Multimedia Tools and Applications (2023) 82:11365–11394
https://doi.org/10.1007/s11042-022-13558-9

1226: DEEP-PATTERNS EMOTION RECOGNITION IN THE WILD

Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models

Swadha Gupta[1] · Parteek Kumar[1] · Raj Kumar Tekchandani[1]

Received: 2 September 2021 / Revised: 14 May 2022 / Accepted: 14 July 2022 /
Published online: 9 September 2022

Abstract

The dramatic impact of the COVID-19 pandemic has resulted in the closure of physical classrooms and teaching methods being shifted to the online medium.To make the online learning environment more interactive, just like traditional offline classrooms, it is essential to ensure the proper engagement of students during online learning sessions.This paper proposes a deep learning-based approach using facial emotions to detect the real-time engagement of online learners. This is done by analysing the students' facial expressions to classify their emotions throughout the online learning session. The facial emotion recognition information is used to calculate the engagement index (EI) to predict two engagement states *"Engaged"* and *"Disengaged"*. Different deep learning models such as Inception-V3, VGG19 and ResNet-50 are evaluated and compared to get the best predictive classification model for real-time engagement detection. Varied benchmarked datasets such as FER-2013, CK+ and RAF-DB are used to gauge the overall performance and accuracy of the proposed system. Experimental results showed that the proposed system achieves an accuracy of 89.11%, 90.14% and 92.32% for Inception-V3, VGG19 and ResNet-50, respectively, on benchmarked datasets and our own created dataset. ResNet-50 outperforms all others with an accuracy of 92.3% for facial emotions classification in real-time learning scenarios.

Keywords Facial expressions · Engagement detection · Emotion detection · Deep learning · Real-time engagement detection · Online learning · Online learner

✉ Swadha Gupta
sgupta_phd18@thapar.edu

Parteek Kumar
parteek.bhatia@thapar.edu

Raj Kumar Tekchandani
rtekchandani@thapar.edu

[1] Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, India

1 Introduction

The digital learning platform has created an affordable learning opportunity for the masses worldwide. It has enhanced the learning process by making educational resources accessible and readily available. After the COVID-19 outbreak, the teaching-learning scenario has dramatically been shifted to digital platforms [40]. With this sudden shift from the physical to virtual classroom around the globe, many challenges have arisen [27]. One such limitation is the lack of physical presence. Due to the absence of face-to-face interaction with the instructor, learners lose motivation and interest, which affects their learning performance [4]. As a result, learners do not complete the online course or leave the online classroom mid-way. So, it becomes important to know how well a learner is engaged in the online learning environment. Therefore, researchers are now working hard to meet the latest challenges faced in online learning [3].

The degree of engagement level is generally reflected by analysing the emotional involvement of learners while studying [56]. Emotions always affect the learning performance of the learners [45]. For instance, positive emotions like joy and curiosity have a positive impact as they facilitate self-regulation and help to focus more on the problem-solving tasks and thus, make learners engaged [26]. On the other hand, negative emotions

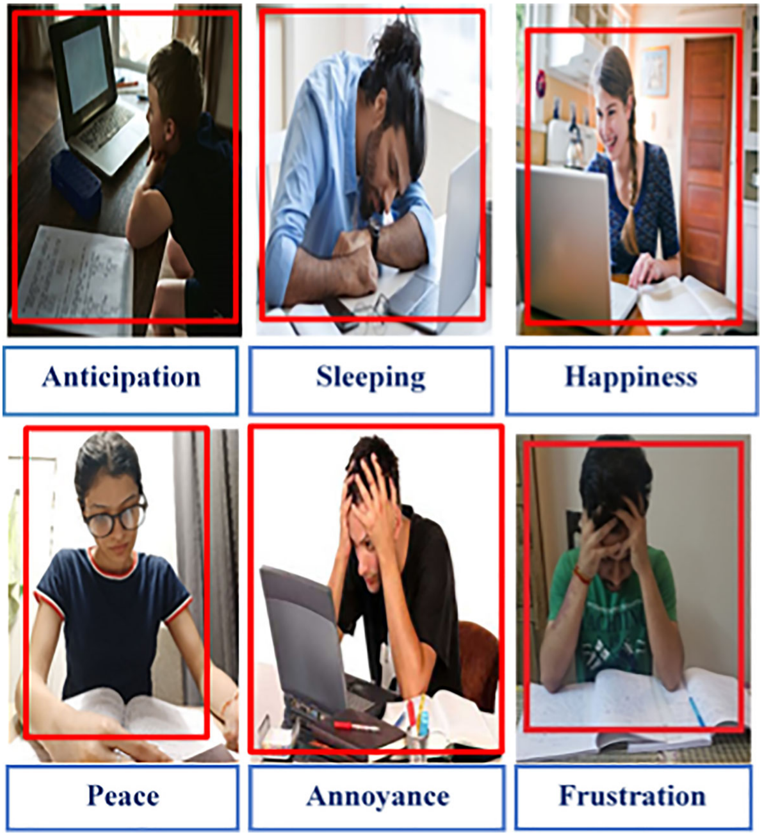


Fig. 1 Different states of mind of human beings. (a) anticipation, (b) sleeping (c) happiness, (d) peace, (e) annoyance and (f) frustration

like boredom and frustration divert the attention of the learners and consume cognitive resources during learning activities and leave the learners disengaged [54]. Thus, the present study focuses on the automatic recognition of facial emotions of online learners. Recognising learners' facial emotions in real-time is quite challenging in an e-learning platform owning to the absence of human supervision. For such a learning environment, learners' facial expressions are analysed in real-time to extract different emotions. Facial expressions are physical muscle movements translated from emotional impulses such as raising eyebrows, wrinkling the forehead, or curling lips. By observing the change of facial expressions automatically, a lot of information about online learner's emotional states can be determined [60]. For instance, see the images of Fig. 1 and estimate their state of mind by observing their expressions.

The anticipatory state can be recognised in Fig. 1(a) as a learner is engrossed in studying. The learners in Fig. 1(b) are sleeping while studying with the laptop on. The learner in Fig. 1(c) is in a happy state while studying. In Fig. 1(d), the learner is peacefully studying. The annoyance state of the learner can be seen in Fig. 1(e) because the learner does not understand the concept. The state of frustration can be seen in Fig. 1(f) as a learner is not able to focus on his studies.

Facial expressions are divided into six basic emotions by the famous psychologist Paul Ekam [12]. Since then, many researchers have universally accepted these basic emotions for FER (Facial expression recognition) research. These six basic emotions are surprise, sadness, happiness, fear, disgust and anger, which are shown in Fig. 2. Therefore, automatically

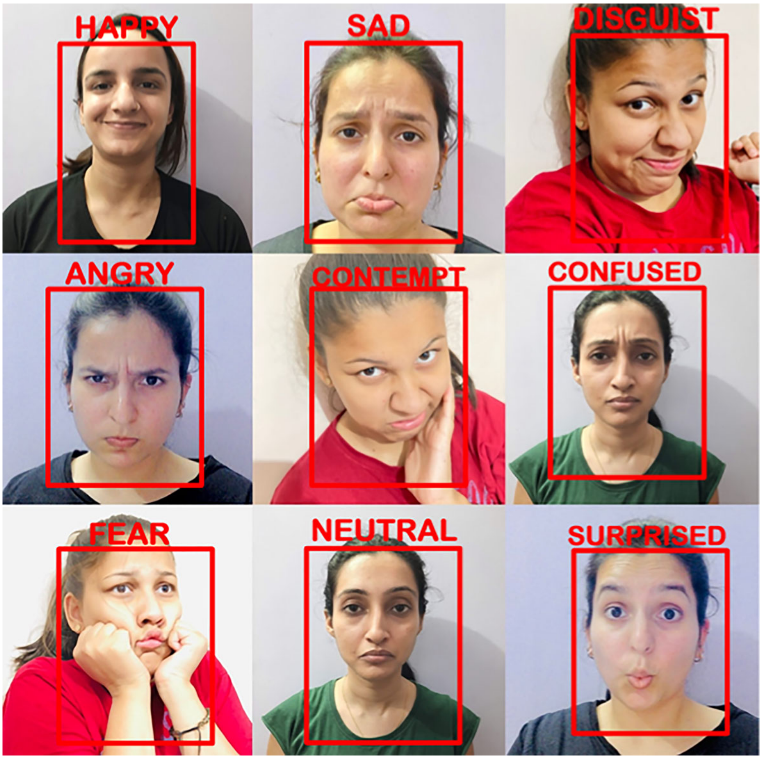


Fig. 2 Facial expressions of different emotions (a) Happy, (b) Sad, (c) Disgust, (d) Angry, (e) Contempt, (f) Confused (g) Fear, (h) Neutral and (i) Surprised

analysing the facial expressions of learners helps in deciding the learner's engagement state in real-time scenarios.

Considering the need to develop a real-time engagement detection system, this paper proposes a novel approach based on deep learning technologies. Our contributions to this paper are as follows.

- An engagement detection system that automatically detects learner engagement in real-time scenarios based on facial emotion recognition is proposed.
- Online learner's engagement is evaluated based on the facial emotion information through real-time facial expression analysis.
- Face detection is done with the help of the pre-trained Faster R-CNN model.
- A modified landmark extractor is proposed, i.e. MFACEXTOR, to extract the 470 key points (face-points).
- Deep learning models such as Inception-V3, VGG19 and ResNet-50 are implemented for real-time learning scenarios to classify student emotions such as angry, sad, happy, neutral etc., with the help of the softmax function.
- An Engagement evaluation algorithm is proposed to calculate the engagement index from facial emotion classification output data.
- Finally, based on the engagement index value, the system decides whether the online learner is engaged or disengaged.

The rest of the paper is organised as follows. Section 2 gives a brief overview of the related work done for engagement detection in an online learning environment. Section 3 discusses the datasets used for the experiment. In Section 4, the methods and proposed system are discussed. Section 5 presents the experimental results achieved after implementing our proposed approach. Section 6 illustrates the comparison of the proposed system with the state-of-the-art models. The visualisation of the engagement detection system is presented in Section 7. Section 8 concludes the paper with the future work.

2 Related work

In recent years, engagement detection during online learning has been gaining attention. Engagement detection is essential in the online classroom environment to engage learners and enhance learning performance, but the studies about it have only been exclusive to the traditional classroom [13, 35]. Many studies reported that faculties who are teaching through virtual mediums believe that they can access learner's understanding better in a face-to-face classroom environment than in an online learning environment [7, 22]. Researchers recently started investigating the impact of monitoring learner's engagement during online content delivery [15, 21]. Various approaches have been implemented for engagement detection, but facial expression is one of the popular and successful methods [6, 44] because it is the visual clue to recognise the emotional state of the learner [36] and face images dataset is easier to collect [38]. Zhang et al. [59] has proposed a multi-task cascaded CNNs based framework. This proposed model used three stages cascaded structure to boost face detection performance.

Most of the recent work on FER (facial expression recognition) performed well on the controlled images dataset but didn't perform well on partial and variation face images. Using facial features, [37] proposed a conceptual framework for the classification of learner engagement. The proposed model included detecting the multiple faces, extracting the facial units, and using SVM (Support vector machine) as a binary classifier model. The proposed model was tested on various datasets, and comparisons with the best-configured models were made. The automated learning system can act as an important tool to identify online learners' attentive and inattentive states. Deep learning networks have been successful in implementing engagement detection using facial expressions [8, 43, 50]. Turabzadeh et al. [55] have researched real-time facial emotion recognition using LBP (Local Binary Point) algorithm. The features were extracted using LBP from the captured video and input into the K-NN (K- Nearest Neighbour) regression with dimension labels. The accuracy recorded using the LBP algorithm was 51.28%. Murshed et al. [46] explored three models, namely network-in-network (NiN-CNN), convolutional network (All-CNN), and very deep convolutional network (VD-CNN) [51]. The advantageous features from these models were extracted, and an improved model was proposed. In the proposed model, a multi-layer perceptron replaces the linear convolutional layer, a further convolutional layer replaces a few max-pooling layers and finally, small (3x3) convolutional filters replace the depth of the network. The engagement detection performance was measured, and the proposed model was compared with three base models performed on the DAiSEE dataset in e-Environments. Li et al. [31] proposed real-time facial emotions recognition system for learners using Xception model. For face detection, Haar-Cascade was used along with the Xception model [19]. The deep separable convolution method was used with pre-activation in the residual block, which reduced the complexity of training the model and reduced the overfitting complication. Altuwairqi et al. [5] proposed a multimodal approach to measure learner's engagement. In this study, three modalities were analysed to represent learner behaviour. Several experiments were conducted to validate the proposed approach, and an accuracy rate of 76.19% was recorded.

Reliable models play a key role in detecting learner engagement during educational activities. Li et al. [30] proposed a multi-kernel convolution block for facial expression recognition. The multi-kernel convolution approach for feature extraction used three depthwise separable convolutions. The multiple size convolution kernels and fuse details are simultaneously obtained along with edge contour details on facial expressions to design the lightweight facial expression network. The experimental results showed an accuracy of 73.3% on FER-2013 and CK+ using the proposed approach. Minaee et al. [39] convolution network with less than 10 CNN layers for emotion detection. The proposed network has been evaluated on JAFFE, CK+, FER-2013 and FERG. The visualisation technique has been used to find important regions of the face for recognising the facial expression of different emotions. It has been concluded from the previous studies that facial expressions play an important role in the emotion recognition process. From prior studies, image-level facial expression recognition is more focused rather than in the real-time environment. There is a lack of work that has been done on detecting engagement using facial expressions. Also, the existing studies for engagement detection were not implemented for an online educational environment. Therefore, the current studies provide a vision to be followed for qualitative and quantitative research for real-time engagement detection. The purpose of the present study is to use deep learning models to automatically detect the engagement states of online learners using facial expressions.

3 Datasets

In this study, the proposed model considered those standard benchmarked and publicly available datasets that work efficiently in real-time scenarios. A brief overview of these datasets is given below.

Table 1 Summary of WIDER face dataset

NAME	YEAR	SIZE (No of images)	FACE LABELS	GRAY/ COLOR	TYPES OF FACE VARIATIONS
WIDER FACE	2016	32,203	393,703	Color	Scale, Pose,Occlusion, Expression, Make up, Illumination

3.1 Wider face

Wider Face dataset [58] is one of the largest datasets for face detection having 32,203 coloured images and 393,703 labelled faces. It is a subset of the publicly available Wider dataset. The summary of the Wider Face dataset is given in Table 1 and its sample images are shown in Fig. 3(a).

3.2 FER-2013 (Facial expression recognition 2013)

FER-2013 (Facial Expression Recognition 2013) is the most popular facial expression dataset introduced in the Representation learning challenge of ICML (Kaggle facial expression recognition challenge) held in 2013 [17]. The summary of FER-2013 dataset is given in Table 2 and its sample images are shown in Fig. 3(b).

3.3 CK+ (extended cohn-kanade dataset)

CK+ (Extended Cohn-Kanade Dataset) is a widely used facial expression recognition dataset and is the extended version of the CK dataset [18]. The summary of CK+ dataset is given in Table 3 and its sample images are shown in Fig. 3(c).

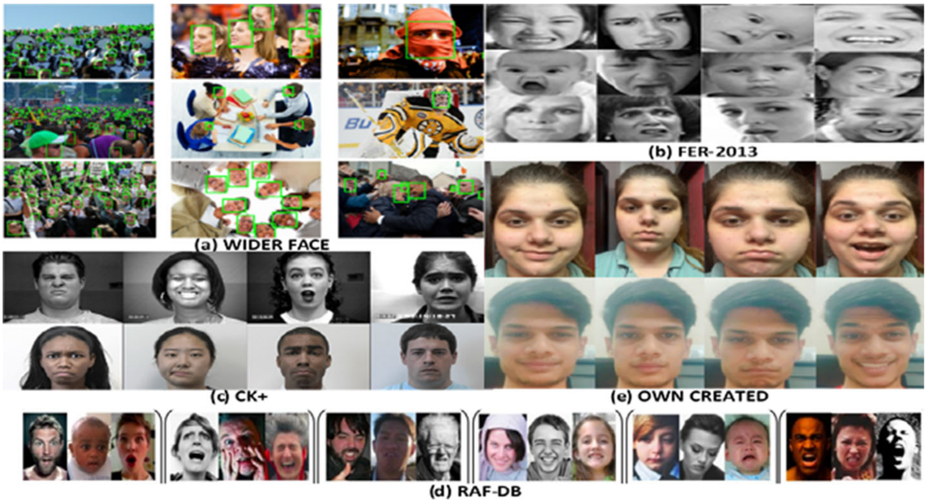


Fig. 3 Examples images from (a) Wider Face (b) FER-2013 (c) CK+ (d) RAF-DB (e) Own dataset

Table 2 Summary of FER-2013 dataset

NAME	YEAR	SIZE (No of images)	GRAY/ COLOR	IMAGE SIZE (In Pixels)	TYPES OF FACE VARIATIONS
FER-2013	2013	35,877	Grey	48 X 48	Angry, Disgust,Frustration, Happy, Sad, Surprise and Neutral

3.4 RAF-DB (real-world affective faces)

RAF-DB [32] is a large scale facial expression recognition dataset with 29672 coloured images. The summary of RAF-DB dataset is given in Table 4 and its sample images are shown in Fig. 3(d).

3.5 Own dataset

We collected and labelled our newly collected dataset for facial expression recognition. The images are labelled with six basic facial expressions. The summary of this dataset is given in Table 5 and and it's sample images are shown in Fig. 3(e).

4 Proposed approach for engagement detection system

This paper proposes an engagement detection system to calculate the online learner's EI (engagement index) using the facial emotion recognition approach to predict the engagement state in an online learning environment. The overview of the proposed engagement detection system is shown in Fig. 4. In the first step, the images are captured using the built-in camera of the device through which the learner is studying the online content. Learner's faces are detected using the Faster R-CNN model [20]. Then, the key points of the facial features are extracted from the detected face using the proposed modified face-point extractor (MFACEXTOR). Individual facial emotion classification is performed using the deep convolutional neural networks (CNNs) from the detected faces and extracted facial key-points information. The predicted emotion from the key image frame is combined to get EI to detect the engagement level of an individual learner in the online learning environment. The detailed description of each step has been further discussed in this section.

Additionally, the modified face-point extractor (MFACEXTOR) is used for face point extraction from the detected face area. Individual facial emotion classification is performed using the deep convolutional neural network (CNN) architecture with the help of proper

Table 3 Summary of CK+ dataset

NAME	YEAR	SIZE (No of images)	GRAY/ COLOR	IMAGE SIZE (In Pixels)	TYPES OF FACE VARIATIONS
CK+	2010	593	Mostly Gray/Color	640 X 480	Neutral, Happy, Angry, Sad, Fear, Surprise, Disgust and Contempt

Table 4 Summary of RAF-DB dataset

NAME	YEAR	SIZE (No of images)	GRAY/ COLOR	IMAGE SIZE (In Pixels)	TYPES OF FACE VARIATIONS
RAF-DB	2017	29672	Color	349 × 349	Neutral,Happy,Angry, Sad, Fear, Surprise, Disgust and Contempt

Table 5 Summary of Own dataset

NAME	SIZE (No of images)	GRAY/ COLOR	IMAGE SIZE (In pixels)	TYPES OF FACE VARIATIONS
Own dataset	1800	Color	48 X 48	Angry, Sad, Happy, Neutral, Surprise and Fear

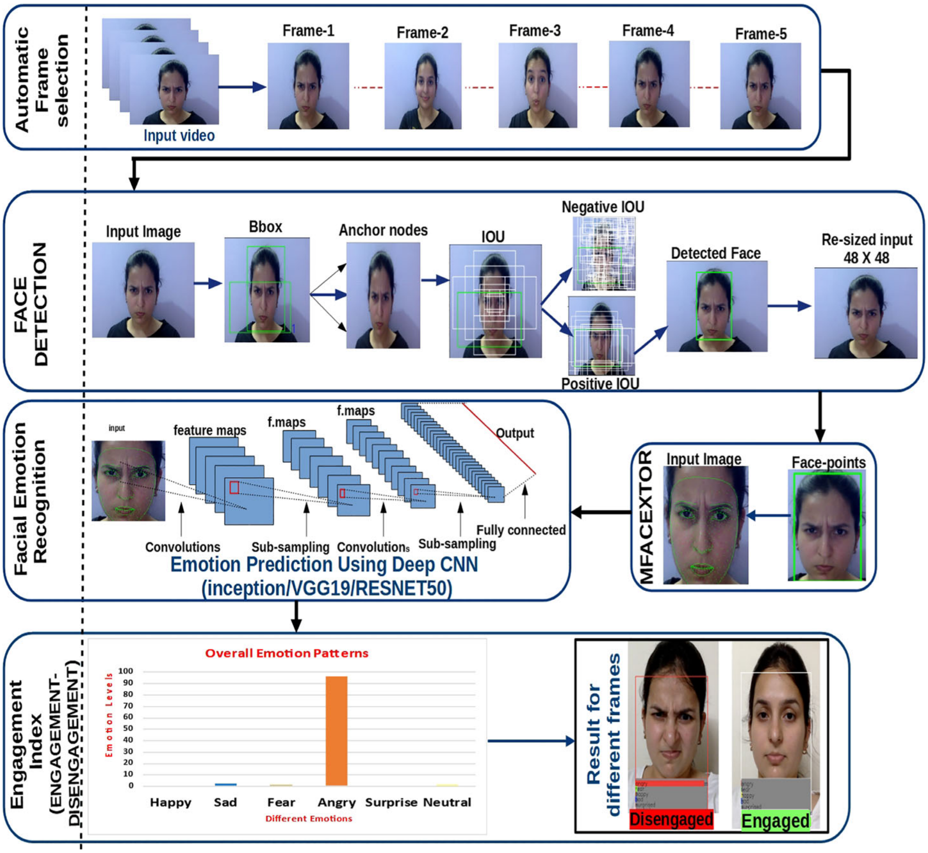


Fig. 4 Proposed framework for engagement detection system

training data. The emotion classification information of the individual learner helps to calculate the engagement index (EI) for the engagement state detection of the learner, i.e. engaged or disengaged state.

4.1 Automatic frame selection

As learners learn through online videos, our input to the proposed system is the video stream from the web camera. To extract discriminative features from video streaming, frame-based processing is used. But, all the frames do not help to detect the face. So for that, frame selection is performed to get the best-suited frames for face detection. The proposed system extracts the images from the video stream after every specific period (i.e., every 20 sec). Extracted images are buffered in the memory and saved with unique frame numbers such as Frame-1 to Frame-n.

4.1.1 Face detection based on faster R-CNN (Region-based convolutional neural network)

The traditional facial detection model involves the problem of low resolution, model complexity, and the complex process of explicit feature extraction for large image datasets [52]. And also, the traditional facial detection model is less tolerant of variations in occlusion, illumination, pose, and expression. But, recently, the Faster R-CNN has overcome these

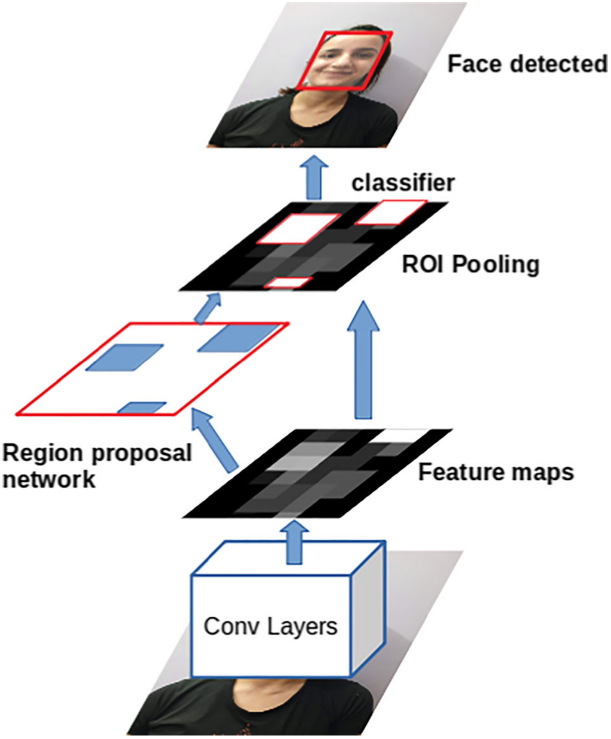


Fig. 5 Building block of Faster R-CNN

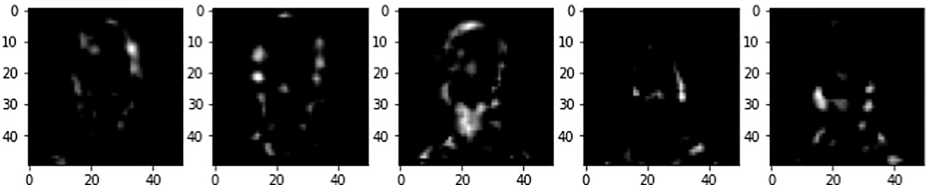


Fig. 6 Visualization of five channels extracted 50 x 50 x 512 feature maps for a input image

issues and shown impressive results for face detection. Therefore, in this paper, Faster R-CNN [48] is proposed for face detection, which consists of two modules, the first module belongs to Regional Proposal Network (RPN), and another one is the detector for refining the proposal. The basic building block of Faster R-CNN is shown in Fig. 5. The two outputs are given for every object: a class label and the bounding box coordinates.

To dive deep insight into Faster R-CNN working, the major part is the Regional Proposal Network (RPN). In the case of the pre-trained model, the RPN consists of convolution layers of 3x3 convolutions. This is further scanned by filters to extract features from an input image and reduces the large spatial window, i.e., 224x224x3 (for VGG19), into a low-dimensional feature vector.

The visualisation of five channels with 50x50x512 features maps is shown in Fig. 6. With the help of the IOU (Intersection over Union) process, the system will extract only the detected region of the face from the image, and the rest of the regions of the face are ignored; the system proceeds with only positive IOU to reduce the congestion of anchors. The first five ROI's (Region of interest) feature maps after the ROI step are shown in Fig. 7.

The hyper-parameters for Faster R-CNN used is represented in Table 6.

The loss function for the Faster R-CNN is as follows.

$$L(\{P_i\},\{\tau_i\}) = \frac{1}{N_{class}}\sum_i L_{class}(\theta_i,\varphi_i^*) + \lambda\frac{1}{N_{regress}}\sum_i \varphi_i^* L_{regress}(\tau_i,\tau_i^*). \qquad (1)$$

Here,
i= index of an anchor for a mini-batch.
θ_i = predicted probability of anchor.
φ_i^* = ground-truth label, Where $\varphi_i^* = \left\{\begin{matrix} anchor\ is\ positive & 1 \\ anchor\ is\ negative & 0 \end{matrix}\right\}$
τ_i =a vector having four parameterized coordinates of the predicted bounding box.
τ_i^*=a ground-truth box associated with a positive anchor.
L_{class} =log loss over two classes.
$L_{regress}$ = regression loss.

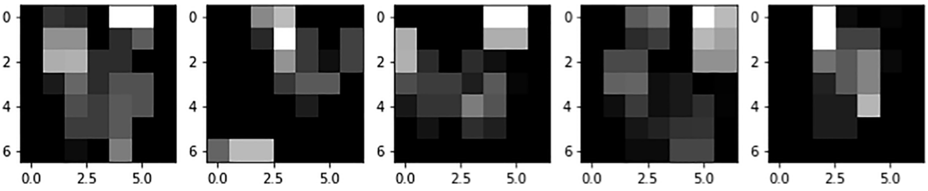


Fig. 7 Visualization of first five ROI's feature maps after ROI pooling

Table 6 Hyper-Parameters for Faster R-CNN model

Hyper-Parameters	Values
Batch size	128
Learning rate	0.001
for SGD	0.9
Weight decay for regularization	0.0001
Threshold for ROI	0.5
Training/Testing Sample	70% / 30%

$\varphi_i^* L_{cls}$ = the regression loss is activated only for positive anchors.

4.1.2 Modified face-points extractor

Facial features are extracted for identifying the corresponding emotion [47]. The facial features define the shape of the face, which consists of different elements such as lips, nose, eyes, and mouth [11]. To extract the facial features in the present work, pre-trained model from MediaPipe Face Mesh is implemented. The received image vector from the Faster R-CNN consists of ROI of the face area. Various key-points from the face image are extracted using the geometric information of facial features which is based on MediaPipe Face Mesh. MediaPipe Face Mesh employs the machine learning based detector to estimate the face geometry. The detector detects the face key-points and its 3D model predicts the face geometry surface. The different facial-features is given in Table 7 and its key-points are shown in Fig. 8.

The aim is to analyse the muscles motions of the face. The facial muscle motions are analysed by extracting the key points from each facial feature using geometry pipeline component of face mesh. The key-points from these Facial features is described by using the (2) below.

$$\{(f1, p1), (f2, p2), \ldots, (f_n, p_n)\}, f = (f1, \ldots, f_n, p1, \ldots, p_n)^T \tag{2}$$

Here, f and p is a 2N-D vector, T denotes transpose to identify the facial pose.

For the facial expression recognition task, many key points are required. The number of key-points for every feature is given in Table 8 and can be seen in Fig. 8(b). In this work, 470 key points are extracted rather than 468 of what is detected in typical MediaPipe Face Mesh implementation, as shown in Fig. 8(b). More key points are detected for better reliability of facial emotion recognition in the online learning environment. The Euclidean distance

Table 7 Facial features Key points

S.No.	Facial Features
FF1	Left Eyebrow
FF2	Right Eyebrow
FF3	Left Eye
FF4	Right Eye
FF5	Nose
FF6	Mouth
FF7	Lips
FF8	Jaw

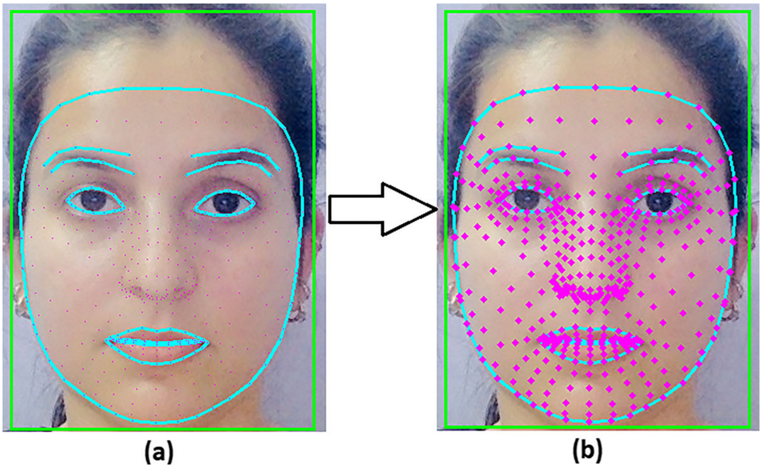


Fig. 8 Facial features with key points

between any two facial features is given in (3) below.

$$\sqrt{(f2 - f1)^2 + (p2 - p1)^2} \tag{3}$$

Here, f and p is a 2n-D vector

The area of key-points whose coordinates are known for every facial feature are shown in Fig. 8(a), which is calculated using the (4):

$$|((f1p2 - p1f2) + (f2p3 - p2f3) + \ldots (f_n p1 - p_n f1))/2| \tag{4}$$

In this way, 470 key points have been extracted from the face image, as shown in Fig. 8(b). These facial features would be given as input to the deep neural network. The neural network would learn from these face-points to decide the final output emotion. The different facial key-point representations of different emotions are shown in Fig. 9.

4.2 Facial emotion recognition using deep learning models

Their emotional state directly influences the learning process of the learner. So, it becomes important for online instructors to gauge their moods using facial indicators. Because in a face-to-face classroom setting, teachers identify the emotional state of the learners by

Table 8 Facial features Key points

Facial Features	Feature points
Left Eyebrow	32
Right Eyebrow	32
Left Eye	61
Right Eye	61
Nose	68
Mouth	74
Lips	48
Jaw	94

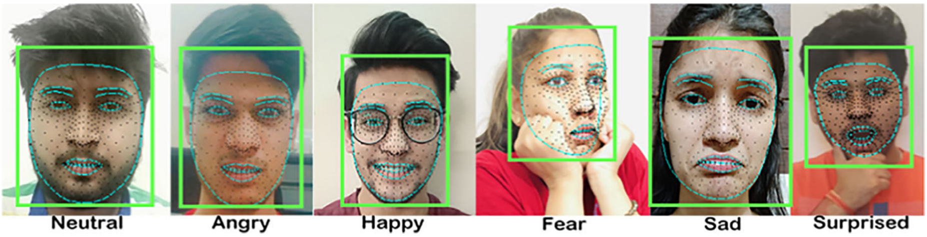


Fig. 9 Key-points representation of different emotions

observing them, but in an online classroom setting, face-to-face observation is not possible due to physical unavailability. Therefore, to assess online learner's frame of mind, automatic recognition of the facial emotion has been focused by most of the researchers [14, 32, 33]. Some methods are based on a set of localised face movements (facial action coding system) to represent facial expression [28]. Recent deep convolutional neural networks (CNNs) identify changes in emotions and perceive the emotional state of online learners while the class is in progress. The present work recognises the facial emotion state by interpreting facial features and analysing expressions. The three deep CNN-based models, such as Inception-V3, VGG19, and ResNet-50, are evaluated one by one in this study, and their architectural representation is shown in Fig. 10. A brief description of each network has been given below.

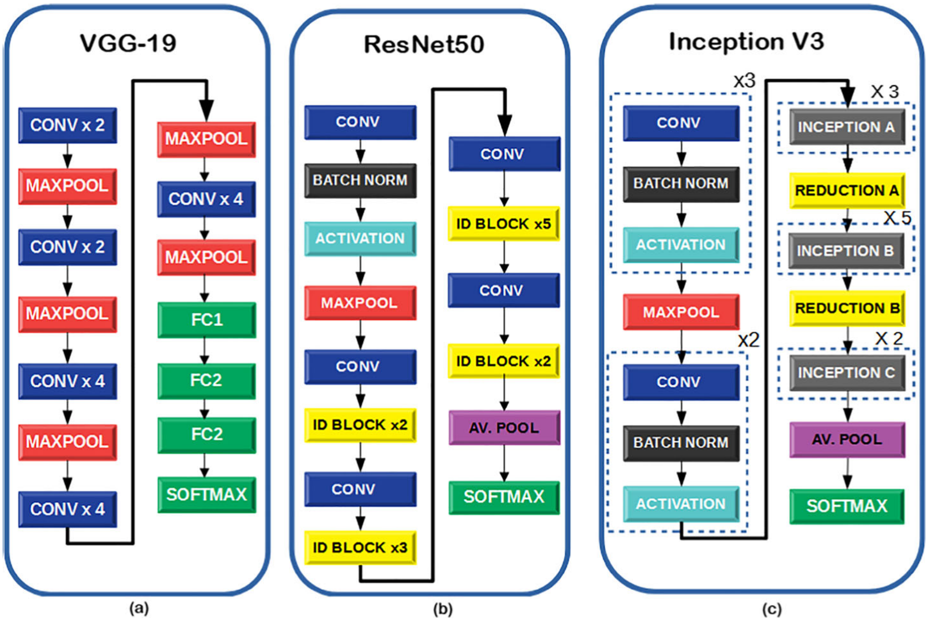


Fig. 10 Architectures of the models used for facial emotion recognition (FER): (a)VGG19 (b)ResNet-50 and (c) Inception-V3

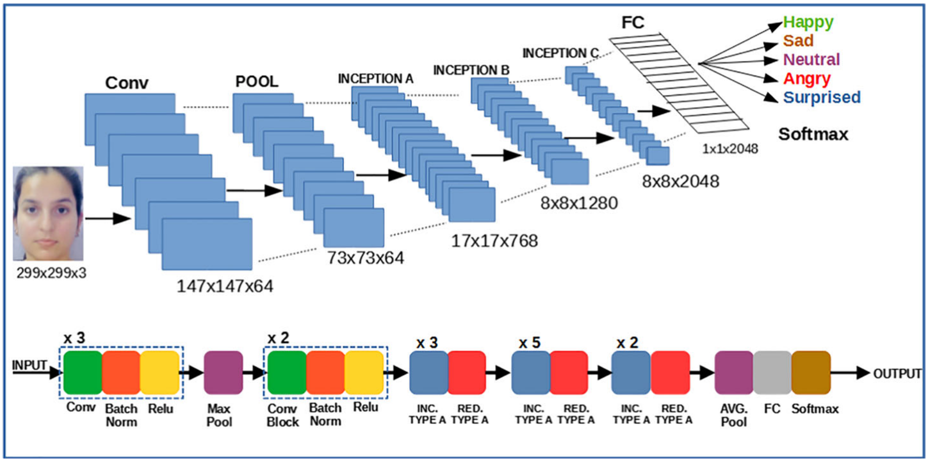


Fig. 11 Network Architecture of Inception-V3

4.2.1 Inception-V3

Inception-V3 is a deep neural network model having 48-layers ([24]). Figure 11 shows the detailed structure of Inception-V3 architecture, which is used in this proposed method. It is developed while keeping in mind the computation power efficiency for assisting in image analysis. The input size of the image in this proposed study is 299x299x3 for Inception-V3. The model in the proposed approach is composed of symmetric and asymmetric building blocks, which consist of five convolution-stem layers. It also compromises the three inception blocks of type A, followed by the reduction block of type A, five inception blocks of type B, followed by the reduction block of type B, and two inception blocks of type C, followed by the reduction block of type C. All these blocks are followed by a layer of average pooling and then the fully connected layer of 2048x1x1 size is the final layer. The factorisation is taken into account to reduce the size of the deep neural network to avoid overfitting. Other techniques are also used to make a more efficient network, such as smaller convolution layers, regularisation, dimension reduction, and parallel computations. Softmax is used in our proposed approach to compute the loss in this model.

4.2.2 VGG19

VGG19 is a deep neural network model having 19-layers [25], and Fig. 12 shows the detailed structure of VGG19 architecture, which is used in this proposed approach. The input image size considered in the proposed method is 224x224x3 for this model. It consists of sixteen layers of convolution followed by a max-pooling layer. It also consists of three fully connected layers. To avoid overfitting, this model used dropout by improving generalisation in the FC (fully connected) layer. The first two and last FC layers comprise 4096, 4096, and 1000 channels. VGG19 uses a convolution kernel of 3x3 size with a stride size of 1-pixel, i.e., 3x3x1. The size of the max-pooling kernel is 2x2 with a stride size of 2-pixel, i.e. 2x2x1. This study aims to classify six basic emotions with an input image size of 48x48. The input image would pass through convolution layers followed by a max-pooling layer for each convolution block. Then after passing through all these sixteen layers, the

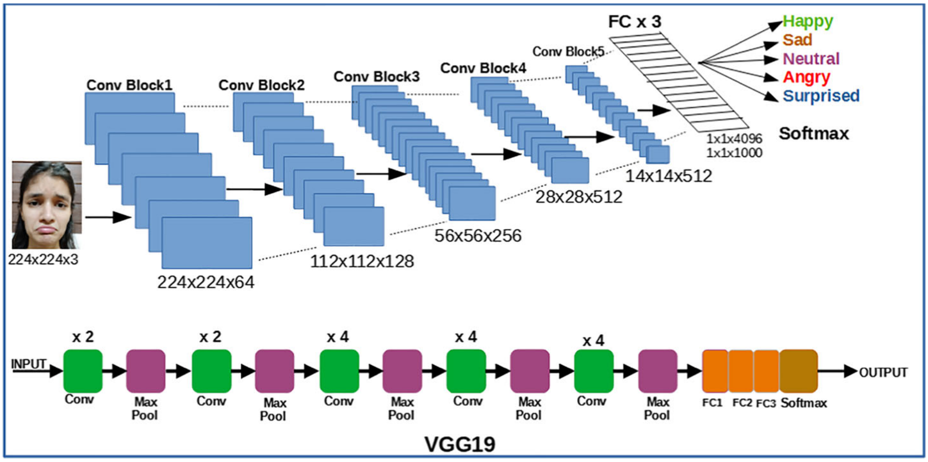


Fig. 12 Network architecture of VGG19

input image would pass through three fully connected layers to perform the classification of facial expressions of different emotions in this proposed study.

4.2.3 ResNet-50

ResNet-50 is a deep neural network model having 50-layers [29] and Fig. 15 shows the detailed structure of ResNet-50 architecture, which is used in this proposed approach. The full form of ResNet is residual networks, and it is called so because the present layer learns from the residual of the past layer. Rather than depending on network depth like most other models, the ResNet-50 relies on the residual from the previous layer to learn more features. In other words, it considers the input value plus current output to make predictions, which

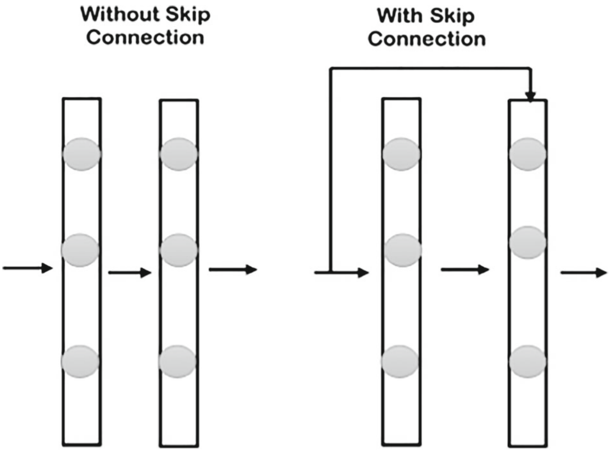


Fig. 13 Skip connection

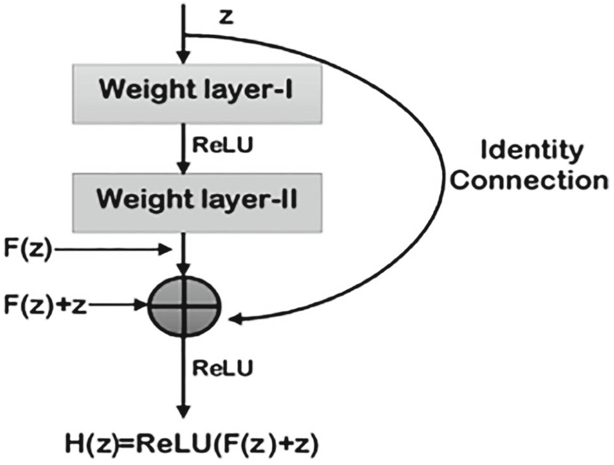


Fig. 14 Single residual block

solves the vanishing gradient problem and improves accuracy. The vanishing gradient problem arises while training the extremely deep networks; accuracy starts degrading as more layers are added. For this, the ResNet-50 has introduced the skip connection concept. The original input is added to the output of the convolution block in the skip connect of ResNet-50 as shown in Fig. 13. The skip connection allows the higher layer not to degrade as it lets the information flows from the lower layers to the higher layers. The input image size is 224x224x3 for this model in the proposed study. ResNet-50 comprises 5 stages. Each stage consists of a convolution block and identity block, and both have three convolution layers for each stage. The three layers have 1x1, 3x3, and 1x1 filters. The kernel of (1x1) is responsible for dimensions reduction and restoration. The kernel size of the convolution layer is 7x7, and the pooling layer is 3x3. The residual block is the main block that makes connections between the original input and the predictions, as shown in Fig. 14. According to

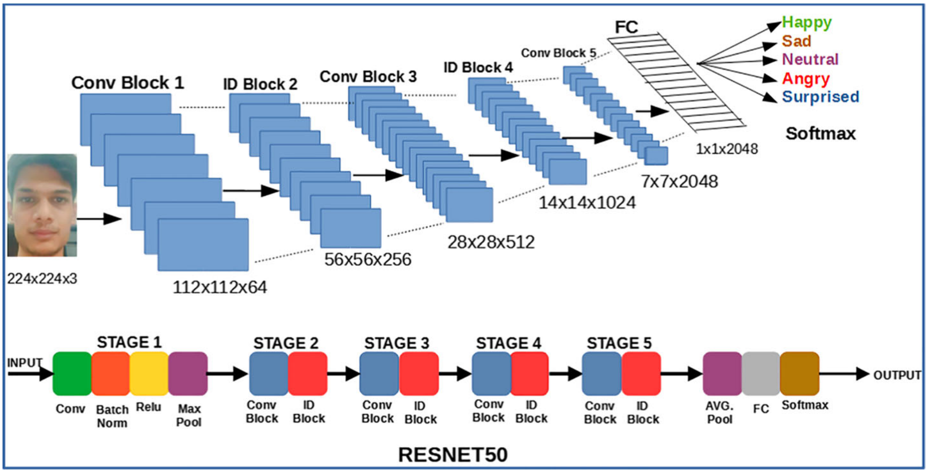


Fig. 15 Network architecture of ResNet-50

Fig. 14, the F(x) is the residual, and the ‘z’ variable is the prediction. When the original input equals ‘z’, the f(z) value is zero. Then, the ‘z’ value is copied by the identity connection. The five stages are followed by the average pooling layer and then the fully connected layer, the final layer. In this study, ResNet-50 is considered the base model to classify the different facial emotions of the online learner for the proposed real-time engagement detection system (Fig. 15).

4.3 Engagement detection system

The proposed system is inspired by the real-time problem of learner’s engagement levels during online studies. The proposed system uses the web camera to get real-time information about the facial emotion of the learners. This information is used to determine the engagement index (EI), which gives output in the form of two states: engaged or disengaged states and the level of engagement (in percentage). In our case, a total of six classes of facial emotions are recognised, and these classes have been further categorised as engaged and disengaged, as shown in Table 9.

The predicted emotion acts as input to decide the engagement states. The engagement index is calculated according to (5) based on the predicted emotion’s value to determine the engagement states.

$$EI = EP \times WE \tag{5}$$

- where EP= Emotion Probability (Emotion=Neutral, Angry, Sad, Happy, Surprised and fear) WE= Weight of corresponding Emotion

The EP (emotion probability) score is generated by a deep CNN classifier along with the corresponding emotion weight. Emotion Weight describes the emotional state’s value that reflects the engagement of a learner at that instant of time. The engagement percentage is calculated based on the engagement index given in (5). The Weight for corresponding emotion is scaled and represented in Table 10. An algorithm to calculate the engagement based on facial emotion recognition is given below.

Table 9 The Engagement categories based on the emotion pattern and EP value from Algorithm 1

Emotion	Engagement
Happy (EP<50)	Engaged
Surprised (EP>60)	Engaged
Neutral (EP>60)	Engaged
Neutral (EP>60)+Angry (EP<20)+Sad(EP>30	Engaged
Angry (EP>20)	Disengaged
Sad	Disengaged
Fear (EP>30)	Disengaged
Angry (EP>20)+Sad (EP>30)	Disengaged
Angry (EP>20)+Sad (EP>30)+Fear (EP>30)	Disengaged
Sad (EP>30) +Fear (EP>30)	Disengaged
Fear (EP>30)+Surprised (EP<60)	Disengaged
Neutral (EP<60)	Disengaged

Table 10 Weight for corresponding emotion table

Emotion	Weight Value
Happy	0.6
Surprised	0.6
Neutral	0.9
Angry	0.25
Fear	0.3
Sad	0.3

```
Declaration:
$EI \leftarrow Engagement\ Index$
$EP \leftarrow Probability\ of\ Emotion$
$WE \leftarrow Weight\ of\ corresponding\ emotion$
$\phi$={neural , happy, surprised, angry, fear, sad}
State = {Engaged, Disengaged}

Begin:

The engagement index is calculated as:

    EI= $(EP_{\phi} \times WE_{\phi} \times 100)$

If $(EP_{Neutral} > 60) \| (EP_{Happy} < 50) \| (EP_{Surprised} > 60)$ then:

        State $\rightarrow$ Engaged

        EI $> 0$

  else if $(EP_{Angry} > 20) \| (EP_{Fear} > 30) \| (EP_{Sad} > 30)$ then:

        State $\rightarrow$ DisengagedEI $> 0$

          EI $> 0$

  else

        State $\rightarrow$ Disengaged

       EI = 0
End
```

Algorithm 1 Engagement evaluation.

This is the detail of the background processing of the proposed system. In the front-end of the proposed system, the engagement state is predicted as either engaged or disengaged with level of engagement (in percentage) and the facial emotion categories are also highlighted based on the learner's facial expressions.

5 Experimental results and analysis

The experimental analysis of the proposed approach is presented in this section to evaluate the proposed system. Firstly, the experimental analysis is done by comparing the performance of the proposed model with previous work by re-implemented their work on different datasets. The effectiveness of the proposed emotion recognition approach is demonstrated by conducting extensive experiments on deep CNN-based models such as Inception-V3, VGG19, and ResNet-50. Then, the visualisation of the real-time learner engagement detection system using the proposed model is provided. Lastly, the comparison of the proposed system with the existing work is also discussed.

5.1 Evaluation metrics

The models have been evaluated on four performance metrics: Accuracy, Precision, Recall, and F1-Score [41, 42]. These metrics are defined in terms of false-negative (A), false-positive (B), true-negative (C), and true-positive (D). The description of performance metrics is given below.

Accuracy

$$\frac{A+B}{A+B+C+D} \tag{6}$$

Precision

$$\frac{A}{A+B} \tag{7}$$

Recall

$$\frac{A}{A+C} \tag{8}$$

F1-Score

$$F1\_Score = \frac{2 \times precision \times recall}{precision + recall} \tag{9}$$

5.2 Result analysis for FER-2013 dataset

The proposed model and previous work are re-implemented with the FER-2013 dataset. The proposed model is able to achieve an accuracy of 73.40%. The accuracy of the previous works is also increased by approx. 1.7% - 1.9% after embedding with MFACEXTOR. The comparison of the proposed model with some of the previous works on the FER-2013 dataset is provided in Table 11. As the proposed model is evaluated with three different models one by one, i.e., Inception-V3 (PROPOSED+Inception-V3), VGG19 (PROPOSED+VGG19), and RESENT50 (PROPOSED+ResNet-50). The accuracy and loss of the proposed model for FER-2013 dataset is shown in Figs. 16(a) and 16(b) respectively.

Table 11 Performance measure of FER-2013 dataset

S.NO.	MODEL	ACCURACY (in %)	Precision(in %)	Recall(in %)	F1_Score(in %)
1.	[23]	64.7	63.23	62.41	62.81
2.	[1]	66.4	65.12	65.01	65.06
3.	[1]	70	68.21	67.32	67.76
4.	**PROPOSED+ResNet-50**	**73.4**	**73.65**	**71.76**	**72.69**

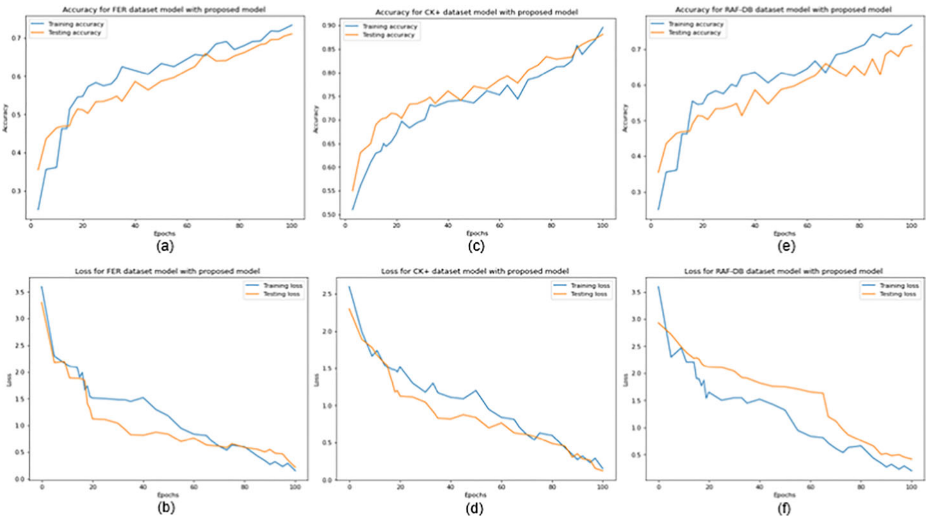


Fig. 16 Analysis of (a) training versus testing accuracy and (b) training versus testing loss for the FER-2013 dataset using the proposed models, (c) training versus testing accuracy and (d) training versus testing loss for the CK+ dataset using the proposed models at different epochs (e) training versus testing accuracy and (f) training versus testing loss for the RAF-DB dataset using the proposed models at different epochs

5.3 Result analysis for CK+ dataset

The proposed model and previous work are re-implemented with the CK+ dataset. 70% of the images are used as the training set, 10% of images as the validation set, and 20% of the images as the testing set. The PROPOSED+ResNet-50 model is able to achieve an accuracy of 89.56% among others. CK+ gives good accuracy because it's a laboratory-controlled dataset. The accuracy of the previous works is also increased by approx. 1.23% - 1.35% after adding MFACEXTOR. The comparison of the proposed model with some of the previous works on the CK+ dataset is provided in Table 12. The accuracy and loss of the proposed model for CK+ dataset is shown in Fig. 16(c) and Fig. 16(d) respectively.

5.4 Result analysis for RAF-DB dataset

The proposed model and previous work are re-implemented with the RAF-DB dataset for training and testing purposes. RAF-DB consists of multi-classes of images. The PROPOSED+ResNet-50 model can achieve an accuracy of 76.72%. The accuracy of the previous works is also increased by approx. 0.47% - 1.87% after adding MFACEXTOR.

Table 12 Performance measure of CK+ dataset

S.NO.	MODEL	ACCURACY (in %)	Precision(in %)	Recall(in %)	F1_Score(in %)
1.	[24]	80.76	79.97	78.48	79.22
2.	[9]	81.99	80.76	79.96	80.36
3.	[53]	83.53	82.93	81.29	82.10
4.	**PROPOSED+ResNet-50**	**89.56**	**88.92**	**88.92**	**88.77**

Table 13 Performance measure of RAF-DB dataset

S.NO.	MODEL	ACCURACY (in %)	Precision(in %)	Recall(in %)	F1_Score(in %)
1.	[16]	67.49	66.79	77.92	76.35
2.	[57]	69.36	68.76	67.86	68.30
3.	[49]	72.96	71.97	70.69	71.32
4.	**PROPOSED+ResNet-50**	**76.72**	**75.96**	**74.25**	**75.09**

The comparison of the proposed model with some of the previous works on the RAF-DB dataset is provided in Table 13. The accuracy and loss of the proposed model for RAF-DB dataset is shown in Fig. 16(e) and (f) respectively.

5.5 Result of own test data

Table 14 shows the evaluation results for Inception-V3, VGG19, ResNet-50 and PROPOSED+Inception-V3, PROPOSED+VGG19 and PROPOSED+ ResNet-50 in terms of accuracy, precision, recall and F1-score. The three deep learning models (Inception-V3, VGG19, and ResNet-50) are evaluated for facial expression recognition. From the results, among all the three models, PROPOSED+ResNet-50 achieved the highest accuracy of 92.32%, followed by PROPOSED+VGG19 with an accuracy of 90.14% followed by PROPOSED+ Inception-V3 with an accuracy of 89.11% for the test data. Similarly, the proposed model also significantly improves the f1-score, precision, and recall performance measures for the PROPOSED+ ResNet-50 model.

The confusion matrices of FER (facial expression recognition) predictions for the PROPOSED+Inception-V3, PROPOSED+VGG19 and PROPOSED+ResNet-50 architectures are shown in Figs. 17(a), 17(b), and 17(c). From the confusion matrix perspective, the best overall results is shown by PROPOSED+ ResNet-50 (Fig. 17(a)), followed by the PROPOSED+VGG19 (Fig. 17(b)) and PROPOSED+Inception-V3 (Fig. 17(c)). The highest scored class, regardless of the network model, is the “happy” class, followed by the “neutral” and “surprise” classes. Predicting “happy” and “neutral” classes is more accurate due to the large number of “happy” and “neutral” images with high variance in the training set. As observed from the confusion matrix, the classes with low variance can be mistaken for each other; for example, “fear” can be mistaken with “sad,” and “sad” can be mistaken with “surprise.” This happened due to less variation in shape of eyebrows and mouth for “fear” and “sad” classes. So, it becomes difficult to distinguish between such classes. This can be improved by adding more images for the same emotion but with variations in expressions. The training and and testing loss as well as accuracy for PROPOSED+Inception-V3, PROPOSED+VGG19 and PROPOSED+ResNet-50 are graphically represented in Figs. 18 (a) to 18(f). Where, Figs. 18(a) and 18(b) is for PROPOSED+Inception-V3 accuracy and

Table 14 Performance measure of Own dataset

S.NO.	MODEL	ACCURACY (in %)	Precision(in %)	Recall(in %)	F1_Score(inb%)
1.	[16]	75.49	75.79	77.92	76.35
2.	[57]	80.36	78.76	79.86	76.30
3.	[1]	79.96	77.97	77.69	78.32
4.	**PROPOSED+ResNet-50**	**90.83**	**90.40**	**90.61**	**92.32**

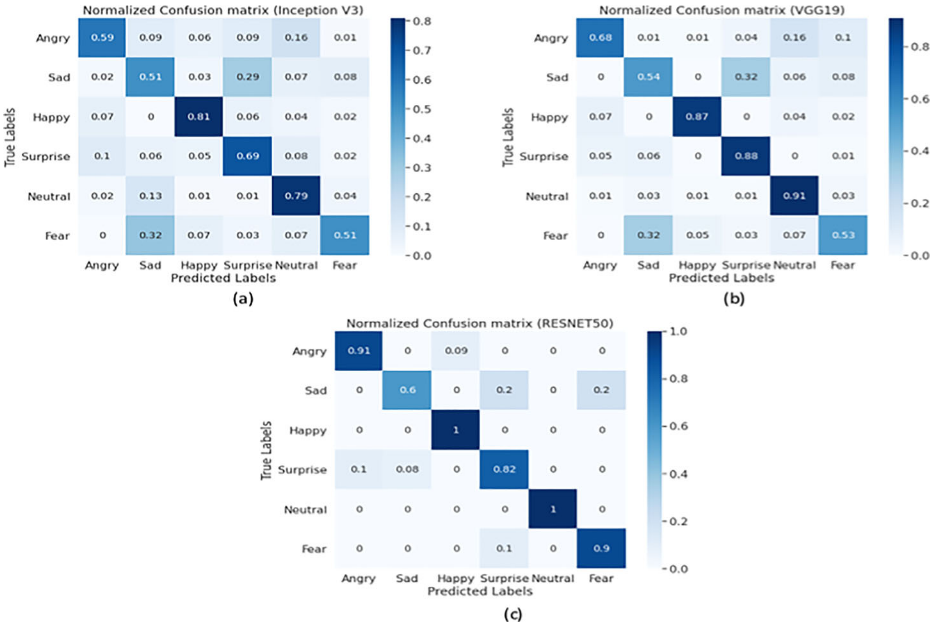


Fig. 17 Normalized confusion matrix for (a) PROPOSED+Inception-V3 (b) PROPOSED+VGG19 (c) PROPOSED+ResNet-50 models which trained with (FER2013+ (CK+)+RAF-DB) and tested on OWN dataset

loss respectively. Similarly, Figs. 18(c) and 18(d) represents accuracy and loss for PROPOSED+VGG19 and Figs 18(e) and 18(f) shows the highest accuracy and lowest loss for own dataset for PROPOSED+ResNet-50 model.

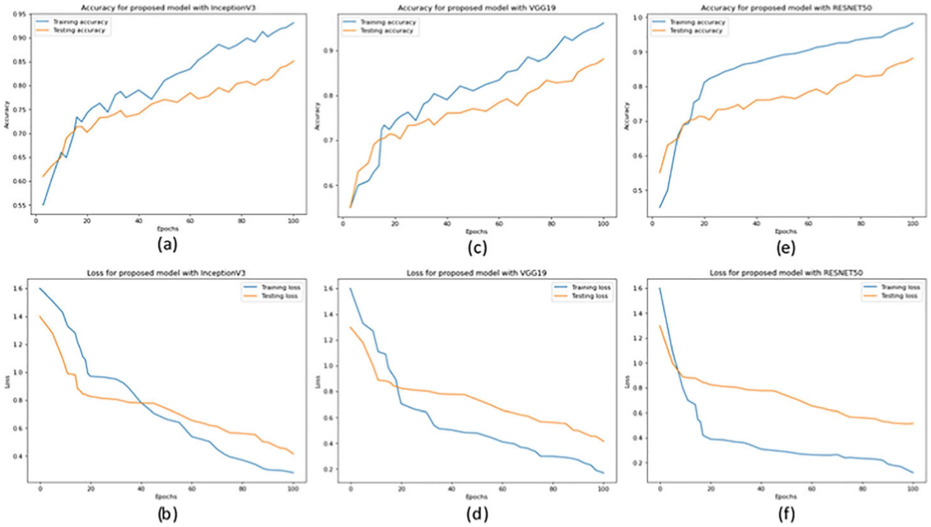


Fig. 18 Analysis of (a) training versus testing accuracy and (b) training versus testing loss for the own dataset using the Inception-V3 model, (c) training versus testing accuracy, and (d) training versus testing loss for the own dataset using the VGG19 model, (e) training versus testing accuracy and (f) training versus testing loss for the own dataset using the RESNET50 model at different epochs

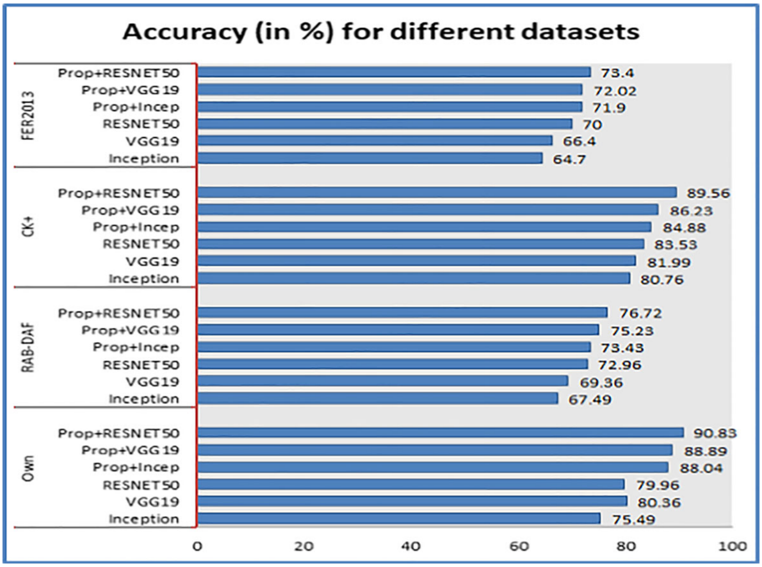


Fig. 19 Graphical representation of overall accuracy evaluation in percentage over different datasets

To summarise the results of different datasets, which are evaluated on different models, are represented in a form graph as shown in Fig. 19. In Fig. 19, it is clear that Proposed+ResNet-50 has out-performed on each dataset among all models. Hence, we will proceed with our next section, i.e., visual demonstration of the proposed engagement detection system with the Proposed+ResNet-50 model only as its accuracy was the highest.

6 Visualisation of the engagement detection system using facial emotion recognition

The trained model is deployed on a real-time system, and its visualisation is shown in Fig. 20. 20 undergraduate learners have participated in this study. Each learner watched an online

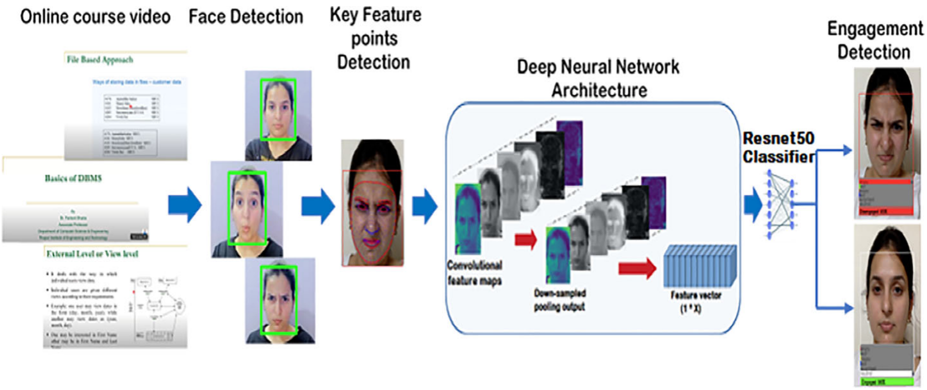
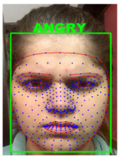
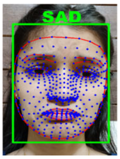
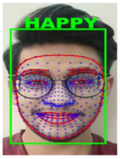
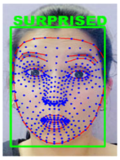
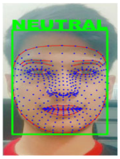
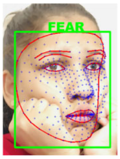

Online course video
Face Detection
Key Feature points Detection
Deep Neural Network Architecture
Engagement Detection
File Based Approach
Basics of DBMS
External Level or View level
Convolutional feature maps
Down-sampled pooling output
Feature vector (1*X)
Resnet50 Classifier

Fig. 20 Visual framework of real-time engagement detection system

learning video of different contents of Basic of DBMS, which is 29 minutes long. The online video consisted of engaging and informative content explained with the help of colourful diagrams and equations. The web camera captures the learner throughout the learning session to monitor the learner's engagement state. The engagement states are continuously predicted and the value of engagement level (in percentage) after every 20 seconds frame.

The confusion matrix for real-time emotion recognition of the different learners are shown in Table 15. The emotion labels are represented by the green boxes over the images.

Table 15 learners' facial emotion recognition results

Emotion	Emotion Detection					
	Angry	Sad	Happy	Surprise	Neutral	Fear
	**95**	1.50	0	0	0	3.5
	5.0	**75**	1.0	2.0	4.0	9.0
	0	0	**99**	0	1.0	0
	0	0.5	0.25	**99**	0	0.25
	0.5	0.5	2.0	0.5	**96**	0.5
	5.5	4.5	0	4.5	0.5	**85**

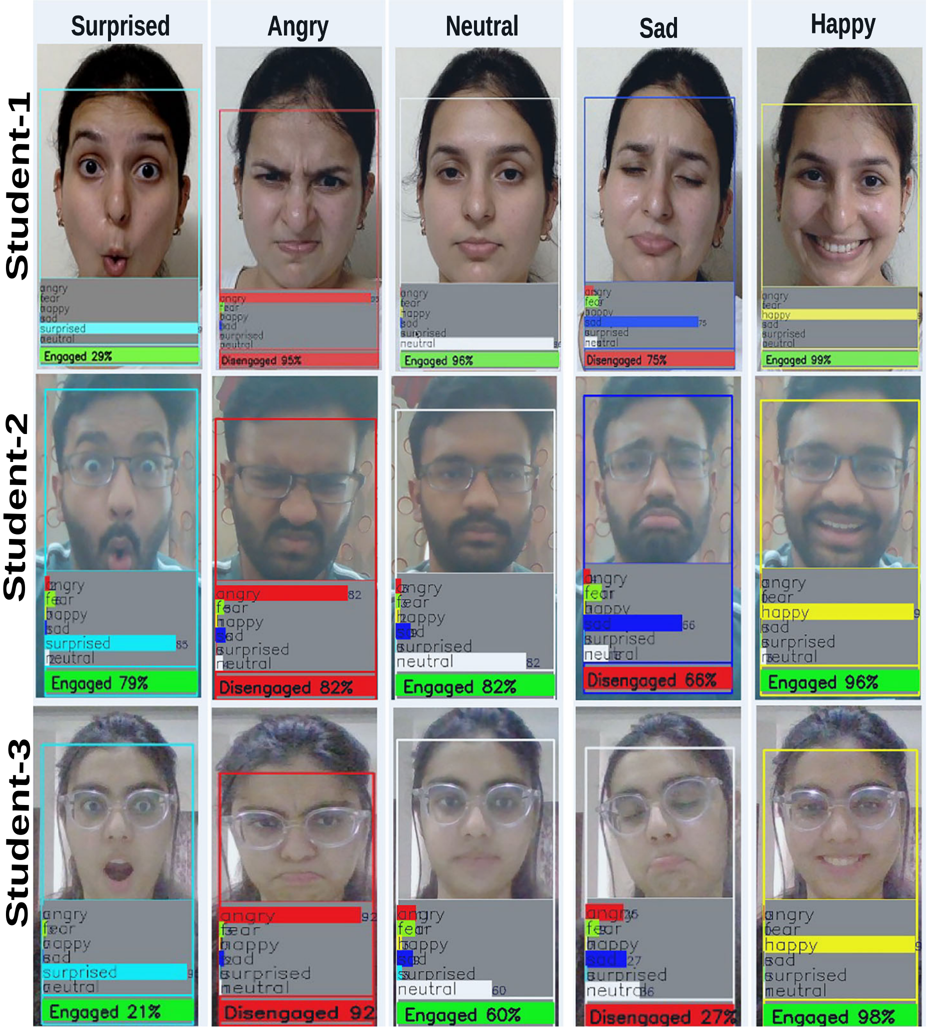


Fig. 21 Example of real-time engagement detection system of learner-1, learner-2 and learner-3

Figure 21 shows snapshot of different students along with the results obtained by proposed real-time learner engagement detection system. Figures 22(a), 22(b)and 22(c) shows the engagement index plots of the three learners for the whole video session. The emotion value (in %) and overall emotion pattern through the learning session is shown in Figs. 23(a) and 23(b) respectively.

7 Comparison with existing systems

The proposed learner engagement detection system using facial emotions has outperformed other existing engagement detection works in terms of performance accuracy as shown in Table 10.

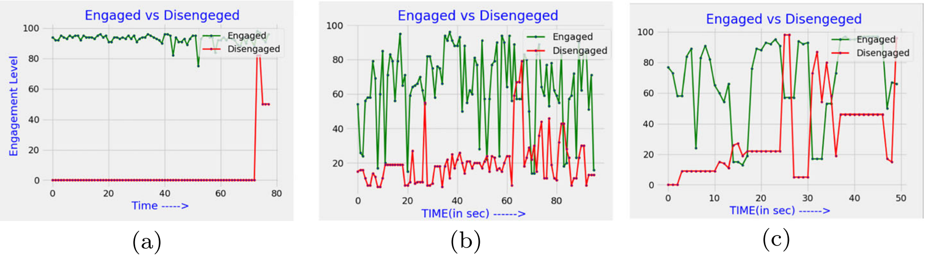


Fig. 22 Plotting of real-time engagement index for (a)learner-1 (b)learner-2 (c)learner-3 over the complete online learning video

Mohamad Nezami et al. [43], have presented an engagement model which considers ER dataset having 4627 total images in grey-scaled type. In this paper, authors have considered CNN models, VGGNET model, HOG+SVM model on Fer-2013, and ER dataset. According to testing results, the engagement model had given an accuracy of 72.38% which is the best among all. In consideration to other papers, [10], have used Local Directional Pattern (LDP), which is a robust and person independent edge feature extraction technique on image dataset. For dimensional reduction, Kernel Principal Component Analysis (KPCA) was applied. After this, a deep belief network (DBN) is implemented for the engagement classification using the features obtained from KPCA. The LDP-KPCA-DBN model shows efficient results on CK(+) dataset [2]. This dataset consists of 568 video snippets, each approximately 10 secs long and 800 pictures. The overall testing accuracy was 87.25%. ResNet+TCN model is proposed by [2] based on end-to-end Neural network architecture. FER-2013 and CK+ dataset was considered for evaluation. The major novelty in this architecture was that they had embedded the RESNET model in each frame and clubbed it at the TCN layer to classify the engagement. The overall accuracy was 63.9%. Liao et al. [34] proposed a model named Deep Facial Spatio-temporal Network (DFSTN) that evaluated the engagement prediction. This model combined two modules; in the first module: SE-ResNet-50 (SENet) was used to extract the spatial features. The second module used the Long Short Term Memory (LSTM) approach to generate global attention. The DFSTN model was performed on the RAF-DB dataset [34], and testing accuracy was 73.6%. These are recent outcomes of online learner engagement systems with different datasets and proposed models. Our proposed model has performed much better than the existing models

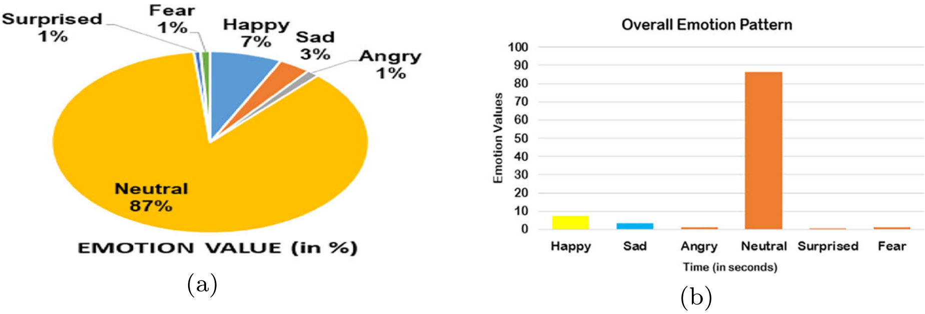


Fig. 23 Plotting of real-time engagement index over the complete video

Table 16 Comparison of the existing models with a proposed system for Real-time Engagement Detection

S.NO.	WORK	DATASET	MODEL	ACCURACY (in %)
1.	[43]	FER-2013	VGGNET	72.38
2.	[10]	CK+	LDP-KPCA-DBN	87.25
3.	[2]	FER-2013+CK(+)	ResNet+TCN	63.9
4.	[34]	RAF-DB	DFSTN	73.6
5.	**Proposed Model**	**FER-2013+ CK(+) + RAF-DB+ OWN Dataset**	**ResNet-50**	**92.32**
			VGG19	**90.14**
			Inception-V3	**89.11**

to counter this. Our proposed model, Fer-2013, CK+ RAF-DB, and our dataset are considered to train the proposed facial emotion recognition model in real-time. We have evaluated the Inception-V3, VGG19, and ResNet-50 as they are well-defined deep CNN models for emotion recognition in real-time. After experimentation, the accuracy for Inception-V3, VGG19, and ResNet-50 is 89.11%, 90.14%, and 92.32%, respectively. One of the main reasons for getting these accuracies is better face detection processing done by the pre-trained Faster R-CNN model on the WIDER face dataset. The input image is processed very efficiently. As background area in the image is ignored, and only the important features are considered for facial expression recognition. The dimension is lower to a certain extent to provide a stream-line face-points encoding with the help of MFACEXTOR. Later, the six emotion classes are classified, and their output information is used to calculate the engagement index to predict an online learner's engagement state. From Table 16, a brief overview can be envisioned for the comparison of existing work with our proposed model. And, the proposed system with ResNet-50 proved to be one of the efficient models for facial emotion recognition based real-time engagement detection system.

8 Conclusion and future scope

With the increasing usage of the digital platform during the COVID-19 pandemic, one of the biggest challenges is: to have a system that determines the engagement of the online learners where no instructor is physically present. This paper proposes a new approach for a real-time engagement detection system based on deep learning models. The learner's engagement is detected by automatically analysing the facial emotions while studying online. The facial emotion state is recognised by observing the change of facial expression in a real-time learning environment. This paper automatically recognises facial expressions during the ongoing learning video session. The system analyses the facial expressions through the built-in web cameras and uses that information to calculate the engagement index (EI). The engagement index gives the output in the form of "engaged" and "disengaged." The six basic emotions contribute to predicting the engagement states, which will act as real-time feedback. This information will help the instructor know about the learner's online learning experience. This will also contribute in making the online learning experience better by supporting the online learners when they are found to be not engaged with the learning content. The proposed system has utilised the Faster R-CNN for face detection and MFACXTOR for face point extraction. The proposed system is trained on FER-2013, CK+, and own dataset and tested on the own created dataset for automatically recognising facial emotion. Deep learning models, namely, Inception-V3, VGG-19, and ResNet-50, has been evaluated to compare their performance measures. The experimental results showed that the ResNet-50 outperformed Inception-V3 and VGG-19 models for classifying the facial emotion in real-time scenarios on their own dataset with an accuracy of 92.32% and other publicly available and benchmarked datasets. The proposed system was tested on 20 learners in an online learning scenario, and it correctly detected the "engaged" and "disengaged" states based on automatic facial emotion recognition. The proposed approach has also outperformed the existing work's methods. In the future, the retrieved information of the proposed system can be combined with the information provided by other sensors such as heart rate and EEG signals. The proposed model can be applied to learners with special needs. Measuring engagement based on eye movements, body movements, and facial emotions can also be done. A large dataset can be created to train and test the proposed approach.

Declarations

Conflict of Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abbassi N, Helaly R, Hajjaji MA, Mtibaa A (2020) A deep learning facial emotion classification system:a vggnet-19 based approach. In: 2020 20Th International conference on sciences and techniques of automatic control and computer engineering STA, IEEE pp 271–276
2. Abedi A, Khan SS (2021) Improving state-of-the-art in detecting student engagement with resnet and tcn hybrid network, CoRR. https://doi.org/1048550/arXiv2104101
3. Adedoyin OB, Soykan E (2020) Covid-19 pandemic and online learning: the challenges and opportunities. Interact Learn Environ, pp 1–13
4. Aguilera-Hermida AP (2020) College students' use and acceptance of emergency online learning due to covid-19. Int J Educ Res Open 1:100,011
5. Altuwairqi K, Jarraya SK, Allinjawi A, Hammami M (2021) Student behavior analysis to measure engagement levels in online learning environments. SIViP, pp 1–9
6. Aneja D, Colburn A, Faigin G, Shapiro L, Mones B (2016) Modeling stylized character expressions via deep learning. In: Asian conference on computer vision, Springer, pp 136–153
7. Bawa P (2016) Retention in online courses: Exploring issues and solutions—a literature review. Sage Open 6(1):2158244015621,777
8. Botelho AF, Baker RS, Heffernan NT (2017) Improving sensor-free affect detection using deep learning. In: International conference on artificial intelligence in education, Springer, pp 40–51
9. Chowdary MK, Nguyen TN, Hemanth DJ (2021) Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Comput Applic, pp 1–18
10. Dewan MAA, Lin F, Wen D, Murshed M, Uddin Z (2018) A deep learning approach to detecting engagement of online learners. In: 2018 IEEE Smartworld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, pp 1895–1902
11. Diego-Mas JA, Fuentes-Hurtado F, Naranjo V, Alcañiz M (2020) The influence of each facial feature on how we perceive and interpret human faces. i-Perception 11(5):2041669520961,123
12. Ekman P, Oster H (1979) Facial expressions of emotion. Annual review of psychology 30(1):527–554
13. Eom SB, Ashill N (2016) The determinants of students' perceived learning outcomes and satisfaction in university online education: An update. Decis Sci J Innov Educ 14(2):185–215
14. Fabian Benitez-Quiroz C, Srinivasan R, Martinez AM (2016) Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5562–5570
15. Fish J, Brimson J, Lynch S (2016) Mindfulness interventions delivered by technology without facilitator involvement:what research exists and what are the clinical outcomes? Mindfulness 7(5):1011–1023

16. Ghosh S, Dhall A, Sebe N (2018) Automatic group affect analysis in images via visual attribute and feature networks. In: 2018 25Th IEEE International conference on image processing (ICIP), IEEE, pp 1967–1971
17. Giannopoulos P, Perikos I, Hatzilygeroudis I (2018) Deep learning approaches for facial emotion recognition:A case study on fer-2013. In: Advances in hybridization of intelligent methods, Springer, pp 1–16
18. Gupta S (2018) Facial emotion recognition in real-time and static images. In: 2018 2nd International conference on inventive systems and control (ICISC), IEEE, pp 553–560
19. Gupta S, Kumar P (2021) Attention recognition system in online learning platform using eeg signals. In: Emerging technologies for smart cities, Springer, pp 139–152
20. Hai L, Guo H (2020) Face detection with improved face r-cnn training method. In: 2020 the 3rd International conference on control and computer vision, pp 22–25
21. Hew KF (2016) Promoting engagement in online courses: What strategies can we learn from three highly rated moocs. Br J Educ Technol 47(2):320–341
22. Huang Q (2016) Learners' perceptions of blended learning and the roles and interaction of f2f and online learning. Ortesol Journal 33:14–33
23. Hung JC, Lin KC, Lai NX (2019) Recognizing learning emotion based on convolutional neural networks and transfer learning. Applied Soft Computing 84:105,724
24. Jmour N, Zayen S, Abdelkrim A (2021) Deep neural networks for a facial expression recognition system. In: Innovative and intelligent technology-based services for smart environments–smart sensing and artificial intelligence, CRC Press, pp 134–141
25. Kim HR, Kim YS, Kim SJ, Lee IK (2018) Building emotional machines:Recognizing image emotions through deep neural networks. IEEE Trans Multimedia 20(11):2980–2992
26. Kiuru N, Spinath B, Clem AL, Eklund K, Ahonen T, Hirvonen R (2020) The dynamics of motivation, emotion, and task performance in simulated achievement situations. Learning and Individual Differences 80:101,873
27. Kundu A, Bej T (2021) Covid-19 response:students' readiness for shifting classes online, Corporate governance: The international journal of business 760 in society
28. Lee J, Kim S, Kiim S, Sohn K (2018) Spatiotemporal attention based deep neural networks for emotion recognition. In: 2018 IEEE International conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 1513–1517
29. Li B, Lima D (2021) Facial expression recognition via resnet-50. Int J Cognit Comput Eng 2:57–64
30. Li M, Li X, Sun W, Wang X, Wang S (2021) Efficient convolutional neural network with multi-kernel enhancement features for real-time facial expression recognition. J Real-Time Image Process pp 1–12
31. Li Q, Liu YQ, Peng YQ, Liu C, Shi J, Yan F, Zhang Q (2021) Real-time facial emotion recognition using lightweight convolution neural network. In: Journal of Physics: Conference Series, IOP Publishing, vol 1827. pp 012130
32. Li S, Deng W, Du J (2017) Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2852–2861
33. Li Y, Zeng J, Shan S, Chen X (2018) Occlusion aware facial expression recognition using cnn with attention mechanism. IEEE Trans Image Process 28(5):2439–2450
34. Liao J, Liang Y, Pan J (2021) Deep facial spatiotemporal network for engagement prediction in online learning. Appl Intell, pp 1–13
35. Liu P, Han S, Meng Z, Tong Y (2014) Facial expression recognition via a boosted deep belief network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1805–1812
36. Long F, Bartlett MS (2016) Video-based facial expression recognition using learned spatiotemporal pyramid sparse coding features. Neurocomputing 173:2049–2054
37. Manseras R, Palaoag T, Malicdem A (2017) Class engagement analyzer using facial feature classification. no November, pp 1052–1056
38. Masi I, Trn AT, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: European conference on computer vision, Springer, pp 579–596
39. Minaee S, Minaei M, Abdolrashidi A (2021) Deep-emotion:facial expression recognition using attentional convolutional network. Sensors 21(9):3046
40. Mishra L, Gupta T, Shree A (2020) Online teaching-learning in higher education during lockdown period of covid-19 pandemic. Int J Educ Res Open 1:100,012
41. Mittal M, Siriaraya CP, Lee abd, Kawai Y, Yoshikawa T, Shimojo S (2019) Accurate spatial mapping of social media data with physical locations. IEEE Int Conf on Big Data (Big Data), pp 4113–4116
42. Mittal M, de Prado R, Kawai Y, Nakajima S, Muñoz-Expósito J (2021) Machine learning techniques for energy efficiency and anomaly detection in hybrid wireless sensor networks. Energies, pp 1–21

43. Mohamad Nezami O, Dras M, Hamey L, Richards D, Wan S, Paris C (2019) Automatic recognition of student engagement using deep learning and facial expression. In: Joint european conference on machine learning and knowledge discovery in databases, Springer, pp 273–289
44. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter conference on applications of computer vision (WACV), IEEE, pp 1–10
45. Mukhopadhyay M, Pal S, Nayyar A, Pramanik PKD, Dasgupta N, Choudhury P (2020) Facial emotion detection to assess learner's state of mind in an online learning system. In: Proceedings of the 2020 5th international conference on intelligent information technology, pp 107–115
46. Murshed M, Dewan MAA, Lin F, Wen D (2019) Engagement detection in e-learning environments using convolutional neural networks. In: 2019 IEEE Intl Conf on Dependable, Autonomic and secure computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), IEEE, pp 80–86
47. Priya RV, Bharat R (2021) A novel geometric fuzzy membership functions for mouth and eye brows to recognize emotions. Concurr Comput Pract. Exp 33(14):e5610
48. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn:Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28:91–99
49. Riaz MN, Shen Y, Sohail M, Guo M (2020) Exnet:an efficient approach for emotion recognition in the wild. Sensors 20(4):1087
50. Rudovic O, Lee J, Dai M, Schuller B, Picard RW (2018) Personalized machine learning for robot perception of affect and engagement in autism therapy. Science Robotics:3(19)
51. Sharma A, Gupta S, Kaur S, Kumar P (2019) Smart learning system based on eeg signals. In: International conference on advances in computing and data sciences, Springer, pp 465–476
52. Sindagi VA, Patel VM (2018) A survey of recent advances in cnn-based single image crowd counting and density estimation. Pattern Recogn Lett 107:3–16
53. Sugianto N, Tjondronegoro D, Tydd B (2018) Deep residual learning for analyzing customer satisfaction using video surveillance
54. Torres II (2020) Emotional needs of online students:a phenomenological study of graduate level nontraditional students. PhD thesis, The Chicago School of Professional Psychology
55. Turabzadeh S, Meng H, Swash RM, Pleva M, Juhar J (2018) Facial expression emotion detection for real-time embedded systems. Technologies 6(1):17
56. Vanneste P, Oramas J, Verelst T, Tuytelaars T, Raes A, Depaepe F, Van den Noortgate W (2021) Computer vision and human behaviour, emotion and cognition detection:A use case on student engagement. Mathematics 9(3):287
57. Wu Y, Zhang L, Chen G, Michelini PN (2021) Unconstrained facial expression recogniton based on cascade decision and gabor filters. In: 2020 25Th international conference on pattern recognition (ICPR), IEEE, pp 3336–3341
58. Yang S, Luo P, Loy CC, Tang X (2016) Wider face:A face detection benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5525–5533
59. Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 23(10):1499–1503
60. Zheng X, Hasegawa S, Tran MT, Ota K, Unoki T (2021) Estimation of learners' engagement using face and body features by transfer learning. In: International conference on human-computer interaction, Springer, pp 541–552

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.