AUTOMATIC FEEDBACK SYSTEM FOR ONLINE VIDEO CLASSES BY USING FACIAL EXPRESSION

R. Soundhara Raja Pandian[1], and C. Christopher columbus[2]

[1]*Research Scholar, Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamilnadu, India (Corresponding Author)*
E-mail:rsrp.iitm@gmail.com

[2]*Professor, Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamilnadu, India*

Abstract:

Recognizing and recommending the teacher capabilities and moving extents of learning is crucial to scheming teaching-learning interventions. Past research in this area are painted the reputation of human facial terminologies in learning-centred sentimental conditions, but tracking human facial poses significant challenges. To avoid the feedback problem, an automatic feedback system is proposed to collect the response of students from their facial expressions and avoid the malpractices during exams to improve the quality of online education and also provide automatic attendance for the exam. For feedback system ORNN classifier is utilized. The proposed method accomplished the maximum sensitivity of 98.2% which is 96.3% for using RNN+GA, 97.2% for using RNN+PSO, and 96% for using RNN + FA.

Keywords: *NPTEL, Facial expression, Classifier, Optimization, Recommendation*

I. Introduction

In recent years, there is a constructive growth in online education and virtual university over the world. Students can listen classes of reputed professors from top institutions around the world through virtual class rooms any time anywhere [1-3]. The quality of education is not homogeneously dispersed in India such as top premier institutions like IISc, IIT's and NIT's provide good teaching resources. NPTEL (National Program on Technology-enhanced Learning) is one of the best e learning platforms to provide huge knowledge resource and proves the rising claim for higher education, which supports formal classroom education [4-6]. Nowadays most of the educational institutions motivate online education rather than classroom education. Using this online class, students can listen classes of reputed professors from top institutions through virtual class room any time anywhere. The advantages of online education system are given below

- This has a huge impact on students, which can be very useful to access the course and take assessments anytime and anywhere
- There are so many open online courses offered in India
- IIT's are offering around 5000 Courses for all streams through NPTEL
- Now IIT madras is offering open Online courses for students and also conducting certification courses through online

However there are some challenges in online education which is explained in this paper. Students who registered for the course may or may not be attending exams. Student who listen the class and submitting assignment may be different from one who is taking the exam; could not able to see the response of the student. Difficult to find what are the problems they have faced in a particular course [7]. We are not able to see the response of students through online learning and the difficulties what they faced in a particular course. To avoid the problems, lot of research has been developed for automatic feedback system for online classes. Chasing Feelings through human Facial Expressions in live systems based on transient human emotion peak introduced. Attitude extraction using facial features to improve learning curves of students in e-learning systems has been proposed [8]. Moreover, Schoolchild emotion recognition system (SERS) for e-learning improvement based on student attention metric is introduced [9-10]. The extracted facial expressions are given the feedback of on-going classes. But these methods can't properly attain the better feedback system [11]. Therefore in this paper efficient automatic feedback system is proposed to collect the response of students from their expressions. This study mainly focus on avoid the malpractices during exams, recommend the good teacher, improve the quality of online education and find out the interest of student from their facial expression and also provide attendance for the exam[12-13]. An ORNN classifier is utilized for automatic feedback system.

Recurrent neural network:

RNN is unique neural system that has been used for overseeing successive data. The main objective of RNN is to predict the name of current time stamp with the logical data of past time stamps. It is a proficient grouping technique yet not broadly used in the literature. The RNN comprise of two phases and three layers. The hidden layer comprise of hidden and context layer. One-step delay in feedback path so that the topology is similar to that of a feed forward network, except that the outputs of the hidden layer are used as the feedback signals[14-15].

II. Automatic Feedback System for Online Video Classes

For on-line feedback system, five types of emotions are identified from the students namely happy, sad, surprise, anger and disgust. Then finally based on the emotion, teachers score value also be calculated. From the score value, teachers are recommended for further classes. The proposed methodology consist of four modules namely, (i) ROI extraction (ii) feature level fusion (iii) emotion classification and (iv) feedback system. The overall diagram of proposed methodology is given in figure 1. ROI extraction is an important process for identifying student feedback about online courses. Basically, for online feedback system, facial expression is mainly used. For emotion identification, initially consider the captured students face image $I_i \ (i = 1,2,....n)$. Then, ROI is extracted from each student face. Then these ROIs are placed in the face region. Each region has its own importance in recognizing an emotion of a face. After ROI extraction process, local and appearance based features are extracted from each face region. These extracted features are fused from feature level fusion module. The fused features are given to the optimized RNN classifier to identify the emotions of students. The classification results are used for improve the teaching quality of NPTEL online class.

ROI Extraction:

ROI extraction process is divided into two stages namely locating eye centre and extraction of ROIs from face region.

A. Locating eye centres:

Locating eye centre is a major part for emotion identification. Initially, the face images are collected and after capturing on-line student face, eye centre is calculated.

A. Extraction of facial ROIs

In this section, facial ROIs are extracted from face image using D and also these seven ROIs extracted using Rule-based technique [1]. The step by step procedure is explained in following steps;

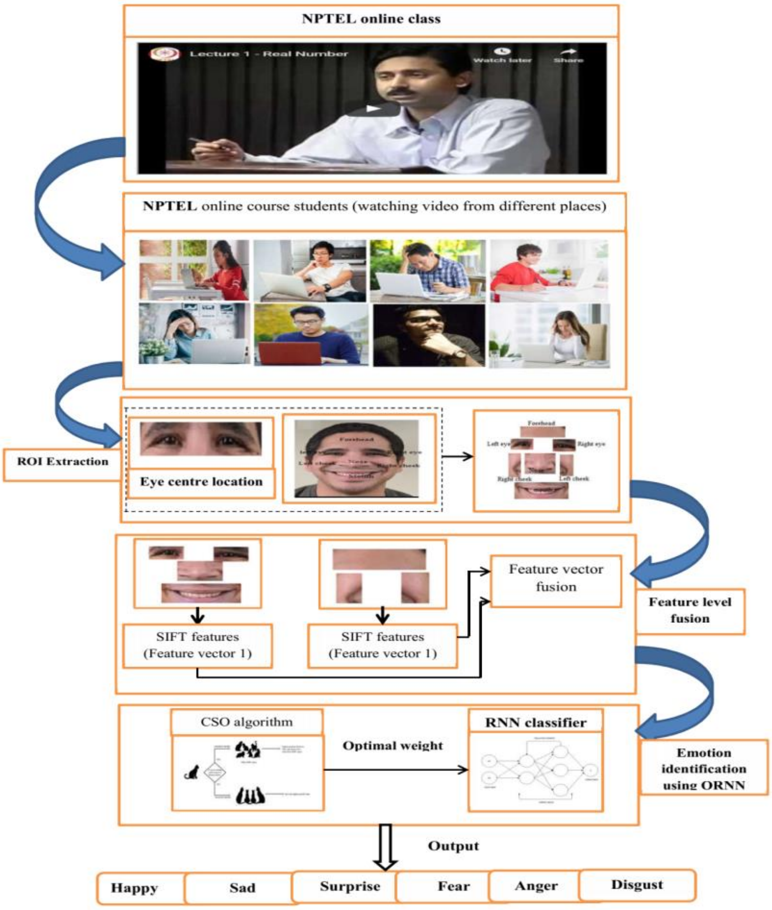


Figure 1: ORNN based Automatic Feedback System

Step 1: Initially, the boundary box (BB) for face area can be located using the value of D. The created BB has width of $3*D$ and height of $2.25*D$.

Step 2: Then, to find the nose region N^O, a new BB is drawn below the L^E. The width of the created BB is D and the height is $2.25*D/4$. This box is drawn 0.15 * D away from the L^E.

Step 3: To find the mouth region $\left(M^O\right)$, a new BB is drawn at the distance of $\frac{2.25*D}{5}$ and $0.2*D$ to the below and left of the L^E, respectively. The created BB has width of $1.4*D$ and height of $\frac{2.25*D}{4}$.

Step 4: Then, above the L^E at the distance of $0.88*D$, forehead region BB is drawn. The Located boundary box has width of D and height of $0.5*D$.

Step 5: To find the right cheek $\left(R^C\right)$, the BB is drawn and the BB width is of $0.5*D$ and height of $0.6*D$.

Step 6: Finally BB is drawn at the distance of $0.15*L$ below the L^E. The created BB width is $0.5*D$ and height is $0.6*D$.

Step 1: Scale space extrema detection:

Scale space detection in the SIFT algorithm is the identification of points or points of interest in the scale space by finding image locations that represent the maximum or minimum of the difference in Gaussian function $G(u,v,\sigma)$. The first stage of the calculation is searching for all scales and image locations. In each candidate location, a detailed sample is appropriate to determine the location and scales. Key points are usually selected based on their stability measures. The scale space of an image is defined as $S(u,v,\sigma)$, which is made from the convolution of the variable-scale Gaussian $C(u,v,\sigma)$, with the input image $I(u,v)$.

$$S(u,v,\sigma)=C(u,v,\sigma)*I(u,v) \tag{1}$$

$$C(u,v,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\left(x^2+y^2\right)/2\sigma^2} \tag{2}$$

Where; σ represent the standard deviation of the Gaussian $C(u,v,\sigma)$.

The Difference of Gaussian function $G(u,v,\sigma)$ calculated from the Gaussian difference of the two scales divided by a factor k can be given as follows;

$$G(u,v,\sigma)=(C(u,v,k\sigma)-C(u,v,\sigma))*I(u,v) \tag{3}$$

$$= S(u,v,k\sigma)-L(u,v,\sigma) \tag{4}$$

Local maxima and minima of $G(u,v,\sigma)$ are calculated depend upon the correlation of the example point and its eight neighbours in the present image as well as the nine neighbours in the scale above and beneath. Normalization of the Laplacian with the factor σ^2 is required for genuine scale invariance. The maxima and minima of $\sigma^2\nabla^2 C$ produce the steadiest image attributes contrasted with a scope of other conceivable image capacities. For example, gradient, Hessian or Harris corner function.

Step 2: Removal of unreliable key points:

At this point, localized points at the lower contrast points and edges are discarded to eliminate noise and instability. To find incredibly important points, the value of $|G(u,v,\sigma)|$ at each candidate's key point is calculated. If the value is below some threshold, the main point is removed because the structure is less differentiable. The Taylor expansion can be converted to quadratic terms of the scale-space function $G(u,v,\sigma)$ to calculate the origin at the sample point.

$$G(u)=G+\frac{\partial G^T}{\partial u}u+\frac{1}{2}u^T\frac{\partial^2 G}{\partial u^2}u \tag{5}$$

$$u=(u,v,\sigma)^T \tag{6}$$

To calculate the location of the extreme $\hat{u}$, the above equation is derivates with respect to u and setting it to zero and is given in equation (15)

$$\hat{u}=-\frac{\partial^2 G^{-1}}{\partial u^2}\frac{\partial G}{\partial u} \tag{7}$$

If the offset $\hat{u}$ is bigger than 0.5 in any measurement, at that point it implies that the extremum lies closer to an alternate sample point. For this situation, the sample point is changed and the interpolation performed about that point. The last offset $\hat{u}$ is added to the area of its sample point to get the interpolated gauge for the area of the extremum. The capacity esteem at the extrema, $G(\hat{u})$, is helpful for dismissing unbalanced extrema with low difference.

$$G(\hat{u})=G+\frac{1}{2}\frac{\partial G^T}{\partial u}\hat{u} \tag{8}$$

After computation, all pixels in the images are normalized between 0 and 1. In this case, all extremes worth less than 0.03 $G(u)$ are discarded. Subsequently, the ratio principal curves of each candidate key point are evaluated to search for poorly defined peaks in the Gaussian function difference. For key points with high edge responses, the primary curve at the edge is much larger than the corresponding main curve. If the ratio is below some threshold (T= 10), the main point is retained, otherwise it will be removed.

Step 3: Orientation assignment

In this stage orientation is calculated for each key point. Orientation is a combination of histogram of gradient orientation $\theta(u,v)$ and gradient magnitude $m(u,v)$. These two are calculated as follows;

$$m(u,v)=\sqrt{(S(u+1,v)-S(u-1,v))^2+(S(u,v+1)-S(u,v-1))^2} \tag{9}$$

$$\theta(u,v)=\tan^{-1}((S(u,v+1)-S(U,V-1))/(S(u+1,v)-S(u-1,v))) \tag{10}$$

Where; S represents the Gaussian smoothed image. By assigning a fixed orientation to each of the key points, the main point descriptor associated with this orientation can be specified so that the variation in image rotation can be achieved.

Step 4: Key point descriptor calculation:

The slope size and orientation at each image point of the 16 x 16 force-point neighbourhood are obtained as shown in Figure 3. The neighbourhood is then weighted by a Gaussian window and accumulated in orientation maps summarizing the contents of adjacent subsets of the 4x4 scale as shown in Figure 2. The length of each arrow is given by Equation (20) Figure 6 corresponds to the sum of the slope sizes near that direction within the region [66].

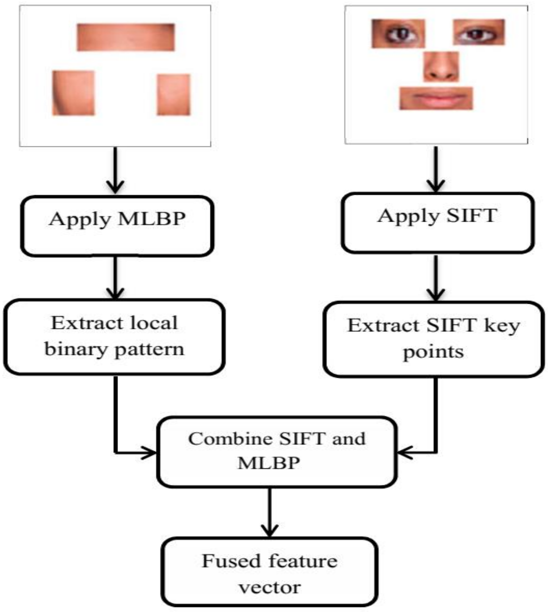


Figure 2: Overall structure of feature level fusion

Since each histogram has 8 pins, each key point description contains 4 x 4 x 8 = 128 elements. The extracted k-points can be used to contest against the key points mined from human facial parts of different subjects. The Feature extraction stages in the LBP may be expressed as follows. The feature vector fusion structure is given in figure 2.

After the feature selection process, the extracted features are given to the classifier to classify an image as sad, happy, anger, disgust and surprise. Nowadays, lot of classifiers are used for classification process. The proposed classification system is the following section.

Proposed ORNN classifier based emotion classification:

RNN is a special type of NN that is primarily used for classification, prognosis and recognition. RNN contain two phases and consists of three layers. Weight principles are improved through the benefit of the CSO algorithm to improve the RNN. The basic structure of the RNN is given in Figure .

RNN is trained using training samples $(a_1,b_1),(a_2,b_2)....(a_n,b_n)$ with $a_i \in R^n$, $b_i \in R^n$ for $1 \le i \le n$. Here, a_i denotes the input vector and b_i denotes the target output.

Step 1: Initially the dimensional reduced dataset is provided for the input of the RNN classifier and assigns their optimal weights.

Step 2: Commonly using the underneath equations, an RNN can be represented as below,

$$x_i(t) = \sum_j y_j(t) w_{ij}(t) \qquad (11)$$

$$y_i(t) = f_i(x_i(t)) \qquad (12)$$

Anywhere, y_i and w_{ij} requires the neuron's beginning method and the weights are updated for further importance. For beginning f_i is based on the input s of the network and context layer inputs.

Step 3: Vector results to determine the activation function of the hidden node, the function are delivered by sigmoid is given in equation (13),

$$f_i = \frac{1}{1+e^{(-x_i)}} \qquad (13)$$

Step 4: Ahead of the dissemination of the practice, the release function of each neuron is calculated as,

$$y_i(t) = f_i(x_i(t), C_i(t)) \qquad (14)$$

$$x_i(t) = \sum_{j\in H} y_j(t) w_{ij} + \sum_{j\in I} x_j(t) w_{ij} + \sum_{j\in C} y_j(t-\tau_{ij}) wij \qquad (15)$$

Where, f, H, I and C represents a neuron activation function of hidden layer values, the values of the input neurons, neuron storing data values in the last phase of the network. Then x_j is j^{th} input neuron and τ_{ij} is an integer value referring to in connection with the series of cases of displacement.

Here, the value t represents the back-propagation network error value for giving the neural calculated

Step 5: The network error is calculated from the equation (15),

$$E_m = Y^{tar} - Y^{act} \tag{16}$$

To minimize the error value, the weights values of RNN are updated with the help of the CSO algorithm.

Recommendation system :

After emotional identification, the best teacher is identified based on emotion. For example, consider an online class of 4 teachers and 100 students. 100 students listen to the speech of four teachers. During lecture, students' facial expressions are automatically identified using the ORNN classifier. Then, the mean of each emotion in a particular teacher class is identified. Happy Emotion is over 75%, teacher teaching quality is considered good, and classes are recommended, otherwise the teacher is not recommended for any other online classes. The level of student satisfaction is very important for online classes. This feedback system improves the quality of online classes.

III. Results and discussion:

The experimental results of proposed methodology are explained in this section. The complete experimentation is done with MATLAB program. The sample images are shown in Figure 3.

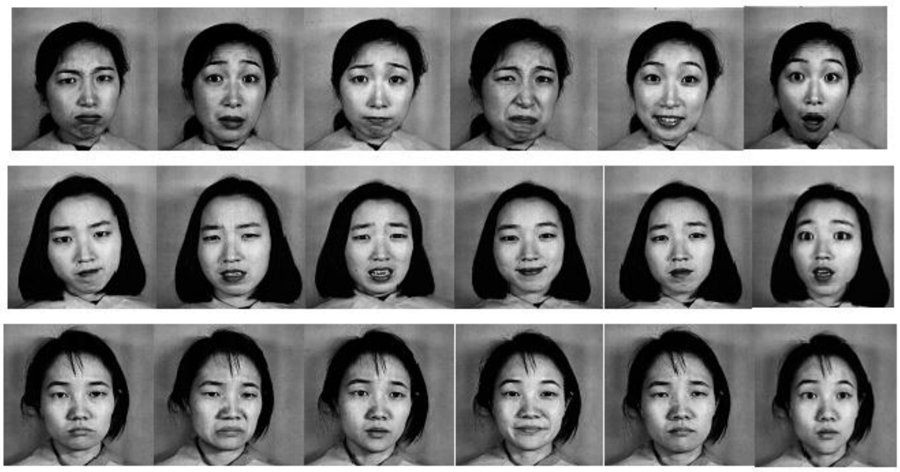


Figure 3: Dataset sample images (a) anger, (b) Disgust, (c) fear, (d) happy, (e) sad and (f) Surprise

The experimnetal results of proposed methodology is listed in this section. The eye center identification is given in figure 4.

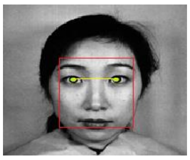
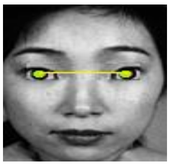
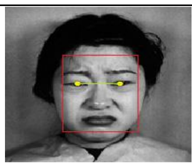
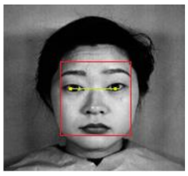
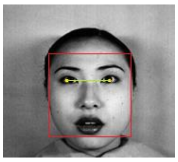
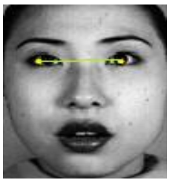
Original image	Eye centers detected	Cropped image of face and located eye centers
		
		
		
		

Figure 4: Results of locating eye centers

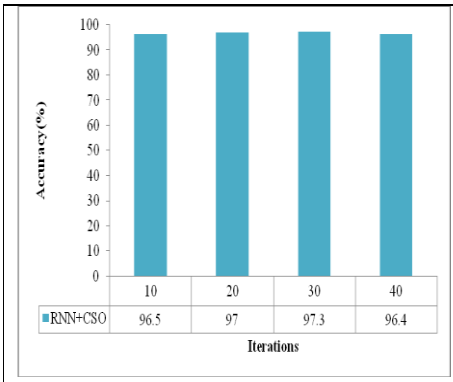


Figure 5: Performance analysis based on accuracy

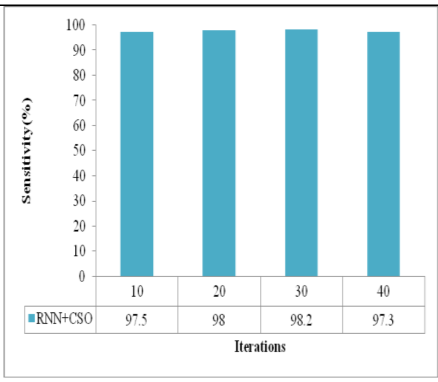


Figure 6: Performance analysis based on sensitivity

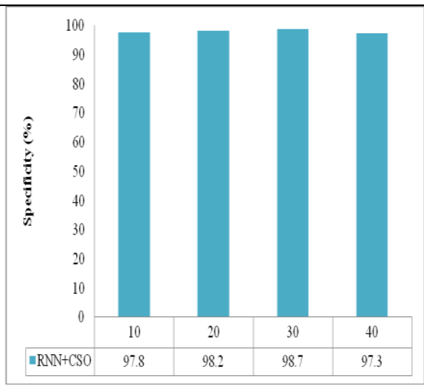


Figure 7: Performance analysis based on specificity

In figure 5, the performance of proposed methodology is analysed in terms of accuracy measure. Here, x-axis represents the iteration and y-axis represents the accuracy measures. When analysing figure 5, the proposed method attained the maximum prediction accuracy of 97.3% for iteration 30. Due to feature level fusion and optimized RNN, the proposed method attained the maximum accuracy. This prediction, helpful for NPTEL online courses teaches to improve the teaching style and methods. In figure 6, the performance of proposed methodology is analysed in terms of sensitivity measures. When analysing figure 6, proposed methodology attained maximum sensitivity of 97.5% for iteration 10, 98% for iteration 20, 98.2% for iteration 30 and 97.3% for iteration 40. From the figure, it’s clearly understand the maximum results came from iteration 30. Similarly, in figure 7 the performance of proposed methodology is analysed in terms of specificity. Here also proposed method attained the maximum results of 98.7%.

Performance comparison based on classification approach:

To prove the effectiveness of proposed methodology, proposed facial expression prediction compared with different method namely, SVM, RNN and ANN. The classification result is given in below figures.

The main objective of this chapter is to predict the NPTEL student feedback based on the facial expression. For this methodology, two main process are introduced namely, feature extraction and prediction. Here feature extraction, combination of SIFT and MLBP are utilized. For prediction process, ORNN classifier is utilized.

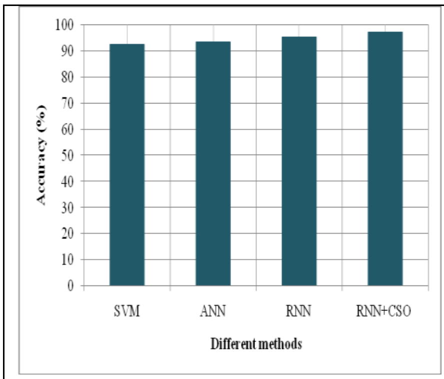


Figure 8: Accuracy comparison of different classifiers

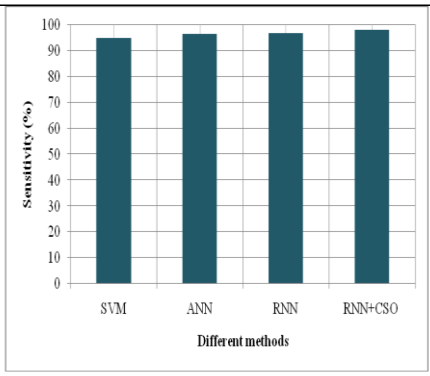


Figure 9: Sensitivity comparison of different classifier

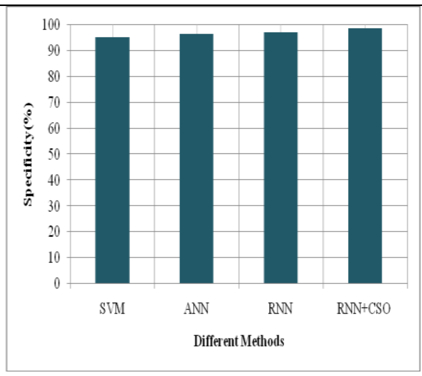


Figure 10: Specificity comparison of different classifier

To enhance the performance of RNN classifier, the weight values are optimally selected with the help of cat optimization. Figure 8 shows the accuracy performance of proposed against existing approach. Here, x-axis represents the different classifiers and y-axis represents the accuracy value. When analysing figure 8, proposed method attain the maximum accuracy of 97.3% which is 92.6% for SVM based prediction, 93.5% for ANN based prediction and 95.4% for RNN based prediction. This is due to effectiveness of ORNN and feature level fusion. The performance of proposed against existing method is analysed in terms of sensitivity is given in figure 9. When analysing figure 9, proposed method attain the maximum precision of sensitivity of 98.2% which is 95.1% for using SVM based prediction, 96.4% for using ANN based prediction and 96.9% for RNN based classification. Similarly in figure 10, the performance of proposed methodology is analysed in terms of specificity. When analysing figure 10, the proposed method attained the maximum specificity of 98.7% which is high compared to existing results. From the results, it's clearly understand the proposed method attained the better results.

Performance comparison based on optimization approach:

To improve the performance of DNN classifier PBOA algorithm is utilized. To prove the effectiveness of proposed algorithm, the performance of different optimization algorithms are compared with proposed algorithm. For comparison, PSO, GA and FA are used. The optimization algorithm based classification result is given in below figures.

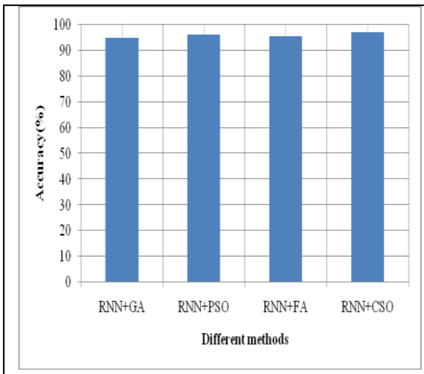


Figure 11: Accuracy comparison by varying optimization approach

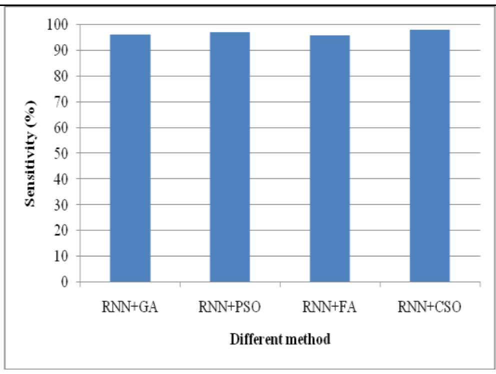


Figure 12: Sensitivity comparison by varying optimization approach

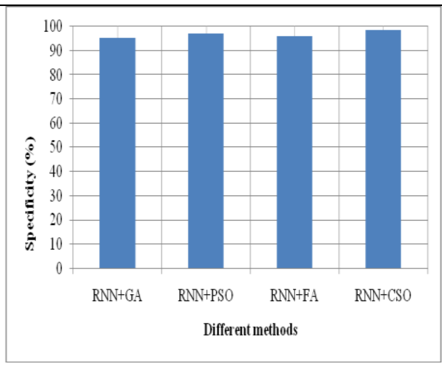


Figure 13: Specificity comparison by varying optimization approach

The performance of proposed approach with different optimization is analysed in terms of accuracy is analysed in figure 11. When analysing figure 11, the proposed method attained the maximum accuracy of 97.3% which is 95.1% for using RNN+GA based NPTEL student feedback prediction, 96.2% for using RNN+PSO based NPTEL student feedback prediction, 95.5% for using RNN+FA based NPTEL student feedback prediction. Moreover, in figure 12, the comparative analysis of proposed against existing is analysed. When analysing figure 12, the proposed method attained the maximum sensitivity of 98.2% which is 96.3% for using RNN+GA, 97.2% for using RNN+PSO, and 96% for using RNN + FA. Similarly, in figure 13, the proposed method attained the maximum specificity compared to existing approach.

Recommendation system :

The main objective of this section is to recommend the efficient teacher based on student facial expression. The above table 1 is given the student satisfaction score.

Teacher / Student	T1		T2		T3		T4	
	Happy	**sad**	**Happy**	**sad**	**Happy**	**Sad**	**Happy**	**Sad**
S1	✓		✓			✓	✓	
S2	✓		✓			✓		✓
S3	✓			✓		✓	✓	
S4	✓			✓	✓		✓	
S5	✓		✓			✓	✓	
S6	✓			✓	✓			✓
Student satisfaction score	**100%**		**50%**		**33.3%**		**66.6%**	

Table 1: Student satisfaction score

The student satisfaction score is analyzed in Table 1. Here, satisfaction level is identified based on student exposure. This study can be helpful in improving the quality of teacher lectures. In this study, a good teacher is recommended for further online classes. When analyzing Table 2, the maximum student satisfaction scores for T1 and the mean score for T4. Teachers T1 and T2 are not recommended for additional classes.

Conclusion:

In this paper, an efficient automatic student feedback prediction based on facial expression has been analysed. Here, mathematical expression of feature level fusion and ORNN classifier has been analysed. Six types of facial expression have been identified. Finally, based on the expression, a teacher has been recommended. The performance of proposed methodology has been analysed in terms of different mathematical expression. The experimental results and comparative analysis has been explained. Through this automatic feedback system we can collect the feedback of students through facial expression. We can classify the facial expression of each student based on time and topic and analyse the students' response of geographical wise. We can also track the online presence of students by calculating the total time and we can provide attendance certificate based on that. It also avoid the malpractices during exams and assignment submission.

References

[1] C. Milligan and A. Little john, “Why Study on a MOOC? The Motives of Students and Professionals", International Review of Research in Open and Distributed Learning, vol. 18, no. 2, April 2017.

[2] R. Firmin, E. Schiorring, J. Whitmer, and . Willet, "Case Study: Using MOOCs for conventional college coursework," Distance Education, vol. 35, no. 2, pp. 178-201, 2014

[3] R.Soundhara Raja Pandian and K.S.Kasiviswanathan " Effective use of CLOUD computing concepts in engineering colleges", 2011 IEEE T4E International Conference pp. 233-236

[4] R. Soundhara Raja Pandian, S. Thangalakshmi, S. Saravanan, " Virtual Learning System: A Conceptual Framework of Network Optimization", Proceedings of the International Conf. on SocProS Springer 2011, AISC 130, pp. 749–755.

[5] P. Dillenbourg, A. Fox, C. Kirchner, J. Mitchell, and M. Wirsing, "Massive open online courses: Current state and perspectives," Dagstuhl Manifestos, vol. 4, no. 1, pp. 1-27, 2014.

[6] M. S. Krishnan, "NPTEL: A programme for free online and open engineering and science education,"in International Workshop on Technology for Education, Bangalore, 2009, pp.1-5.

[7] Warriem, Jayakrishnan Madathil, "NPTEL Local Chapters: Facilitating Mainstreaming of MOOCs in Higher Education", In 2018 IEEE Tenth International Conference on Technology for Education (T4E), pp. 89-92. IEEE, 2018.

[8] Jayasinghe, Udeni, and AjanthaAtukorale, "Tracking Emotions through Facial Expressions in Online Education Systems Based on Transient Emotion Peak", world conference on future of education, 2019

[9] Al-Awni, A. "Mood extraction using facial features to improve learning curves of students in e-learning systems." International Journal of Advanced Computer Science and Applications 7, no. 11 (2016): 444-453.

[10] Krithika, L. B., and Lakshmi Priya GG. "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric." Procedia Computer Science 85 (2016): 767-776.

[11] Liu, Yongna, Lirong Wang, and Wanping Li. "Emotion analysis based on facial expression recognition in virtual learning environment." International Journal of Computer and Communication Engineering 6, no. 1 (2017): 49.

[12] Vasanth, P. C., and K. R. Nataraj. "Facial Expression Recognition Using SVM Classifier." Indonesian Journal of Electrical Engineering and Informatics (IJEEI) 3, no. 1 (2015): 16-20.

[13] Kalsum, Tehmina, Syed Muhammad Anwar, Muhammad Majid, Bilal Khan, and Sahibzada Muhammad Ali. "Emotion recognition from facial expressions using hybrid feature descriptors." IET Image Processing 12, no. 6 (2018): 1004-1012.

[14] EbrahimiKahou, Samira, Vincent Michalski, Kishore Konda, Roland Memisevic, and Christopher Pal. "Recurrent neural networks for emotion recognition in video." In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 467-474. 2015.

[15] Chu, Shu-Chuan, Pei-Wei Tsai, and Jeng-Shyang Pan. "Cat swarm optimization." In Pacific Rim international conference on artificial intelligence, pp. 854-858. Springer, Berlin, Heidelberg, 2006.