MicroFlow: Advancing Affective States Detection in Learning Through Micro-Expressions

Mohammad Rahul Islam and Sang Won Bae

November 23, 2023

MicroFlow: Advancing Affective States Detection in Learning through Micro-expressions

***Abstract*—Gaining a deep understanding of student engagement is essential for designing effective learning experiences. In this study, we proposed the MicroFlow framework inspired by the concept of micro-expressions, to advance detecting learners' affective states in learning. We collected data from 19 students (54 sessions) during Python programming. We found that micro-expression features, Inter Vector Angles (IVA) combined models demonstrated the highest performance in detecting anxiety and flow state. The AUC for flow state improved by 10% (reaching 84%) compared to the AU model. For anxiety and boredom, we achieved AUC values of 71% and 70%, respectively. We highlighted the feasibility of our framework as a cost-effective tool that enable educators to create a more engaging learning environment by adjusting the complexity level of learners tasks, ultimately improve learning outcomes.**

***Index Terms*—Flow Theory, Micro-expression Theory, Facial Expression, Emotion, Education, Passive Sensing**

I. Introduction

One crucial factor that affects the effectiveness of the learning process is individuals' emotional response to a given task [1]. Students may find easy tasks boring or feel anxiety and frustration when tackling highly complex assignments. The ultimate goal for learners is to attain a state of flow, characterized by deep engagement, focus, and enjoyment, as it facilitates effective learning [2, 3]. Understanding learners' emotional responses enables instructors to optimize educational materials by adjusting the task difficulty and creating an engaging learning environment. According to the Flow Theory, achieving a state of flow involves balancing skill level and task complexity [2].

Facial expressions are commonly used to detect learners' emotional responses, and facial expression recognition (FER) algorithms have been applied to reveal affective states such as boredom, anxiety, and flow [4, 5]. While macro-expressions have been used to recognize emotions during studying, the complexity of human emotions necessitates the inclusion of micro-expressions. Combining Inter Vector Angles (IVA) features for micro-expression detection with Action Unit (AU) features for macro-expression recognition can enhance the detection of concealed and unconsciously produced emotions [6, 7]. By incorporating both IVA and AU features, our proposed approach improves the accuracy of detecting flow, anxiety, and boredom in the learning context [8].

II. Background

A. Methods for Capturing Students' Learning Experience

Individual learning experiences are influenced by various factors, including emotions, cognitive skills, language proficiency, and prior knowledge [9]. This study focuses on learning-centered affective states such as boredom, anxiety, and flow, which reflect learners' emotional responses during their educational journey. Emotional experiences in learning have been linked to academic performance [10, 11, 12], and task complexity can impact learners' emotions [13]. Traditional survey-based approaches have limitations in capturing real-time experiences and may not be feasible in a classroom setting [14]. Alternative methods like eye-tracking analysis [15, 16] or wearable sensing technology such as electroencephalogram [17] offer advanced ways to identify affective states but may be costly and challenging to implement. In contrast, computer vision technology, specifically facial expression recognition, provides a cost-effective and high-performance approach [18]. This study explores the effectiveness of this approach in analyzing macro- and micro-expressions in the learning process, leveraging its advantages for educational contexts.

B. The Potential of Micro-expressions in Detecting Learning Affective States

Micro-expressions refer to the rapid and involuntary facial contractions and relaxations that typically occur within a short duration, often within a 500 ms window [8]. These subtle facial expressions are characterized by distinctive features, such as the presence of an apex phase, which is the most intense moment of the micro-expression [8]. Micro-expressions have been found to be reliable indicators of genuine emotions, as they occur spontaneously and are difficult to control consciously. Existing studies have shown micro-expressions' efficacy in capturing human emotions [18, 19, 20].
However, a research gap exists in applying facial micro-expression recognition to track learners' affective states, particularly in education. Our study addresses this gap by exploring the application of micro-expression analysis to identify learners' emotional responses in an educational context.

III. Method

A. Dataset

We developed a novel facial behavior sensing system (a.k.a FacePsy) equipped with state-of-the-art Facial Expression Recognition (FER) modules to collect data. The dataset collection pipeline involved real-time video capture from a camera, followed by face detection using the dlib face detector [21]. For each frame, if a face bounding box was detected, the face was cropped using the bounding box coordinates. The cropped face was then processed using a dlib shape detector to extract 68 facial landmark points, which were utilized to calculate head pose representation (yaw, pitch, and roll). Additionally, the cropped face was used to estimate different Action Unit

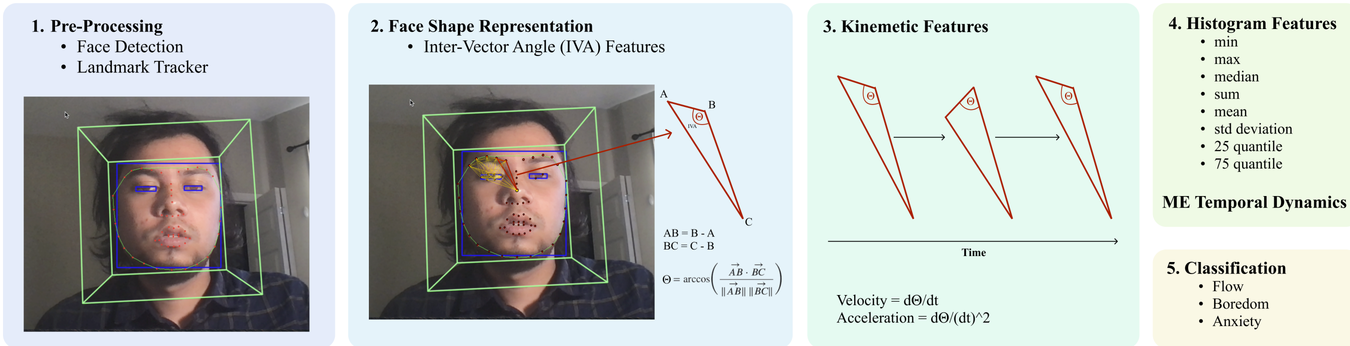

Fig. 1. Systematic Diagram of the Proposed MicroFlow Framework

(AU) intensities. Our dataset comprised feature values of the 68 landmarks, head pose, and various AU intensities.

During the spring semester of 2021, we collected data from 19 students, resulting in 54 sessions in engineering courses. A session refers to the duration in which a student dedicated time to complete a specific Python programming assignment. While the majority of students adhered to the study protocol by completing both data collection and surveys, a few only completed the end-of-session survey. Overall, 19 completed surveys were obtained, with sensor data missing from 3 students. Sessions lasting less than 5 minutes were excluded, leading to the removal of 4 sessions from 4 students. After these exclusions, the dataset consisted of 31 sessions from 12 students. On average, each participant had 2.5 sessions. The mean and median session lengths were computed as 73.02 minutes and 31.31 minutes, respectively.

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

B. Feature Extraction

The IVA (Inter Vector Angles) method is employed to capture facial expressions by analyzing the movement of facial landmarks through the contraction and expansion of facial muscles. It achieves this by segmenting the face into small triangular regions. The IVA feature incorporates the scale-invariant property of angles. In our study, we consider the nose center as the face's centroid for computing IVA features (Figure 1. The face is further divided into six regions: nose center, nose, right eye, left eye, jawline, and mouth. By permuting all possible combinations of the remaining 67 face landmarks with the centroid and the individual parts of the face, we compute a total of 464 triangles. Specifically, 11 landmarks are used for the left eye, 11 for the right eye, 8 for the nose, 20 for the mouth, and 17 for the jawline. After computing the IVA features, our feature space comprises a total of 1392 values.

For spatial processing, we employed 1392 IVA features for Face Shape representation and 3 features (Yaw, Pitch, and Roll) for Head Pose representation. These features captured the spatial information of the micro-expressions. To address the high dimensionality of the IVA features, we utilized Principal Component Analysis to reduce them to 10 features. To analyze the temporal dynamics of the micro-expressions, we computed velocity and acceleration on the dimension-reduced IVA features and the head pose representation (yaw, pitch, and roll) data. Inspired by Micro-expression theory, we computed histogram features, such as minimum, maximum, median, mean, standard deviation, quartile 1, and quartile 3, at intervals of 500ms. This choice of a 500ms window allowed us to capture the rapid occurrence of micro-expressions, enabling us to spot important characteristics such as the min/max for the apex phase accurately.

Our framework introduces a novel method for extracting micro-expression features that captures the spatial-temporal deformation of facial muscle movement, incorporating both temporal dynamics and geometry-based face shape and head pose representation.

C. Machine Learning Modeling

In our study, we used LightGBM [22], a Boosting Trees implementation, to build the detection model. Gradient boosting is a type of boosting method that iteratively learns from weak learners to create a strong model. To accommodate our substantial dataset (n=31 sessions), we evaluated the model using leave-one-session-out (LOSO) cross-validation. The session data from each participant, starting 15 minutes after the session's onset, was paired with ground truth flow state values. We employed LightGBM Classifier to train and predict the low/high flow states. LOSO cross-validation ensured that our flow model remained effective when multiple coding sessions from a subject were included. For hyperparameter tuning of the LightGBM classifier, we utilized Optuna [23] with 1000 iterations.

IV. 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 I shows, we conducted experiments to understand what categories of facial behavior features, macro expression (AU), micro-expression (IVA), or combinations achieve the best performance.

A. Using AU Features

AU features are widely used for macro-expression recognition and have showcased state of the art performance in the same [24, 25]. 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 I demonstrate the results.

B. 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.

C. 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 I
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

V. DISCUSSION AND LIMITATION

The study demonstrated the feasibility of extracting temporal dynamics from micro-expression sequences to enhance the predictability of algorithms in detecting learners' affective states during online Python programming in the context of the COVID-19 pandemic. We confirmed the presence of micro-expressions as responses to the learning process, which can be utilized to identify learners' affective states. Our hypothesis was supported for anxiety and flow prediction, showing superior performance with the micro and macro-expression features-combined models. However, boredom exhibited a different pattern, performing better with the macro- or micr-expression features only models. This suggests that different affective states may require distinct approaches, as their expressions can vary.

The dataset was relatively small, consisting of only 12 participants and 31 sessions. To mitigate this limitation, we employed leave-one-out cross-validation to assess the generalizability of the model. The data was collected at a frame rate of 7 frames per second, potentially limiting the capture of certain micro-expression characteristics. Despite these limitations, our findings complement prior research by exploring the use of micro-expressions in a learning context. The key advantage of our method is its compatibility with built-in or web cameras commonly available on most computers, eliminating the need for additional devices or hardware and making it cost-effective for educational institutions. Our approach does not involve explicit face recording, addressing participants' privacy concerns and facilitating the inclusion of more participants in research studies.

VI. CONCLUSION AND FUTURE WORK

We propose the MicroFlow framework, which combines micro- and macro facial expressions to enhance the understanding of affective states in the learning context. In this study, we explored the use of facial expression recognition as a means to identify specific affective learner states, including boredom, anxiety, and flow. To capture micro-expression, we utilized IVA features, and for macro-expressions, we employed AU features, drawing inspiration from micro-expression theory. While individual models showed improved performance for detecting boredom, the combined use of IVA and AU features achieved the highest accuracy for anxiety and the flow state. We highlight the potential of the MicroFlow framework which help educators and researchers adjust a level of complexity of assignments or lecture materials and create more engaging learning environments which potentially led to improved learning outcomes. To ensure applicability and generalizability, we plan to test our framework in diverse learning environments and larger classroom settings.

REFERENCES

[1] E. Hewson, "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*, vol. 1, no. 10, July 2018. [Online]. Available: https://eprints.leedsbeckett.ac.uk/id/eprint/5070/

[2] M. Csikszentmihalyi, "Flow: The psychology of optimal experience," 2009.

[3] "Flow and optimal experience," in *Encyclopedia of Creativity (Second Edition)*, second edition ed., M. A. Runco and S. R. Pritzker, Eds. San Diego: Academic Press, 2011, pp. 522–528.

[4] A. Sarrafzadeh, H. G. Hosseini, C. Fan, and S. P. Overmyer, "Facial expression analysis for estimating learner's emotional state in intelligent tutoring systems," in *Proceedings 3rd IEEE International Conference on Advanced Technologies*. IEEE, 2003, pp. 336–337.

[5] P. Ekman, "Emotions revealed," *Bmj*, vol. 328, no. Suppl S5, 2004.

[6] A. M. Perkins, S. L. Inchley-Mort, A. D. Pickering, P. J. Corr, and A. P. Burgess, "A facial expression for anxiety." *Journal of personality and social psychology*, vol. 102, no. 5, p. 910, 2012.

[7] R. Islam, K. Ahuja, S. Karmakar, and F. Barbhuiya, "Sention: A framework for sensing facial expressions," *arXiv preprint arXiv:1608.04489*, 2016.

[8] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.

[9] J. F. Kisiel, "Teachers, museums and worksheets: A closer look at a learning experience," *Journal of Science Teacher Education*, vol. 14, no. 1, pp. 3–21, 2003. [Online]. Available: https://doi.org/10.1023/A:1022991222494

[10] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld, and R. P. Perry, "Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq)," *Contemporary Educational Psychology*, vol. 36, no. 1, pp. 36–48, 2011, students' Emotions and Academic Engagement.

[11] I. Bretherton, J. Fritz, C. Zahn-Waxler, and D. Ridgeway, "Learning to talk about emotions: A functionalist perspective," *Child Development*, vol. 57, no. 3, pp. 529–548, 1986. [Online]. Available: http://www.jstor.org/stable/1130334

[12] R. E. Mayer, "Searching for the role of emotions in e-learning," *Learning and Instruction*, vol. 70, p. 101213, 2020, understanding and measuring emotions in technology-rich learning environments.

[13] P. Robinson and R. Gilabert, "Task complexity, the cognition hypothesis and second language learning and performance," vol. 45, no. 3, pp. 161–176, 2007. [Online]. Available: https://doi.org/10.1515/iral.2007.007

[14] S. B. Feichtner and E. A. Davis, "Why some groups fail: a survey of students' experiences with learning groups," *Organizational Behavior Teaching Review*, vol. 9, no. 4, pp. 58–73, 1984. [Online]. Available: https://doi.org/10.1177/105256298400900409

[15] K. Sharma, H. S. Alavi, P. Jermann, and P. Dillenbourg, "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*, ser. LAK '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 417–421. [Online]. Available: https://doi.org/10.1145/2883851.2883902

[16] Y. Hu, B. Wu, and X. Gu, "An eye tracking study of high- and low-performing students in solving interactive and analytical problems," *Journal of Educational Technology Society*, vol. 20, no. 4, pp. 300–311, 2017. [Online]. Available: http://www.jstor.org/stable/26229225

[17] X. Li, B. Hu, T. Zhu, J. Yan, and F. Zheng, "Towards affective learning with an eeg feedback approach," in *Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning*, ser. MTDL '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 33–38. [Online]. Available: https://doi.org/10.1145/1631111.1631118

[18] V. Bruni and D. Vitulano, "A fast preprocessing method for micro-expression spotting via perceptual detection of frozen frames," *Journal of Imaging*, vol. 7, no. 4, p. 68, 2021.

[19] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, p. 5, 1978.

[20] W.-W. Yu, J. Jiang, and Y.-J. Li, "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*, 2021, pp. 4745–4749.

[21] D. E. King, "Dlib-ml: A machine learning toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[22] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017.

[23] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[24] N. Hussain, H. Ujir, I. Hipiny, and J.-L. Minoi, "3d facial action units recognition for emotional expression," *arXiv preprint arXiv:1712.00195*, 2017.

[25] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 97–115, 2001.