MicroFlow: Detecting Learning Affective States Using Micro-expression Theory

Our work proposed the MicroFlow framework motivated by micro-expression theory to detect learners' affecting states such as boredom, anxiety, and flow. We tested whether macro-expression combined with micro-expression will improve the model performance in detecting flow and affective states. We found that the model combining AU and IVA features performed the best for anxiety and the state of flow detection, while individual models performed better for boredom prediction. We achieved 0.84 of AUC for the flow state improved by 10% compared to the AU model. For anxiety and boredom we achieved 0.71, and 0.70 respectively. Thus, we obtained feasibility of our framework that can be used to detect learners' affective states. Our study adopted micro-expressions in the learning context by proposing a cost-effective tool that could be used in a learning context by educators to create more engaging learning environment by adjusting the complexity level of a given task.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**.

Additional Key Words and Phrases: Flow, Emotion, Facial Behavior Markers, Passive Sensing, Machine learning

ACM Reference Format:

. 2022. MicroFlow: Detecting Learning Affective States Using Micro-expression Theory. 1, 1 (January 2022), 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

One of the key factors affecting learning process effectiveness is people's emotional response to a given task [15]. Students may be bored with easy-to-complete tasks or experience anxiety and frustration working on highly complex assignments. The main goal for learners is to achieve a state of flow that provides deep involvement, focus, and enjoyment in the educational process. According to Csikszentmihalyi, the state of flow is the most optimal affective state for effective learning [3, 9]. Knowing learners' emotional responses, instructors can optimize educational materials by readjusting the level of a given assignment and providing an engaging learning environment. According to the Flow Theory, the state of flow may be achieved by the right balance of skill level and task complexity as shown in Figure 1 [9].

Facial expression is commonly used to detect learners' emotional responses and is easily captured via web camera during the learning process. Since the facial expression is one of the most important cues for understanding human emotions, facial expression recognition (FER) algorithms can be successfully used to reveal learners' affective states (such as boredom, anxiety, and flow). Prior work has demonstrated that macro-expression (distinguishable facial representation of emotions) allows recognizing some of the emotions that occur while studying [11]. However, our emotions are complex and not all of them can be accurately identified using FER relying only on macro-expressions. Therefore, facial micro-expressions can be used as an alternative or together with macro-expressions to provide a significant boost in detecting concealed and unconsciously produced emotions [29]. While some of the prior studies explored the

Author's address:

Manuscript submitted to ACM

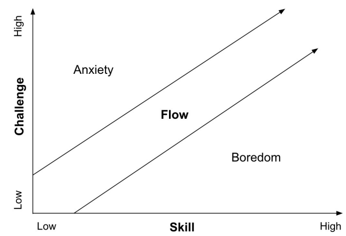

Fig. 1. Flow State Model [9]

usage of IVA features to capture micro-expression for FER task, none of those investigated the combination of both IVA and AU features to capture macro- and micro-expression respectively in the learning context. We hypothesize that that flow, anxiety, and boredom can be detected using FER via the combination of both IVA and AU features. IVA capturing micro-expressions complements AU capturing macro-expressions by providing information about subtle muscle movement which increases the recognition accuracy [38].

Research Questions:

- RQ1: Can a student's flow state be detected using IVA features that capture micro-facial expressions in a real-time online learning context?
- RQ2: Can the AU-based IVA combined model capturing macro-expressions and micro-facial expressions enhance the detection performance of a flow state?

In this study, we propose a flow detection framework based on micro-expression to identify the state of flow in a learning context. Our approach employs spatial-temporal deformations properties of micro-expression to model this framework. We aim to test our method on our in the wild dataset and discuss the applicability of our method for real-world settings, limitations, and potential implications. The motivation of this study is to advance the understanding of emotions learners experience and provide guidance about how to increase their engagement and educational performance using facial micro-expressions recognition.

Our Contribution:

- We propose the framework motivated by micro-expression theory for extraction of temporal dynamic information during the learning process.
- We introduce a model for detecting facial macro- and micro-expressions using AU and IVA features to detect learners' affecting states including boredom, anxiety, and flow.
- We extended the application of facial micro-expressions recognition into an educational domain to provide suggestions aimed to advance learners' performance.

The rest of the paper discusses relevant literature, the method and the dataset we used to develop our framework, results evaluating models performance, the discussion of our results, limitation, and future work.

2 RELATED WORK

In this section, we review existing work about previously used methods for capturing learners' emotional experience algorithmically, the potential use of AU and IVA features in detecting learners' affective states via macro- and micro-expression, and psychological studies on facial expressions of boredom, anxiety, and the state of flow that often occur as a response to a learning experience.

2.1 Methods for Capturing Students' Learning Experience

The learning experience is a unique process for different individuals as it may be affected by multiple factors such as current emotions, cognitive skills, language proficiency, familiarity or previous experience with the topic, and more [20]. Our study focuses on learning-centered affective states such as boredom, anxiety, and flow as a proxy for emotional responses learners experience through their educational process. The rationale for this assumption is multiple prior works suggesting that students' performance and their academic achievements are associated with their learning-related emotional experiences [5, 25, 28]. For example, positive emotions (e.g., excitement) are more likely to be the indicator of better performance, while negative emotional experiences(e.g., anxiety, and boredom) are associated with lower educational achievements [37]. Additionally, it is been shown that most of the learners' emotions are related to task complexity [31], which opens the room for interventions to adjust the task difficulty and create a more comfortable learning environment. Yet it is a challenging task to determine individual learning experiences without an explicit questionnaire with a student which may not always be feasible within a classroom environment [13]. Along with the simple survey-based approach used previously [13], there were some advanced methods for identifying learning-centered affective states including gazed-base eye-tracking analysis [16, 34] and wearable sensing technology such as electroencephalogram [22]. While the survey-based method is accessible and widely used, it limits the understanding because the data is usually collected at retrospective after educational task and doesn't capture real-time experience which limits potential interventions [13]. Other methods are capable of solving this problem but require expensive hard- and software equipment implementation. The main limitation of these methods is scaling because the implementation in real-world settings may hardly be feasible for educational institutions. Another alternative is the use of computer vision technology (facial expression recognition capturing macro- and micro-expressions) that allows to accurately determine affective state and require only camera-based modality [6]. However, prior research exploring this method has been limited in terms of educational context. Thus, our purpose is to test the effectiveness of macro- and micro-expressions analysis for the learning process because it represents a trade-off between low cost and high performance.

2.2 Micro-expressions Mechanism

In this section, we introduce the terms used to describe the principles of macro- and micro-expressions recognition. Firstly, we refer to the Facial Action Coding System (FACS) which is known as a taxonomy of facial muscle movements that correspond to expressed emotions [14]. Secondly, we use AU as building blocks of the FACS which are often used as a standard method of studying emotions that occurs on the entire face [11]. Each AU describe facial changes based on macro-expressions that occurs due to a facial muscle movement. However, some of the emotions may not be explicitly expressed on the face. These emotions may involve only subtle and hardly visible movements. This type of facial expression is called Micro Expression(ME). MEs usually occur for a brief period of time, typically lasting for less than 500ms [38] and are involuntary in nature, i.e. they naturally occurs when a person either deliberately or unconsciously conceals genuine emotions [6]. MEs are also signs of rapidly processed but unconcealed emotional states

[24]. MEs are known to appear in depressed inpatients [12] and in situations when individuals attempt to deceit their emotional expressions [30]. Capturing facial micro-expressions becomes more important in capturing boredom, anxiety, and flow because these emotions are not accompanied by a distinct combination of AUs [4, 29, 33].

Micro expression spotting involves automated search for temporal dynamics of MEs by detecting the temporal interval of facial muscles micro-movement in sequence of video frames [27]. Temporal dynamics of ME can be characterized in five phases: neutral (ME absent), onset (ME starts), apex (ME reaches its maximum expression), offset (ME begins to dissipate), and neutral (ME disappears) [6]. In the scope of our study, we explore the current trends in ME spotting research to advance the accuracy of face recognition algorithm. Prior works have tackled the problem in posed ME databases such as SMIC [23], SAMM [40], and CASME [39]. Most of the ME spotting algorithms work by computing the feature difference between frames. One of the recent studies, used both geometric features and appearance features based on dynamic image for capturing ME motion samples [36].

2.3 The Potential of Micro-expressions in Detecting Learning Affective States

In our study, we focus on three components of flow theory such as boredom, anxiety, and the state of flow [9] since successful detection of these emotions may assist in adjusting the complexity level of the task and creating an engaging learning environment.

2.3.1 Boredom. Boredom is one of the most frequent emotions occurring within the learning process and may be characterized by sudden and substantial movements which indicate discomfort [26]. This can be explained by the fact that dissatisfaction and low arousal cause the combination of prolonged periods of sitting still and sudden substantial movements such as stretching or switching pose [21]. These actions may not involve facial micro-expression, it can rather be expressed with the head movement [14, 41] and may possibly be captured with the use of macro-expressions.

2.3.2 Anxiety. Anxiety is another emotional response occurring in the learning process which may be reflected on learners' faces and is accompanied by specific facial expressions [7, 29]. Learning anxiety may be caused by many factors such as ambiguous terminology, unfamiliar topics, complicated formulas, and unclear annotations [32]. It is been shown that confusion caused by highly complex tasks may cause anxiety and reduce learning performance [32]. Prior work has also proved that anxiety may be successfully detectable by facial expression recognition algorithms [7]. It was shown that anxiety can be described as environmental-scanning behaviors and expressed via head swivels and eye darts [29].

2.3.3 Flow. Lastly, we define flow as a state of mind when the learner becomes fully immersed in the task [10]. Students' flow state is a less common metric, however, it was successfully measured during learning activities in prior showing promising opportunities and benefits of applying Flow Theory within educational settings [2]. The state of flow usually involves full involvement, focus, and enjoyment in the process of the activity [10]. The concentration causes reduced head movement and may also be captured by face recognition algorithms [8].

While multiple studies have confirmed the effectiveness of micro-expression for capturing human emotions [6, 14, 41], there is a gap in prior work in facial micro-expressions application for tracking learners' affective states. Inspection of earlier work led to the conclusion that there are no existing studies investigating the application of facial micro-expression recognition within an educational context with a focus on learners' affective states. This gap limits the understating of levering of this cost-effective method for enhancing the learning experience. Therefore, our work extends previous research on the micro-expression application within a new domain and tests the effectiveness of the facial micro-expression recognition method for identifying learners' emotional responses to a given learning task.

3 METHOD

3.1 Data Collection

3.1.1 Dataset. The dataset collection pipeline begins with detecting the user's face from the real-time video captured by the camera. Next, for each frame, the dlib face detector [19] is used to detect face bounding box. If no bounding box is found, the frame is dropped. Following this, faces from the original frames are cropped using these face bounding boxes. The cropped face is finally fed to a dlib shape detector to extract 68 facial landmark points. These facial landmarks are further used to calculate head pose representation(i.e yaw, pitch, and roll). The cropped face is also used to estimate different AU intensities. Our dataset contained features values of 68 landmarks, head pose and different AU intensities.

The data collected by our facial behavior sensing framework throughout the 2021 spring semester represented 54 sessions from 19 students. While most of the students complied with the study, completing both data collection and survey, a few only completed the end session survey. Overall, the study had 19 completed surveys with sensor data missing from 3 students. Moreover, the sessions with lengths less than 5 mins were dropped (4 sessions from 4 students). After applying these exclusions, the dataset had 31 sessions from 12 students. The number of sessions across all participants averaged to 2.5 sessions per participant. The mean and median of the session lengths were also calculated (73.02 minutes = mean and 31.31 minutes = median).

The ground truth for anxiety, flow, and boredom scores was collected via survey at end of each session. We used median score of individual class as the threshold to label them into high/low state of anxiety, flow, and boredom.

3.2 Framework

In this section, we describe our framework for micro-expression detection. To explain the details our framework we refer to three components including face shape representation, kinematic features extraction, and machine learning modeling.

3.2.1 Face Shape Representation. Face shape representation plays an important role in tasks such as psychological diagnosis, lie detection, facial expression recognition, Action Unit detection detection, and security systems. The foremost task for facial shape representation is landmark detection. But these facial landmarks may be distorted by head pose changes. To address such issues the IVA are used. These are a scale invariant geometric features computed on facial landmarks for the purpose of facial shape representation. IVA features are previously been used for task of FER and AU recognition. Facial expressions are usually manifested by contraction and expansion of facial muscles which alters the position of facial landmarks. To capture these changes the IVA features segment the face into small parts in shapes of triangle's angles. IVA geometric feature imbibes the scale invariant property of angles. Hence, the face Region of Interest no longer needs to be taken into normalized space leading to lossy down-sampling as is the case with most appearance and geometric features which depend upon the size of the image.

Motivated by prior work we consider nose centre as centroid of the face to compute IVA features. We segment the face into 6 regions namely nose centre, nose, right eye, left eye, jawline, and mouth. In total we compute 464 numbers of triangles by taking permutation of all possible triangles from the centroid to the individual parts of the face (right eye, left eye, nose, mouth and jaw line) from remaining 67 face landmarks. Of these landmarks, 11 for left eye, 11 for

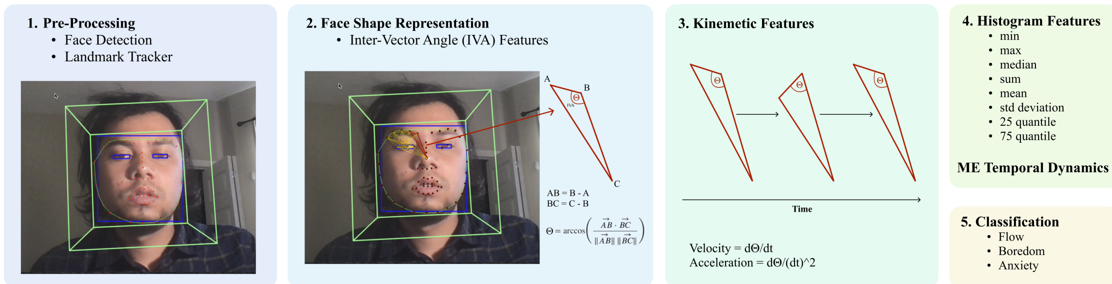

Fig. 2. Systematic diagram of proposed MicroFlow framework

right eye, 8 for nose, 20 for mouth, and 17 for jawline are used. After IVA features computation our feature space totals to 1392. Please refer Fig. 2 for visualization all the processes.

3.2.2 Kinematic Features Extraction. For spatial processing, we had 1392 IVA features for Face Shape representation and 3 features (i.e Yaw, Pitch, and Roll) for Head Pose representation. These features serves as a way to represent spatial information of the micro-expressions. Since the dimensionality of the IVA features(i.e 1392 features) is very high we used dimensionality reduction method, Principal Component Analysis (PCA) to reduce the dimensionality of the IVA to 10 features. To compute the temporal dynamic of the micro-expressions, we then compute velocity and acceleration on dimension reduced IVA features and head pose representation(i.e yaw, pitch, and roll) data-frame. Finally motivated by Micro-expression theory we computed histogram feature: min, max, median, mean, std deviation, quartile 1 and quartile 3 at an interval of 500ms. Since micro-expression go on and off the face in a fraction of a second we consider 500ms as window to compute histogram features. We hypothesised that 500ms interval combined with histogram features should be able to capture characteristics of temporal dynamics of Micro-expression such min/max for apex phase requiring accurate spotting of Micro-expression.

Our framework contributes by proposing micro-expression feature extraction method that model spatial-temporal deformation of facial muscle movement during micro-expression. Our framework imbibes the temporal dynamics of micro-expression combined with geometry based face shape and head pose representation.

3.2.3 Machine Learning Modeling. In our study, we used the LightGBM [18] an implementation of Boosting Trees to build the detection model. Gradient boosting falls under the category of boosting methods, which iteratively learns from multiple weak learners to build a strong model. Due to the sheer size of our dataset (n=31 sessions), we decided to evaluate our model over leave-one-session out(LOSO) cross validation. We took the entire session data with 15 minutes from the onset of the session of each participant and mapped it to ground truth values of flow state. Then a LightGBM Classifier, an implementation of boosted trees was used to train and predict low/high states of flow. In our study context, validating the model with LOSO cross validation means our flow model still works if multiple coding sessions from a subject come to the flow algorithm.

For the purpose of hyperparameter tuning of our LightGBM classifer we employ Optuna [1] over 1000 iterations. The hyperparamter tuning setting listed in Table 1 to ensuring reproducibility and explaining the ranges we explored.

Table 1. LightGBM parameters after applying Optuna Hyperparameter tuning

Parameter name	**Search space**	**Description**
lambda_l1	(1e-8, 10.0)	L1 regularization
lambda_l2	(1e-8, 10.0)	L2 regularization
num_leaves	(2, 256)	Number of branches in the tree
feature_fraction	(0.0, 1.0)	Used to speed and suppress overtraining of the learning
bagging_fraction	(0.0, 1.0)	Used to speed and suppress overtraining of the learning
min_child_samples	(5, 200)	Minimum number of data points needed in a child (leaf) node
learning_rate	(0.0, 1.0)	Learning rate

4 RESULTS

In this section, we describe our MicroFlow framework for the prediction of low/high states of flow across the three different models namely AU, IVA, and lastly AU and IVA combined model. To compare the model performance among different model we report accuracy, precision, recall, F1 and AUC score of each model. As Table 2 shows, we conducted experiments to understand what categories of facial behavior features, macro expression (AU), micro-expression (IVA), or combinations achieve the best performance.

4.1 Evaluation

4.1.1 Using AU Features. AU features are widely used for macro-expression recognition and have showcased state of the art performance in the same [17, 35]. We learn AU features with LightGBM classifier to predict flow, anxiety, and boredom and report individual performance. Using AU features we achieve 0.68 accuracy, 0.67 precision, 0.67 recall, 0.67 F1 and 0.74 AUC for flow model. For Boredom, we achieve 0.61 accuracy, 0.78 precision, 0.41 recall, 0.54 F1 and 0.70 AUC. For anxiety model, we achieved 0.71 accuracy, 0.71 precision, 0.90 recall, 0.79 F1, and 0.70 AUC. LightGBM classifer was able to learn AU features for flow, anxiety, and boredom model to an acceptable AUC. Table 2 demonstrate the results.

4.1.2 Using IVA Features. We test the performance of our framework using IVA features as described in section 3.2. with LightGBM as classifier. Our method focuses on capturing spatial-temporal deformation information during micro expression sequence. Using IVA features we achieve 0.61 accuracy, 0.64 precision, 0.47 recall, 0.54 F1 and 0.82 AUC for flow class. Our IVA model achieves 8% boost AUC from the AU only model. For boredom, we achieve, 0.58 accuracy, 0.70 precision, 0.41 recall, 0.52 F1, and 0.70 AUC. For anxiety model, we achieved 0.58 accuracy, 0.64 precision, 0.74 recall, 0.68 F1, and 0.64 AUC.

4.1.3 Combined Model: AU + IVA. We combine temporal dynamics information of micro-expression extracted using IVA features with AU features, to create a combined model which is an amalgam of both micro and macro-expression based features. We achieved an AUC of 0.84 for flow class, 10% improvement over AU only model. For same model we achieved accuracy 0.74, precision 0.81, recall 0.60 and F1 0.70. This model got the best accuracy, precision, and F1 score for flow class. However, the same trend was not observed in the boredom class. For boredom, AUC dropped to 0.66 and achieved 0.64 accuracy, 0.75 precision, 0.53 recall and 0.63 F1. For anxiety class we achieved an AUC of 0.71, 1% improvement over AU only model. We achieved auccuracy of 0.65, precision 0.65, recall 0.90, and F1 0.76. For flow and anxiety class combined model got the best performance.

Table 2. Performance Metrics with Head Pose

Class	Model	Acc	Precision	Recall	F1	AUC
Flow	AU	0.68	0.67	0.67	0.67	0.74
	IVA	0.61	0.64	0.47	0.54	0.82
	AU+IVA	0.74	0.81	0.60	0.70	0.84
Boredom	AU	0.61	0.78	0.41	0.54	0.70
	IVA	0.58	0.70	0.41	0.52	0.70
	AU+IVA	0.64	0.75	0.53	0.62	0.66
Anxiety	AU	0.71	0.71	0.90	0.79	0.70
	IVA	0.58	0.64	0.74	0.68	0.64
	AU+IVA	0.65	0.65	0.90	0.76	0.71

5 DISCUSSION AND LIMITATION

We trained and cross-validated our framework for flow, boredom, and anxiety on our in the wild dataset. We achieved an AUC of 0.84 for flow and 0.71 for anxiety respectively using the combination of AU and IVA features. For boredom, individual models using AU features and IVA features on their own performed better and achieved the same AUC score of 0.70. The study shows the viability of extracting temporal dynamics of the micro-expression sequence and learning it to predict affective states. The study confirms the existence of micro-expression as a response to the learning process which may be used to identify affective states. Our hypothesis was confirmed for anxiety and flow prediction showing the best performance with the combined model, however, the boredom did not follow the same trend and performed better using individual models. Thus, we concluded that different affective states may require different approches since they may be expressed differently.

Our study involved several limitations. First, our dataset was small including only 12 participants and 31 sessions in total. To compensate for this limitation we perform leave one out cross-validation on our dataset to check the generalizability of the model. Our data set was also collected with a frame rate of 7 frames per second which maybe not be frequent enough to capture all more characteristics of micro-expressions sequence.

Despite the limitations, our findings complemented prior work by investigating the use of micro-expressions in a learning context. The advantage of our method is that our framework can be used with the in-build/web camera available for most of the commodity computers. It does not require any additional devices or hardware which makes it affordable for educational institutions. This method also does not involve explicit face recording addressing participants privacy concerns meaning that engaging more participants in research study.

6 CONCLUSION AND FUTURE WORK

Our study tested the application of facial expressions recognition via AU features capturing macro-expressions, and IVA features capturing micro-expressions, and the combination of these models for affective learners states recognition including boredom, anxiety, and flow. Our results demonstrated the best performance of combined models was achieved for anxiety and the state of flow, while individual models performed better for boredom. We propose the MicroFlow framework based on micro-expression theory in a learning context that may be used by educators to adjust assignments' complexity level and create an engaging environment that may potentially contribute to achieving a flow state by learners and better educational outcomes. Going forward, we suggest using more data from a larger group of participants to collect data to make the model more efficient.

REFERENCES

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2623–2631.

[2] Karen Stansberry Beard. 2015. Theoretically Speaking: An Interview with Mihaly Csikszentmihalyi on Flow Theory Development and Its Usefulness in Addressing Contemporary Challenges in Education. *Educational Psychology Review* 27, 2 (Jun 2015), 353–364. https://doi.org/10.1007/s10648-014-9291-1

[3] M. Biasutti. 2011. Flow and Optimal Experience. In *Encyclopedia of Creativity (Second Edition)* (second edition ed.), Mark A. Runco and Steven R. Pritzker (Eds.). Academic Press, San Diego, 522–528. https://doi.org/10.1016/B978-0-12-375038-9.00099-6

[4] Isabelle Blanchette, Anne Richards, and Adele Cross. 2007. Anxiety and the interpretation of ambiguous facial expressions: The influence of contextual cues. *Quarterly Journal of Experimental Psychology* 60, 8 (2007), 1101–1115.

[5] Inge Bretherton, Janet Fritz, Carolyn Zahn-Waxler, and Doreen Ridgeway. 1986. Learning to Talk about Emotions: A Functionalist Perspective. *Child Development* 57, 3 (1986), 529–548. http://www.jstor.org/stable/1130334

[6] Vittoria Bruni and Domenico Vitulano. 2021. A Fast Preprocessing Method for Micro-Expression Spotting via Perceptual Detection of Frozen Frames. *Journal of Imaging* 7, 4 (2021), 68.

[7] Robbie M. Cooper, Angela C. Rowe, and Ian S. Penton-Voak. 2008. The role of trait anxiety in the recognition of emotional facial expressions. *Journal of Anxiety Disorders* 22, 7 (2008), 1120–1127. https://doi.org/10.1016/j.janxdis.2007.11.010

[8] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine* 18, 1 (2001), 32–80. https://doi.org/10.1109/79.911197

[9] Mihaly Csikszentmihalyi. 2009. Flow: The psychology of optimal experience.

[10] Mihaly Csikszentmihalyi. 2014. *Flow and Education*. Springer Netherlands, Dordrecht, 129–151. https://doi.org/10.1007/978-94-017-9094-9_6

[11] Paul Ekman. 2004. Emotions revealed. *Bmj* 328, Suppl S5 (2004).

[12] Paul Ekman and Wallace V Friesen. 1974. Nonverbal behavior and psychopathology. *The psychology of depression: Contemporary theory and research* (1974), 3–31.

[13] Susan Brown Feichtner and Elaine Actis Davis. 1984. Why Some Groups Fail: a Survey of Students' Experiences with Learning Groups. *Organizational Behavior Teaching Review* 9, 4 (1984), 58–73. https://doi.org/10.1177/105256298400900409 arXiv:https://doi.org/10.1177/105256298400900409

[14] E Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.

[15] ERF Hewson. 2018. Students? emotional engagement, motivation and behaviour over the life of an online course: reflections on two market research case studies. *Journal of Interactive Media in Education* 1, 10 (July 2018). https://eprints.leedsbeckett.ac.uk/id/eprint/5070/

[16] Yiling Hu, Bian Wu, and Xiaoqing Gu. 2017. An Eye Tracking Study of High- and Low-Performing Students in Solving Interactive and Analytical Problems. *Journal of Educational Technology Society* 20, 4 (2017), 300–311. http://www.jstor.org/stable/26229225

[17] Norhaida Hussain, Hamimah Ujir, Irwandi Hipiny, and Jacey-Lynn Minoi. 2017. 3D facial action units recognition for emotional expression. *arXiv preprint arXiv:1712.00195* (2017).

[18] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017), 3146–3154.

[19] Davis E King. 2009. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research* 10 (2009), 1755–1758.

[20] James F. Kisiel. 2003. Teachers, Museums and Worksheets: A Closer Look at a Learning Experience. *Journal of Science Teacher Education* 14, 1 (2003), 3–21. https://doi.org/10.1023/A:1022991222494 arXiv:https://doi.org/10.1023/A:1022991222494

[21] Stefan Kroes. 2005. Detecting boredom in meetings. *University of Twente* (2005), 1–5.

[22] Xiaowei Li, Bin Hu, Tingshao Zhu, Jingzhi Yan, and Fang Zheng. 2009. Towards Affective Learning with an EEG Feedback Approach. In *Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning* (Beijing, China) *(MTDL '09)*. Association for Computing Machinery, New York, NY, USA, 33–38. https://doi.org/10.1145/1631111.1631118

[23] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen. 2013. A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg)*. IEEE, 1–6.

[24] David Matsumoto and Hyi Sung Hwang. 2011. Psychological Science Agenda| May 2011. *Psychological Science* (2011).

[25] Richard E. Mayer. 2020. Searching for the role of emotions in e-learning. *Learning and Instruction* 70 (2020), 101213. https://doi.org/10.1016/j.learninstruc.2019.05.010 Understanding and measuring emotions in technology-rich learning environments.

[26] Stefanie Obergriesser and Heidrun Stoeger. 2020. Students' emotions of enjoyment and boredom and their use of cognitive learning strategies – How do they affect one another? *Learning and Instruction* 66 (2020), 101285. https://doi.org/10.1016/j.learninstruc.2019.101285

[27] Yee-Hui Oh, John See, Anh Cat Le Ngo, Raphael C-W Phan, and Vishnu M Baskaran. 2018. A survey of automatic facial micro-expression analysis: databases, methods, and challenges. *Frontiers in psychology* 9 (2018), 1128.

[28] Reinhard Pekrun, Thomas Goetz, Anne C. Frenzel, Petra Barchfeld, and Raymond P. Perry. 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology* 36, 1 (2011), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002 Students' Emotions and Academic Engagement.

[29] Adam M Perkins, Sophie L Inchley-Mort, Alan D Pickering, Philip J Corr, and Adrian P Burgess. 2012. A facial expression for anxiety. *Journal of personality and social psychology* 102, 5 (2012), 910.

[30] Stephen Porter and Leanne Ten Brinke. 2008. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological science* 19, 5 (2008), 508–514.

[31] Peter Robinson and Roger Gilabert. 2007. Task complexity, the Cognition Hypothesis and second language learning and performance. 45, 3 (2007), 161–176. https://doi.org/doi:10.1515/iral.2007.007

[32] Edgar H Schein. 2002. The anxiety of learning. Interview by Diane L. Coutu. *Harvard business review* 80, 3 (March 2002), 100—6, 134. http://europepmc.org/abstract/MED/11894378

[33] Hadas Shahar and Hagit Hel-Or. 2019. Micro Expression classification using facial color and deep learning methods. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 0–0.

[34] Kshitij Sharma, Hamed S. Alavi, Patrick Jermann, and Pierre Dillenbourg. 2016. A Gaze-Based Learning Analytics Model: In-Video Visual Feedback to Improve Learner's Attention in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics Knowledge* (Edinburgh, United Kingdom) *(LAK '16)*. Association for Computing Machinery, New York, NY, USA, 417–421. https://doi.org/10.1145/2883851.2883902

[35] Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence* 23, 2 (2001), 97–115.

[36] Thuong-Khanh Tran, Quang-Nhat Vo, and Guoying Zhao. 2021. DynGeoNet: Fusion Network for Micro-expression Spotting. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 745–749.

[37] Keith Trigwell, Robert A. Ellis, and Feifei Han. 2012. Relations between students' approaches to learning, experienced emotions and outcomes of learning. *Studies in Higher Education* 37, 7 (2012), 811–824. https://doi.org/10.1080/03075079.2010.549220 arXiv:https://doi.org/10.1080/03075079.2010.549220

[38] Wen-Jing Yan, Qi Wu, Jing Liang, Yu-Hsin Chen, and Xiaolan Fu. 2013. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior* 37, 4 (2013), 217–230.

[39] Wen-Jing Yan, Qi Wu, Yong-Jin Liu, Su-Jing Wang, and Xiaolan Fu. 2013. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*. IEEE, 1–7.

[40] Chuin Hong Yap, Connah Kendrick, and Moi Hoon Yap. 2020. Samm long videos: A spontaneous facial micro-and macro-expressions dataset. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 771–776.

[41] Wang-Wang Yu, Jingwen Jiang, and Yong-Jie Li. 2021. LSSNet: A Two-stream Convolutional Neural Network for Spotting Macro-and Micro-expression in Long Videos. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4745–4749.