International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-10, Issue-5; Jul-Aug, 2024
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.105.2

Classifying Emotional Engagement in Online Learning Via Deep Learning Architecture

Prisha Jain[1], Chaya Ravindra[2]

[1]Neerja Modi School, India
Email: prishajain.1709@gmail.com
[2]Department of Engineering and Technology, REVA University, India
Email: chaya.ravindra@gmail.com

Received: 08 May 2024; Received in revised form: 16 Jun 2024; Accepted: 24 Jun 2024; Available online: 02 Jul 2024

***Abstract*— *The world has seen a phenomenal rise in online learning over the past decade, with universities shifting courses to online modes, MOOCs(Massive Open Online Course) emerging and laptop and tab-based initiatives being extensively promoted. However, educators face significant challenges in analyzing learning environments due to issues like lack of in-person cues, small video size, etc. To address these challenges, it is crucial to analyze the engagement levels of online classes. Out of the various subcategories of engagement, emotional engagement is one that is overlooked, but integral to analysis and deterministic in its approach. In response, we developed a deep learning architecture to analyze emotional engagement in online classes. Our method utilizes a ResNet50-based algorithm, refined through experimentation with various techniques such as transfer learning, optimizers, and pre-trained weights. The model adds a unique layer to the analysis of different algorithms used for engagement detection in academia while also achieving stellar rates of 81.34% validation accuracy and 81.04% training accuracy. Unlike other models, our approach employs high-quality image data for training, ensuring more reliable results. Moreover, we constructed a novel framework for applying emotional engagement to real-world scenarios, thus bridging the pre-existing gap between implementation and academia. The integration of this technology into online learning has immense potential, and can bring with it a shift in the quality of education. By fostering a safe and healthy learning space for every student, we can significantly enhance the effectiveness of online education systems.*

***Keywords*— deep learning, emotional engagement, engagement, framework, online learning, ResNet-50**

I. INTRODUCTION

Education stands as a fundamental pillar of modern- day society and one of the most influential developments in this field is online learning. Over the past decade, online learning has rapidly gained popularity and usage (Mukhopadhyay et al., 2020), with the COVID 19 pandemic greatly catalyzing its implementation into society (Gupta & Kumar, 2022). For instance, many universities and institutes have shifted onto virtual platforms. MOOCs (Massive Open Online Course) have emerged, dramatically changing the education landscape, with over 150,000 being available in 2023 (Pickard et al., 2023). Multiple laptop and tab-based initiatives have been promoted by schools and governments globally (Clarke & Svanaes, 2014; Fuhrman, 2014). While online learning is advantageous due to how ubiquitously and flexibly it can be used along with the increased course variety it provides, it still lacks in many aspects, including teacher- student interaction and practical education provision (Das & Paris, 2022).

One key challenge with online classes is analyzing learning environments. This is due to multiple reasons, including the absence of non-verbal and in-person cues, the miniscule size of videos which makes it impractical to assess students' reactions and teach simultaneously, the necessity of muting student microphones which hinders interactive feedback, etc. Therefore, teachers tend to teach without a complete understanding of whether or not students are concentrating on and comprehending the material, as has been proved in multiple studies

(Sobieszczuk-Nowicka et al., 2018; Mashoedah et al., 2018). It is also difficult for educators to understand the class dynamics and environment in online modes. As a result, students' emotional well-being can't be catered to. In turn, since students' participation is highly impacted by the direct attention and support they obtain from teachers, students are prompted to leave the class or disengage from lessons (Azlan et al., 2020).

To initiate change, it is necessary to systematically analyze online classes. The principal approach for analyzing learning environments is to monitor student engagement levels. Engagement can be defined as "the interaction between the time, effort and other relevant resources invested by both students and their institutions to optimize the student experience while also enhancing the learning outcomes and development of students as well as the performance of the institution" (Trowler, 2010). There are multiple types or sub-categories of engagement within the educational setting. Researchers agree that cognitive, emotional and behavioral engagement are the most deterministic. Cognitive engagement refers to the willingness and effort to grasp more difficult concepts and try challenging puzzles, behavioral engagement refers to concentration and attention on the material, and emotional engagement refers to the presence of positive emotion such as interest and enthusiasm in regards to the material being taught (Hasnine et al., 2023).

This paper has limited its scope to emotional engagement due to its comprehensiveness and significance, along with the elusiveness of its quantifiability in pre-existing frameworks. According to Patrick et al, the premise is simple: "the more emotionally involved students are with their environment while studying a subject, the more engaged they are, and the more support students get with managing their emotional states, the more they can pay attention in classes" (Patrick et al., 2007). In other words, student engagement is directly proportional to their achievement (Skinner et al., 1998). Hence, it is crucial for achieving learning goals and receiving quality education.

Many methods are used to gauge emotional engagement. Traditionally, educators rely on quizzes and questionnaires at the end of sessions, but this is prone to demand characteristics and is susceptible to the student's angle of analysis (McCambridge et al., 2012). It also requires a lot of effort from both the students and the educators. Hence, automation has been brought into the limelight, significantly shifting the potential scope of emotional engagement analysis. Our research delves into the field of automated analysis through the usage of deep learning.

Deep Learning (DL) is a subset of machine learning that utilizes multi-layered neural networks called deep neural networks to imitate the intricate decision-making capability of the human brain. These deep neural networks are trained on vast amounts of data to enable them to identify phenomena, observe patterns in information, and make predictions and decisions. They only need to be trained once, however, after which they can efficiently be used for purposes ranging from medical diagnosis to voice-enabled machinery (Goodfellow et al., 2016). Many deep learning algorithms are used to create neural networks. This paper focuses on ResNet-50, which was developed by Microsoft researchers in 2015. It was designed to enable better performance through its residual connections. Interestingly, its name was derived from its characteristic feature of having 50 layers in its network.

One particular machine learning technique that we will use in the study is transfer learning. Regarding theoretical context, transfer learning can be defined as a method where a model trained on one task is used as the starting point for a model on a second task. By using the learned features from the first task, the model can work more efficiently and quickly even with a small amount of data (Ali et al., 2023).

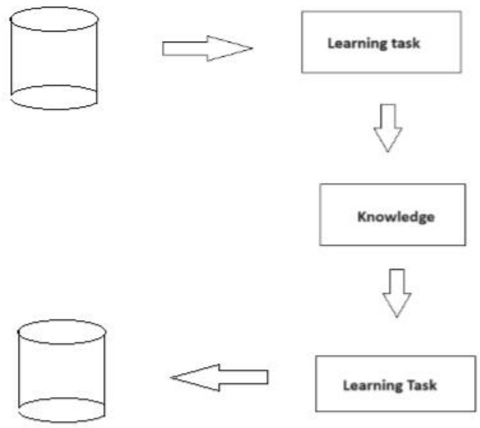

Fig. I The process of transfer learning

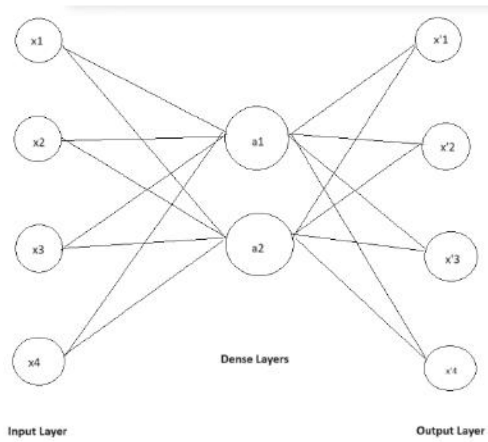

Fig. II Deep learning architecture

The key contributions of this paper are:

- This paper proposes a model that has been trained to detect the emotional engagement levels of students in real-time with a stellar accuracy of 81.34% val. accuracy and 81.01% test accuracy
- This paper adds on to the plethora of research done in this field by methodologically experimenting with 4+ datasets, 3+ algorithms, and a wide range of machine learning techniques to determine which is more lightweight and yields better results, along with the learning rate and epoch number at which it does so
- The model uses high quality data, a feature of datasets that is rarely seen in research in this field
- This paper also aims to provide a modified framework that prioritizes privacy by analyzing student videos on their own devices and provides visual, easy to navigate, graphical summaries to educators. It will also enable a student support system to assist students with dire emotional states

In terms of potential limitations in our research, a prevalent issue is the scarcity of available high-quality data, which reduces the accuracy of models and their ability to learn relevant features. Moreover, there may be biases due to deep learning models mirroring the innate biases of the training data. For example, cultural accessories such as bindis and headscarves may not be properly identified by the model and hence may create discrepancies. The model may also have difficulty interpreting mixed emotions, since it is trained on artificially emotive images.

II. METHOD

This research was carried out on Google Collab software with T4 GPU, using the highly-acclaimed python libraries of Keras and Tensorflow. The dataset was uploaded to Google Drive, where file paths were used to reference the images and train the model on them. Initially, the employed system underwent training with the FER-2013 dataset, which contains 30,000+ images of people of different cultures and ages. However, due to low image quality and lack of color, the Facial Expressions Training Data was chosen instead. This dataset is a high-quality, coloured dataset consisting of 29,000+ (96 by 96 pixels) images. It was taken from Kaggle, a public dataset publishing platform.

To pre-process the data, multiple steps were taken. The labeled data was first sorted into its respective emotion class folders, and split into validation, training and testing data by a 10-80-10 split. Training and validation data was shuffled to ensure random selection.

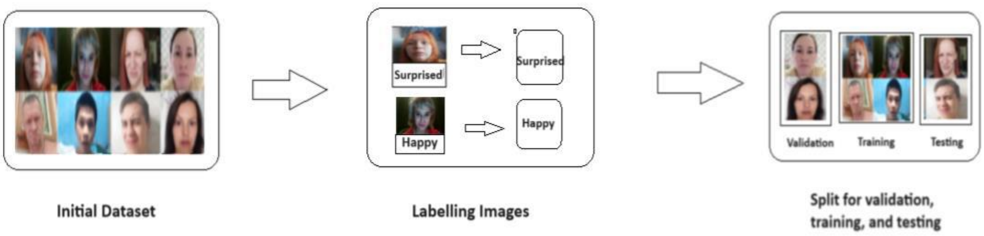

Fig. III Process of data cleaning

The employed CNN (Convolutional Neural Network) architecture was integral to this study. We experimented with MobileNet, ResNet-50 and EfficientNet, evaluating which would be better for the chosen objective. While all of them converged as epochs increased, ResNet-50 had the best overall performance since it gave higher accuracies even at smaller epochs. Additionally, transfer learning proved to be a crucial technique to increase the speed and accuracy of the model. We used pre-trained a ResNet50 model from Keras Applications.

To construct the architecture, we removed the fully connected layers at the top of the pre-trained models to enable customization of layers. 8 output classes were added, namely ‘Happy’, ‘Sad’, ‘Contempt’, ‘Surprised’, ‘Neutral’, ‘Fear’, and ‘Anger’.

In terms of the layers in the models, the functional transfer learning layers were followed by alternating flatten and dense layers. These dense layers were composed of 2048 neurons. For activation, ReLu was used to prevent gradients from saturating and hence solve the issue of vanishing gradients. In the final layer, Softmax was used, which helped training converge at a faster rate.

Moreover, the model weights pre-trained on the standard ImageNet dataset were used. These weights were locked into the models to ensure learned representations are not lost. After the convolutional layers, global average pooling was used to reduce the amount of computation required while retaining important features. In terms of optimizers, we initially implemented Adam, which is a standard method to help the model converge faster. However, upon analysis, we deemed SGD (Stochastic Gradient Descent) to be better suited due to how well it converged to more optimal solutions.

In this study, loss calculation was done through sparse categorical cross entropy. In comparison to other methods, it saves time in memory as well as computation. The key metric we used to measure the success of the model was training accuracy, which estimates the potential of a model.

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2048)	23587712
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 1024)	2098176
flatten_1 (Flatten)	(None, 1024)	0
dense_1 (Dense)	(None, 1024)	1049600
flatten_2 (Flatten)	(None, 1024)	0
dense_2 (Dense)	(None, 1024)	1049600
dense_3 (Dense)	(None, 8)	8200

Total params: 27793288 (106.02 MB)
Trainable params: 4205576 (16.04 MB)
Non-trainable params: 23587712 (89.98 MB)

Fig. IV Model structure summary

III. RESULTS

i) Model

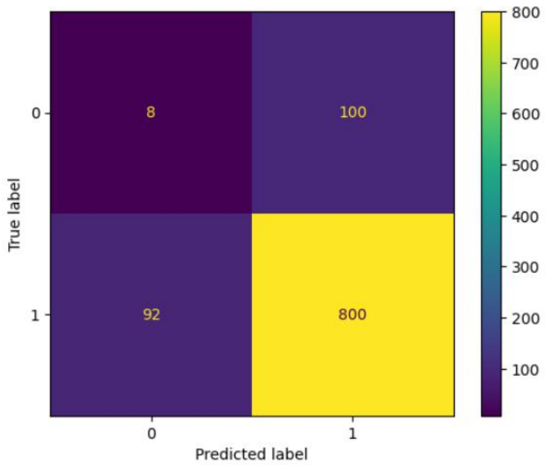

Fig. V: Confusion matrix

In summary, the confusion matrix indicates that the model has a high accuracy (80.8%) and performs well in terms of precision (88.9%) and recall (89.7%) for class 1. However, it has a relatively high false positive rate (92.6%) and a low false negative rate (10.3%).

The final model is an 8-layer sequential classification model, composed of a pre-trained layer along with alternating dense and flattening layers. The usage of high-quality data and experimentation with parameters has resulted in lightweight yet high performance structuring. In essence, this model analyses student expressions to accurately classify their emotional engagement states.

ii) Framework

This framework is designed to be an extension app in online learning platforms such as Zoom. Currently, the market does not host any such platforms, with the closest alternative being Engagement Hub, an extension on Zoom Marketplace that allows users to automatically transcribe and analyze meeting recordings. This lack of implementation may be a result of how restricted engagement analysis via deep learning architecture is to academia.

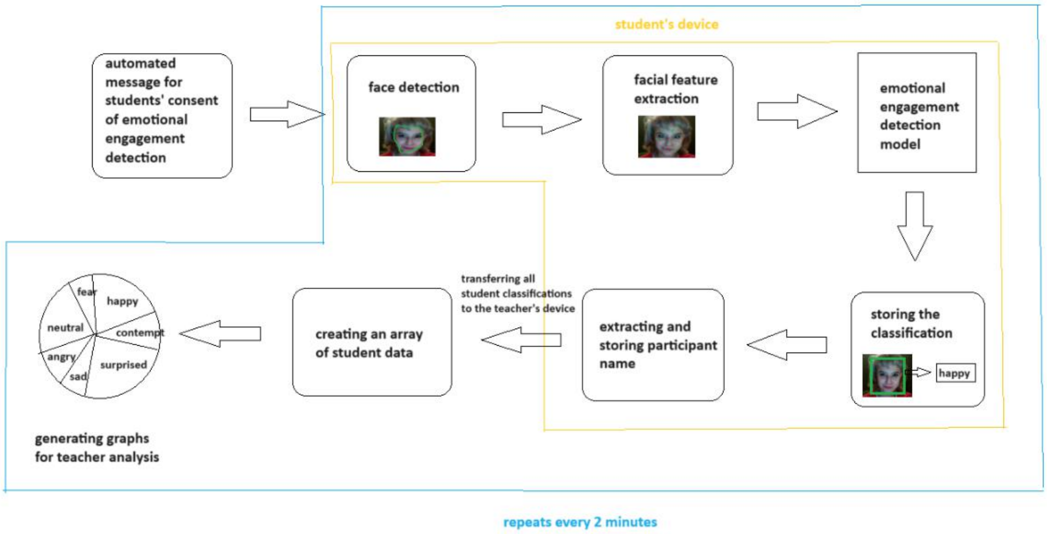

Fig. VI Process of emotional engagement analysis

Following are the steps of the devised framework-

- At the start of any session, an automated message will be displayed on all student devices to notify them that they are being recorded and analyzed. This will be similar to the pre-existing feature on zoom that notifies participants when screen recording is turned on by a user. Through this feature, the privacy rights of students will be protected.
- The cycle of emotional engagement analysis will repeat in a set interval of time, for example, every 2 minutes. On each student's device, their camera will be connected to the framework and a screenshot will be taken.
- Through a basic AI (Artificial Intelligence) algorithm, the student's face will be detected. Then, facial features of the image will be extracted by mapping of facial points. For both face detection and feature extraction, the OpenCV library will be used, which provides ready-to-use methods with advanced capabilities.
- The extracted image will then be run through an emotional engagement detection model, where it will be pre-processed and then analyzed. Through methods like transfer learning, optimization, pooling layers, etc. the model is fine tuned to accurately predict the emotion of the student.
- The student's name is then extracted from their name label. The name and its associated emotion classification is encrypted and sent to the teacher's device.
- At the teacher's device, all data is decrypted and entered into an array. This process will run in the backend, where it can't be accessed by the teacher.
- The emotion classification data of the array will then be used to generate a pie-graph. This will be an easy to read, understandable format for educators to quickly access and analyze. The graph will be available during screen sharing and be readily movable across the educator's page. This will ensure ease and efficiency.

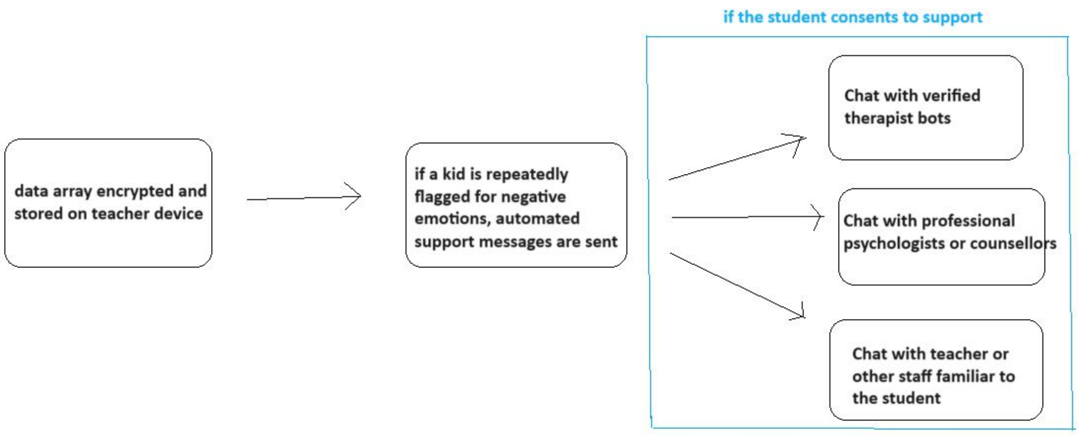

Fig. VII Process of student support

Following are the steps for student support in the devised framework-

- For long-term courses that engage with students for more than 3 sessions, educators can turn on settings to enable student support. Data arrays from each session will be automatically stored on the teacher's device. This data will be encrypted to prevent privacy invasion. It will be loaded back onto the streaming platform architecture in the backend when the next session starts.
- The data arrays in the backend will be analyzed and if a student is flagged to have shown negative emotions such as 'sad', 'angry', 'contempt', 'fear', etc. repeatedly (i.e. more than 7 times in an average session of 30 minutes), their device will be contacted. An automated support message will be sent asking for their consent to take further action. Moreover, if all students show negative emotions consistently, this will be an indication to the educator to make their sessions more engaging.
- If the student gives consent, they will be prompted to take one or more of 3 actions-
    - They can contact their teacher or other trusted staff, with whom they can then share their concerns. This method would be best suited for issues with the learning style or course load.
    - They can contact professional therapists or psychologists. We will suggest trained experts they can reach out to. This method is suited for personal issues, such as mental health disorders, financial issues, health-related challenges, etc.
    - We can also collaborate with high-quality therapist AI bots. This would work best for students who have minor problems, and aren't willing to spend a lot or aren't comfortable with professional therapy

IV. DISCUSSION

For the purpose of this study, we developed a ResNet50-based classification model, with the aim of analyzing different architectures, datasets, parameters, etc. to develop the most accurate and efficient version. This was accompanied with constructing a framework which detailed the real-time process of image extraction, feature detection, emotion classification and data storage. The student support system is unique from pre-existing research through its ability to actually utilize the data emotion classification data to assist students that are struggling.

This study's results are promising, both in terms of model analysis and framework development. The high training accuracy reflects that the model architecture and hyperparameters are well-suited to the task. The ResNet50 model is particularly noteworthy due to its performance and lightweight characteristics. Additionally, the features in the dataset are highly predictive of the target variable. While the research objectives were met, it's essential to consider the limitations as well. Due to the lack of available data, the potential of this model was stunted to some extent. With resources like more computational power, it could have had better performance. As predicted, the model also may have difficulties in real- world scenarios, where lighting, angles, accessories, etc may distort faces in images and lead to inaccurate predictions. The training data may also be artificial in its expression of specific emotions, leading to disparity with real-life analysis scenarios since students don't portray singular emotions in real life but rather have mixed emotions that the model may get confused with.

The framework is well developed through its cumulation of the ideations of notable researchers and creative addition of more unique features. It is also realistic in terms of implementation and ethically sound. While it prompts the shifting of academia into practical usage, limitations such as the lack of resources meant that this study was unable to fully implement the framework. Moreover, the immediate concern of users not willing to share their data or permit to be recorded persists and can only be resolved with a change in ideology towards sharing data.

In response to these limitations, a strategic procedure is necessary for future developments. Most importantly, gathering sufficient resources is required, since only with more data and computational power can the model be made to classify all variations of emotive images. Data can furthermore be augmented to increase both the amount of data and the symmetry of the amount per class of emotions. This will also reduce any chances of overfitting. Researchers that aim to create an optimal solution should also specifically use datasets that have extracted data from online learning sessions to guarantee the training data is similar in characteristics to real world emotive data in online classes. In terms of future steps with this research, this model exhibits noteworthy scalability. We can make the analysis mechanism more multifaceted, with inclusion of behavioral engagement analysis and chat analysis, hence improving not only the accuracy of the model but also the reliability of its analysis.

V. CONCLUSION

This study presents a ResNet50-based model for real-time emotion classification of students, achieving validation and test accuracies of 81.34% and 81.01%, respectively. The research evaluates various architectures, datasets, parameters, and machine learning techniques to optimize performance. It uniquely employs high-quality data and considers privacy by processing videos on student devices, offering visual summaries for educators and support for emotionally distressed students.

Limitations include limited data and computational power, potential inaccuracies in diverse real-world conditions, and reluctance from users to share data. Enhancing performance requires more data, improved computational resources, and augmenting datasets to balance emotional classes and prevent overfitting. The model has significant scalability potential. Future enhancements could include analyzing behavioral engagement and chat interactions, which would increase both the accuracy and reliability of the model's emotional engagement analysis.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to Ndeavors, which managed the communication and collaboration between the authors of this paper, and to Neerja Modi School, which sponsored and supported the paper.

REFERENCES

[1] Ali, A. H., Yaseen, M. G., Aljanabi, M., & Abed, S. A. (2023). Transfer learning: A new promising technique. Mesopotamian Journal of Big Data, 2023, 29-30.

[2] Azlan, C. A., Wong, J. H. D., Tan, L. K., Huri, M. S. N. A., Ung, N. M., Pallath, V., ... & Ng, K. H. (2020). Teaching and learning of postgraduate medical physics using Internet-based e-learning during the COVID-19 pandemic–A case study from Malaysia. Physica Medica, 80, 10-16.

[3] Das, I., & Paris, K. (2022). A Deep Learning-Based Approach for Adaptive Virtual Learning with Human Facial Emotion Detection. Journal of Student Research, 11(3).

[4] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[5] Gupta, S., Kumar, P., & Tekchandani, R. K. (2022). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimedia Tools and Applications, 82(8), 11365–11394. https://doi.org/10.1007/s11042-022-13558-9

[6] Hasnine, M. N., Nguyen, H. T., Tran, T. T. T., Bui, H. T., Akçapınar, G., & Ueda, H. (2023). A real-time learning analytics dashboard for automatic detection of online learners' affective states. Sensors, 23(9), 4243.

[7] M. Mashoedah, M. Hartmann, H. D. Surjono, Z. Zamroni. The Vocational High School Teachers 'Awareness Level and Implementation of the Students' Learning Style Assessment. Jurnal Pendidikan Teknologi dan Kejuruan. 24, 91-101 (2018).

[8] McCambridge, J., De Bruin, M., & Witton, J. (2012). The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. PloS one, 7(6), e39116.

[9] Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P. (2020). Facial emotion detection to assess learner's state of mind in an online learning system (ICIIT'20). Association for Computing Machinery. https://doi.org/10.1145/3385209.3385231

[10] Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. Journal of educational psychology, 99(1), 83.

[11] Pickard, L. (2024, April 30). Massive List of MOOC Platforms Around the World in 2024 — Class Central. The Report by Class Central. https://www.classcentral.com/report/mooc-platforms/

[12] Shah, D., Pickard, L., & Ma, R. (2022). Massive list of MOOC platforms around the world in 2023. Class Central's MOOC Report.

[13] Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). Individual differences and the development of perceived control. Monographs of the society for Research in Child Development, i-231.

[14] Sobieszczuk-Nowicka, E., Rybska, E., Jarmużek, J., Adamiec, M., & Chyleńska, Z. (2018). Are We Aware of What Is Going on in a Student's Mind? Understanding Wrong Answers about Plant Tropisms and Connection between Student's Conceptions and Metacognition in Teacher and Learner Minds. Education Sciences, 8(4), 164.

[15] Trowler, V. (2010). Student engagement literature review. The higher education academy, 11(1), 1-15.

[16] A mobile initiative that's more than just a tablet handout -- campus technology. (2014, March 20). Campus Technology. https://campustechnology.com/articles/2014/03/20/a-mobile-initiative-thats-more-than-just-a-tablet-handout.aspx