FACIAL EXPRESSION ANALYSIS FOR ACADEMIC ENGAGEMENT MONITORING WITH KRIGING GEOMETRY AND REGION PROPOSAL NETWORK DEEP TRANSFER LEARNING

NOORA C.T, P.TAMIL SELVAN

Department Of Computer Science Karpagam Academy Of Higher Education, Coimbatore, India

22drcs001@kahedu.edu.in ,*tamilselvancs@kahedu.edu.in*

ABSTRACT

Over the past few years, contemporary educational movements have made it feasible to apprehend students' facial expressions during erudition to identify learners' poignant status. A growing insistence has been exhibited on utilizing deep learning and the Internet of Things (IoT) to algorithmically identify and elucidate facial changes in various settings owing to technological advancements. Analyzing facial expressions desires to recognize human emotions by exploring visual face information. , This work proposes a method called Kriging Face Geometry and Region Proposal Network-based Deep Transfer Learning (KFG-RPNDTL) to gauge students' level of involvement in the classroom by analyzing their faces and gestures. First, Face Geometry Point Extractor (i.e., face-related feature extraction) is performed to obtain or acquire face-related feature points in an extensive manner. The proposed KFG-RPNDTL method is based on the Euclidean Polar Coordinate Distance Circumplex two-dimensional models of emotions, and it uses the kriging predictor of Best Linear Unbiased Predictors, which minimizes the prediction error. The classification problem related to academic engagement monitoring by analyzing facial expressions has been formulated and solved. The relationship of different emotions is evaluated by plotting different emotions as the points on the plane. The objective is to arrive at an estimate of picture emotion on the plane by kriging and determining which emotion is identified as the closest one. Six basic emotions (Boredom, Confusion, Drowsiness, Engaged, Frustration and Neutral) have been selected. The proposed KFG-RPNDTL method recommends that the Deep Transfer Learning basis of a multimodal scheme precisely find out student academic rendezvous. Experimental research attained an accuracy of 98.85% as well as demonstrated which student academic engagement method is considerably better than existing methods in various metrics through minimal error rate.

***Keywords**: Internet Of Things, Facial Expression, Deep Transfer Learning, Kriging Face Geometry, Region Proposal Network, Student Engagement Management*

1. INTRODUCTION

Digital data traces from different sources covering divergent facets of student life are stockpiled daily in most contemporary university premises. Nevertheless, it remains demanding to integrate these data to acquire a comprehensive perspective of a student, exploit the data to predict academic performance accurately and utilize such prediction results to contribute to positive student engagement. In the case of educational research, the analyzed students' states can be used by faculty members as feedback to refine their teaching approach and mechanism so that the learning rate of the students can be boosted.

A hybrid of ResNet50, CBAM and TCNs was proposed in [1], exploiting the facial expression recognition mechanism to keep an eye on the advancement of students within the classroom environment. Using this integrated mechanism improved robustness by acquiring temporal dynamics and boosting feature relevance that, in turn, ensured valuable understandings for educators to strengthen both teaching procedures and student results with improvements in precision, recall, and accuracy. However, different emotions like, boredom, engagement, frustration, drowsiness, neutral and confusion to gauge students' level of involvement still lack.

A study focusing at students' engagement prediction automatically from continuous video streaming employing ensemble of LSTM was designed in [2]. The objective of an ensemble of LSTM remains to apprehend the efficiency for automatic prediction of engagement from different facial expressions and make in-depth comparisons with the recognized engagement level. Moreover, the engagement tendency in the learning activity using slides and video-lecture were also included to assess student behavior therefore ensuring good level of accuracy. Nevertheless, temporal dimension i.e., analyzing students' performance across time were not analyzed. Emotion is pivotal at inter subjective association, understanding as well as further directions of existence. Comprehending emotions become paramount for humans' everyday working due to the reason that acquisition of emotion as well as incidents are crucial for intercommunication at societal settings. Review of emotional recognition with ML algorithms was reviewed [3]. However, in real time it is found to be highly complicated in analyzing facial factors. To address on this gap, a novel facial expression recognition method was proposed in [4] to acquire emotional changes in timely manner. Yet another dynamic refinement model to monitor students' progress in classroom employing convolutional attention mechanism [5] was presented in [6] therefore improving recognition accuracy. Artificial Intelligence (AI) based method in learning sciences research for accurate collaborative environmental monitoring was proposed in [7]. Engagement level measurement between participants in a meeting is pivotal for estimating collective understanding. Engagement is a paramount element in strengthening a students' learning level. Owing to the significance of engagement in learning efficiency, numerous studies have explored mechanisms in estimating student academic engagement during lectures.

A lightweight deep learning employing CNN was proposed in [8] for accurate emotion analysis. Yet another custom lightweight CNN was applied in [9] and was applied in public datasets for performance analysis. Also by applying this lightweight CNN aided in improvement of overall accuracy. However, the samples involved in analyzing expression was not considered. To mention on this problem, a four layer CNN was designed [10]. Here by employing four layers easily fit the model and sensed emotions in an accurate manner. Despite improvement in accuracy the optimality was not arrived at. To center on this feature, deep learning basis of method using MobilNetV2 [11] was designed that not only improved accuracy but also aided in achieving optimality in an extensive manner.

To sum up the discussion, in spite of different methods were involved in analyzing engagement monitoring in classroom, these methods lacks in analyzing different emotion classes with minimal training time and prediction error owing to distinct geometry patterns involved. It is also inferred that the materials and methods that employed deep learning algorithms struggles in attaining improved precision and recall rate. This motivated to perform research work by proposing a method called, Kriging Face Geometry and Region Proposal Network based Deep Transfer Learning (KFG-RPNDTL) for academic engagement monitoring. This method efficiently addresses the accurate and precise academic engagement monitoring with different emotion classes from the raw sample face images and make easier improved detection rate. The study takes into consideration the MAAED dataset to access the performance of proposed method in monitoring academic engagement. Initially, the raw facial input images obtained from MAAED dataset is applied with Euclidean Polar Coordinate Distance to extract face related features. Then, region of interest is arrived at using Intersection over Union-based Region Proposal Network model that has the potentiality of detecting ROI using aspect ratio at different time intervals. Further, the face related features along with the ROI are transferred to deep learning employing Kriging Best Linear Unbiased Predictor for academic engagement monitoring. Finally, by involving dissimilar performance parameters, result of KFG-RPNDTL technique is analyzed. Moreover, to validate efficiency, the present method, KFG-RPNDTL is compared with other existing methods.

1.1 Contributions Of Paper

- To improve the academic engagement monitoring accuracy and, therefore, the overall prediction rate of the students involved in learning, the KFG-RPNDTL are developed based on three major processes: facial, feature extraction, obtaining region of interest and transferring the results to the prediction model.
- First, Euclidean Polar Coordinate Distance based face related feature extraction model is proposed in the KFG-RPNDTL method for extracting the face-related features using geometric information via the

Euclidean Polar Coordinate Distance functions, therefore improving the overall precision and recall involved in the analysis of academic engagement monitoring using six different emotion classes of students in the classroom.

- To improve accuracy and minimise training time, generate object proposals or regions of interest (ROI) employing Intersection over Union-based Region Proposal Network.
- To reduce the prediction error involved, the KFG-RPNDTL method uses the Kriging Best Linear Unbiased Predictor, with which the facial features extracted and region of interest obtained are transferred for analyzing academic engagement monitoring.
- Lastly, complete experimental evaluation is performed through different performance metrics to exemplify performance enhancement of the proposed KFG-RPNDTL method over traditional methods in accuracy, training time and prediction error.

1.2 Paper Organization

Remainder of manuscript is structured as follows. Section 2 appraisals the literature reviews in area of academic engagement monitoring employing ML and DL methods. Section 3 gives brief explanation of KFG-RPNDTL method with diagrammatical demonstration. Section 4 explains experimental setup followed by performance metrics in Section 5. Section 6 provides with a detailed analysis of discussion and at last, Section 7 summarizes the manuscript.

2. LITERATURE REVIEW

The current technological and technical advancements in the education domain have opened up exciting possibilities. One such prospect is the use of facial expression recognition to gauge students' emotional states during learning. This focus on facial expressions can provide a deeper understanding of the emotional basis of learning, empowering educational professionals to develop more responsive instructional offerings on a larger scale. The potential of this technology to enhance student learning and academic performance is significant and should not be overlooked.

In [12], facial expressions in an educational setting modelled a detailed ethical consideration. A method to focus on the hyperparameter tuning using an optimization technique was designed in [13] with improved face detection and minimal error. Deep learning technique using CNN for facial expression recognition was presented in [14] to focus on the accuracy aspect. Despite improvements in precision and accuracy, the computational complexity involved was not focused. To address on this aspect, Deep CNN using Galactic Swarm Optimization was proposed in [15]. Using this optimization technique, an optimized hyperparameter was employed for detection, reducing the overall computational complexity. An in-depth academic performance prediction employing LSTM as well as ML basis of the classifier was employed in [16] that, in turn, predicted student performance prediction with high accuracy.

A systematic review of facial expression recognition employing machine learning techniques in educational mining was investigated in [17]. A complete study was performed in [18] to examine feeling of 67 students while providing a lecture on essential information technologies. However due to the presence of occlusions emotion detection is still considered a significant task in analyzing the student academic performance. Only with harmonious mutual learning students' attentiveness can be improved that in turn would assist in improving the overall academic performance. In [19], the spatial attention module and convolutional block attention module were presented to perform class semantic interaction. Students' behavior through e-learning session was designed in [20] using deep learning technique, improving overall accuracy extensively.

A holistic review of student engagement and classroom self-efficacy for engagement monitoring was investigated in [21]. Yet another deep learning framework involving HopeNet for monitoring audience engagement was presented in [22]. A state-of-the-art method for analyzing automated facial expression recognition using CNN was proposed in [23]. Deep learning insights and visual narratives involving engagement and emotion were designed in [24].DL has extensive application in emotion detection, but owing to inadequate training information, pre-trained methods are constrained. To address this gap, transfer learning carried out through fine-tuning pr-etrained methods was proposed in [25] for improved accuracy.

Recent advancements in Facial Emotion Recognition (FER) reflect various methodologies and the persistent challenges within the field. Desai et al[26] developed a transfer learning framework that effectively leverages pre-trained Convolutional Neural Networks (CNNs) for identifying basic emotions; however, it falls short in capturing nuanced emotional expressions and lacks cultural and demographic diversity, which may hinder its generalizability in varied contexts. Similarly, Betageri and Yelamali [27] utilized the FER2013 dataset, criticized for its limited diversity and focus solely on facial expressions, neglecting other nonverbal cues such as body language and vocal tone, thus narrowing its applicability to real-world scenarios.Akinduyite et al [28] introduced a Facial Emotion Recognition System tailored for educational settings, emphasizing the importance of emotional analysis to adapt teaching strategies. However, this approach faces challenges related to real-time processing capabilities and the limited range of emotions, primarily focusing on seven basic expressions. Goel et al. [29] addressed the absence of real-time emotional feedback in online education, developing a system that recognizes students' emotional states to enhance engagement and learning outcomes. Yet, they acknowledge technological limitations and privacy concerns that could impede widespread adoption.

Shan and Eliyas[30] further advanced subtle emotion detection using De-expression Residue Learning, employing Conditional Generative Adversarial Networks (cGANs) to analyze students' emotional states. While their model demonstrated improved performance across multiple datasets, it faced limitations concerning computational complexity and the need for diverse training data. Bakyt et al. [31] also utilized CNNs for student emotion analysis, enhancing accuracy through techniques like data augmentation; however, their reliance on curated datasets raised concerns regarding the representation of various emotional expressions and cultural variations.Yuan et al. [32] proposed a Vision Transformer equipped with a hybrid local attention mechanism to tackle occlusion, and head poses variability, demonstrating impressive results but not fully addressing the challenges associated with the size of the datasets required for effective training. Xing and Wang [33] introduced a double-branch fusion learning network to improve accuracy by integrating both shallow and deep features; however, they noted that the model's computational complexity could limit its real-time applications, and its performance may vary across different datasets.Qiancheng et al. [34] presented DDFNet-A, an attention-based dual-branch feature decomposition fusion network aimed at enhancing the quality of fused infrared and visible images, filling significant gaps in existing image fusion methodologies but facing challenges in extreme lighting conditions and dynamic scenes. Finally, Yan-Ping et al. [35] developed a dual-branch structure that combines enhanced relation-aware attention with a cross-feature fusion transformer, effectively addressing inter-class similarity, intra-class differences, and facial occlusion issues. Despite demonstrating significant performance improvements over traditional methods, further work is needed to optimize real-time performance and broaden its applicability across various contexts.

Overall, these studies collectively underscore the pressing need for more inclusive datasets, improved model efficiency, and the enhancement of real-world applicability to advance the effectiveness of FER systems, ultimately paving the way for more sophisticated emotional recognition technologies across diverse applications.Table 1, given below, lists the overview of existing studies.

Table 1 Overview on existing studies

Ref	Work	Methodology	Advantages	Drawbacks	Dataset used
[5]	Facial expression recognition using deep learning	ResNet-50	Accuracy	Training time	FER2013 expression datasets
[6]	Facial expression recognition using online collaborative learning	Online collaborative learning	Precision, recall and accuracy	Computational complexity	FER dataset
[7]	Facial expression recognition using hybrid CNN and ConvLSTM	CNN and ConvLSTM	Accuracy	Training time	SAVEE, CK +, and AFEW
[8]	Light weight deep learning for facial emotion recognition	Light weight CNN	Accuracy	Prediction error	Indian Spontaneous Expression Dataset (ISED)

[9]	Facial emotion recognition through custom light weight CNN	Custom lightweight CNN	Accuracy	Training time and error	RAF-DB dataset
[13]	Optimal deep CNN based emotion recognition	ODCNN-FDER method	Accuracy, precision, recall	Lack optimization principle	FER2013 expression datasets

3. KRIGING FACE GEOMETRY AND REGION PROPOSAL NETWORK BASED DEEP TRANSFER LEARNING (KFG-RPNDTL)

Studies in Facial Expression Analysis for Academic Engagement Monitoring have suffered from the emotions obtained from multiple faces in a single frame for training deep CNN methods that contain outcome at overfitting. To work around this issue, deep transfer learning extensively utilized for facial expression analysis towards academic engagement monitoring. On the contrary, our research gear additional demanding transfer learning to analyze a classroom video, analyze emotions of multiple faces in a single frame and determine group engagement of a class that form of DTL. Specially, we gear issue of learning demonstration in source domain for the task of facial expression using Face Geometry Point Extractor-based face related feature extraction and transferring this learned representation via Intersection over Union-based Region Proposal Network to the task of academic engagement monitoring in a controlled environment. Finally, the proposed method adapted pre-trained network to novel domain through aligning aspect depiction using Kriging predictor function of the source and target domains towards academic engagement monitoring. Figure 1 show the proposed Kriging Face Geometry and Region Proposal Network based Deep Transfer Learning (KFG-RPNDTL) to analyze facial expressions for academic engagement monitoring.

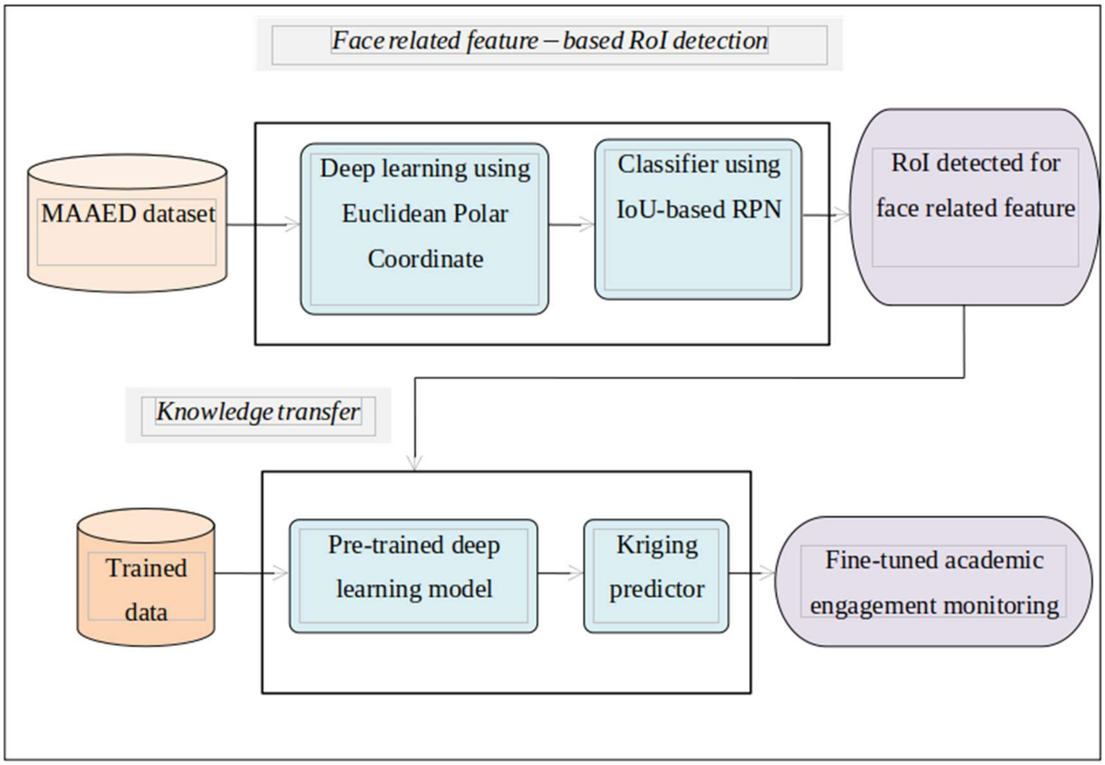

Face related feature – based RoI detection
MAAED dataset
Deep learning using Euclidean Polar Coordinate
Classifier using IoU-based RPN
RoI detected for face related feature
Knowledge transfer
Trained data
Pre-trained deep learning model
Kriging predictor
Fine-tuned academic engagement monitoring

Figure 1 Block diagram of Kriging Face Geometry and Region Proposal Network based Deep Transfer Learning (KFG-RPNDTL)

As illustrated in the above figure, the steps involved in the design of the KFG-RPNDTL method to analyze facial expressions for academic engagement monitoring .initially, the facial expression is analyzed employing the facial landmark technique (i.e., pre-trained model) via Euclidean Polar Coordinate Distance function. The KFG-RPNDTL method is trained and tested on the MAAED datasets for facial expression analysis towards academic engagement monitoring. The method to analyze facial expressions for academic engagement monitoring utilizes the knowledge learned from one task (i.e., Euclidean Polar Coordinate Distance based face related feature extraction and Intersection over Union-based Region Proposal Network [Region of interest detection]) to enhance the result of one more connected task (i.e., academic engagement monitoring).

The input image is fed into the Euclidean Polar Coordinate Distance using Face Geometry called Euclidean Polar Coordinate distance-based related feature extraction to extract face-related features. The Face Point Extractor, using Face Geometry, provides aspect maps at the input image. For example, filters study the outline of student countenance for personal face images, which subsists during training.

The final layer of the pre-trained method is followed by an Intersection over the Union-based Region Proposal Network to every point of facet maps to generate object proposals. It performed by locating '$M - anchors$' boxes with respect to 'N' samples at every position of aspect maps at input image.Here, inferior layers of N, namely Region Proposal Network, learn common aspects as superior layers study task-definite features (i.e., automatic engagement monitoring using the Kriging predictor function). Third, fine-tuning is performed by applying the Best Linear Unbiased Predictor by training it on the new test dataset. Finally, the results of the fine-tuned method are employed to forecast novel information. Through fine-tuning weights of superior layers, academic engagement monitoring is performed precisely and accurately with minimal training time and error.

3.1 Dataset Description

The MAAED dataset creation procedure consists of the employment of five different available datasets, namely, the BAUM-1 dataset, the Yawning Detection Dataset (YawDD), the Many Faces of Confusion in the Wild dataset (MFC-Wild) dataset, the Daisee dataset and the FER dataset respectively. Although the MAAED datasets included four distinct emotions, like drowsiness, frustration, confusion and boredom, two more classes, neutral and engaged, incorporated positive emotions alongside negative emotions to provide a comprehensive understanding of students' affective states in the classroom. Negative emotion monitoring in students' affective states is mandatory for mental health. Moreover, by monitoring negative emotions, students in emotional distress can be identified by the educators and impart relevant assistance and arbitrations. In this manner, it can also ensure personalized assistance with creating a positive learning environment that encourages academic success and overall well-being. The dataset description utilized in this work, including sample sizes and emotion classes, is provided in Table 2. With the above set of emotion recognition datasets, an overview of the dataset involving six different classes, namely, boredom, confusion, drowsiness, engaged, frustration and neutral split into training, testing as well as validation set are given Table 3.

Table 2 Emotion recognition datasets for constructing MAAED dataset

Dataset	Number of samples	Emotion classes
YawDD	351	Normal, talking and yawning
Daisee dataset	9068	Boredom, confusion, frustration, engagement
MFC-Wild dataset	1000	Confusion, anger, distrust
FER dataset	35887	Anger, distrust, fear, happiness, sadness, surprise, neutral
BAUM-1 dataset	1519	Happiness, anger, sadness, disgust, fear, surprise, boredom, contempt, confusion, concentration

Table 3 Dataset overview

Engagement	Training	Testing	Validation	Total
Boredom	4179	1394	1393	6966
Confusion	3428	1106	1195	5729
Drowsiness	3623	1121	1108	5852
Engaged	4159	1399	1381	6939
Frustration	3901	1300	1302	6503
Neutral	4116	1904	1904	7924
Total	23406	8224	8283	39913

3.2 *Euclidean Polar Coordinate Distance Based Face Related Feature Extraction*

In this section face related aspects are extorted for acquiring equivalent academic engagement monitoring. Facial features define face shape that comprises of distinct elements like, lips and so on. To extort facial aspects different key-points as of face picture are extorted employing geometric data of facial aspects or simply exploit face geometry. With this model the face related features or face key points are detected via Face Geometry function. It represented as follows.

$$F = \{(F_1, P_1), (F_2, P_2), (F_n, P_n)\} \quad (F_1, F_2, \dots, F_n, P_1, P_2, \dots, P_n)^T \qquad (1)$$

From the above equation (1), 'F_n' represents the facial features and its corresponding feature points 'P_n' respectively. The facial features with '$n = 6$' along with its points are listed in below Table 4.

Table 4 Features with points

S. No	Features	Points
1	Left eye	61
2	Right eye	61
3	Left eyebrow	32
4	Right eyebrow	32
5	Nose	68
6	Mouth	74
7	Lips	48
8	Jaw	94
Total		470

In this manner 470 features points are extracted and provided as input to the proposed KFG-RPNDTL method..In our work the Circumplex model is employed wherein the facial expression analysis can be made as a linear combination of the dimensions of arousal and valence (i.e., boredom, drowsiness, frustration and neutral). However for better reliability, in our work more key-points are detected employing the Euclidean Polar Coordinate Distance function. The polar coordinate system employed here refers to two-dimensional coordinate scheme at that every point on facial feature is computed through distance as of reference point as well as angle from reference direction. Let us assume polar coordinates of 'a' as '(x, α)' and polar coordinates of 'b' as '(y, β)', then, the distance is stated based on the law of cosines as given below.

$$Dis(a, b)[F] = \sqrt{x^2 + y^2 - 2xycos(\alpha - \beta)}[F] \qquad (2)$$

According to the above considerations as given in equation (2) the radial coordinate is represented by 'α' whereas angular coordinate is represented by 'β' respectively. The angle 'β' starts from '0°' at a reference direction as given in the below figure. The figure below shows the symbolic representation of a polar grid representation-based face recognition network. The polar grid is formed with several angles, each denoting the emotions of engagement in a classroom to gauge students' level of emotion. As illustrated in Figure 2 above, with the aid of the polar grid representation equivalent Circumplex-based face recognition network, 68 facial landmark points for analyzing academic engagement monitoring using facial expressions are obtained for further processing.

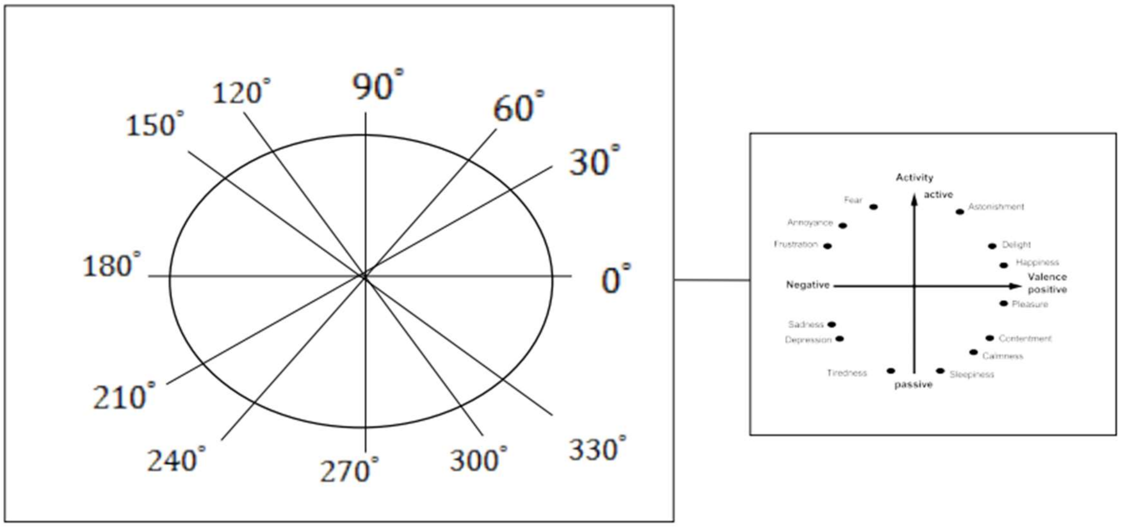


Figure 2 Polar grid representation equivalent Circumplex-based face recognition network

3.3 Intersection over Union-based Region Proposal Network

Followed by which the Intersection over Union-based RPN is utilized to every point of aspect maps with the purpose of generating object proposals or regions of interest (ROI). It performed through locating '$M - anchors$' boxes with respect to 'N' samples at every position of aspect maps at input image.

In current learn value of '$M - anchors$' boxes employed are '6' i.e. '$M = 6$'. By employing Region Proposal Network (RPN) the potential candidates of objects are extracted in the image in an accurate manner. Here, the anchor boxes are generated on the basis of the aspect ratio and image scaling.

From the above equations (6) and (7) 'i' refers to the index of anchor (i.e., 6 different classes) '$Prob_i$' representing the probability of being an object (i.e., parts of face) or not '$Prob_i'$' denoting the target with 'N_{Cl}' and 'N_{Reg}' forming the normalized class and regression values respectively. To be more specific, the Intersection over Union-based Region Proposal Network obtains a group of boxes which categorized as either foreground or background with which foreground anchor boxes moreover fine tune to generate RoI. On one hand aspect ratio refers to the ratio of image width to height and on the other hand scaling denotes the image size. With the aid of the eye coordinate point aspect ratio and scaling are computed. Eye width is represented through computing distance among points 'P_1' as well as 'P_4' (i.e., horizontal line) and on the other hand, eye height is represented through determining distance among middle points 'P_2, P_3' and 'P_5, P_6' respectively. Then, eye aspect ratio to position the anchor boxes across entire feature map is mathematically formulated as given below.

$$EAR = \frac{(P_2 - P_6) + (P_3 - P_5)}{(P_1 - P_4)} \quad (3)$$

Then, average aspect ratio to predict whether the box has an object or not are determined as given below.

$$Avg_{rat} = \frac{EAR(l) + EAR(r)}{2} \quad (4)$$

From the above equation (4) the average aspect ratio for left '$EAR(l)$' as well as right eye

‘$EAR(l)$’ are averaged. Finally, with the above-average aspect ratio, foreground and background labels are differentiated based on Intersection over Union (IoU). The IoU here refers to the ratio of the area of intersection among the anchor box and box through RoI to the area of the union of two boxes. This is mathematically formulated as given below.

$$IoU = \frac{\text{Area of overlap}}{\text{Area of union}} [\text{Avg}_{\text{rat}}] \quad (5)$$

From the above formulates, we can only obtain if the anchor box arrived through IoU in RPN does not let us know whether the parts are either eyes or nose. It only tells us if the part is either foreground or background. Now, finally, the offset thus obtained is applied to get RoI using the overall loss function and is formulated as given below.

$$Loss = Loss_{Cl} + Loss_{Reg}[IoU] \quad (6)$$

$$Loss(Prob_i, T_i) = \frac{1}{N_{Cl}} \sum Loss_{Cl} (Prob_i, \\ + \lambda \frac{1}{N_{Reg}} \sum Loss_{Reg} (Prob_i, Prob_i' \quad (7)$$

Structure of Intersection over Union-based Region Proposal Network is given in Figure 3

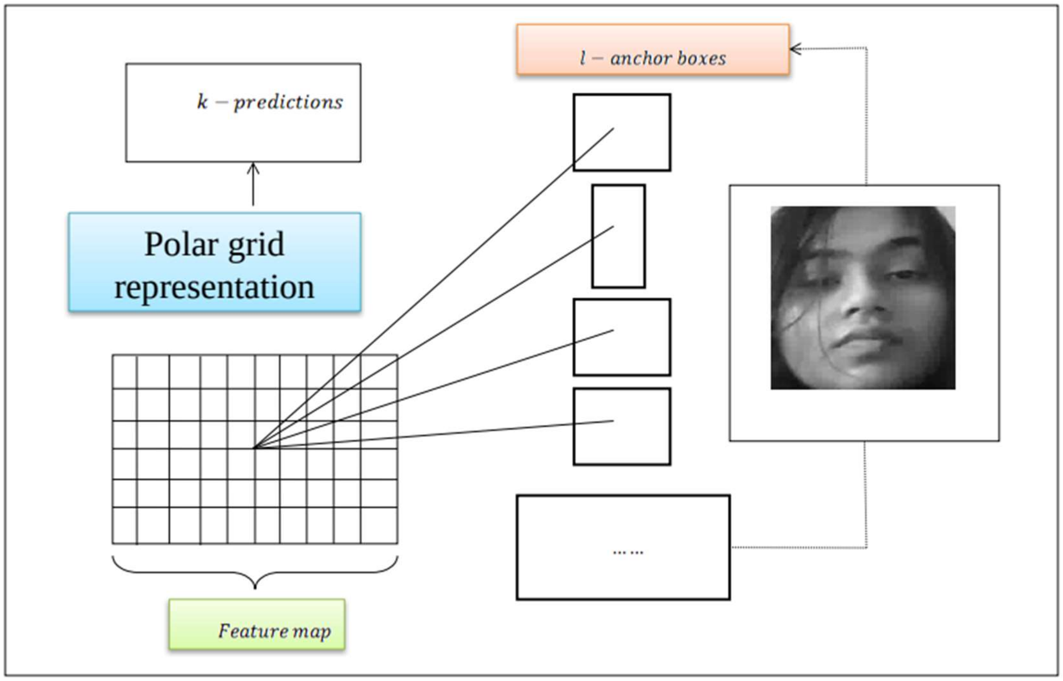


Figure 3 Structure of Intersection over Union-based Region Proposal Network

3.4 Kriging Best Linear Unbiased Predictor-Based Academic Engagement Monitoring

Let us consider that the analyzed dataset ‘ $F = (F_1, F_2, \dots, F_N)$ ’ comprises of ‘ $Ndimensional$ ’ data points ‘ $F_i = (F_{i1}, F_{i2}, \dots, F_{\square})$ ’ where ‘ $i = 1,2, \dots, N, F_i \in R^n$’. The data point or the facial feature point ‘F_i’ corresponds to the ‘ith’ picture in the facial picture set. The Kriging Predictor- these pictures display seven classes (Boredom, Confusion, Drowsiness, Engaged, Frustration and Neutral). Since the two-dimensional Polar grid representation equivalent Circumplex-based face recognition network is used for academic engagement monitoring by analyzing facial expressions, every emotion class is represented as a point possessing dual coordinates: valence and arousal.

Let us assume that the set ‘ $F = (F_1, F_2, \dots, F_N)$’ of mutually disjoint vectors ‘$F_i \in R^N$’, where each vector denotes one facial picture of student in the classroom and data of measurement '$Res = (Res_1, Res_2, \dots, Res_N)^T$' representing the emotions at facial feature points of 'F' is known, '$Res_i = Kf(F_i, \omega)$'. The matrix of fractional Euclidean Polar Coordinate Distance is computed as given below.

$$DV = Dis(a,b)F \tag{8}$$

Owing to the reason that the picture emotion is known priori, each facial feature point 'F_i' is analogous to an emotion point '$Res_i = (Res_{i1}, Res_{i2})$' that give details of the '$i-th$' facial feature picture emotion. Six different combinations of 'Res_{i1}, Res_{i2}' are obtained. To be more specific, 'Res_{i1}' and 'Res_{i2}' refers to the valence and arousal coordinates, respectively, of the '$i-th$' facial feature picture emotion in two-dimensional circumplex emotion space. Then, for the entire region of interest detected facial features of students in a classroom 'F', two column vectors '$Res_1 \wedge Res_2$', column vector 'Res_1' consists of the valence coordinates of the facial feature picture emotion points '$Res_i, i = 1,2, \dots, N$' and the column vector 'Res_2' consist of the arousal coordinates of these facial feature picture emotion points, i.e. '$Res_1 = (Res_{11}, Res_{21}, \dots, Res_{N1})^T$' and '$Res_2 = (Res_{12}, Res_{22}, \dots, Res_{N2})^T$', respectively.

Finally, with the obtained face-related features extracted and fine-tuned regions of interest, the Kriging predictor function is applied to analyze both classroom video and the emotions of multiple faces in a single frame to determine group engagement in a class. The advantage of using the Kriging predictor function to analyze facial expressions for academic engagement monitoring is that it utilizes a restricted set of sampled data points to evaluate the variable value over a continuous spatial and temporal field.

Here, the temporal field denotes the students' affective state in the classroom, which is obtained by considering the sample frames, both single and multi frames, at different time intervals. To be more specific, the kriging weights are evaluated in such a manner so that points nearby to the region of interest are given more weight than those. The Kriging predictor is an optimal linear predictor, and each interpolated value is measured so that the prediction error is said to be reduced.

Then interpolated values with BLUPs are formulated as given below.

$$Kf(F) = Res^T DV^{-1} \left(DV + uCV \left(\frac{1 - uCV^T DV^{-1} DV}{uCV^T DV^{-1} uCV} \right) \right) \tag{9}$$

From the above equation the kriging predictor function for the corresponding facial feature points '$Kf(F)$' results are arrived at by utilizing the distance vector 'DV' denoting the distance between testing data points and training data points, unit column vector 'uCV', of size '$[N * 1]$' respectively. The weights derived from the covariance structure are used in interpolating values for un-sampled points across spatial and temporal domains. Then, the covariance matrix for kriging prediction for academic engagement monitoring is mathematically represented below.

$$\beta(F) = \beta \left(DV^T DV^{-1} DV - \frac{(1 - uCV^T DV^{-1} DV)^2}{uCV^T DV^{-1} uCV} \right) \tag{10}$$

Finally kriging prediction employing Best Linear Unbiased Predictor of a testing facial image picture emotion is performed by utilizing a posteriori expectation as given below.

$$Kf_1 = Res_1^T DV^{-1} \left(DV + uCV^{-1} \frac{1 - uCV^T DV^{-1} DV}{uCV^T DV^{-1} DV} \right) \tag{11}$$

$$Kf_2 = Res_2^T DV^{-1} \left(DV + uCV^{-1} \frac{1 - uCV^T DV^{-1} DV}{uCV^T DV^{-1} DV} \right) \tag{12}$$

From the above equations (11) and (12), 'DV^{-1}' denotes the inverse matrix of distance vector 'DV' with 'uCV' representing the unit column vector and 'DV' representing the distance vector, i.e. distances between testing data point (i.e. whose emotion is to be predicted) and all the training data points (i.e. whose emotions are known in advance). On the other hand, outputs 'Kf_1' (i.e., valence) and 'Kf_2'

(i.e., arousal) refer to the first and second prediction factors, in the two-dimensional circumplex space model. The pseudo code representation of Kriging Face Geometry and Region Proposal Network based Deep Transfer Learning (KFG-RPNDTL) for gauging students' level of involvement in classroom using facial expressions is provided below.

```
Input: Dataset '$DS$', Class '$Cl = \{Cl_1, Cl_2, \dots, Cl_M\}$', Samples '$S = \{S_1, S_2, \dots, S_N\}$'
Output: Precise and accurate academic engagement monitoring
Step 1: Initialize '$M = 6$', '$N = 39913$'
Step 2: Begin
Step 3: For each Dataset '$DS$' with Samples '$S$'
    //Initialize pre-trained model on a large dataset
    Step 4: Obtain facial features and its corresponding feature points as given in equation (1)

    //Replace final layer(s) of pre-trained model with new, untrained layers
    // First step: Euclidean Polar Coordinate Distance based face related feature extraction (i.e., extracting face related feature)
    Step 5: Evaluate distance based on the law of cosines as given in equation (2)
    Step 6: Return (i.e., extract) face related features '$Dis(a,b)[F]$'

    //Second step: Intersection over Union-based Region Proposal Network (i.e., region of interest detection)
    //Classify as either foreground or background
    Step 7: Evaluate aspect ratio as given in equation (3)
    Step 8: Evaluate average aspect ratio as given in equation (4)
    Step 9: Evaluate of Intersection over Union (IoU) as given in equation (5)

    //Fine tune to generate the regions of interest
    Step 10: Formulate loss function as given in equations (6) and (7)
    Step 11: Return ROI '$IoU$' with minimal loss

    //Third step: Predictor-based Academic Engagement Monitoring
    Step 12: Formulate fractional Euclidean Polar Coordinate Distance for each facial feature points as given in equation (8)
    Step 13: Evaluation kriging predictor function with the Best Linear Unbiased Predictors (BLUPs) as given in equation (9)
    Step 14: Obtain covariance matrix for kriging prediction for academic engagement monitoring as given in equation (10)
    Step 15: Evaluate posterior expectation resultant values as given in equations (11) and (12)
    Step 16: Return emotions of multiple faces in a single frame and group engagement of class
Step 17: End for
Step 18: End
```

Algorithm 1 Kriging Face Geometry and Region Proposal Network based Deep Transfer Learning

As given in the above algorithm, the overall facial expression analysis for academic engagement monitoring is split into four steps. First, the pre-trained model using Euclidean Polar Coordinate Distance face-related feature extraction (i.e., extracting face-related features) is initialized with weights using the samples from the raw MAAED dataset. Second, the prê-trained models are replaced using Intersection over Union-based Region Proposal Network (i.e., region of interest detection). Third, the weights of pre-trained models are trained by utilizing the Kriging predictor function through the Best Linear Unbiased Predictors. The above three-step process is repeated to analyze the emotions of multiple faces in a single frame and also group engagement of class both spatially and temporally.

4. EXPERIMENTAL SETUP

The results produced by the computation of proposed Kriging Face Geometry and Region Proposal Network-based Deep Transfer Learning (KFG-RPNDTL) for academic engagement monitoring are provided in this section. In addition to exploratory data analysis as mentioned above, performance metrics and experimental results of the proposed KFG-RPNDTL method are included in this section. The process is executed in Python. To project the efficiency of the present method, a comparison of existing methods, ResNet50, CBAM, and TCNs [1], and the Ensemble of LSTM neural networks [2] with the proposed KFG-RPNDTL method are also provided.

Table 5 Python high-level general-purpose programming language Requirements

Software requirements	**Hardware requirements**
O/S: Windows 10 and above	System: Pentium I3 processor
Language: Python 3.12	Hard disk: 512 GB
	Mouse: Logitech
	Keyboard: 110 keys enhanced
	RAM: 4GB

5. EVALUATION METRICS

By analyzing various parameters, the efficiency of the proposed Kriging Face Geometry and Region Proposal Network-based Deep Transfer Learning (KFG-RPNDTL) for acquiring students' level of involvement in the classroom by analyzing facial expressions are evaluated in this section. Precision, a key metric in interpreting facial expressions for monitoring academic engagement, is the first parameter. It gives the number of facial expressions correctly identified among the overall facial expressions, as follows.

$$Pre = \frac{TP}{TP + FP} \tag{13}$$

From equation (13), precision rate 'Pre' is calculated employing true positive rate 'TP' (i.e., sample facial expressions with boredom retrieved as boredom) and false positive rate 'FP' (i.e., sample facial expressions with neutral retrieved as boredom) respectively. Recall rate on the other hand, gives the ratio of facial expressions correctly identified among the total sample facial expressions provided as input, which is represented as follows.

$$Rec = \frac{TP}{TP + FN} \tag{14}$$

From equation (14), recall rate 'Rec' is estimated by 'TP' (i.e., facial samples with neutral retrieved as neutral) and 'FN' (i.e., facial samples with neutral as boredom) respectively. Next, the accuracy involved for academic engagement management using facial expression is measured. The accuracy rate is formulated as follows.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

From equation (15) accuracy 'Acc' is calculated. It is computed in percentage (%). Fourth, the training time utilised to measure academic engagement monitoring is discussed. To be more specific, training time refers to the time used to extract precise facial features along with the region of interest. The lower the training time involved in extracting accurate and precise facial features along with the area of interest, the more significant the method is said to be because irrelevant features can be discarded, and only relevant facial features can be used for further monitoring of academic engagement, therefore ensuring earlier remedial actions. The training time is expressed as below.

$$TT = \sum_{i=1}^{n} S_i * Time(IoU) \tag{16}$$

From equation (16) training time 'TT' is measured by considering the face samples involved in the monitoring of academic engagement of students' in classroom 'S_i' and time utilized ay extracting region of interest '$Time(IoU)$'. It is calculated in milliseconds (ms). Finally prediction error is measured being samples provided for simulation as well as samples predicted correctly.

$$PE = \sum_{i=1}^{N} \frac{S_{\wp}}{S_i} \tag{17}$$

From equation (17) prediction error 'PE' is measured by employing the actual facial sample images provided as input 'S_i' and the samples wrongly predicted '$S_{\wp}$'. It is calculated in percentage (%). Precision is a performance metrics that evaluates the ratio of TP forecast between every positive forecasts made through academic engagement monitoring process. It is evaluated as follows.

$$Pre = \frac{TP}{TP + FP} \qquad (18)$$

From equation (18) 'Pre', is evaluated by considering 'TP' and 'FP' into consideration. Recall referred to sensitivity or TP rate, evaluates ratio of TP predictions between every actual positive instances. It is formulated as follows.

$$Rec = \frac{TP}{TP + FN} \qquad (19)$$

From equation (19) recall rate 'Rec', is calculated by considering 'TP' and false negative rate 'FN' into consideration. F1-score is compute of predictive result and is estimated through considering 'Pre and Rec to account.F1-score is mathematically formulated as given below.

$$F_1\text{score} = \frac{2 \cdot Pre \cdot Rec}{Pre + Rec} \qquad (20)$$

From the above equation (20), the F1-score results are arrived at based on the precision 'Pre' and recall 'Rec' rate.

6. RESULTS AND DISCUSSION

The comparative analysis of precision, recall, and accuracy for five different classes achieved by the proposed KGHF-CCDTL method, along with the existing two methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2], are discussed in this section. Figure 4 below shows five different emotion classes for a student in a classroom obtained at various time intervals split into five frames. The figure below shows the proposed method to analyze a classroom video for multiple faces (i.e., for simplicity, two students' facial expressions for academic engagement monitoring). Figure 5 below shows two different emotion classes for a student in a classroom obtained at various time intervals split into five frames.

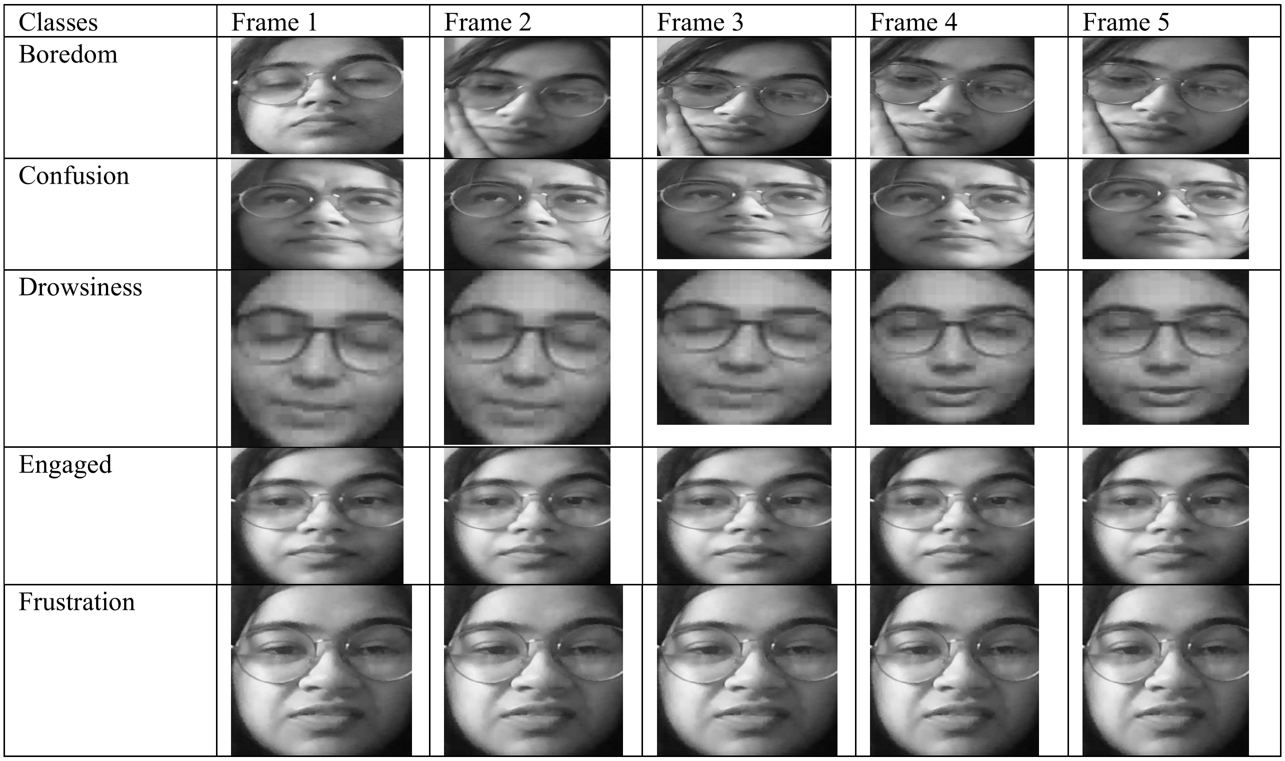


Figure 4 Sample students involved in classroom academic engagement monitoring

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

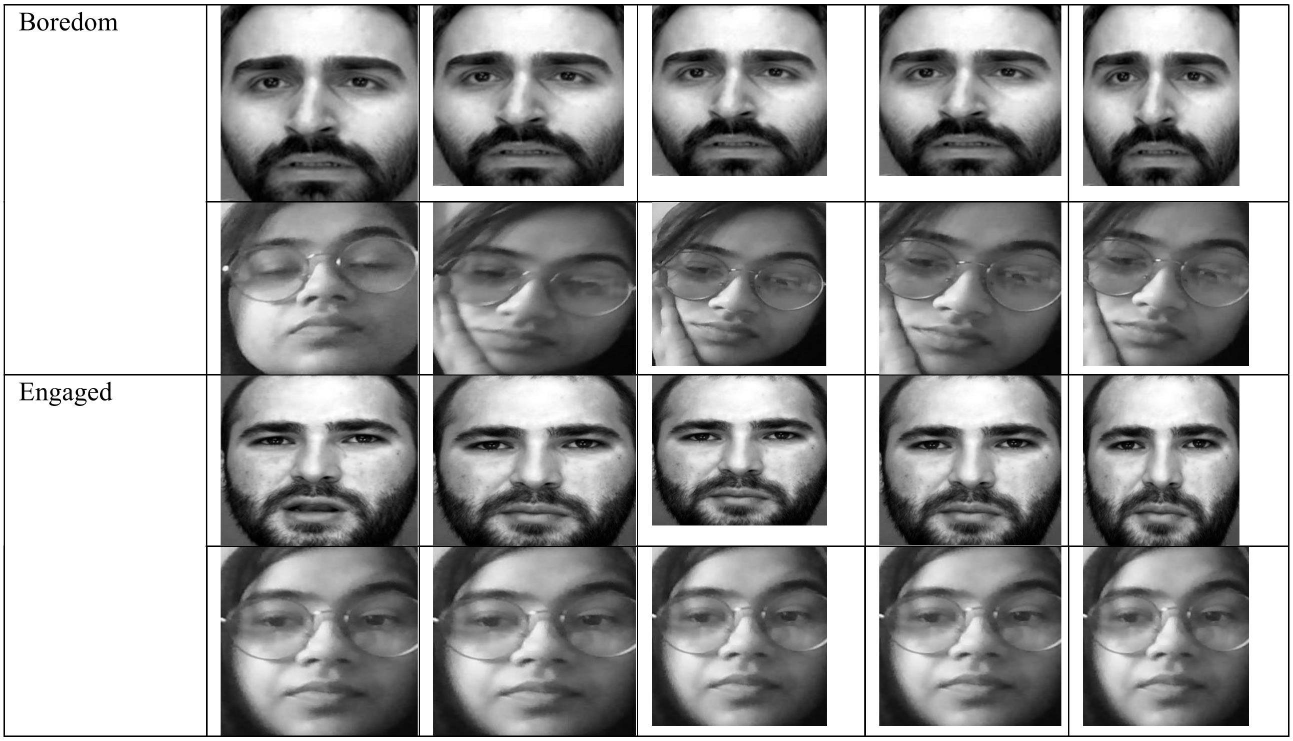


Figure 5 Sample students involved in classroom academic engagement monitoring with two different emotions (i.e. engaged and boredom)

In a similar manner as given above, a classroom video was analyzed for 50 different samples. Also, emotions of multiple faces in a single frame were employed for validation and analysis. Different time intervals were employed to acquire multiple frames for different students' facial expression recognition.

6.1 Case Scenario 1: Precision, Recall And Accuracy Analyses

In this section, first, six classes of emotions, including both positive and negative emotions, are analyzed by making in-depth comparisons between the proposed KFG-RPNDTL method and two existing methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]. Table 6, given below, lists the results of the analysis of precision, recall, and accuracy.

Table 6 Comparative analysis of precision, recall and accuracy for six different engagement categories

Methods	Boredom			Confusion			Frustration			Drowsiness			Neutral			Engaged		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy
KFG-RPNDTL	0.96	0.96	0.97	0.98	0.94	0.97	0.97	0.97	0.97	0.95	0.94	0.95	0.96	0.97	0.98	0.97	0.96	0.97
CBAM-TCN	0.94	0.94	0.95	0.95	0.91	0.94	0.93	0.93	0.93	0.91	0.92	0.93	0.93	0.94	0.95	0.94	0.92	0.95
Ensemble of LSTM neural networks	0.92	0.92	0.93	0.93	0.89	0.90	0.90	0.90	0.89	0.89	0.89	0.90	0.90	0.91	0.92	0.90	0.88	0.91

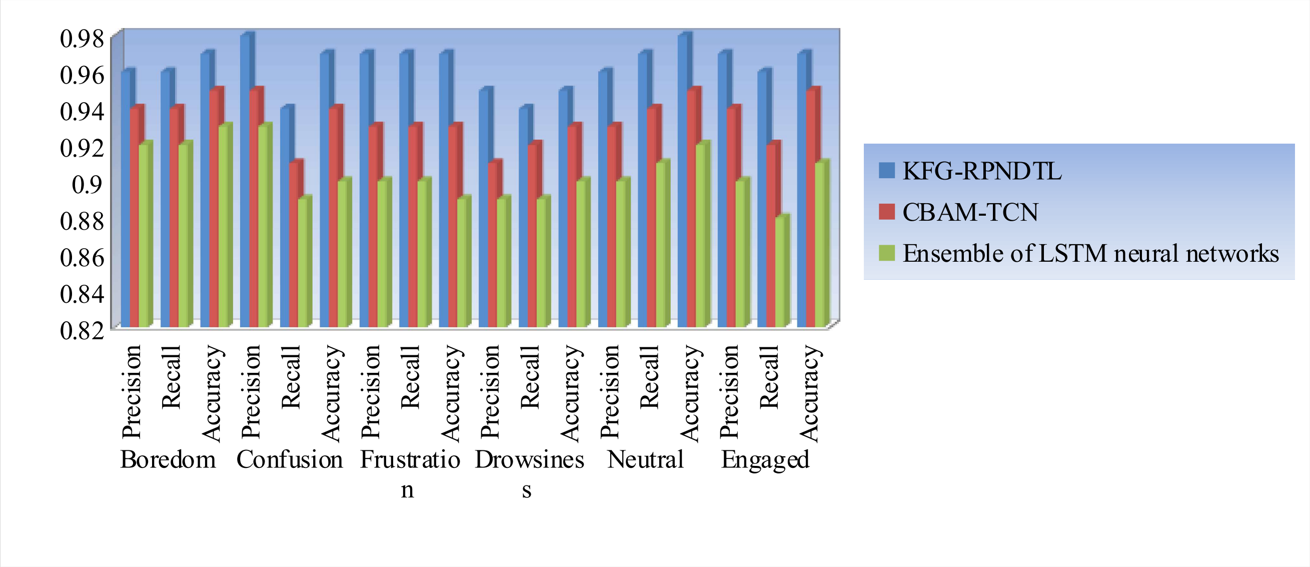


Figure 6 Performance analyses of precision, recall and accuracy using KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]

Figure 6 show graphical representations of Pre, recall and accuracy using KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for six different classes (boredom, confusion, drowsiness, frustration, neutral and engaged). From the above figure, two inferences are made. First, the confusion emotion class of students is highly in the increased state, therefore providing a comprehensive understanding of students' affective states in the classroom. Hence, from this analysis, the teachers can conclude that the covered topic needs to be clearly understood by the students in the classroom while performing academic engagement monitoring. On the other hand, the drowsiness state of activity was less. Second, the precision, recall and accuracy rate of the proposed KFG-RPNDTL method was comparatively higher than the [1], [2]. The cause behind the enhancement was owing to the relevance of the Euclidean Polar Coordinate Distance-based face-related feature extraction model. By applying this model, related features were extracted by employing geometric information about facial features. Moreover, more key points are detected using the Euclidean Polar Coordinate Distance function for better reliability. This, in turn, improved the precision, recall and accuracy of the proposed KFG-RPNDTL method overall with 0.98, 0.94 and 0.97. In contrast, using the existing methods [1] and [2], it was found to be 0.95, 0.91, 0.94 and 0.93, 0.89, 0.90 respectively.

6.2 Case Scenario 1: Precision, Recall And Accuracy Analyses

In this section, first, six classes of emotions, including both the positive and negative emotions (i.e. neutral, engaged, boredom, confusion, frustration and drowsiness), are analyzed by making comparisons between the proposed KFG-RPNDTL method and existing methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]. Table 7, given below, lists the accuracy of the analysis results.

Table 7 Tabulation of accuracy using two classes neutral and engaged for KFG-RPNDTL, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]

Samples	Accuracy (%) – Neutral			Accuracy (%) – Engaged			Accuracy (%) – Boredom			Accuracy (%) – Confusion			Accuracy (%) – Frustration			Accuracy (%) – Drowsiness		
	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]
5	98.8	92	89.1	98.3	94	92	98	95	90	97.9	87.3	82.1	98.1	93.2	90	98.0	94.2	89.0

	5		5	5							5	5	5	5		5	5	5
10	93.15	88.25	80.15	95.35	85.25	80.15	95.15	92.25	87.45	97	86.25	81.35	97.35	92.45	89.25	97.35	93.15	88.25
15	90.45	85.35	78.35	94.15	84.15	79.35	93	90.35	85.35	95	85.45	80.45	96	90	87	95.35	91.25	86.65
20	88.15	83.15	75.25	92	82	77	92.45	89.15	84.25	91	81.35	76.55	95.35	92.15	89.45	93	89	84
25	90.35	85.55	77.15	90.15	80.35	75.25	91	88.35	83.15	90	80.45	75.35	93.15	90.45	87.65	90	86	81
30	92	87.25	79.25	88.35	78.45	73.45	90	87.45	82.45	85	75.35	70.25	90.25	87.55	84.25	88	84	79
35	94.15	89.25	81.35	86	76.25	71.55	88.25	85.15	80.35	82	72.25	67.35	87.55	84.35	81.35	86.35	82.15	77.56
40	95	90.35	82.45	85.15	85.45	80.25	83	80.35	75.25	78	68.35	63.45	90	87	84	85	81	76
45	92.35	87.65	79.55	88	78	73	85.25	82.45	77.35	80	70	65	92.45	89.35	86.25	87	83	78
50	96.15	91.25	83.25	90.35	80.35	75.65	87	83	78	83	73	68	93.55	90.1	87	89	84.45	79.25

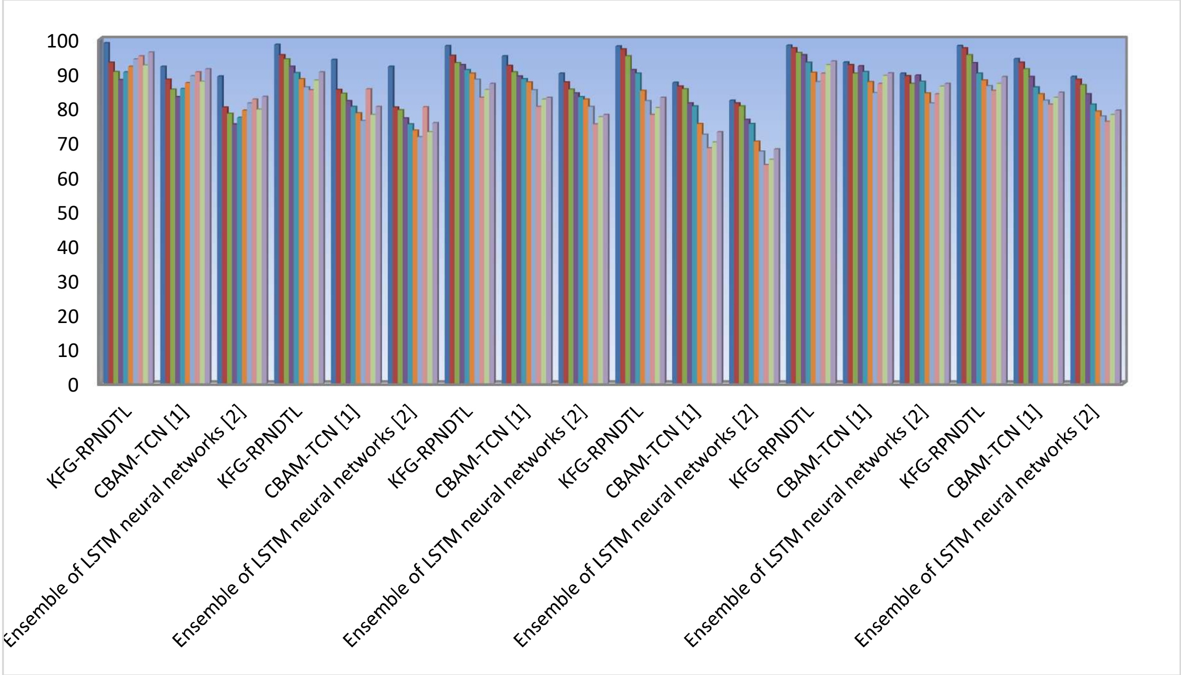


Figure 7 Performance analyses of accuracy using KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for neutral, engaged, boredom, confusion, frustration and drowsiness classes

Figure 7 given above illustrates the accuracy measure in terms of percentage for 50 different samples using the proposed KFG-RPNDTL method and two existing methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2], for boredom, confusion, frustration, drowsiness, engaged and frustration classes. Three inferences are observed from the above results. First, the accuracy aspect for engaged classes during academic engagement monitoring was found to be comparatively better than five different classes, namely, boredom, confusion, frustration, neutral and drowsiness classes for the three methods. From this, it is inferred that in the classroom session, the students are attentive while performing these 50 samples when applied with three methods. Second, the accuracy of the proposed KFG-RPNDTL method was found to be comparatively better than [1] and [2]. Third, increasing the sample size did not influence the accuracy or, to be more specific, increasing the sample size neither increased nor decreased the accuracy rate. The reason for accuracy improvement using the proposed KFG-RPNDTL method was owing to the relevance of Intersection over the Union-basis of Region Proposal Network. By applying this technique, potential candidates of objects were extracted based on the aspect ratio. Next, foreground and background labels are differentiated depending on the IoU. This, in turn, generates the regions of interest accurately, therefore improving the overall accuracy while measuring emotions of engagement in a classroom. This, in turn overall improved of KFG-RPNDTL technique with 98.85%, 92% [1], 89.15% [2] for neutral class, whereas of KFG-RPNDTL technique was found to be 98.35%, 94% and 92% for engaged classes, 98%, 95% and 90% for boredom classes, 97.9%, 87.35% and 82.15% for confusion classes, 98.15%, 93.25% and 90% for frustration classes and 98.05%, 94.25% and 89.05% for drowsiness classes respectively.

6.3 Case Scenario 3: Prediction Error

In this section, six classes of emotions, including both the positive and negative emotions (i.e. boredom, confusion, drowsiness, frustration, engagement and neutral), are analyzed by making comparisons between the proposed KFG-RPNDTL method and existing methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] in prediction error.

Table 8 Tabulation of prediction error using six classes boredom, confusion, drowsiness, frustration, engaged and neutral for KFG-RPNDTL, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]

Samples	Prediction error (%) – Neutral			Prediction error (%) – Engaged			Prediction error (%) – Boredom			Prediction error (%) – Confusion			Prediction error (%) – Frustration			Prediction error (%) – Drowsiness		
	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]	KFG-RPNDTL	CBAM-TCN [1]	Ensemble of LSTM neural networks [2]
5	1.35	1.45	1.6	1.05	1.32	1.65	1.15	1.3	1.45	1.2	1.28	1.35	1.25	1.4	1.55	1.3	1.45	1.6
10	1.55	1.6	1.75	1.35	1.55	1.9	1.28	1.35	1.5	1.28	1.4	1.49	1.33	1.55	1.7	1.35	1.55	1.75
15	1.85	2.15	2.35	1.8	1.95	2.35	1.35	1.42	1.65	1.33	1.45	1.55	1.38	1.65	1.85	1.55	1.8	1.95
20	2	2.35	2.45	2	2.35	2.75	1.45	1.55	1.8	1.4	1.52	1.6	1.45	1.7	1.93	1.7	1.95	2.05
25	2.35	2.55	2.7	2.15	2.55	2.85	1.55	1.68	1.92	1.48	1.58	1.75	1.55	1.85	1.99	1.85	2.05	2.25
30	2.45	2.9	3	2.45	2.9	3.35	1.7	1.75	2.15	1.55	1.6	1.82	1.65	1.93	2.15	1.93	2.15	2.35
35	2	3	3.25	3	3.35	3.85	1.78	1.92	2.25	1.6	1.75	1.95	1.85	2	2.35	2	2.25	2.55
40	1.75	2.75	3	2.65	3	3.35	2	2.35	2.48	1.7	1.82	2.05	2	2.15	2.45	2.05	2.35	2.68
45	1.45	2.45	2.75	2.35	2.75	3	1.75	2	2.25	1.55	1.6	1.85	1.75	1.85	2.25	1.75	2.15	2.55
50	1.55	2.6	3	2	2.55	2.8	1.55	2.25	2.4	1.48	1.76	1.72	1.9	1.98	2.35	1.55	2.35	2.34

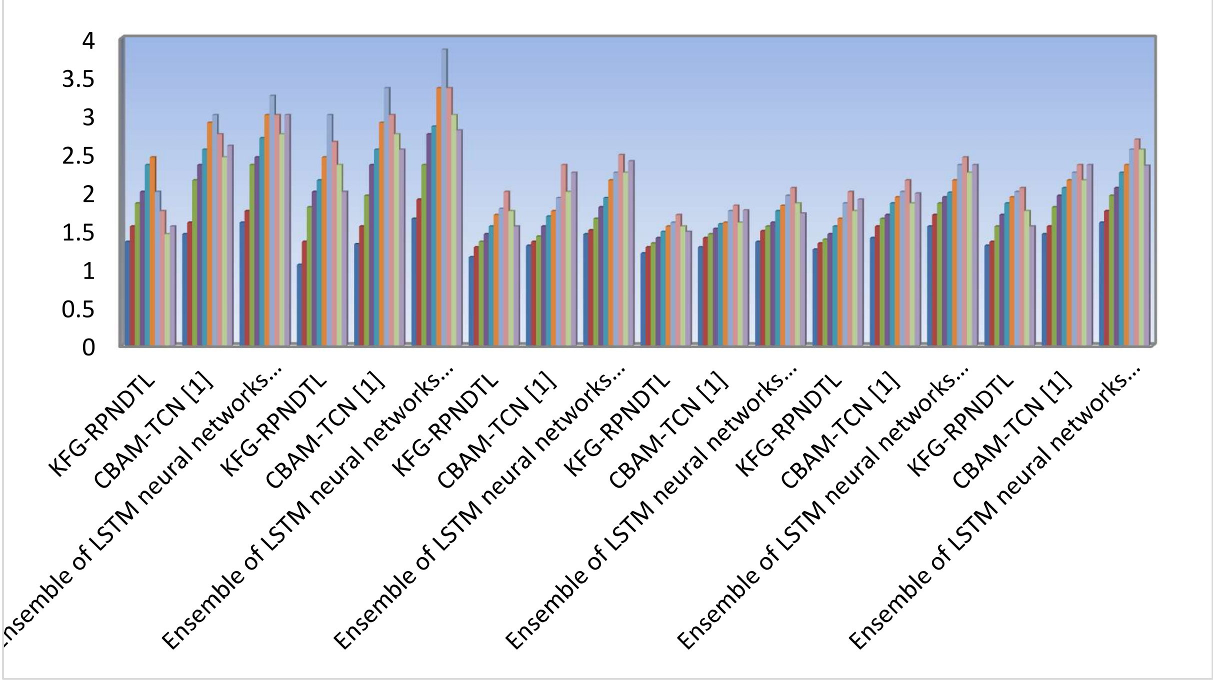


Figure 8 Performance analyses of prediction error using KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for six classes, boredom, confusion, drowsiness, frustration, engaged and neutral

Finally, figure 8, given above, illustrates the graphical representation of prediction error using three methods for six different emotion classes. From the above figure, a set of 50 different samples was applied as a simulation to measure the prediction error using the three methods, KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for academic engagement monitoring with six different classes, boredom, confusion, drowsiness, frustration, engaged and neutral. From the above graphical representation, it is first inferred that the prediction error of the engaged class of emotion is found to be reduced upon comparison to the prediction error of neutral classes. From this result, it is inferred that more focus should be placed on neutral classes of emotion than engaged. Second, neither increase nor decrease is found to be increasing the sample size. Third, the prediction error when applied with the KFG-RPNDTL technique was less than [1], [2]. The cause behind prediction error reduction was the relevance of the Kriging Face Geometry and Region Proposal Network-based Deep Transfer Learning method. Through using this method, a pre-trained model using Euclidean Polar Coordinate Distance based face related feature extraction was initialized with weights. Next, the prê-trained models were replaced by employing Intersection over the Union-based Region Proposal Network. Finally, the weights of pre-trained models were trained using the Kriging predictor function with the Best Linear Unbiased Predictors. This, in turn, aided in minimizing the prediction error using the proposed KFG-RPNDTL method upon comparison to [1] and [2].

6.4.Case Scenario: Precision, Recall And F1-Score

In this section the academic engagement monitoring results are arrived at based on three different parameters, precision, Recall and F1-score. Table 9 lists tabulation of both the positive and negative emotions (i.e. boredom, confusion, drowsiness, frustration, engaged and neutral) are analyzed by making comparisons between proposed KFG-RPNDTL method and existing methods, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] in Pre, $Rec$and F1-score.

Table 9 Tabulation of precision, recall and F1-score using six classes boredom, confusion, drowsiness, frustration, engaged and neutral for KFG-RPNDTL, CBAM-TCN [1] and Ensemble of LSTM neural networks [2]

Samples	Neutral			Engaged			Boredom			– Confusion			Frustration			Drowsiness		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KFG-RPNDTL	0.97	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.96	0.97	0.97
CBAM-TCN [1]	0.93	0.92	0.94	0.94	0.92	0.95	0.93	0.92	0.94	0.94	0.92	0.95	0.94	0.93	0.95	0.92	0.93	0.94
Ensemble of LSTM neural networks [2]	0.90	0.87	0.91	0.89	0.87	0.90	0.91	0.89	0.92	0.90	0.89	0.91	0.90	0.90	0.91	0.87	0.84	0.90

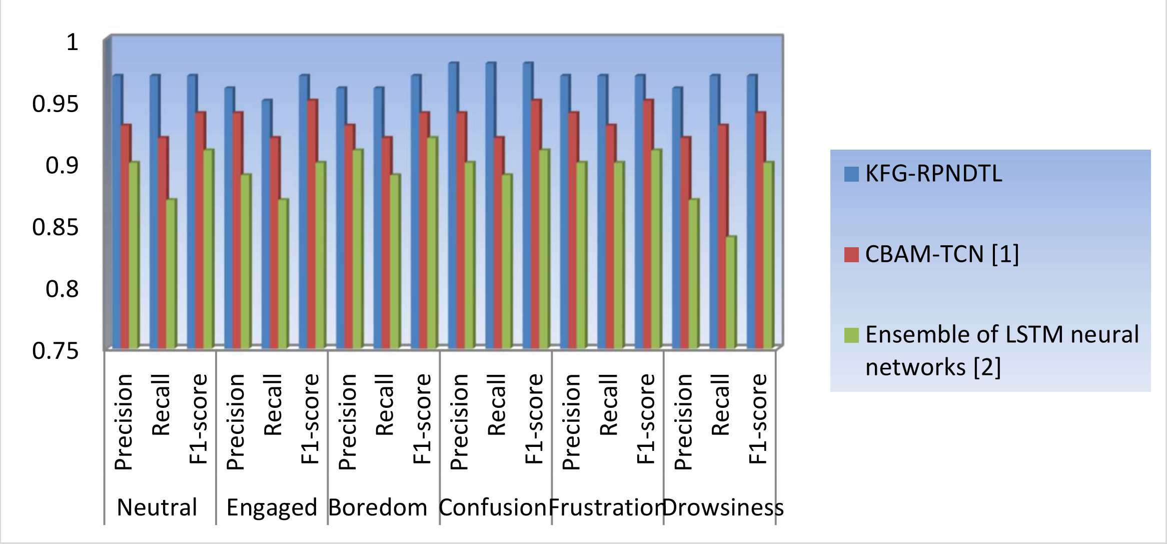


Figure 9 Performance analyses of precision, recall and F1-score using KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for six classes, boredom, confusion, drowsiness, frustration, engaged and neutral

Figure 9 depicts the results of Precision, Recall, and F1-score values using three different methods, the KFG-RPNDTL method, CBAM-TCN [1] and Ensemble of LSTM neural networks [2] for six classes, boredom, confusion, drowsiness, frustration, engaged and neutral. From the above results, two inferences are made. First, three parameters, precision, recall and F1-score, are superior by the KFG-RPNDTL method than the CBAM-TCN [1] and the Ensemble of LSTM neural networks [2]. The second precision, recall and F1-score value of the confused state were found to be higher than the other five classes. From this analysis, it is inferred that the sample instance students considered for simulation are more confused than other aspects. So, more productive results can be achieved by only eradicating the confusion state.

6.5 Case Scenario: Computational And Time Complexity

Time complexity and computational complexity involved in the design of the algorithm are presented. On one hand, computational complexity refers to the amount of resources required to run, and on the other hand, time complexity denotes the computation time required to be performed. Despite deep being complicated and necessitating widespread computational resources as well as elevated time complexity, employing DTL in our work reduces the necessity of elevated computational sources through pre-trained methods with no negotiating a result. Also as the amount of resources necessitated to run the Kriging Face Geometry and Region Proposal Network-based Deep Transfer Learning algorithm differs with the size of the sample input acquired for simulation, the complexity (i.e. computational complexity) is expressed as a function with average-case complexity ‘’ (i.e., the average amount of resources over all sample inputs of size ‘’) where ‘’ refers to the sample input size. Next, the time complexity consumed in our work is in the form of logarithmic expressed as ‘’ due to the reason that the time increases slowly and hence is said to exhibit logarithmic time complexity.

The computational complexity is measured in floating-point operations per second (FLOPS), which is a unit that measures how many floating-point calculations are performed in a second. FLOPS are a helpful way to recognize the computational complexity of deep learning models. A detailed explanation is given below in Table 10

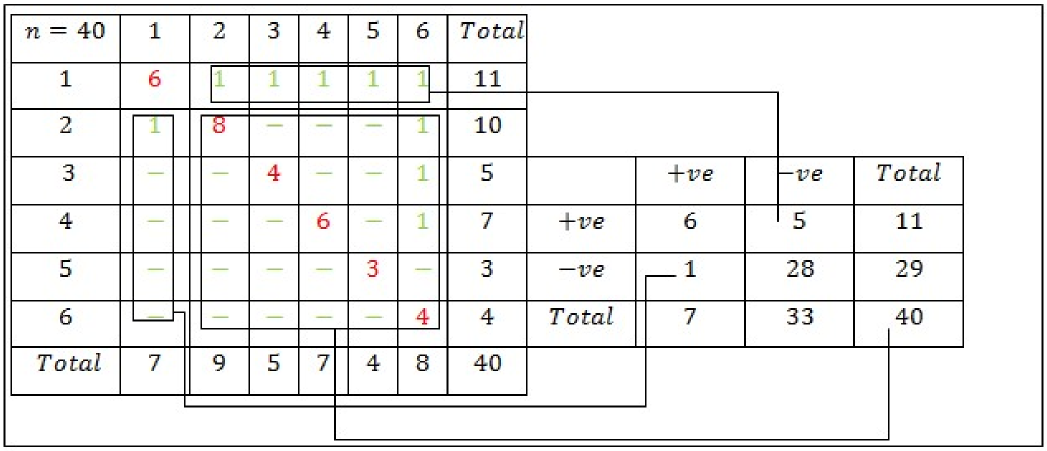
Table 10. Comparison of computational complexity in terms of FLOPS

Number of Samples	Computational complexity		
	CBAM and TCNs	Ensemble of LSTM neural networks	KFG-RPNDTL
5	0.008	0.006	0.004
10	0.012	0.010	0.006
15	0.016	0.012	0.009
20	0.013	0.009	0.005
25	0.019	0.015	0.010
30	0.021	0.018	0.013
35	0.017	0.013	0.011
40	0.023	0.017	0.015
45	0.015	0.011	0.008
50	0.014	0.008	0.005

In Fig 10, describe the comparison of computational complexity in terms of floating point operations per second Vs number of samples. In above fig 10, complexity was found to be minimized in proposed KFG-RPNDTL method than the CBAM-TCN [1] and Ensemble of LSTM neural networks [2]. Therefore, the results of computational complexity using KFG-RPNDTL method is reduced by 32% and 16% when compared to existing CBAM-TCN [1] and Ensemble of LSTM neural networks [2] methods.

6.6 Case Scenario: Confusion Matrix Analysis

When solving a human emotion recognition issue, confusion matrix is standard technique to examine model’s results. This approach is utilized for binary as well as multiclass recognition issues. Confusion matrix illustrates distribution of human emotion recognitions across different six classes (i.e., neutral [1], confusion [2], drowsiness [3], frustration [4], engaged [6] and boredom [6]) permitting for exact mapping among method’s human emotion recognitions as well as innovative class labels of sample instances. Numerous assessment parameters employed to estimate proposed KFG-RPNDTL method’s results calculated through aid of confusion matrix. Confusion matrix is employed to show human emotion recognitions result depend on values ‘TP’, ‘FP’, ‘TN’ and ‘FN’. Concept of the multi-class (i.e. six different classes) confusion matrix is comparable to binary-class matrix. Here, columns denote expected class allocation as well as rows indicate output allocation through classifier.

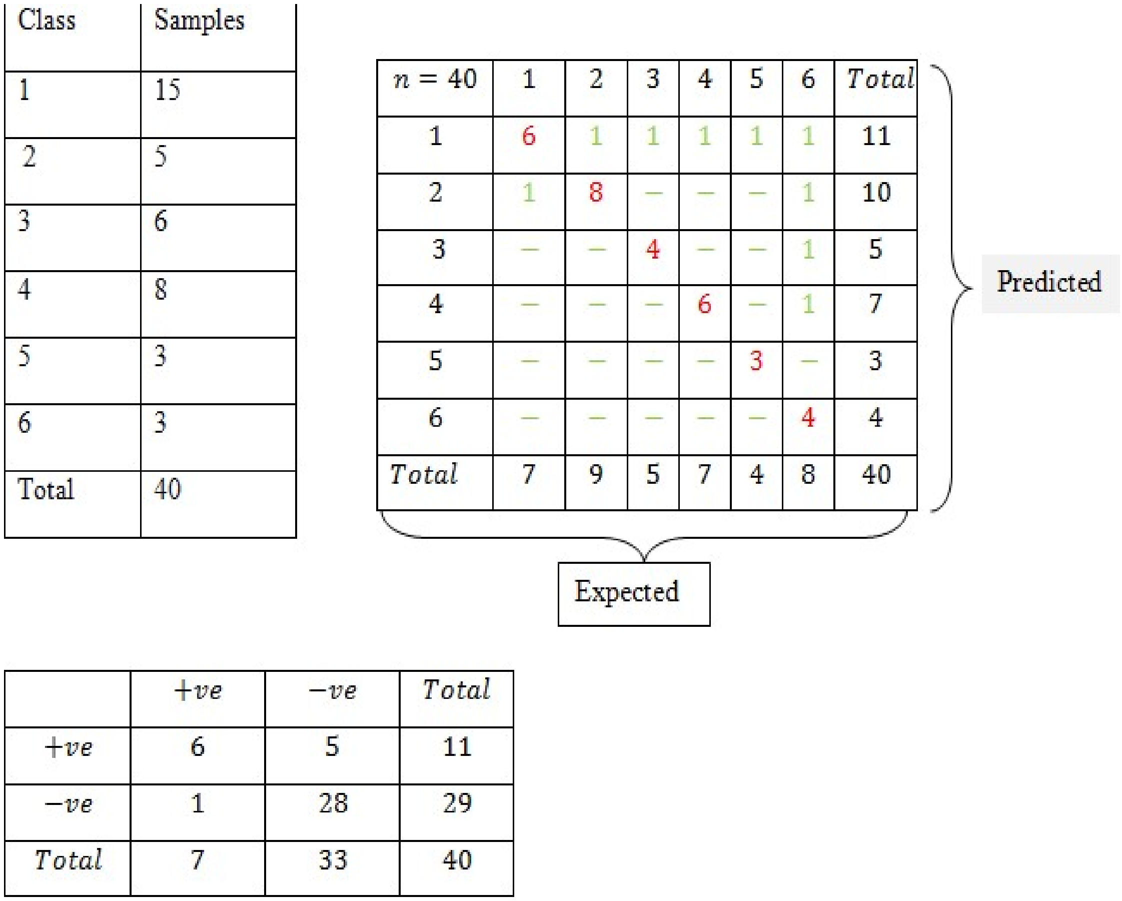


$n = 40$	1	2	3	4	5	6	$Total$
1	6	1	1	1	1	1	11
2	1	8	—	—	—	1	10
3	—	—	4	—	—	1	5
4	—	—	—	6	—	1	7
5	—	—	—	—	3	—	3
6	—	—	—	—	—	4	4
$Total$	7	9	5	7	4	8	40

	$+ve$	$-ve$	$Total$
$+ve$	6	5	11
$-ve$	1	28	29
$Total$	7	33	40

Fig 10. Comparison of computational complexity in terms of FLOPS

Table 12 Confusion matrix for multi class emotion



Class	Samples
1	15
2	5
3	6
4	8
5	3
6	3
Total	40

$n = 40$	1	2	3	4	5	6	$Total$
1	6	1	1	1	1	1	11
2	1	8	—	—	—	1	10
3	—	—	4	—	—	1	5
4	—	—	—	6	—	1	7
5	—	—	—	—	3	—	3
6	—	—	—	—	—	4	4
$Total$	7	9	5	7	4	8	40

Predicted
Expected

	$+ve$	$-ve$	$Total$
$+ve$	6	5	11
$-ve$	1	28	29
$Total$	7	33	40

From the above confusion matrix a lot of information are said to be arrived at.

- As customary, diagonal factors are properly forecasted samples. Sum of 31 samples were properly forecasted out of total 40 samples. Therefore, overall Acc' is 75.92%.
- To enhance model's result, one center on prognostic outcomes at class-6 (i.e. boredom). A total of 4 samples miscategorized through classifier, that is maximum misclassification rate between every class.

7. CONCLUSION

This study introduces the KFG-RPNDTL method, a significant advancement in academic engagement monitoring through recognising six distinct student-specific emotions. The method's integration of Euclidean Polar Coordinate Distance-based feature extraction and the Kriging Best Linear Unbiased Predictor has successfully reduced training time and computational complexity, making the system suitable for real-time classroom use. Additionally, the Intersection over the Union-based Region Proposal Network enhances the accuracy of region identification, further contributing to the model's performance.

A key achievement of this work is its focus on student-specific emotions, offering a more tailored approach to engagement monitoring compared to traditional models that rely on basic emotions. This refinement allows for more precise insights into students' emotional states, helping educators implement strategies that improve classroom engagement.

However, while the method demonstrates clear improvements in accuracy and efficiency, the reliance on a fixed set of six emotions may limit detecting more nuanced emotional states. Additionally, this study does not consider external factors such as class timing, subject matter, and student demographics—each potentially influencing emotional responses. Furthermore, facial occlusion, typical in real-world scenarios, remains unaddressed, potentially affecting the model's practical application.

In conclusion, the KFG-RPNDTL method offers a promising foundation for real-time emotion recognition in educational settings. Future work should focus on expanding the emotional range, incorporating external influences on engagement, and addressing facial occlusion to enhance the system's robustness. These improvements will further strengthen its application in monitoring and improving student learning experiences.

REFERENCES

[1] Mohammed Aly, "Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model", Multimedia Tools and Applications, Springer, May 2024 [ResNet50, CBAM (Convolutional Block Attention Module), and TCNs (Temporal Convolutional Networks)

[2] Paolo Buono, Berardina De Carolis, Francesca D'Errico, Nicola Macchiarulo, Giuseppe Palestra, "Assessing student engagement from facial behavior in on-line learning", Multimedia Tools and Applications, Springer, Feb 2023 [Ensemble of LSTM neural networks]

[3] Naveed Ahmed, Zaher Al Aghbari, Shini Girija, "A systematic survey on multimodal emotion recognition using learning algorithms", Intelligent Systems with Applications, Elsevier, Jan 2023

[4] Junge Shen, Haopeng Yang, Jiawei Li, Zhiyong Cheng, "Assessing learning engagement based on facial expression recognition in MOOC's scenario", Multimedia Systems, Springer, Feb 2022

[5] Mohammed Aly, Abdullatif Ghallab, Islam S. Fathi, "Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model", IEEE Access, Nov 2023

[6] Duong Ngo, Andy Nguyen, Belle Dang, Ha Ngo, "Facial Expression Recognition for Examining Emotional Regulation in Synchronous Online Collaborative Learning", International Journal of Artifcial Intelligence in Education, Springer, Nov 2023

[7] Rajesh Singh, Sumeet Saurav, Tarun Kumar, Ravi Saini, Anil Vohra, Sanjay Singh, "Facial expression recognition in videos using hybrid CNN & ConvLSTM", International Journal of Information Technology, Springer, Apr 2023

[8] Ruhina Karani, Jay Jani, Sharmishta Desai, "FER-BHARAT: a lightweight deep learning network for efcient unimodal facial emotion recognition in Indian context", Discover Artificial Intelligence, Feb 2024

[9] Mustafa Can Gurseli, Sara Lombardi, Mirko Duradoni, Leonardo Bocchi, Andrea Guazzini,

Antonio Lanata, “Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets”, IEEE Access, Apr 2023

[10] Tanoy Debnath, Md. Mahfuz Reza, Anichur Rahman, Amin Beheshti, Shahab S. Band, Hamid Alinejad-Rokny, “Four-layer ConvNet to facial emotion recognition with minimal epochs and the signifcance of data diversity”, Scientific Reports, Jun 2022

[11] Ko Watanabe, Tanuja Sathyanarayana, Andreas Dengel, Shoya Ishimaru, “EnGauge: Engagement Gauge of Meeting Participants Estimated by Facial Expression and Deep Neural Network”, IEEE Access, Jun 2023

[12] Allison Macey Banzo , Jonathan Beever, Michelle Taub, “Facial Expression Recognition in Classrooms: Ethical Considerations and Proposed Guidelines for Affect Detection in Educational Settings", IEEE Transactions on Affective Computing, Vol. 15, No. 1, Jan 2024

[13] Ambika G. N., Yeresime Suresh, “Optimal Deep Convolutional Neural Network Based Face Detection and Emotion Recognition Model”, International Journal of Intelligent Systems and Applications in Engineering, Mar 2023

[14] Prof. M. P. Nerkar, Apurva Sawant, Sayali Jawade, Rutuja Shinde, Rushikesh Thakur, “Automatic Recognition of Student Engagement using Deep Learning and Facial Expression”, International Engineering Research Journal (IERJ), Oct 2020

[15] Rana H. AL-Abboodi, Ayad A. AL-Ani, “Facial Expression Recognition Based on GSO Enhanced Deep Learning in IOT Environment”, International Journal of Intelligent Engineering & Systems, Mar 2024

[16] Liang Zhao, Kun Chen, Jie Song, Xiaoliang Zhu, Jianwen Sun, Brian Caulfield, Brian Mac Namee, “Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data”, IEEE Access, Jan 2021

[17] Bei Fang, Xian Li, Guangxin Han, Juhou He, “Facial Expression Recognition in Educational Research From the Perspective of Machine Learning: A Systematic Review”, IEEE Access, Oct 2023

[18] Güray Tonguç, Betul Ozaydın Ozkara, “Automatic recognition of student emotions from facial expressions during a lecture”, Computers & Education, Elsevier, Jan 2020

[19] Yanling Gan, Luhui Xu, Haiying Xia, Gan Liu, “Harmonious Mutual Learning for Facial Emotion Recognition”, Neural Processing Letters, Springer, Mar 2024

[20] Swadha Gupta, Parteek Kumar, Rajkumar Tekchandani, “A multimodal facial cues based engagement detection system in e-learning context using deep learning approach”, Multimedia Tools and Applications, Springer, Feb 2023

[21] Ting-Ting Wu, Hsin-Yu Lee, Wei-Sheng Wang, Chia-Ju Lin, Yueh-Min Huang, “Leveraging computer vision for adaptive learning in STEM education: efect of engagement and self-efcacy”, International Journal of Educational Technology in Higher Education, Nov 2023

[22] Alexandros Vrochidis, Nikolaos Dimitriou, Stelios Krinidis, Savvas Panagiotidis, Stathis Parcharidis, Dimitrios Tzovaras, “A Deep Learning Framework for Monitoring Audience Engagement in Online Video Events”, International Journal of Computational Intelligence Systems, Apr 2024

[23] Gerard Pons and David Masip, “Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis”, IEEE Transactions on Affective Computing, Nov 2017

[24] Jayasankar Santhoshm, Akshay Palimar Pai, Shoya Ishimaru, “Toward an Interactive Reading Experience: Deep Learning Insights and Visual Narratives of Engagement and Emotion”, IEEE Access, Jan 2024

[25] Dung Nguyen, Duc Thanh Nguyen, Sridha Sridharan, Simon Denman, Thanh Thi Nguyen, David Dean, Clinton Fookes, “Meta-transfer learning for emotion recognition”, Neural Computing and Applications, Springer, Jan 2023

[26] Desai, Keya, et al. "A Transfer Learning Framework for Facial Emotion Recognition: Leveraging Pre-Trained Convolutional Neural Networks." 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT). IEEE, 2024.

[27] Betageri, Deepa, and Vani Yelamali. "Detection and Classification of Human Emotion Using Deep Learning Model." 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT). IEEE, 2024.

[28] Akinduyite, Olanike Christianah, et al. "Facial Emotion Recognition Using Convolutional Neural Network in a Learning Environment." 2024 International Conference on Science,

Engineering and Business for Driving Sustainable Development Goals (SEB4SDG). IEEE, 2024.

[29] Goel, Mehak, Akshat Mittal, and Mitu Sehgal. "Emotionally Intelligent Edtech: A CNN Perspective." 2024 International Conference on Computational Intelligence and Computing Applications (ICCICA). Vol. 1. IEEE, 2024.

[30] Shan, J., and Sherin Eliyas. "Exploring AI Facial Recognition for Real-time Emotion Detection: Assessing Student Engagement in Online Learning Environments." 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT). IEEE, 2024.

[31] Bakyt, Ersultan, uly, Aitzhan. "Determining emotions from students' facial expressions using CNN." Қ.А. Ясауи атындағы Халықаралық қазақ-түрік университетінің хабарлары, 25 (2023).:49-59. doi: 10.47526/2023-2/2524-0080.057.

[32] Yuan, Tian., Jingxuan, Zhu., Yao, Huang., D., Y., Chen. "Facial Expression Recognition Based on Vision Transformer with Hybrid Local Attention." Applied Sciences, 14 (2024).:6471-6471. doi: 10.3390/app14156471

[33] Xing, Liu., Hui, Wang. "An Expression Recognition Method Based on Feature Double-Branch Fusion." Advances in transdisciplinary engineering, null (2024). doi: 10.3233/atde2403999.

[34] Qiancheng, Wei., Ying, Liu., Xiaoping, Jiang., Ben, Zhang., Qiya, Su., Muyao, Yu. "DDFNet-A: Attention-Based Dual-Branch Feature Decomposition Fusion Network for Infrared and Visible Image Fusion." Remote sensing, null (2024). doi: 10.3390/rs1610179510.

[35] Yan-Ping, Dong., Ting, Wang., Yanfeng, Pu., Jian, Gao. "Facial Expression Recognition with Enhanced Relation-Aware Attention and Cross-Feature Fusion transformer." null (2024). doi: 10.21203/rs.3.rs-3948258/v1