Use of the Student Engagement as a strategy to optimize online education, applying a supervised machine learning model using facial recognition

Noboa Andrés[1 [0000−0001−7118−620X]], Gonzalez Omar[1 [0000−0002−8086−2422]], and Tapia Freddy[1 [0000−0001−9591−3563]]

Universidad de las Fuerzas Armadas ESPE, Sangolqui, Ecuador
Departamento de Ciencias de la Computación
{aenoboa1,ofgonzalez1,fmtapia}@espe.edu.ec

Abstract. The engagement of a student is a vital part of an online learning environment, however, because of the spatial separation between instructors and students in an online environment, it is often very difficult to measure the level of engagement in online learning. Higher levels of engagement are often related to a better sense of well-being, and more related to emotions like happiness. In this paper, two state-of-the-art models that can help measure the engagement and emotions are tested, in this case, we investigated the suitability of two popular models: XCeption Architecture proposed for the DAiSEE dataset, and the DeepFace Emotion Recognition model. Both models are then applied in a real-life test, using real students in an online class, we measured their emotions using the models and then compared the results to obtain the effectiveness of correctly measuring the engagement and emotions of the students. We interpret the findings from our experimental results based on psychology concepts in the field of engagement.

Keywords: Engagement · Machine Learning · Online Environment · XCeption · DAiSEE · DeepFace · emotion recognition

1 Introduction

People's life has changed quickly, and they need to keep learning new knowledge, and new skills. In this case, traditional school education and traditional classroom teaching cannot meet their needs fully. With the development of computer technology, network technology, and multimedia technology, online learning has emerged. Different from traditional classroom learning, online learning break through the limit of time and space, with its flexibility, mobility, and convenience, thus becoming an important way to learn. However, from the data obtained from MOOC (Massive Open Online Course) platforms, the completion rate of the courses is less than 10% and the participation of the students is relatively low [6].

In the words of Hu et al [6] , one of the most important characteristics of an online learning environment is the Engagement of a student. It's composed of 3 dimensions: behavioral engagement, cognitive engagement and emotional engagement. In the online learning environment, the relationship and communication between teachers and students are not enough to get emotional engagement from the student. Therefore, it is crucial to carry out an analysis of the 'engagement' of a student in the online environment; for this, the use of a supervised model of machine learning is proposed, which allows measuring with certainty how engaged a student really is.

Nowadays, Facial emotion recognition fundamentally identifies the emotion and reaction based on situations as well as the environment to which they belong [12].In addition, a technology that goes hand in hand with the previous term and is presented in the proposed articles [4] is machine learning, which is one of the emerging technologies that is considered to have an impact of 90% in the next 4 years[11].

Taking this background into account, the present work focuses on developing a supervised machine learning model for facial emotion recognition to measure 'Student Engagement' in students who are in an online learning environment. The processes for the execution of the project will be to perform the training of a Convolutional Neural Network (CNN) model using images of faces that are in a freely accessible data set to perform the classification of engagement, Dataset for Affective States in E-Environments (DAiSEE), we will also test a commonly used open-source tool "Deep Face" to measure other relevant emotions, then we validate the classification results obtained by both models using common validation techniques such as the matrix of confusion, we then apply the models to generate results based on the emotions of the students. Finally, we generate conclusions from the data obtained with the intervention of two expert psychologists.

The following are key highlights of this paper:

- Measure the level of engagement of a student from the facial emotions presented in a real environment using an Xception neural network.
- Use of techniques applied in psychology such as the EVEA test to fact-check the moods of students in a real environment.
- Develop an application that allows the inference of the emotional states of a student using their web cam applying both Xception and DeepFace.
- Compare the predicted results with the emotions obtained by the EVEA test using common evaluation techniques such as the Confusion Matrix.

This article is composed of five sections: the first section, related work, the second section architecture and implementation, four-section evaluation and results, and finally, conclusions and future work.

2 Related work

Three relevant databases were consulted for the collection of related works on the subject. Association for Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), and ScienceDirect. For data collection, the following search string was proposed (Machine Learning or ML) and (Emotion Recognition) and (Facial Recognition) and (Online Environment or Online Education) and (DaisEE or Student Engagement dataset) with the following inclusion criteria: journals, early access articles, and research articles between the years 2018-2021. Courses, standards, and Conferences were not taken into account. For the present work, the 12 most relevant and important works have been considered on the subject in the last 4 years.

The fields of Face Recognition (FR) and Facial Expression Recognition (FER) in Computer Vision are still evolving because of different challenges that are posed when recognizing an Image: pose variation, poor lightning, or even movement can play an important factor. In the case of FR , the most prevalent algorithms being: Multi Support Vector Machine(Multi- SVM), and Convolutional Neural Networks (CNN).

Tamil et Al[16] intends to tackle these challenges with a proposed Hybrid Robust Point Set Matching Convolutional Neural Network (HRPSM_CNN) to effectively recognize faces from data , this method works by detecting faces using the Viola-Jones algorithm, then features are extracted and detected by the HRPSM_CNN , this proposed method proves to be more efficient than traditional ones with an accuracy rate of 97% for easy to identify conditions , and an accuracy of 95 % in the case of harder situations like poor lighting and weather.

As a way to easily apply FER in supervised learning domains , DeepFace, proposed by Serengil et Al. [15] offers an open-source framework that can help with the easier testing and implementation of FR models , the study discusses the implementation of many state-of-the-art algorithms like: VGG-Face , FaceNet , OpenFace, DeepFace , DeepID and Dlib, this helpful tool can aid with an easy way to switch up FR models on the fly.

Dewan et al [3] demonstrates the feasibility of using Deep Learning methods to predict the different levels of engagement using facial features, by applying Local Directional Patterns (LDP) and KPCA to capture different features, they achieve a really high accuracy of 90.89% on two-level engagement detection and 87.25% on three-level engagement detection using the DAiSEE data set, it is mentioned how Engagement and affects can be linked, moreover, the work mentions increased importance of recognizing the levels of user engagement, with many actual useful applications, like being able to adjust the teaching strategy and obtaining real-time feedback, further optimizing the use of online education for instructors.

In the article by Kuruvayil et al [9], they chose to use the meta-learning concept since they mention that there is a lack of sufficient samples with real conditions, such as partial occlusions, different head positions, and inadequate lighting conditions to be able to perform facial emotion recognition. Meta-learning using prototypical networks (metric-based meta-learning) has been proven to be well-fit for few-shot problems without severe overfitting. They used the CMU Multi-PIE dataset which contains images with partial occlusions, varying head poses, and illumination levels for training and evaluating the model. The proposed method is called ERMOPI (Emotion Recognition using Meta-learning across Occlusion, Pose, and Illumination). The proposed method achieved 90% accuracy for CMU Multi-PIE database images and 68% accuracy for AffectNet database images.

Abedi et al[1] presents and mentions the existence of psychological evidence for the incorporation of affect states in engagement level detection, a proposed Temporal Convolutional Network (TCN) serves to analyze affect and behavioral features which are concatenated to one feature vector for all the frames of a video , with this Architecture, they accomplish an Mean Squared Error (MSE) of 0.0708 on the DAiSEE with their proposed method, there is also a mention of how difficult it is to measure disengagement as an anomaly. Also worthy of note, the actors highlight the use of the circumplex model [Fig 1] of both positive and negative values of valence and arousal, can help represent if a person is being engaged in a certain task.

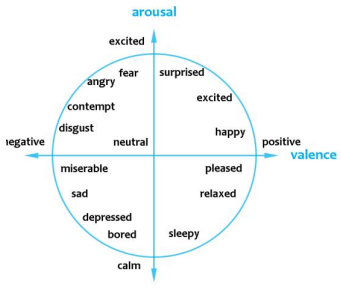

Fig. 1: Circumplex Model of Affect with both positive and negative emotional states [1]

In the article by Pronav et al [11] they focused on developing a Deep Convolutional Neural Network (DCNN) , model that classifies 5 human emotions (happy, angry, neutral, sad, and surprised) the proposed CNN consists of 4 layers composed of convolutional, pooling dropout and fully connected to extract the features from input data. The model uses an Adam optimizer to reduce the

loss function and is tested to have an accuracy of 78.04%.

As a point of comparison, the potential for shallower neural networks in FER tasks is tested by [17], achieving an accuracy of 93.4 % on the DAiSEE dataset, using a shallow depth of 50 layers, the method helps avoiding the vanishing gradient problem and the inclusion of residual layers helps to reduce the high loss found in deeper networks.

It is worth noting, that the related works found didn't measure engagement in a real-life scenario, instead of using preexisting data sets for engagement measurement.

3 Architecture and Implementation

This section presents the architectural proposal and describes the aspects of the development of the application using the Deep Face system, which allows deep learning facial recognition.In addition, we also applied the **XCeption** Architecture [2] a model trained on the DAiSEE Dataset through deep learning with an XCeption network [Fig 2], this model applies the techniques of transfer learning, taking advantage of what XCeption has to offer, this will allow measuring the engagement of real students in 4 distinct labels identically as presented in the DAiSEE dataset: with 0 being a "Very Low" level of engagement, "1" being Low, 2 as High and finally "3" to represent the highest possible level of Engagement.

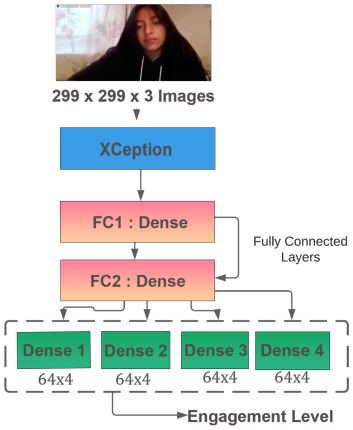

Fig. 2: XCeption Architecture

The proposed architecture is based on Yogesh et al. [8] Xception implementation, removing the top classification layers of XCeption , with a new classification head, with two Fully Connected (FC) layers between the average pooling layer and classification.

The Exit flow of the architecture is as follows [Fig 3] , the final (Dense) layer of the model serves as the output of the engagement classification prediction.

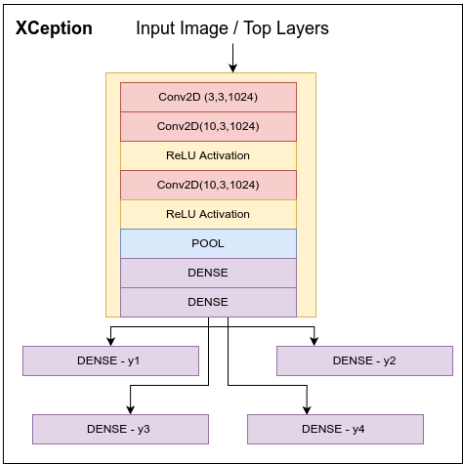

Fig. 3: XCeption on DAiSEE exit flow

Therefore, it is indicated which tools will be used, identifying the details related to the architecture of the system and its implementation.

3.1 Tools

The main language used was Python, which has several libraries suitable for the development of machine learning tools. In addition, Raschka's article [13] mentions that Python is the most preferred language for scientific computing, data science, and machine learning, boosting both performance and productivity by enabling the use of low-level libraries and clean high-level APIs. For proper training of the data from the Daisee dataset, the Tensor Flow library was used since it has benefits such as creating and training models through intuitive high-level APIs such as Keras, with immediate execution. In addition, OpenCv was used, which is an open-source library that can be used to perform tasks such as face detection, object tracking, and landmark detection.

3.2 Architecture

The DeepFace framework allows the possibility of analyzing facial features with the facial expression module, of the available emotions: **happy, neutral, sur-**

prise, sad, angry and fear and **disgust**, we tested the feasibility of this module within a real-life environment.

As presented in Figure 4, the architecture is composed of two sections user layer and the application layer: i) the user layer, where it is necessary to interact with the interface of the web application developed with the React framework [Figure 5 and 6] in order to access the webcam, ii) the application layer,composed of flask as its main back-end, is located in a container that will receive the frames of the user's face through the webcam. The responses sent by the server to the user will originate from the feeding of the Daisee data to the development platform, where through the TensorFlow, OpenCV libraries, and the DeepFace and XCeption models, they will be sent in JSON format the percentages of the possible emotions of the user, to the backend in Flask.

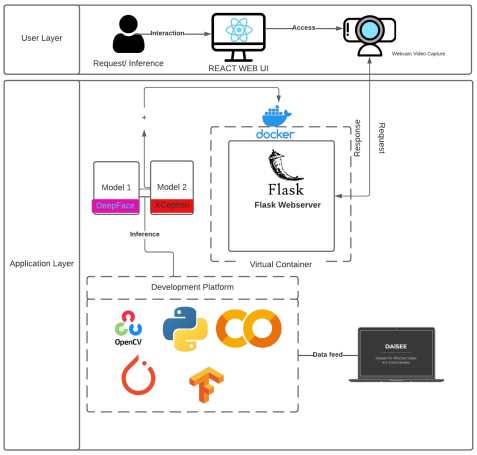

Fig. 4: End to End Architecture of the Proposed Application

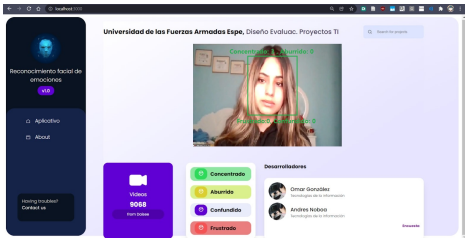

Fig. 5: Front-End of the Proposed Application

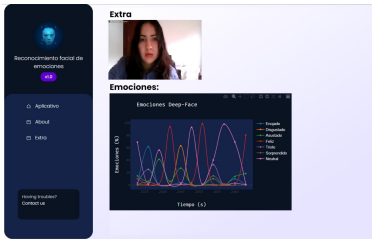

Fig. 6: Front-End of the Proposed Application

4 Evaluation

This section describes the parameters that were considered to carry out the evaluation (demographics, time, tools), the proposed scenarios, and each proposed session, in order to implement and subsequently analyze the results, with the aim of obtaining data to measure the engagement of students through their emotions.

We will work with a number of 18 students between 16 and 19 years of age belonging to the Unidad Educativa Dr Arturo Freire.

4.1 Scenario Definition

Scenario 1 The study of the first scenario will be done to verify the functionality of the models applied within the application , therefore the following sessions will be defined :

- **High School Students between the Age range of 16-19 years old** : Students will engage in an online lecture, then they will complete a survey based on a rating scale of their frame of mind "Escala de Valoracion del Estado de Animo (EVEA)" [14] to verify which emotion they had during the class.

  The objective will be to measure their engagement with the lecture. Various emotions that could affect how they perceive the lecture will be reviewed, such as sad, happy, disgust, anger, fear, neutral, and finally engagement itself. Said emotions will be verified and compared with the result of a Deep Face Emotion Model and an XCeption model, following that, we will draw conclusions based on the EVEA test that the students will complete.

  To predict the various emotions from video, we made use of 10-second clips for each student, as is mentioned by White Hil et Al [18] that labeling videos of this length proves to be a reliable predictor. For the experiment we extracted 10 consecutive frames per clip using the open-source tool "FFmpeg", we then applied the Viola-Jones algorithm for face extraction, and finally, we predicted the emotion for each clip using both models.

5 Results

5.1 DeepFace Results

The results generated from the DeepFace model [Fig 7] indicate that the dominant emotion in the classroom is the neutral emotion with 38.9%, the happiness emotion with 27.8%, sadness with 22.2 %, fear with 5.6 %, and finally anger with 5.6%

The results generated from the survey indicate that the dominant emotion in the classroom is the neutral emotion with 83.3%, the happiness emotion with 5.6%, disgust with 5.6%, and finally sadness with 5.6%.

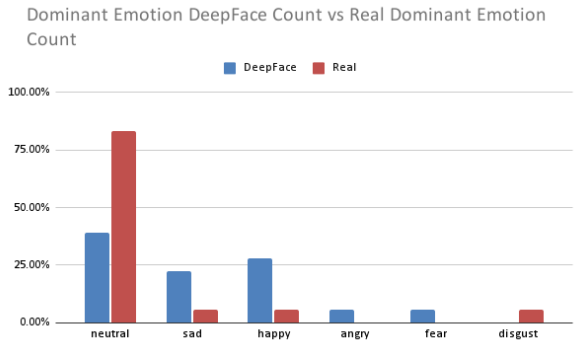

Fig. 7: Dominant Emotion DeepFace vs Real Dominant Emotion

Comparing the predicted vs. actual results, using a confusion matrix [Table 1] , it is noted that the most favorable result was obtained for the neutral emotion, with an accuracy of 33.33%. The way for which the DeepFace model failed to predict different results could be due to the different changes in illumination and camera quality.

		Predicted					
		0	**1**	**2**	**3**	**4**	**5**
Actual	**0**	0.0	0.0	0.0	5.56	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.0	0.0	5.56	0.0
	3	0.0	5.56	0.0	0.0	22.22	0.0
	4	0.0	0.0	0.0	0.0	33.33	5.56
	5	0.0	0.0	0.0	0.0	22.22	0.0

Table 1: Confusion Matrix Deep Face Test for various emotions (in Percentages).

5.2 XCeption vs Real Engagement Results

The results generated from the XCeption model, which measures engagement on a scale of 0 to 3, indicate that engagement levels 2 and 3 present 44.4%, while scales 0 and 1 present a minimum percentage of 5.56%, we can see the parallel between both real(surveyed) values and the predicted Engagement values [Fig 8 and 9].

In the case of the Real Levels, the results indicate a very similar occurrence of high levels of engagement , it is difficult, however, to make a discernible difference in the occurrences between levels 2 and 3 .

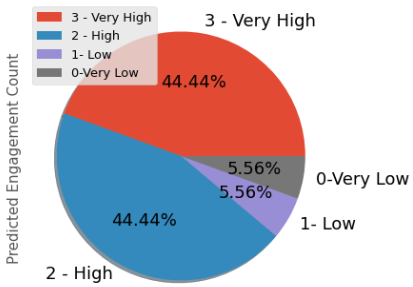

Fig. 8: Predicted Levels of Engagement

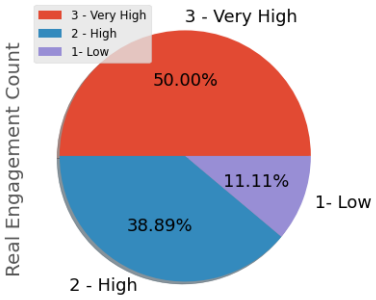

Fig. 9: Real Levels of Engagement

Both the predictions and real levels surveyed levels of engagement show a most common level of high and very high presence [Fig 10], this is in line with the work of Geng et Al. [5], and the DaiSEE data set itself since an overwhelming majority of the videos are labeled level 2 and 3 for high and very high respectably, creating class imbalance. This suggests a higher level of engagement for online learning is present in the real world, at least when the students are asked to turn on their cameras.

The confusion matrix for the predictions goes as follows [Table 2] , the most favorable results were from the engagement labels 2 (16.67%) and 3 (27.78%) respectively. Being very difficult to differentiate the two in a real-life classroom environment. It is worth noting that doing this task as a binary classification could be much more doable, as noted by Liao et Al [10].

		Predicted			
		Very Low	**Low**	**High**	**Very High**
Actual	**Very Low**	0.0	0.00	5.56	0.00
	Low	0.00	0.00	5.56	0.00
	High	0.00	5.56	16.67	22.22
	Very High	0.00	5.56	11.11	27.78

Table 2: Confusion Matrix for student's Engagement levels (in Percentages).

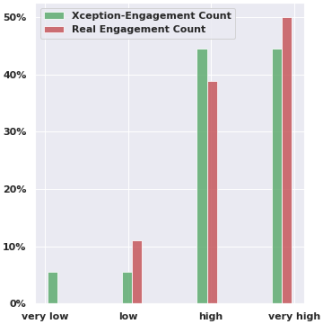

Fig. 10: Dominant Engagement Level Count (In Percentages)

5.3 EVEA Test

In the EVEA test section, the results indicate that the level of anxiety among the students is 3.083/10, anger-hostility 1.069/10, sadness and depression 1.94/10 and finally, happiness 4.22/10, these results were obtained by taking into account the averages of the EVEA test, it can be seen that the levels of happiness [Fig 11] are the highest compared to the other emotions, according to the psychologists who worked on the project, the level of happiness may have a relationship with the level of engagement as mentioned in the article [7]. Therefore, the students had an almost regular level of engagement in an online environment.

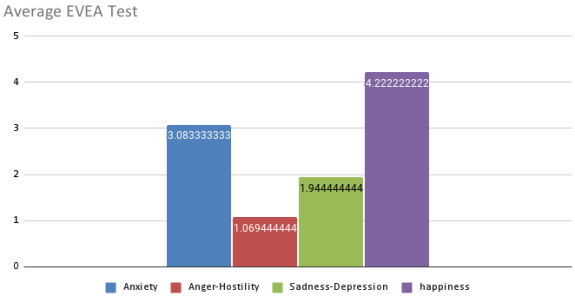

Fig. 11: Average Results for the Evea Test

We also measured the Spearman correlation coefficient as a way to obtain a measure of the affiliation that can exist between variables, in this case, the different affect emotions present in the EVEA test and the relations that can exist with engagement values, this coefficient has been used to identify strong relationships in emotion intensity predictions, such as in the work of Xie et al [19] . The Spearman correlation is defined as :

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{1}$$

Where ρ is the Spearman coefficient in (1), it can be $+1$ or -1 depending on how strong a relationship between variables can be in our data, measuring ρ between the results in the test and the engagement predictions and the results of the EVEA test is as follows, Where the asterisks represent the value p for statistical significance, this is indicated in [Table 3]

	Engagement	**Anxiety**	**Anger-hostility**	**Sadness-depression**	**Happiness**
Engagement	1***	0.44**	0.15	0.26	0.15
Anxiety	0.44**	1***	0.63	0.58*	0.43
Anger-Hostility	0.15	0.63***	1***	0.87***	0.37
Sadness-depression	0.26	0.58***	0.87***	1***	0.41***
Happiness	0.15	0.43*	0.37	0.41*	1***

Table 3: Correlation table with asterisks (*) indicating statistically significant scores for the student group ($n = 18$). * **p<.1** ; ** **p<.05** ; *** **p<0.01**.

5.4 Future Work

In future work, it would be feasible to determine what factors can influence the recognition of students' emotions. In addition, the recognition of other emotions could be added depending on the area where the model is applied. Finally, the inclusion of state-of-the-art FER and FR models with different architectures like shallow networks, can be applied to increase the precision when recognizing emotions.

6 Conclusions

At the end of the analysis of the emotions of the students, through the Xception generated model, the DeepFace model, it was determined that the majority of the students present a neutral emotion, and this has a relationship with the surveys carried out directly to the students. It should be noted that the other emotions in the models vary, this is due to the accuracy of the models with 33.3% and 27.7%, respectively.

Regarding the EVEA Test, the results indicate that the students presented a moderate percentage of happiness and low anxiety during class, and this is directly related to the engagement level where 88.9% of the student sample presents a level 2 - 3 of engagement and 11.1% presents a level 1, according to the psychologists who participated in the project. The online environment will be adequate only if there is a healthy environment taking into account factors such as interpersonal interactions between peers and teaching strategies.

References

1. Abedi, A., Khan, S.: Affect-driven engagement measurement from videos. arXiv preprint arXiv:2106.10882 (2021)
2. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
3. Dewan, M.A.A., Lin, F., Wen, D., Murshed, M., Uddin, Z.: 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). pp. 1895–1902. IEEE (2018)
4. Dhall, A., Sharma, G., Goecke, R., Gedeon, T.: EmotiW 2020: Driver Gaze, Group Emotion, Student Engagement and Physiological Signal Based Challenges, p. 784–789. Association for Computing Machinery, New York, NY, USA (2020), https://doi.org/10.1145/3382507.3417973
5. Geng, L., Xu, M., Wei, Z., Zhou, X.: Learning deep spatiotemporal feature for engagement recognition of online courses. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 442–447. IEEE (2019)
6. Hu, M., Li, H., Deng, W., Guan, H.: Student engagement: One of the necessary conditions for online learning. In: 2016 International Conference on Educational Innovation through Technology (EITT). pp. 122–126 (2016). https://doi.org/10.1109/EITT.2016.31
7. Joo, B.K., Lee, I.: Workplace happiness: work engagement, career satisfaction, and subjective well-being. Evidence-based HRM: a Global Forum for Empirical Scholarship **5**(2), 206–221 (Jan 2017). https://doi.org/10.1108/EBHRM-04-2015-0011, https://doi.org/10.1108/EBHRM-04-2015-0011, publisher: Emerald Publishing Limited
8. Kamat, Y.: Edumeet. https://github.com/yogesh-kamat/EduMeet (2020)
9. Kuruvayil, S., Palaniswamy, S.: Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. Journal of King Saud University - Computer and Information Sciences (2021). https://doi.org/https://doi.org/10.1016/j.jksuci.2021.06.012, https://www.sciencedirect.com/science/article/pii/S1319157821001452
10. Liao, J., Liang, Y., Pan, J.: Deep facial spatiotemporal network for engagement prediction in online learning. Applied Intelligence **51**(10), 6609–6621 (2021)
11. Pranav, E., Kamal, S., Satheesh Chandran, C., Supriya, M.: Facial emotion recognition using deep convolutional neural network. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). pp. 317–320 (2020). https://doi.org/10.1109/ICACCS48705.2020.9074302

12. Prospero, M.R., Lagamayo, E.B., Tumulak, A.C.L., Santos, A.B.G., Dadiz, B.G.: Skybiometry and affectnet on facial emotion recognition using supervised machine learning algorithms. In: Proceedings of the 2018 International Conference on Control and Computer Vision. p. 18–22. ICCCV '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3232651.3232665, https://doi.org/10.1145/3232651.3232665
13. Raschka, S., Patterson, J., Nolet, C.: Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. Information **11**(4), 193 (Apr 2020). https://doi.org/10.3390/info11040193, https://www.mdpi.com/2078-2489/11/4/193, number: 4 Publisher: Multidisciplinary Digital Publishing Institute
14. Sanz Fernandez, J., Gutierrez, S., Garcia Vera, M.P., Sanz Fernandez, J., Gutierrez, S., Garcia Vera, M.P.: Propiedades psicométricas de la Escala de Valoración del Estado de Ánimo (EVEA): una revisión (2014), https://eprints.ucm.es/id/eprint/58409/, publisher: Sociedad Española para el Estudio de la Ansiedad y el Estrés (SEAS)
15. Serengil, S.I., Ozpinar, A.: Hyperextended lightface: A facial attribute analysis framework. In: 2021 International Conference on Engineering and Emerging Technologies (ICEET). pp. 1–4. IEEE (2021). https://doi.org/10.1109/ICEET53442.2021.9659697, https://doi.org/10.1109/ICEET53442.2021.9659697
16. Tamilselvi, M., Karthikeyan, S.: An ingenious face recognition system based on hrpsm_cnn under unrestrained environmental condition. Alexandria Engineering Journal **61**(6), 4307–4321 (2022)
17. Thiruthuvanathan, M.M., Krishnan, B., Rangaswamy, M.: Engagement detection through facial emotional recognition using a shallow residual convolutional neural networks. International Journal of Intelligent Engineering and Systems **14**(2) (2021)
18. Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A., Movellan, J.R.: The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing **5**(1), 86–98 (2014)
19. Xie, H., Feng, S., Wang, D., Zhang, Y.: A novel attention based cnn model for emotion intensity prediction. In: CCF International Conference on Natural Language Processing and Chinese Computing. pp. 365–377. Springer (2018)