ISSN 2347 - 3983
Volume 8. No. 9, September 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter204892020.pdf
https://doi.org/10.30534/ijeter/2020/204892020

Deep Learning-Based Emotion-Sensitive Learning Cognitive State Analysis

S. Aruna[1], Swarna Kuchibhotla [2]
[1] Research Scholar, CSE Department
Koneru Lakshmaiah Education Foundation
Guntur, Andhra Pradesh, India,s.aruna@staff.vce.ac.in
[2] Associate Professor,CSE Department
Koneru Lakshmaiah Education Foundation Guntur,
Andhra Pradesh,India, drkswarna@kluniversity.in

ABSTRACT

The cognitive state of a student is unique and exciting**.** The current developments in deep learning gives an unparalleled chance to analysts to assess the cognitive state. But most of the existing cognitive state strategies centre around attention, overlooking the importance of emotions in humans. Human feelings play a huge role in the computer vision industry and numerous researches are performed with its assistance. Hence, our aim is to propose a cognitive state investigation system which is emotion sensitive. It will consequently assess the students' attention based on head posture and emotion recognized from the face detection in an unobtrusive way. The system presented is an implementation of multiple tasks learning cascaded with a convolutional neural network (CNN), introduced for detecting expression, locating landmarks, and estimating head pose all at a time. The expression detection and landmark location help in alignment of the face. The estimation of head pose and face alignment are further used to evaluate the leaner's attention. Exploratory outcomes show that this technique can obtain students' emotion with an accuracy of 94%.

Key words: Cognitive state, Head pose estimation, Landmark location.

1. INTRODUCTION

The crucial factor that decides learning adequacy in a class is learning their intellectual state, along with investigation of students' cognitive or psychological state which has been a great hurdle for researchers to overcome [1]. Learning cognitive state can be recognized by a great progress by different social sign investigation, such as attention and feeling. Attention is the only significant pointer for analysis of psychological state. A person's attention is straightforwardly identified by estimating his head pose. It works as the true proportion of attention and mirrors the intellectual or psychological state in computer-human associations [2].

Apart from attention, emotion has a significant job in the field of human learning. Emotion impacts the capacity of processing information and disrupt events. Past researches display the positive feelings along the learning procedure, for example, enthusiasm and happiness can advance students' cognitive state, while negative feelings, for example, sad and angry can upset students' cognitive state [3]. Most existing cognitive state analysis strategies anyway centre around attention, while emotion is to a great extent overlooked. Subsequently, it implies to include analysis of emotions to learn psychological state of learner [4].

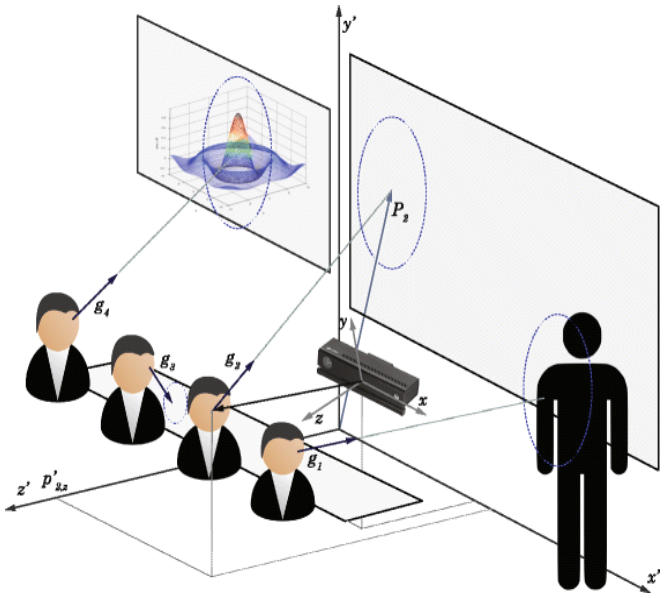


Figure 1: A Classroom Approximation Scenario

The fig 1. depicts a generic classroom scenario which has many students with varied cognitive states. Typically, we can see that some students are focused, and others are unfocused which can be determined by their head pose, keeping the black board as a reference point. Another major factor being considered is recognizing the facial expression of each student and classifying it amongst six basic expression variants: anger, fear, disgust, surprise, sadness and happiness

[5]. The combined effect of these modules will ultimately be resolved to the Boolean value of whether the student is attentive in the classroom.

2. RELATED WORK

Various strategies for analysing emotion are created and they are generally partitioned into pair of classes based on: biological signals [6,7,8] and facial expressions [9,10]. Various signs of these methods such as electrothermal action, blood pressure (BP) and action of Electroencephalogram is gathered to investigate the students' cognitive state [11]. In any case, it is hard to apply these techniques in reality since these methods require complex wearable gadgets to gauge these physiological signals. On the other hand, as detecting expression is among the most remarkable signals for individuals to display feelings and attitude, so that can be utilized to analyse feelings. A study by Mehrabian has shown around 55% of feelings are derived from facial expression, 38% is derived from vocals and the remaining 7% is derived through words [12]. Hence, facial expressions analysis (FEA) is a possible answer for comprehending emotions.

The analysis framework of facial expressions comprises of 2 primary stages: pre-processing and analysing the facial expressions. The target of pre-processing is just modifying and to standardize the head with a progression of tasks including identifying expressions, aligning a face and locating landmarks. As of late, many multitask CNN [13] strategies have been introduced for combining facial expressions and aligning them, for example extended MTCNN [14] and hyperface [15]. By mutually practising numerous assignments, multi-task learning can efficiently solve the issue caused due to small number of labelled examples. FEA comprises of classifications of facial expressions and expression intensity calculation.

Off late, there is an expanding enthusiasm for a more detailed examination, to be specific, estimating the intensity of facial expressions. It is introduced to rank facial expression with unique levels of intensity. Anyhow, labelling the intensity of expression for deep learning is exceptionally tedious. To tackle this issue, various techniques look at the estimation of intensities of expressions as a logistic classification issue which knows the brief structure of information from a process underneath an independent framework.

To flatten the control of built features, a couple of works presented a technique based on deep learning to look at the problem of logistic classification. Consider, Chen et al. [16] presented the CNN which is a ranking model that consolidates numerous parallel CNN for measuring age, that altered the logistic classification to a progression of parallel classifications to mini tasks. Anyhow, it grasped the age model from the facial expression samples that are labelled**.** Dong et al. [17] proposed a technique based on artificial intelligence called RankCNN which prepares the information to exhibit the fundamental method of ranking analysis.

We might know that there isn't an unsupervised strategy to rank intensity of expression using CNN. Similar techniques for analysing students' state are introduced based on deep learning. Fan et al. [18] decrypted mind state with functional magnetic resonance imaging pictures which prepared the sorter of SVM along many particular characteristics, picked using a method of hybrid features selection.

There are some other emotion recognition algorithms which use multi SVNN classifiers (Multiple Support Vector Neural Network) [19] for identifying the emotions and sentiment analysis [20]. The significance of obtaining student intellectual states information in an unobtrusive way had incited the headway of various methodology to propel different procedures to advance the result in estimating head posture and analysis of facial expressions. [21] Introduced a complete assessment of LBP which include extensions of that idea are explained. As a normal usage of the LBP approach, LBP-primarily based facial picture examination is widely evaluated, at the same time as its fruitful expansions, which manipulate extraordinary assignments of facial photograph research. By applying principal component analysis algorithms to find duplication of applicant's face [22]. By applying Image Processing Techniques to Identify and Recognize Facial Expressions [23].

3. METHODOLOGY

The proposed framework is an implementation of multiple tasks method cascaded with a CNN i.e., MTCNN. It is an algorithm comprising of 3 phases, that identifies the bounding boxes of faces in a picture along with the facial landmarks. Each stage is improved step by step by passing its inputs through the CNN, which returns candidate bounding boxes with their scores, followed by non max. suppression. The extended multi-task convolutional neural network is a three-stage cascaded algorithm it performs the task in more refined way.

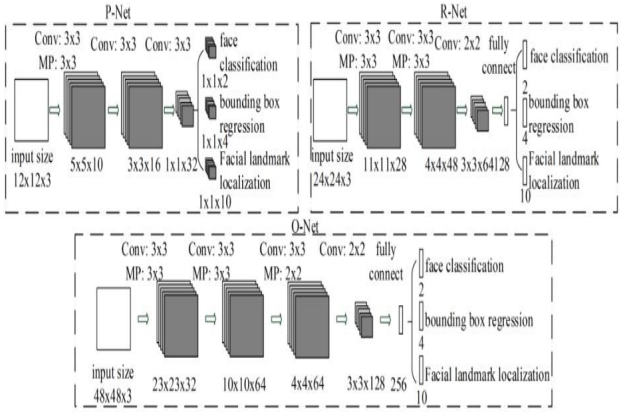


Figure 2: A MTCNN Model with Three Separate Nets: P-Net, R-Net, O-Net

At the start, many pyramids of images are built from the existing pictures which is fed to the P-net to deliver the regression vectors for bounding box. To guarantee more effective output it embraces an extremely thin convolutional network. The user window is fed to R-net in the subsequent stage, to dismiss many non-face candidates. At the last, R-net embraces a progressively unpredictable convolutional neural network. In the next step, the processed output is fed to O-net to explain the features of faces in detailed way.

Consequently, third net has a comparable structure to second network yet has another CNN. Not the same as multi-task CNN, here the third network structure deals with 3 jobs (detection of expression, regression of bounding boxes, and locating landmarks), estimation of head pose is also included in this method. The proposed method comprises of three different nets: The P-Net, the R-Net, and the O-Net are shown in fig 2.

Facial emotion recognition is the process of detecting human emotions from facial expressions. The classification of facial expressions is done into six basic expressions as surprise, disgust, happiness, anger, fear and sadness. A feature set is extracted from the training data to obtain the essential and distinct characteristics of speech signals. A training model is then constructed by feeding pairs of feature sets and the target values of emotion categories into the learning algorithm of support vector machines (SVMs). Once the expression is derived, the landmarks are calculated using Histogram of Oriented Gradients (HOG) feature along with an image pyramid, linear classifier and a sliding window detection scheme.

The proposed system is as appeared in the below fig. 3. This study centres around the advancement of a cognitive state analysis system sensitive to emotion. To comprehend student's cognitive state in an unobtrusive way, it is important to get the student's emotion and attention utilizing posture of the head and facial expression. Hence, the more interest to comprehend the feelings, it becomes important to explore the appearance of face and intensity estimation of feeling at the same time. A deep learning based multi task learning CNN is introduced to address the overfitting problem and perform these multiple tasks simultaneously. The MTCNN uses a 5 points facial landmark for facial landmark detection. The five points on face are two near the eyes, one on the nose tip and other two for the edges of the mouth. We can determine the roll, pitch and yaw angles of the human head corresponding to its head pose as shown in fig.4. As there is a change in the movement of head, the values of these angles are also changed respectively.

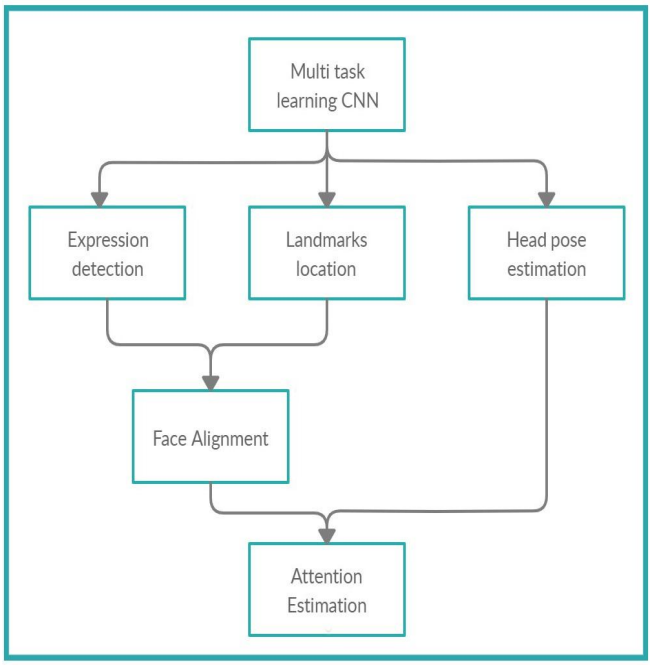


Figure 3: Structure of the Proposed System

We normally use facial landmarks to distinguish one person's face from another. Suppose, a person may have a big nose or a wide set of eyes, the landmarks located help to differentiate the faces. For performing affine transformations, we simply select the operation we want, we obtain the corresponding transformed matrix and perform a dot product on the original image. The best way that the black gaps can be filled in easily is by using the nearest algorithm. The numerical representation of points (landmarks) is nothing but someone's face transformed into a matrix. The next logical step of facial recognition would be to calculate some kind of landmark distance by comparing these landmarks. But these are just the landmarks that make sense to humans.

Although this method is useful to capture different parts of a person's face, they might not be best suited to tell us definitely if the photos are of the same person. The process of transforming a face into a numerical representation can be done by recent machine learning techniques. Suppose a facial vector representation of a person has a large value in some point of the 128 points, it may not mean anything to us. We can use the Euclidian distance to understand how different a vector is, by comparing it with some previous vector, after we get a numerical representation of the face. We do not have any magical distance that can be used to tell whether the two people are same. We can only calculate distances and test the accuracy. The face mark detector will work around the faces which are detected, beginning with the bounding boxes. We use the standard trained classifier on frontal faces.

After acquiring the facial landmarks, we make an attempt to get the dimension of the face. The 2D landmarks located on the face basically adjust to the shape of the head. Thus, we can get approximately corresponding 3D points for almost all the landmarks when given a 3D model of a generic human head.

Facial landmark points capture rigid and non-rigid disfigurement of faces in an extremely conservative depiction and are in this way important for a wide range of tasks involving facial analysis. Temporal stability of landmarks is an issue becoming increasingly important with more applications shifting from single images to videos. There are only a few datasets with ground truth landmarks for videos. This is with facial landmark detection.

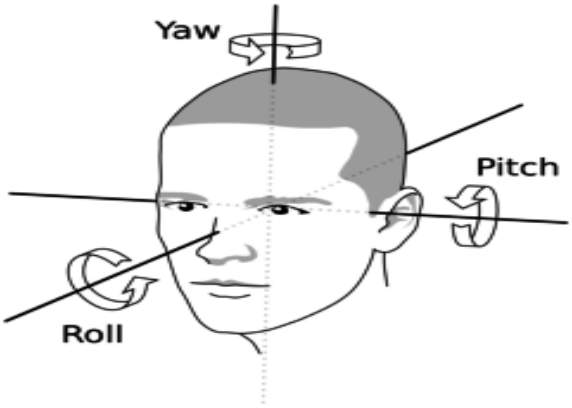


Fig. 4 Roll, Pitch and Yaw Angles in Human Head Motion

4. EXPERIMENTAL RESULTS

The proposed system is trained with CK+ dataset [24], the input used while testing is live feed from the system's camera. The emotion of the student is detected from the facial expression and the facial description. The facial description for each emotion and also description of the facial muscles involved in emotions are as considered by Darwin universal theory of emotions. The images in the form of pixels is stored as arrays and is represented as the 24 x 24 kernel. We need to make a pyramid of images so as to distinguish face of every single distinctive size. At the end, we need to make various duplicates of the images in different sizes to find distinct sized face within the images. Sometimes, an image may consist only a piece of a face peeping into the frame from the corner. In such cases, the net(network) might result a bounding box that is slightly out of the frame. For each bounding box, we make an array of a fixed size, and note the pixel values to a separate array. Since we have many 24 x 24 arrays of images, we resize the boxes to 48 x 48 pixels and further reshaping the bounding boxes to a square. The output of detection of facial expressions is as shown in fig. 5.

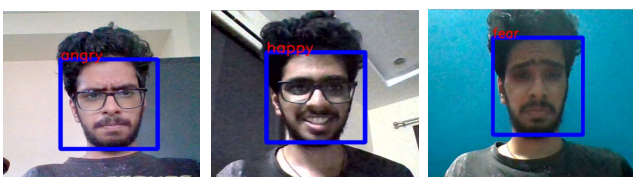


Figure 5: Facial Expression Detection

The MTCNN uses a five points facial landmark for facial landmark detection. The five points on face are two near the eyes, one on the nose tip and other two for the edges of the mouth. We normally use facial landmarks to distinguish one person's face from another. Suppose, a person may have a big nose or a wide set of eyes, the landmarks located help to differentiate the faces. After acquiring the facial landmarks, we make an attempt to get the dimension of the face. The 2D landmarks located on the face basically adjust to the shape of the head. Thus, we can get approximately corresponding 3D points for almost all the landmarks when given a 3D model of a generic human head. Though the head pose might be obtained by landmarks, the learning algorithm sensibly leverages the associations among multiple tasks, typically prompting to the development of individual performance. The output of head pose estimation as shown in fig. 6.

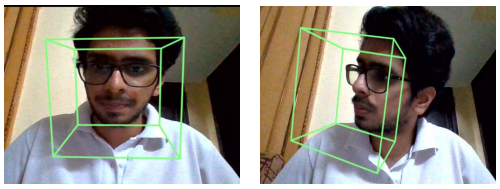


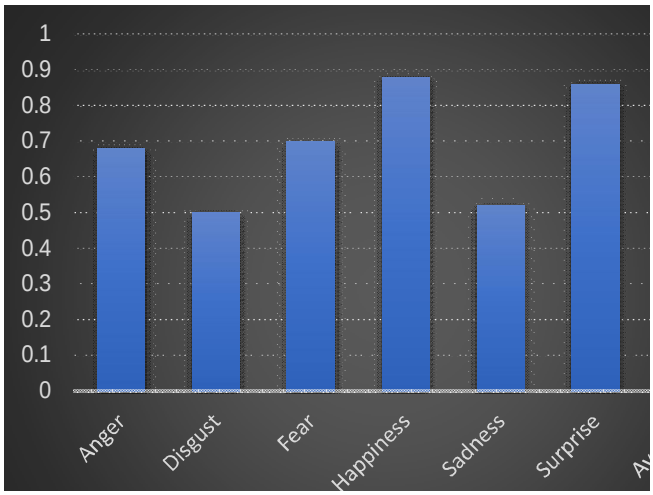
Figure 6 : Head Pose Estimation

Pitch, yaw and roll are the three dimensions of movement of the students' head. Further the roll, pitch and yaw angles of the head are calculated and then the learners' attention is estimated as shown in Fig.7. A metric is considered using these angles to determine if the student visual focus of attention is on the whiteboard. If the emotion is positive and the calculated head pose is in range considered as metric then the attention is high. If the emotion is negative and the calculated head pose is in range considered as metric then the attention is moderate. If the head pose calculated is not matching the metric then the attention is low irrespective of students' emotion.

Figure 7 : Attention Estimation

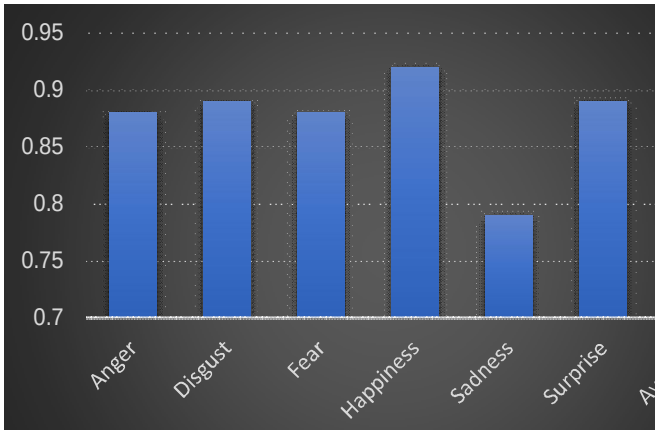
Extended Cohn–Kanade (CK**+**) dataset is the extended version of the Cohn-Kanade (CK) dataset which contains 593 video sequences and stationary images containing six basic emotions and an additional neutral emotion. The stationary images and the videos are shot in a lab environment. 123 subjects are selected for capturing these videos and images whose age varies from 18 to 30 years. The resolution for each image is 640 pixels × 490 pixels and 640 pixels × 480 pixels, and the grey value is 8-bit precision. The accuracy of detecting the basic expressions for different methods on CK+ dataset is shown in below graphs.

The accuracy of the proposed method is compared with the existing methods such as AdaSVM [25], Adaboost [26], Rankboost and RegRankboost [27].



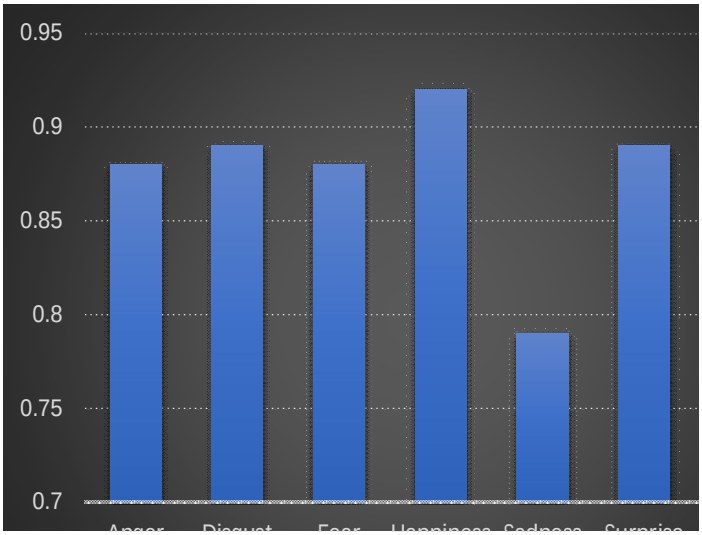
Graph 1 Expression Detection Rate of AdaSVM on CK+ Dataset

Exploring trained SVM classifiers on the features selected by Adaboost which were trained on the threshold outputs of the selected Gabor features. Anyhow, we trained SVM's on the continuous outputs of the selected filters. We generally call these combined classifiers AdaSVM. The rate of detecting disgust expression of AdaSVM is very low as shown in graph 1.



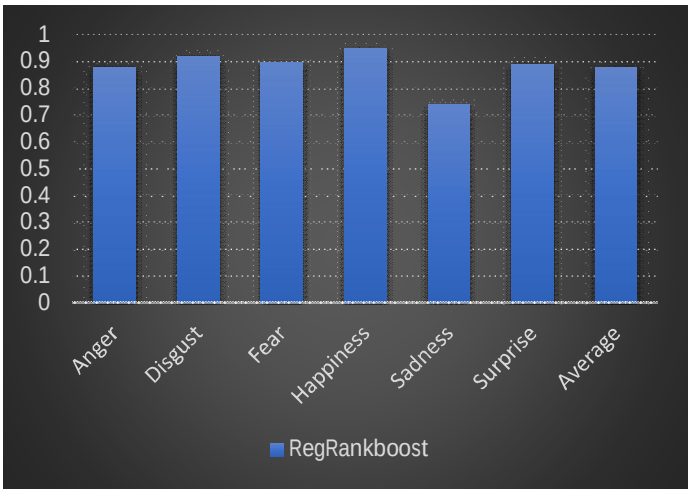
Graph 2 Expression Detection Rate of Adaboost on CK+ Dataset

Adaboost, along with being a feature selection method is also a fast classifier. Its advantage is that the features are selected dependent upon the features that have already been selected. Adaboost outperforms AdaSVM for almost all expressions but is nearly the same for sadness expression. Graph 2 shows expression detection rate of Adaboost on CK+ dataset.



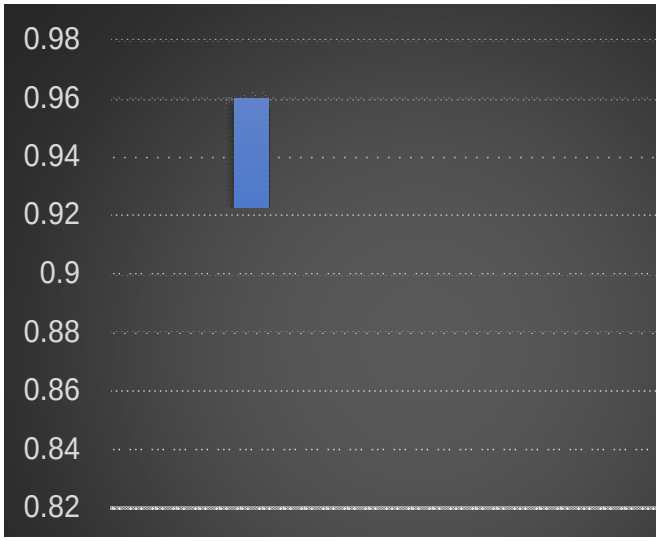
Graph 3 Expression Detection Rate of Rankboost on CK+ Dataset

The Rankboost method on CK+ dataset showed a far better accuracy in recognising sad expression when compared to AdaSVM and Adaboost as shown in graph 3. The recognition rate of expressions happy and surprise was almost same as AdaSVM and Adaboost.



Graph 4 Expression Detection Rate of RegRankboost on CK+ Dataset

Overfitting is expected in supervised learning due to many input features. It is notable that sample complexity increases proportionally with the VC dimension while utilizing unregularized discriminative models to fit the samples by training error minimization. Thus, the l1 regularization is adopted to additionally improve the exhibition of the Rankboost method. The accuracy in recognizing expressions is further increased compared to Rankboost as shown in graph 4.

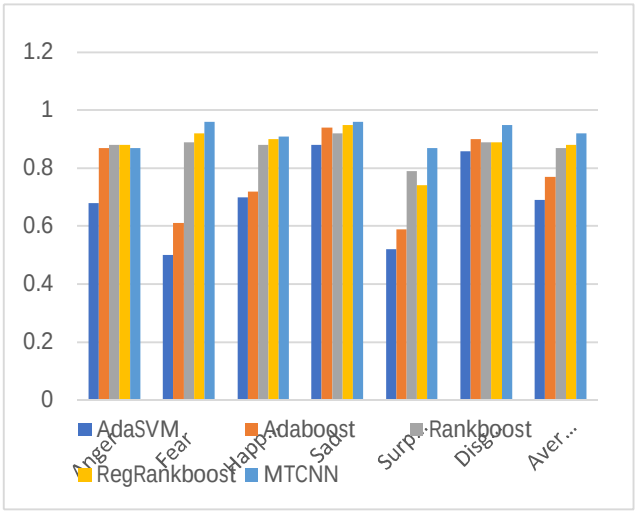


Graph 5 Expression Detection Rate of MTCNN on CK+ Dataset

The graph 5 is about MTCNN's expression detection rate on CK+ dataset and average recognition rate of MTCNN method is 94%, which is highest when compared to remaining methods. Some methods recently have achieved 98% accuracy; however, they use only peak expression (apex frame) when testing the performance.

5. Comparative Analysis

Here, the analysis of proposed system for learning cognitive state which is sensitive to emotions is completely assessed along with various pictures from openly accessible databases [28] such as the CCNU dataset which is a classroom dataset. This dataset was obtained by a installing a CCTV video surveillance in the classroom and it consists of students' pictures with head poses ranging from −90◦ to +90◦ along with instant expressions. Graph 6 depicts the comparison of expression detection rate of different methods on CK+ dataset.



Graph 6 Expression Detection Rate of Different Methods on CK+ Dataset

6. Conclusion and Future Scope

The analysis of the cognitive state of learner in an unobtrusive way is a difficult job, the related works of estimating emotion and attention are two significant challenges for the researchers. To solve this issue, we present a cognitive state investigation system that is sensitive to emotions. The proposed framework embraces a multi-task implementation method which improves the original MTCNN for detecting expression, locating landmarks and estimating head posture. The landmarks located are utilized to prepare the face for the analysis of facial expressions. The estimated head posture is utilized to identify the students' visual focus of attention. The correctness in estimating head posture can be increased with the underlying work of detecting facial expression and locating landmarks. The facial expressions are analysed for the analysis of students' feelings while concentrating in the classroom which determines the learners' attention. In future, we will further improve the accuracy and address the efficiency of the proposed method for other datasets. The proposed model can serve other purposes also other than in intelligent classrooms like, it can be used where facial recognition is used for security lock, as the emotion is associated along with the face, it cannot be opened easily. The security is increased and also it reduces the threat to user's privacy. It does not require any highly sophisticated devices so it can be used easily.

REFERENCES

1. Siau K, ShengH,NahFH(2006) **Use of a classroom response system to enhance classroom interactivity**. IEEE Trans Educ 49(3):398–403
2. Odobez JM, Ba S (2007) **A cognitive and unsupervised map adaptation approach to the recognition of the focus of attention from head pose**. In: IEEE international conference on multimedia and expo, pp 1379–1382
3. Ashby FG, Isen AM, TurkenAU(1999) **A neuropsychological theory of positive affect and its influence on cognition**. Psychol Rev 106(3):529–550
4. P.Vishal, L.K.Snigdha,Shahana Bano "**An Efficient Face Recognition System using Local Binary Pattern**" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.
5. 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,10.1007/s10772-020-09726-7
6. Chen D, Hu Y, Wang L, Zomaya AY, Li X (2017) H-PARAFAC**: Hierarchical parallel factor analysis of multidimensional big data**. IEEE Trans Parallel Distrib Syst 28(4):1091–1104
7. Tang Y, Chen D,Wang L, ZomayaAY,Chen J, LiuH(2018) **Bayesian tensor factorization for multi-way analysis**

of multi-dimensional EEG. **Neurocomputing** 318:162–174

8. Kuchibhotla,S. & Niranjan,M.S.R 2018, "**Emotional classification of acoustic information with optical feature subset selection methods".** International journal of Engineering and technology(UAE), vol. 7, no. 2, pp.39-43.
9. Puja, Er. Rachna Rajput "**Feature Extraction in Face Recognition using SVM-LBP Detection Technique"** International Journal of Innovative Research in Computer and Communication Engineering. International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 10, October 2016.
10. Kiran P. Gaikwad, C. M. Sheela Rani, S. B. Mahajan, P. Sanjeevi kumar, "**Dimensionality Reduction of facial features to recognize emotion state**", Lecture Notes in Electrical Engineering, (LNEE)-Advances in Systems, Control and Automation", Vol.442, pp-719-725, Dec 2017.https://link.springer.com/chapter/10.1007/978-98 1-10-4762-6_69.
11. Shen L, Wang M, Shen R (2009) **Affective e-learning: using "emotional" data to improve learning in pervasive learning environment.** J Educ Technol Soc 12(2):176–189
12. Mehrabian A, Wiener M (1967) **Decoding of inconsistent communications**. J Pers Soc Psychol 6(1):109–114
13. Narsingarao, M. R, Venkatesh prasad ,V.,Sai teja,P.,Zindavali,M. & Phani reddy,O.2018,"**A survey on prevention on overfitting in convolution neural networks using machine learning techniques",** International journal of engineering and technology(UAE),vol.7,no.2.32 Special Issue 32,pp.177-180.
14. Zhang K, Zhang Z, Li Z, Qiao Y (2016) **Joint face detection and alignment using multitask cascaded convolutional networks**. IEEE Signal Process Lett 23(10):1499–1503
15. Ranjan R, Patel VM, Chellappa R (2016) **Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition**. IEEE Trans Pattern Anal Mach Intell 41:121–135
16. Chen S, Zhang C, Dong M, Le J, Rao M (2017) **Using ranking-CNN for age estimation.** In: IEEE conference on computer vision and pattern recognition, pp 742–751
17. Dong Y, Huang C, Liu W (2014) **RankCNN: when learning to rank encounters the pseudo preference feedback**. Comput Stand Interfaces 36(3):554–562
18. FanY, ShenD,Davatzikos C (2006) **Detecting cognitive states from FMRI images bymachine learning and multivariate classification.** In: Conference on computer vision and pattern recognition workshop, 2006. CVPRW '06, p 89
19. Kasiprasad Mannepalli, Panyam Narahari Sastry, Maloji Sumana "**Emotion recognition in speech signal using optimization based multi-SVNN classifier**" Journal of King Saud University - Computer and Information Sciences (DOI:10.1016/j.jksuci.2018.11.012).
20. Pradeepini,"**Performing the sentimental analysis on BLOG data"**, G (2017) International Journal of Pure and Applied Mathematics 115-6-253-259.
21. Anusha M, K Karthik, P Padmini Rani, VSrikanth "**Prediction of Student Performance using Machine Learning"** International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019.
22. Dr Nishad Nawaz,"**Artificial Intelligence Face Recognition for applicant tracking system"**, International Journal of Emerging Trends in Engineering Research (IJETER),Volume 7, No. 12 December 2019, ISSN 2347 – 3983
23. J.D.Pujari, Dashrath.K.Bhadangkar and Rajesh. Yakkundimath, "**Identification and Recognition of Facial Expressions Using Image Processing Techniques: A Survey"**, International Journal of Emerging Trends in Engineering Research (IJETER), Volume 5, No.5 May 2017, ISSN 2347 - 3983
24. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, **The Extended Cohn-Kanade Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression,** in *3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis*, 2010
25. Littlewort G, Bartlett MS, Fasel I, Susskind J, Movellan J (2006) **Dynamics of facial expression extracted automatically from video**. Image Vis Comput 24(6):615–625
26. Koelstra S, Pantic M (2008**) Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics**. IEEE Trans Med Imaging 18(8):712–721
27. Yang P, Liu Q, Metaxas DN (2010) **Rankboost with l1 regularization for facial expression recognition and intensity estimation.** In: IEEE international conference on computer vision, pp 1018–1025
28. Kiran Gaikwas, C.M.Sheela "**Comparative Analysis of Emotion states Based on Facial Expression Modality"** Journal of Advanced Research in Dynamical and Control Systems 11(1):462-466.