitive state assessment framework. The proposed system includes a multitasking strategy, the first Multi-Task Cascaded Convolutional Neural Networks to advance expression recognition, searching for points of interest, and assessing head poses. The points of interest found are used to plan conflicts for examining facial expressions. The evaluated head pose is used to distinguish the student's visual display center. You can extend the accuracy of head pose assessment to include the basic task of recognizing facial expressions and finding points of interest. Analyze facial expressions to determine student feelings and focus on the classroom to determine student attention. In the future, we will encourage greater accuracy and address the effectiveness of the proposed strategies for other datasets. The proposed show could also be used for purposes other than smart classrooms, such as when facial recognition is used as a security bar. Emotions are related to conflict and cannot be opened effectively. Security is further enhanced and the risk to user safety is reduced. You don't need overly sophisticated equipment to use it effectively. This model can be further developed attention and engagement of not only for students, but also for others for advertisements in malls, roads or conferences to identify what catches the eyes of a customer basically it can be used for situations where there a need to determine any person's attentiveness. Future research can be incorporated usingmultimodal data fusion or exploring deep reinforcement learning techniques for adaptive learning environments.

Author contributions

Conceptualization SA, SK; methodology, SA, SK; software, SA, SK; validation SA, SK; formal analysis, SA, SK; investigation, SA, SK; resources, SA, SK; data curation, SA, SK; writing—original draft preparation, SA, SK; writing—review and editing, SA, SK; visualization, SA, SK; supervision, SA, SK; project administration, SA, SK. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Ekatushabe M, Nsanganwimana F, Muwonge CM, Ssenyonga J. The relationship between cognitive activation, self-efficacy, achievement emotions and (meta) cognitive learning strategies among ugandan biology learners. *African Journal of Research in Mathematics, Science and Technology Education* 2021; 25(3): 247–258. doi: 10.1080/18117295.2021.2018867
2. Obergriesser S, Stoeger H. Students' emotions of enjoyment and boredom and their use of cognitive learning strategies—How do they affect one another? *Learning and Instruction* 2020; 66: 101285. doi: 10.1016/j.learninstruc.2019.101285
3. D'errico F, Paciello M, De Carolis B, et al. Cognitive emotions in e-learning processes and their potential relationship with students' academic adjustment. *International Journal of Emotional Education* 2018; 10(1): 89–111.
4. Raytchev B, Yoda I, Sakaue K. Head pose estimation by nonlinear manifold learning. In: Proceedings of the 17th International Conference on Pattern Recognition; 26 August 2004; Cambridge, UK. pp. 462–466.
5. Langton SRH, Honeyman H, Tessler E. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics* 2004; 66: 752–771. doi: 10.3758/BF03194970
6. Baxter RH, Leach MJV, Mukherjee SS, Robertson NM. An adaptive motion model for person tracking with instantaneous head-pose features. *IEEE Signal Processing Letters* 2014; 22(5): 578–582. doi: 10.1109/LSP.2014.2364458
7. Shenoi VV, Kuchibhotla S, Kotturu P. An efficient state detection of a person by fusion of acoustic and alcoholic features using various classification algorithms. *International Journal of Speech Technology* 2020; 23: 625–632. doi: 10.1007/s10772-020-09726-7
8. Wu S, Wang B. Facial expression recognition based on computer deep learning algorithm: Taking cognitive acceptance of college students as an example. *Journal of Ambient Intelligence and Humanized Computing* 2021; 13(1): 45. doi: 10.1007/s12652-021-03113-z
9. Han ZM, Huang CQ, Yu JH, Tsai CC. Identifying patterns of epistemic emotions with respect to interactions in massive online open courses. *Computers in Human Behavior* 2021; 122: 106843. doi: 10.1016/j.chb.2021.106843
10. Karagiannopoulou E, Milienos FS, Rentzios C. Grouping learning approaches and emotional factors to predict students' academic progress. *International Journal of School & Educational Psychology* 2022; 10(2): 258–275. doi: 10.1080/21683603.2020.1832941
11. Zhan Z, Shen T, Jin L, et al. Research on evaluation of online teaching effect based on deep learning technology. In: Proceedings of 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC); 12–14 March 2021; Chongqing, China.
12. Myers MH. Automatic detection of a student's affective states for intelligent teaching systems. *Brain Sciences* 2021; 11(3): 331. doi: 10.3390/brainsci11030331
13. Jyotsna C, Amudha J. Eye gaze as an indicator for stress level analysis in students. In: Proceedings of 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 19–22 September 2018; Bangalore, India. pp. 1588–1593.
14. Hasnine MN, Bui HTT, Tran TTT, et al. Students' emotion extraction and visualization for engagement detection in online learning. *Procedia Computer Science* 2021; 192: 3423–3431. doi: 10.1016/j.procs.2021.09.115
15. Indhumathi R, Geetha A. Survey on recognition of head movements and facial emotions in e- learning system. *International Journal of Scientific Research in Computer Science Applications and Management Studies* 2018; 7(4).
16. Kalliatakis G, Stergiou A, Vidakis N. Conceiving human interaction by visualizing depth data of head pose changes and emotion recognition via facial expressions. *Computers* 2017; 6(3): 25. doi: 10.3390/computers6030025
17. Shivaranjani M. Emotion recognition framework using EEG signals for music persuaded activity. *International Journal of Innovations in Scientific and Engineering Research* 2020; 6(6): 75–82.

18. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: Taxonomy, review and current challenges. *Sensors* 2020; 20(8): 2384. doi: 10.3390/s20082384
19. Kadam S, Dhawale V, Patil S. Facial gesture detection and eye tracking during virtual interview. *Pramana Research Journal* 2019; 9(6): 170.
20. Priyanka KS, Ravikumar G. Fake biometric detection applied to iris, fingerprint, and face recognition by using image quality assessment. *International Journal of Innovations in Scientific and Engineering Research* 2015; 2(3): 57–72.