Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 24, Issues 4, pp. 957–970, DOI 10.12694/scpe.v24i4.2470

DETECTING ACADEMIC AFFECTIVE STATES OF LEARNERS IN ONLINE LEARNING ENVIRONMENTS USING DEEP TRANSFER LEARNING

PURUSHOTTAMA RAO K * AND B JANET †

Abstract. Online Learning Environments (OLEs) have become essential in global education, especially during and after the COVID-19 pandemic. However, OLEs face a challenge in recognizing student emotions, hindering educators' ability to provide effective support. To address this issue, researchers emphasize the importance of a balanced dataset and a precise model for academic emotion detection in OLEs. However, the widely-used DAiSEE dataset is imbalanced and contains videos captured in well-lit environments. However, real-time observations reveal students' diverse lighting conditions and proximity to cameras. Consequently, models trained on DAiSEE dataset exhibit poor accuracy. In response, this work suggests a customized DAiSEE dataset and proposes the Xception-based transfer learned model and AffectXception model. Our customization process involves selectively extracting single-label frames with intensity levels 2 or 3 from the original DAiSEE dataset. To enhance dataset diversity and tackle the issue of dataset imbalance, we meticulously apply data augmentation techniques on these extracted frames. This results in frames that showcase variations in lighting, both low and high, as well as diverse camera perspectives. As a result, the customized DAiSEE dataset is now well-balanced and exceptionally suitable for training deep learning models to detect academic emotions in online learners. Then we trained and tested both proposed models on this dataset. The AffectXception model outperforms existing models, achieving significant improvements. For Boredom, Engagement, Confusion, and Frustration, it attains accuracy rates of 77%, 79.28%, 83.76%, and 91.87%, respectively. Additionally, we evaluate the AffectXception model on the Online Learning Spontaneous Facial Expression Database (OL-SFED), obtaining competitive results across various emotion classes. This work empowers educators to adjust their content and delivery methods based on learners' emotional states, resulting in more effective and informative online sessions. As OLEs continue to play a crucial role in education, our approach enhances their capacity to address students' emotional needs.

Key words: Academic Affective States, Affective Computing, Deep Learning, Fine-tuning, Online Learning Environment, Transfer Learning

1. Introduction. Education is crucial for the overall development of an individual and plays a vital role in shaping up their future. It provides individuals with the necessary knowledge, skills, values, and attitudes to succeed in life [1, 2]. The classroom is a crucial setting for education as it provides a structured environment for learning and interaction with teachers and peers. It also plays a vital role in promoting critical thinking, problem-solving, and collaboration skills [3, 4]. So it is the responsibility of the learners to be engaged and attentive in the class time to gain more knowledge and to achieve course outcomes [5]. In the class time, students show positive expressions or emotions such as attention, engagement, and understanding to indicate that he or she is comprehending the material. They express Negative emotions such as confusion, frustration, or boredom to indicate that the student is struggling and may need additional support or clarification [6, 7]. Table 1.1 gives the cases when students deliver either positive or negative emotions. Therefore it concludes that emotions play a crucial role in the learning process and it is important for a teacher to understand the emotions of their students to make the class effective and to improve the academic performance of learners [8, 9]. A seasoned teacher is capable of identifying the understanding level of all students by observing their affective states during the class in traditional classroom environment and it is one of the main reasons for the success of these offline classroom environments [10]. Teachers can use this emotion feedback to make adjustments such as slowing down or speeding up the pace, repeating the subject, or changing their way of delivering the concept by using innovative teaching methods and active based methods, which all assist to keep the session interesting and lively [4].

*Department of Computer Applications, National Institute of Technology Tiruchirappalli, Tamilnadu - 620015, India. (k.purushottamaa@gmail.com). Questions, comments, or corrections to this document may be directed to that email address.
†Department of Computer Applications, National Institute of Technology Tiruchirappalli, Tamilnadu - 620015, India. (janet@nitt.edu).

Table 1.1: Cases when students deliver either positive or negative emotions

Students express Positive emotions when they:	Students express Negative emotions when they:
Adhere to the lecture	Cannot keep up with the lecture
Satisfied with the lecture	Confused and want the instructor to repeat it
Capable of understanding the lecturer's thoughts	Try to get the lecturer's assistance
Would like to emphasize how well the lecture was received	Unable to keep up with the lecturer's pace

However, with the COVID-19 epidemic and also the enrichment and development of Internet technology the learning environment has shifted from purely traditional classroom to a hybrid learning environment (HLE) [11, 12]. Online learning environments (OLEs) have made it possible for students all around the world to pursue their education during and after the COVID-19 pandemic. Google Meet, WebEx, MS Teams, Zoom are some of the examples for online learning environments. OLEs are widely accessible, adaptive, and versatile due to their non-physical nature [12, 13]. Though these are having some benefits over offline classrooms, E-learning, unlike traditional classrooms, lacks the ability to capture students' emotions and dynamically take steps to improve their engagement. It can be challenging for teachers using an e-learning system to identify students' emotions and issues, particularly those brought on by uncertainty or apathy. With the help of prior research, it has been determined that e-learning systems typically ignore the Emotion, a crucial part of the learning process [14, 15]. As a result, even the important lectures or courses frequently results in the student's disinterest. Furthermore, during lectures, e-learning students have been shown to feel distinct negative emotional states such as apathy, indifference, drowsiness, and confusion [16]. As a result, the instructor's and instructional aids' efforts are rendered ineffective [17].

It is vital for online educators to accurately and efficiently detect their online learners' engagement status in order to provide personalized pedagogical support. However, teachers are finding it difficult to gauge pupils' engagement without having direct interaction or eye contact with them. As a result, a model that detects students' affective states in virtual classes is needed. The primary contributions of this paper are: (1) We create a customized DAiSEE dataset, by extracting single-label frames (frames that only belong to one class) from original DAiSEE dataset [18] and applied data augmentation techniques to make the new dataset balanced; (2) Using the customized DAiSEE dataset and the OL-SFED dataset [33, 34], which are more appropriate for academic affective states, we evaluate the suggested AffectXception model in detecting academic emotions; (3) and we compare the proposed model's performance against that of existing methods on original DAiSEE dataset. Our obtained results demonstrate that, compared to previous efforts, our technique classifies different academic emotional states more accurately.

The remaining portions of the paper are organized as follows: The background and a review of relevant existing literature are provided in Section 2. The details of the dataset modified and the model's training procedure for identifying students' academic affective states are presented in Section 3. Section 4 of the paper details the experiments and their results. The last section summarizes the findings and potential directions for future research.

2. Related Work. The works on (1) Correlation between Students' Emotions and their Engagement, (2) Students' Emotion Recognition in E-Learning Platforms, and (3) Academic Affective States are reviewed in this section.

According to existing studies, there is a positive correlation between learners' emotional status, their attention and their academic performance. According to the existing research, ensuring learner engagement is one of the most important aspects of quality online education [19]. As per [20], one of the parameters to determine students' engagement during lecture is their emotions. Emotions can have a significant impact on a student's engagement and academic achievement by influencing their interest and motivation in a lecture or course [21, 22]. It has been demonstrated that pleasant feelings such as happiness and neutrality can have favorable benefits on students, whereas unpleasant emotions such as sadness, anger, and boredom can have detrimental consequences [21]. Authors [23] demonstrated the importance of knowing emotional feedback to both students and teachers in another study. The authors also show that students are more motivated, engaged, and perform well in assessments while experiencing pleasant emotions. Students, on the other hand, are unmotivated, disengaged, and underperform on assessments with negative emotions.

Previous studies have proved that there is a need to recognize the students' emotions in online classrooms in order to find out the comprehension level of both individual student and the overall classroom. One of the primary means of recognizing and interpreting human emotions is facial expressions [24]. In order to detect the emotional states of students during lessons, majority of researchers have utilized seven basic emotions namely happy, anger, fear, sadness, joy, disgust, and surprise [5, 25]. The authors used deep learning-based algorithms to detect student involvement by detecting at their facial expressions. Students expressing 'Sad' emotions were found to be the least interested, while those displaying 'Happy' and 'Neutral' emotions were found to be the most involved [25]. Authors [26] developed a framework for evaluating students' emotions using facial expressions in an online lecture. They have presented the overall class feedback towards the lecture based on students' emotions. Authors have proposed a CNN based model to detect the emotions of students in E-learning platforms. Then they have used another module in their proposed architecture to send this emotional feedback of students during the class to the corresponding teacher that helps to engage the class in a more productive way. Authors have used single face images in their proposed work [15]. By combining global and local face features, [27] established a framework called Multi-region Attention Transformation to recognize facial expressions.

Few scholars have stated that the fundamental emotions for human being play a minimal role when we want to detect the emotional state of learners in educational settings [28]. Authors found that non-basic emotions including engagement, boredom, confusion, and frustration happened five times more frequently than basic emotions in classroom. Based on the results it was suggested that these non-basic affective states are more suitable for detecting learners' academic affective states [29]. In another work [30], authors have also stated that these affective academic states are more relevant than basic emotions to detect students' emotional state. In another work [20], authors proved that the predominant emotions experienced by students were curiosity, frustration, boredom, confusion, happiness, and anxiety, with contempt, anger, sadness, fear, disgust, eureka, and surprise being less frequently reported. Therefore, academic emotion is a term used to describe a variety of emotional reactions that students experience in academic activities like learning [32].

It is important to have a relevant dataset in order build a robust model that meets our requirements. But unfortunately there is a lack of academic affective states datasets. Then "Dataset for Affective States in E-Environments" (DAiSEE), a dataset created by [18], was released in 2015. It is the first multi-label video classification dataset and contains 9068 video clips from 112 users. It was created to identify the affective states of boredom, confusion, engagement, and frustration in users "in the wild." Each of the four affective states in the dataset has a label at one of four different levels: very low, low, high, and very high. They have used InceptionNet, C3D, C3D with fine tuning, LRCN and EmotionNet models to analyse and give benchmark results for DAiSEE dataset. Results have showed that there is a huge scope to get even better accuracies. An online learning spontaneous facial expression database (OL-SFED) [33, 34] for academic emotions is created in another study to address the lack of academic emotions datasets. This dataset contains 30,184 images and 1274 videos with 224 x 224 and 1280 x 720 resolutions, respectively. Authors have used different CNN models to give benchmark results on the created dataset.

Authors [35] mentioned the difficulties online teachers face in identifying the comprehension level of students during the lecture. He proposed a model to find the facial expressions including Enlightened, Confused and Bored expressions of students and a method to transfer this emotional feedback to the corresponding teacher. Another work uses a deep learning system to evaluate academic emotions of distance learners using data from online learning behavior. The multimodal weighted feature fusion algorithm is applied to extract the data, and an academic cognition motivation model and online learning emotion measurement framework are constructed. This study finds a positive correlation between distance learners' academic emotions and their learning outcomes [7]. To assist teachers in real-time monitoring of student engagement, authors suggested a framework for an academic emotions monitoring system. In order to train a CNN model, they built a dataset containing six academic emotional states from classroom lecture videos and incorporated relevant samples from the BAUM-1, DAiSEE, and YawDD databases in the dataset. Their approach analyzed student facial expressions captured in lecture recordings to evaluate the level of participation of the entire class [36]. The authors [37] opted for academic affective states instead of basic emotions due to their significance in detecting students' involvement

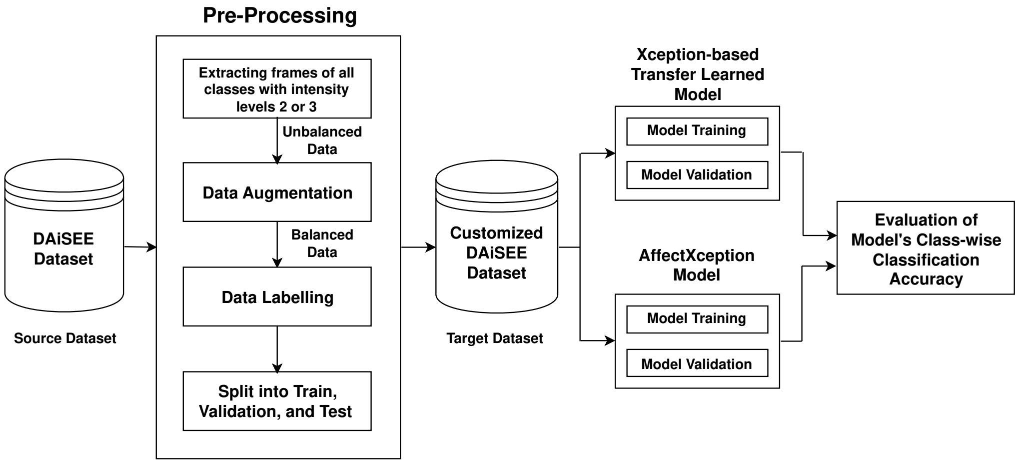

Fig. 3.1: Workflow of Proposed method

in videos in learning settings. They employed Conv3D, VGG16, ResNet50, and LSTM models, and trained them with the DAiSEE dataset. Results showed that Conv3D has achieved better results than the remaining models. The authors in [38] proposed a method for automatically estimating student engagement in online class by analyzing videos. This approach took into account facial expressions, head pose, and gaze movements to determine engagement levels. The authors suggested a model that predicts engagement through an LSTM network, trained using the EmotiW 2019 challenge dataset.

3. Proposed Work. In this section, we will discuss the issues with the existing dataset, creation of target dataset, and the workflow of our proposed method. Figure 3.1 shows the architecture diagram of the proposed work.

3.1. Dataset. In this work, we utilized the publicly accessible dataset named "Dataset for Affective States in E-Environments" (DAiSEE) [18] to train and evaluate our deep learning model. As its name implies, this dataset is well-suited for training models to detect affective states in e-learning environments. It includes 9068 labelled video clips of size 15 GB featuring 112 students participating in online lectures while seated in front of a webcam. The dataset encompasses four ideal classes for our work: Boredom, Engagement, Confusion, and Frustration, which are frequently expressed by learners during lectures. Each affective state is labelled with one of four intensity levels: 0 (very low), 1 (low), 2 (high), and 3 (very high), reflecting the observed level of emotion in the given video.

Based on the literature review, it has been discovered that the majority of researchers focusing on academic affective states utilized the DAiSEE dataset to train their models in order to determine student engagement in learning environments [5, 6, 8, 9, 12, 36, 37]. A commonly noted challenge among researchers working with the DAiSEE dataset is the issue of dataset imbalance [10, 14, 30, 37]. Furthermore, a common limitation observed in existing works that utilized the DAiSEE dataset is the accuracy of the models employed.

3.2. Pre-processing for Customized DAiSEE Dataset Creation. The pre-processing procedure seeks to enhance the data's quality and produce customised datasets from the original dataset in accordance with requirements in order to get better outcomes [39, 40, 41]. Our aim in this work is to use single-label classification method to improve the robustness of the deep learning model and attain higher accuracy. The dataset includes two categories of frames: single-label frames (frames that belong to only one class, such as [0, 2, 0, 0]), and multi-label frames (frames that belong to multiple classes, such as [2, 1, 0, 1]). See Fig. 3.2 for sample multi-label images in the dataset. To apply single-label classification method, we have only considered single-label frames that have high (2) or very high (3) intensity levels. Two diagrams Fig. 3.3 and Fig. 3.4 below present the number of single-label frames and single-label frames with high (2) or very high (3) intensity

Fig. 3.2: Sample multi-label images with annotations

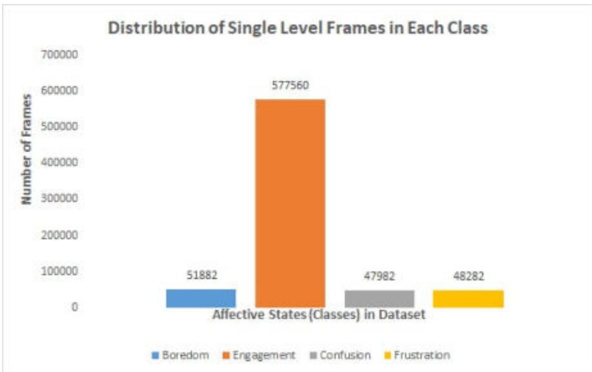

Fig. 3.3: Number of Single-label Frames in Each Class in DAiSEE Dataset

levels per class in the dataset.

It is evident from the information above that the DAiSEE dataset is significantly unbalanced. If a deep learning model is trained with the current unbalanced dataset, it can lead to biased results. This means that the model may perform well on the majority class but poorly on the minority class, as it has seen more examples of the majority class during training. This can result in poor accuracy or incorrect predictions for the minority class, and the overall performance of the model will be affected. Therefore, a balanced dataset is crucial for training an effective model, which requires presenting the model with roughly equal amounts of data samples for each class.

This work's important module is determining what kind of data must be extracted from the videos in order to have related and significant features for the detection of academic affective states in learning contexts. First we have extracted the frames with intensity level High (2) or Very High (3) of each class. We have noticed that most of the participants in online class videos are either in bright or dim environments, and are either up close or far away from the camera. See below Fig. 3.5 for reference.

So we have chosen and applied Brightness, Zoom, and Horizontal Flip data augmentation techniques on available images of classes Boredom, Confusion, and Frustration to make sure our new customized dataset contains the images with above said conditions and to make the dataset balanced. And we have applied Under-sampling technique on Engagement class since it is having huge number data samples. After data augmentation

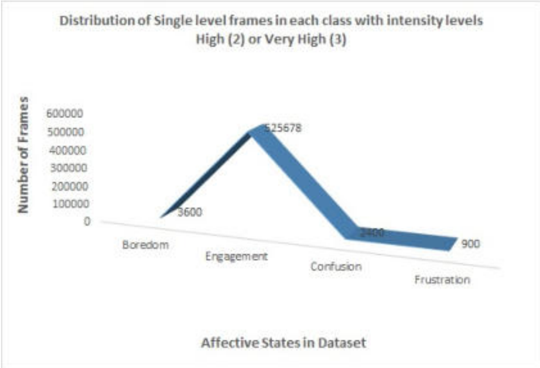

Fig. 3.4: Number of Single-label Frames in Each Class with intensity levels High (2) or Very High (3) in DAiSEE Dataset

Fig. 3.5: Samples to show online class participants in different conditions. (a) and (b) in high light environment; (c) and (d) shows in low light; (e) and (f) are far away from camera; and (g) and (h) are close to the camera

we have partitioned data into Train, Validation, and Test with 70:20:10 ratio. The number of frames in each of the four classes after balancing the dataset is given in below Table 3.1. There are 40,320 images in the training data where each class contains 10,080 images, 11,520 images in the validation data, and 5,760 images in the test data. Then we relabeled the dataset with values 1 for the specific class and 0 for all remaining classes.

The algorithm 1 for creating customized DAiSEE dataset from the original DAiSEE dataset is given below:

3.3. Methodology. Here, authors used Xception [42], a deep learning model for image classification, to detect learners' academic affective states in online classrooms. The Xception model uses depth wise separable convolutions to learn complex representations of input data while reducing computational cost by applying a single filter to each channel of the input feature map. The model also includes residual connections, global

Table 3.1: Class-wise distribution of customized DAiSEE dataset samples into train, validation, and test sets

Class	Train	Validation	Test	Total
Boredom	10080	2880	1440	14400
Engagement	10080	2880	1440	14400
Confusion	10080	2880	1440	14400
Frustration	10080	2880	1440	14400

Algorithm 1 Customized DAiSEE Dataset

```
dataset = load_DAiSEE_dataset() {Load the DAiSEE dataset}
frames = convert_videos_to_frames(dataset) {Convert the videos in the dataset into frames}
selected_frames = select_frames_classwise_with_levels_2_or_3(frames) {Select class wise frames with levels 2 or 3}
Class_distribution = count_classwise_frames(selected_frames) {Find the class distribution from selected_frames}
augmented_frames = apply_data_augmentation(frames, techniques=[”brightness”, ”zoom”, ”horizontal_flip”]) {Apply data augmentation techniques to address data imbalance}
train_data, validation_data, test_data = split_dataset(augmented_frames, ratio=(0.7, 0.2, 0.1)) {Split the selected frames into train, validation, and test sets with an 70:20:10 ratio}
train_labels = label_frames(train_data) {Label the frames in the train set}
validation_labels = label_frames(validation_data) {Label the frames in the validation set}
test_labels = label_frames(test_data) {Label the frames in the test set}
```

average pooling, and fully connected layers. The Xception model consists of an input layer, entry flow (Fig. 3.6), middle flow (Fig. 3.7), exit flow (Fig. 3.8), and output layer, with each flow containing a series of depth wise separable convolutions, activation functions, and residual connections. The output layer is a fully connected layer that outputs the predictions for the task at hand using the Softmax activation function.

4. Experimental Results. In order to assess Xception-based transfer learning model and AffectXception model performance on a customized DAiSEE dataset, an experimental study of the suggested approach is described in this section.

4.1. Xception-Based Transfer Learned Model. We used the four-class customized dataset mentioned above to train the Xception deep learning model to predict learners' affective states. Figure 4.1 shows the workflow of Xception based transfer learned model. There are 40,320 images in the training data where each class contains 10,080 images. The input dimension was 299 × 299 × 3, where 299 × 299 represents the image resolution, and 3 represents the three channels of RGB color image. First we have trained the lower layer of the model by freezing the remaining layers with the specified parameters. The model is trained with 10 epochs by setting a learning rate of 0.0001 to learn the parameters. To train the model authors have used the loss function by using Eq. 4.1:

$$Loss = \Sigma_1^C log(f(s)_i) \tag{4.1}$$

where C denotes the total number of classes, y_i represents ground truth value, and $f(s)_i$ indicates predicted value.

A model is generated following the training procedure and is predicted to be able to categorize images of test data. The metric we have used to assess the model performance is Accuracy. Refer Eq. 4.2. Accuracy is defined as the number of correct predictions made by the model as a percentage of the total number of predictions.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{4.2}$$

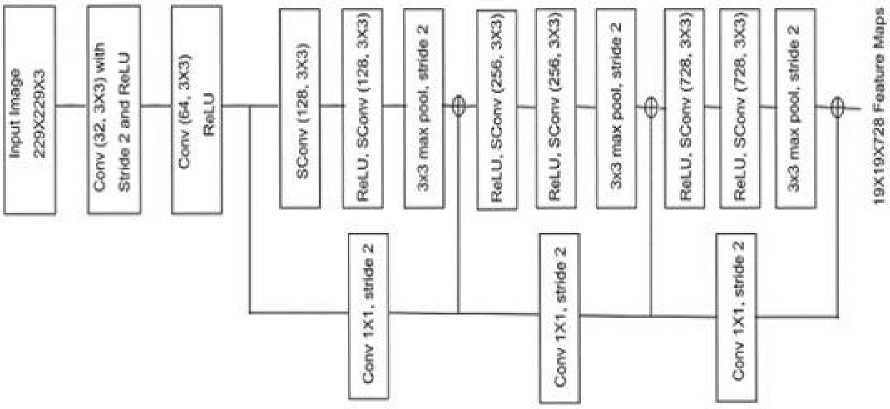

Fig. 3.6: Architecture diagram of Entry Flow

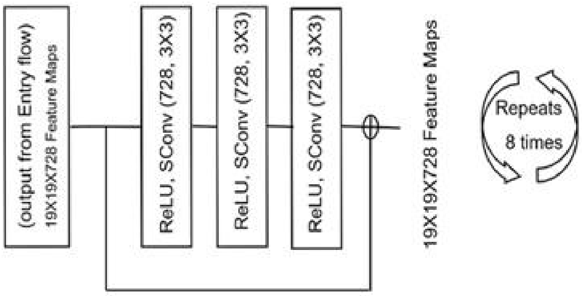

Fig. 3.7: Architecture diagram of Middle Flow

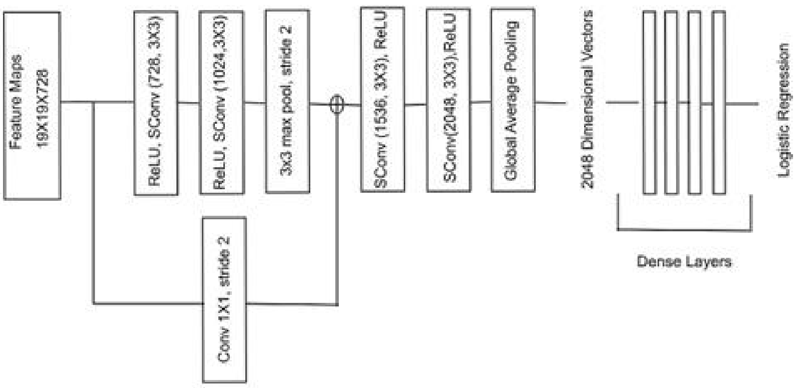

Fig. 3.8: Architecture diagram of Modified Exit Flow

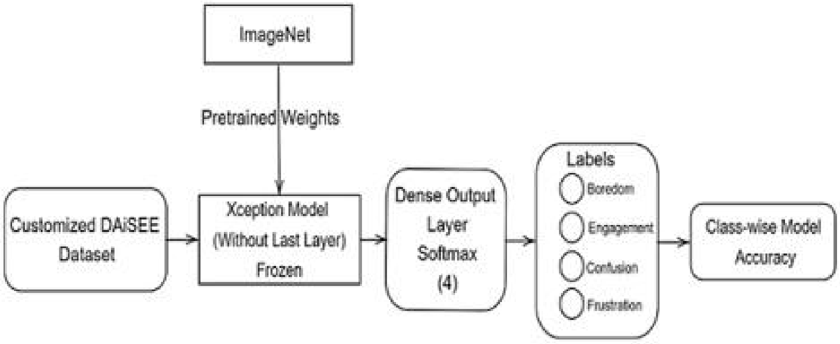

Fig. 4.1: Workflow of Xception-based transfer learned model

Table 4.1: Class-wise accuracies of Xception-based Transfer Learned Model

Class Name	Obtained Class-wise Accuracy
Boredom	72%
Engagement	75.6%
Confusion	75%
Frustration	68.38%

Class wise results obtained for Xception-based transfer learning model are given in Table 4.1 below. Please refer Table 4.3 for a comparison of these results with prior works.

4.2. AffectXception Model. By seeing the above unimpressive results of the Xception-based transfer learning model, we thought that while it is possible to employ a pre-trained deep learning model like Xception trained on the ImageNet dataset for face emotion recognition, doing so may not be the best option. The reason is that the ImageNet dataset is designed for object recognition, and the categories it contains (such as dogs, cats, and cars) may not necessarily be relevant for academic emotion detection. Therefore, a model trained on ImageNet may not be optimized to detect affective states of a person or learner. Affective computing requires a model that is trained specifically on facial images datasets like fer-2013, AffectNet datasets for basic emotions, and DAiSEE, OL-SFED datasets for academic affective states. Training a deep learning model on datasets like DAiSEE, OL-SFED will result in a more accurate and specialized academic affective states detection model as it will have learned to recognize the specific features that are indicative of different emotions of learners during lectures. So, we chose fine tuning as a solution to this problem and to achieve better results. In this instance, we unfreeze the entire model and retrain it using the new customized DAiSEE dataset with a learning rate of 0.00001 and 20 epochs rather than freezing all the layers except last one. Results obtained with the AffectXception model are given below. Table 4.2 gives the obtained class-wise accuracies of AffectXception model. Figure 4.2 shows the comparison of both the models' results.

4.3. Comparison with Existing Methods. With the AffectXception model trained on our customized DAiSEE dataset, we got the better results over the existing methods to detect academic affective states in E-learning platforms on DAiSEE dataset. The results and comparison of the proposed AffectXception model trained on customized DAiSEE dataset with the most recent state-of-the-art approaches are shown in the Table 4.3 below. Here, we utilized Accuracy as a metric for comparing our results to prior studies.

4.4. Performance Analysis of OL-SFED Dataset. The proposed AffectXception model's performance is also assessed on Online Learning Spontaneous Facial Expression Database (OL-SFED) [33, 34]. A total of 30,184 facial expression images with a resolution of 224X224 and 1274 videos with a resolution of 1280X720

Table 4.2: Class-wise accuracies of AffectXception model

Class Name	Obtained Class-wise Accuracy
Boredom	77%
Engagement	79.28%
Confusion	83.76%
Frustration	91.87%

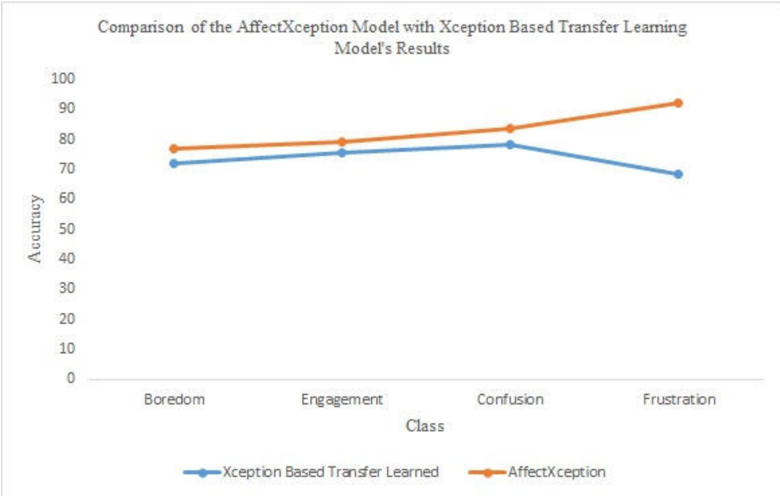

Fig. 4.2: Comparison of Xception-Based Transfer learned and AffectXception models results

are included in the OL-SFED. There were 82 healthy students who voluntarily took part in the experiment, including 53 female and 29 male students. Five significant academic emotions have been chosen in the database as labels for facial expressions: Confusion, Distraction, Happy, Neutral, and Tired. The details of the number of images that each class contains is given in the below Fig. 4.3.

The above figure makes it clear that the OL-SFED dataset is considerably imbalanced. But it is required to have a balanced dataset if we train an efficient model to detect academic affective states of students' in online learning environments. And as seen in Fig. 4, most participants in online class recordings are either in bright or dark settings, and they are either close to or far from the camera. To make sure our dataset comprises of the images with the aforementioned conditions and to ensure the dataset is balanced, we have chosen and used Brightness, Zoom In, Zoom Out, and Horizontal Flip data augmentation techniques on available photographs of each class. The number of images after applying data augmentation techniques on OL-SFED dataset are given in Table 4.4.

Then we have trained the AffectXception model with a learning rate of 0.00001, batch size of 32, and with 20 epochs. Class-wise results obtained by AffectXception model on OL-SFED dataset are given Table 4.5 below.

5. Conclusion. Because teachers can analyze students' facial expressions to determine how attentive they are and change their teaching style appropriately, traditional classrooms continue to be a successful educational setting. The accessibility, affordability, and adaptability of online learning, particularly in the wake of the Covid-19 outbreak, have increased its popularity. One significant distinction between traditional classroom learning and online learning which is hybrid now is that during virtual sessions, teachers are unable to gauge participants' levels of participation in real-time. As a result, it could be more challenging to tell whether or not students are paying attention.

In this work, we have worked on Dataset for Affective States in E-Environments (DAiSEE) to build an

Table 4.3: Performance comparison with existing methods. Best results are given in bold

Method	Academic Affective States			
	Boredom	Engagement	Confusion	Frustration
DenseAttNet [5]	54.27%	63.59%	69.22%	78.58%
GaussianNB Classifier [30]	69%	**84%**	81%	84%
InceptionNet Frame Level [18]	36.5%	47.1%	70.3%	78.3%
InceptionNet Video Level [18]	32.3%	46.4%	66.3%	77.3%
C3D [18]	47.2%	48.6%	67.9%	78.3%
C3D Fine Tuning [18]	45.2%	56.1%	66.3%	79.1%
LRCN [18]	53.7%	57.9%	72.3%	73.5%
Multi-level Classification Approach [43]	63.82%	77.28%	81.95%	86.6%
Xception-based Transfer Learned Model (Ours)	72%	75.6%	75%	68.38%
AffectXception (Ours)	**77%**	79.28%	**83.76%**	**91.87%**

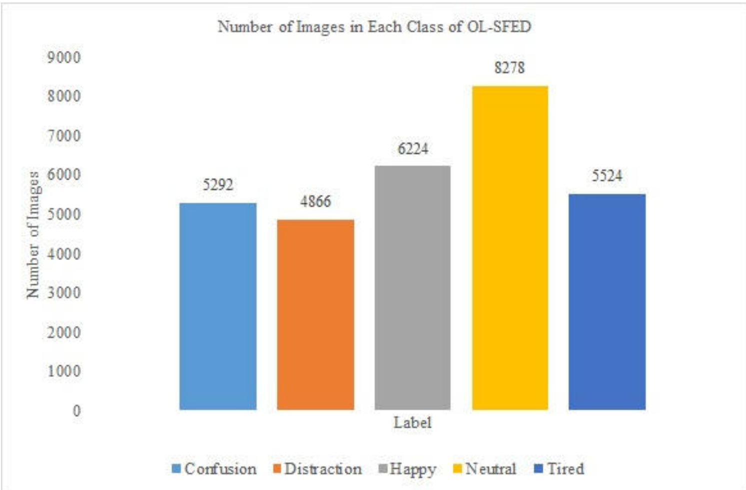

Fig. 4.3: Number of Images in each academic emotion label of OL-SFED

Table 4.4: Class-wise distribution of balanced OL-SFED images into train, validation, and test sets

Class	Train	Validation	Test	Total
Confusion	10500	3000	1500	15000
Distraction	10500	3000	1500	15000
Happy	10500	3000	1500	15000
Neutral	10500	3000	1500	15000
Tired	10500	3000	1500	15000

Table 4.5: Class-wise accuracies of AffectXception model on OL-SFED

Class Name	Obtained Class-wise Accuracy
Confusion	80.01%
Distraction	80%
Happy	80.24%
Neutral	93.01%
Frustration	70.2%

effective model that can detect the academic affective states of online learning learners. We have customized the dataset by extracting the frames those represents a single emotion strongly. Then we have applied Brightness, Zoom In, Zoom Out, and Horizontal Flip data augmentation techniques on extracted frames to make the dataset balanced and more suitable for academic emotion detection. To identify the academic affective states of learners in online learning environments, we trained both the Xception based transfer learned model and the AffectXception model on a bespoke DAiSEE dataset. In terms of class-wise performance, the AffectXception model outperformed the Xception-based transfer learnt model and cutting-edge techniques. Findings from the AffectXception model for boredom, engagement, confusion, and frustration are 77%, 79.28%, 83.76%, and 91.87%, respectively. The performance of the AffectXception model on the online learning spontaneous facial expression database was then examined (OL-SFED). For the classes of Confusion, Distraction, Happy, Neutral, and Tired, the model has scored 80.01%, 80%, 80.24%, 93.01%, and 70.2%, respectively.

Though the obtained results are better than existing works on DAiSEE dataset, there is still a chance to improve the results. Future work includes developing comprehensive academic emotions dataset, detecting individual student academic affective states during class, and applying our learners' emotion detection models with edge devices in real-time classroom scenarios.

REFERENCES

[1] Raghupathi, Viju, and Wullianallur Raghupathi, *The influence of education on health: an empirical assessment of OECD countries for the period 1995–2015*, Archives of Public Health, vol. 78, no. 1, pp. 1-18, 2020.

[2] Idris, Fazilah, Zaharah Hassan, Azizah Ya'acob, Saran Kaur Gill, and Noor Aziah Mohd Awal, *The role of education in shaping youth's national identity*, Procedia-Social and Behavioral Sciences, vol. 59, pp. 443-450, 2012.

[3] Rao, K. Purushottama, and B. Janet, *Teacher Assistance System to Detect Distracted Students in Online Classroom Environment*, 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, 2022, pp. 1743-1749.

[4] Umarale, Dhaval, Shyam Sodhani, Amodh Akhelikar, and Reeta Koshy, *Attention Detection of Participants during Digital Learning Sessions using Edge Computing*, International Conference on IoT based Control Networks and Intelligent Systems (ICICNIS 2020), 2020.

[5] Mehta, Naval Kishore, Shyam Sunder Prasad, Sumeet Saurav, Ravi Saini, and Sanjay Singh, *Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement*, Applied Intelligence, vol. 52, no. 12, pp. 13803-13823, 2022.

[6] Thiruthuvanathan, M., Balachandran Krishnan, and M. A. D. Rangaswamy, *Engagement detection through facial emotional recognition using a shallow residual convolutional neural networks*, International Journal of Intelligent Engineering and Systems, vol. 14, no. 2, 2021.

[7] Sun, Yang, and Shaoze Wang, *Study on the Distance Learners' Academic Emotions Using Online Learning Behavior Data*, Mobile Information Systems, vol. 2022, 2022.

[8] Hu, Yan, Zeting Jiang, and Kaicheng Zhu, *An Optimized CNN Model for Engagement Recognition in an E-Learning Environment*, Applied Sciences, vol. 12, no. 16, pp. 8007, 2022.

[9] Batra, Sarthak, Hewei Wang, Avishek Nag, Philippe Brodeur, Marianne Checkley, Annette Klinkert, and Soumyabrata Dev, *DMCNet: Diversified model combination network for understanding engagement from video screengrabs*, Systems and Soft Computing, vol. 4, pp. 200039, 2022.

[10] Ai, Xusheng, Victor S. Sheng, and Chunhua Li, *Class-attention Video Transformer for Engagement Intensity Prediction*, 2022, arXiv preprint arXiv:2208.07216.

[11] Samuel Kolawole Ajiboye, David Obafemi Adebayo, Saidu Abubakar, *Teachers' Assessment of Inattentive Classroom Behavior Among Primary School Students in Ilorin City, Kwara State, Nigeria*, Mimbar Sekolah Dasar, vol. 7, no. 2, pp. 172-183, 2020, doi: 10.17509/ mimbar-sd.v7i2.24101

[12] Ma, Xiaoyang, Min Xu, Yao Dong, and Zhong Sun, *Automatic student engagement in online learning environment based on neural turing machine*, International Journal of Information and Education Technology, vol. 11, no. 3, pp. 107-111, 2021.

[13] Abedi, Ali, Chinchu Thomas, Dinesh Babu Jayagopi, and Shehroz S. Khan, *Bag of States: A Non-sequential Approach to Video-based Engagement Measurement*, 2023, arXiv preprint arXiv:2301.06730.

[14] Abedi, Ali, and Shehroz S. Khan, *Detecting disengagement in virtual learning as an anomaly*, 2022, arXiv preprint arXiv:2211.06870.

[15] Sun, Ai, Ying-Jian Li, Yueh-Min Huang, and Qiong Li, *Using facial expression to detect emotion in e-learning system: A deep learning method*, Emerging Technologies for Education: Second International Symposium, SETE 2017, Held in Conjunction with ICWL 2017, Cape Town, South Africa, September 20–22, 2017, Revised Selected Papers 2, pp. 446-455. Springer International Publishing, 2017.

[16] M. Ali Akber Dewan, Mahbub Murshed and Fuhua Lin, *Engagement detection in online learning: a review*, Smart Learning Environments, vol. 6, no. 1, pp. 1-20, 2019, https://doi.org/10.1186/s40561-018-0080-z.

[17] Krithika.L.B, Lakshmi Priya GG, *Student Emotion Recognition System (SERS) for e-learning improvement based on*

learner concentration metric, International Conference on Computational Modeling and Security (CMS 2016), Procedia Computer Science, vol. 85, pp. 767-776, 2016.

[18] Gupta, A., A. D'Cunha, K. Awasthi, and V. Balasubramanian, *Daisee: Towards user engagement recognition in the wild*, 2016, arXiv preprint arXiv:1609.01885.

[19] Yousra Banoor Rajabalee, Mohammad Issack Santally, Frank Rennie, *A study of the relationship between students' engagement and their academic performances in an eLearning environment*, E-Learning and Digital Media, vol. 17, no. 1, pp. 1–20, 2020, doi: 10.1177/2042753019882567.

[20] Dixson, Marcia D., *Creating effective student engagement in online courses: What do students find engaging?*, Journal of the Scholarship of Teaching and Learning, pp. 1-13, 2010.

[21] Pekrun, R., and Linnenbrink-Garcia, L., *Academic Emotions and Student Engagement*, Handbook of Research on Student Engagement, pp. 259–282, 2012.

[22] Khawlah Altuwairqi, Salma Kammoun Jarraya, Arwa Allinjawi, Mohamed Hammami, *A new emotion–based affective model to detect student's engagement*, Journal of King Saud University – Computer and Information Sciences, vol. 33, no. 1, pp. 99-109, 2021.

[23] Marta Arguedas, Thanasis Daradoumis and Fatos Xhafa, *Analyzing How Emotion Awareness Influences Students' Motivation, Engagement, Self-Regulation and Learning Outcome*, Educational Technology & Society, vol. 19, no. 2, pp. 87–103, 2016, ISSN 1436-4522 (online) and 1176-3647 (print).

[24] Ram Krishn Mishra, Siddhaling Urolagin, J. Angel Arul Jothi, Pramod Gaur, *Deep hybrid learning for facial expression binary classifications and predictions*, Image and Vision Computing, vol. 128, pp. 104573, 2022, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2022.104573.

[25] Prakhar Bhardwaj, P.K. Gupta, Harsh Panwar, Mohammad Khubeb Siddiqui,Ruben Morales-Menendez, Anubha Bhaik, *Application of Deep Learning on Student Engagement in e-learning environments*, Computers and Electrical Engineering, vol. 93, pp. 107277, 2021.

[26] Weiqing Wang, Kunliang Xu, Hongli Niu, and Xiangrong Miao, *Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation*, Complexity, vol. 2020, pp. 1-9, 2020, https://doi.org/10.1155/2020/4065207.

[27] Yifei Guo, Jian Huang, Mingfu Xiong, Zhongyuan Wang, Xinrong Hu, Jihong Wang, Mohammad Hijji, *Facial expressions recognition with multi-region divided attention networks for smart education cloud applications*, Neurocomputing, vol. 493, pp. 119-128, 2022, https://doi.org/10.1016/j.neucom.2022.04.052.

[28] Wei, Qinglan, Bo Sun, Jun He, and Lejun Yu, *BNU-LSVED 2.0: Spontaneous multimodal student affect database with multi-dimensional labels*, Signal Processing: Image Communication, vol. 59, pp. 168-181, 2017.

[29] D'Mello, Sidney, and Rafael A. Calvo, *Beyond the basic emotions: what should affective computing compute?*, in CHI'13 extended abstracts on human factors in computing systems, 2013, pp. 2287-2294.

[30] Joshi, Nidhi, Niklas Beecken, Hawa Bah, Frank Steinicke, and Juliane Degner, *Advanced Emotion Analytics of Virtual Group Meetings involving Intelligent Virtual Agents*, 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE, 2022, pp. 344-350.

[31] D'Mello, Sidney K., Blair Lehman, and Natalie Person, *Monitoring affect states during effortful problem solving activities*, International Journal of Artificial Intelligence in Education, vol. 20, no. 4, pp. 361-389, 2010.

[32] Pekrun, R., Goetz, T., Titz, W.; Perry, R.P., *Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research*, Educational Psychologist, vol. 37, no. 2, pp. 91–105, 2002.

[33] Bian, Cunling, Ya Zhang, Deliang Wang, Yi Liang, Bo Wu, and Weigang Lu, *An academic emotion database and the baseline evaluation*, 13th International Conference on Computer Science & Education (ICCSE), IEEE, 2018, pp. 1-6.

[34] Bian, Cunling, Ya Zhang, Fei Yang, Wei Bi, and Weigang Lu, *Spontaneous facial expression database for academic emotion inference in online learning*, IET Computer Vision, vol. 13, no. 3, pp. 329-337, 2019.

[35] Haobang Wu, *Real Time Facial Expression Recognition for Online Lecture*, Wireless Communications and Mobile Computing, vol 2022, Article ID 9684264, 2022.doi: https://doi.org/10.1155/2022/9684264

[36] Pabba, Chakradhar, and Praveen Kumar, *An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition*, Expert Systems, vol. 39, no. 1, pp. e12839, 2022.

[37] Werlang, Pablo S., and Patrícia A. Jaques, *Student Engagement Recognition from Videos: A Comparison Between Deep Learning Neural Network Architectures*, IEEE Computer Society Technical Community on Learning Technology, vol. 21, no. 3, pp. 7-12, 2021.

[38] Buono, Paolo, Berardina De Carolis, Francesca D'Errico, Nicola Macchiarulo, and Giuseppe Palestra, *Assessing student engagement from facial behavior in on-line learning*, Multimedia Tools and Applications, pp. 1-19, 2022.

[39] Mishra, Ram Krishn and Raj, Harshit and Urolagin, Siddhaling and Jothi, J Angel Arul and Nawaz, Nishad, *Cluster-based knowledge graph and entity-relation representation on tourism economical sentiments*, Applied Sciences, vol. 12, no. 16, pp. 8105, 2022, https://doi.org/10.3390/app12168105.

[40] Mishra, Ram Krishn and Jothi, J Angel Arul and Urolagin, Siddhaling and Irani, Kayan, *Knowledge based topic retrieval for recommendations and tourism promotions*, International Journal of Information Management Data Insights, vol. 3, no. 1, pp. 100145, 2023, https://doi.org/10.1016/j.jjimei.2022.100145.

[41] Ram Krishn Mishra, *Deep Learning Model for Multiclass Classification of Diabetic Retinal Fundus Images Using Gradient Descent Optimization*, Advances in Signal Processing and Communication Engineering: Select Proceedings of ICASPACE 2021, pp. 27-35, 2022.

[42] Chollet, François, *"Xception: Deep learning with depthwise separable convolutions*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251-1258.

[43] Solanki, Naman, and Souraj Mandal, *Engagement Analysis Using DAiSEE Dataset*, 17th International Conference on

Control, Automation, Robotics and Vision (ICARCV), IEEE, 2022, pp. 223-228.

Edited by: Rajni Mohana
Special issue on: Sentiment Analysis and Affective computing in Multimedia Data on Social Network
Received: Jul 31, 2023
Accepted: Sep 26, 2023