The 3rd Challenge on Human Behavior Analysis for Emotion Understanding (MiGA) 2025: From Recognition to Emotion Understanding

Haoyu Chen[1], Björn W. Schuller[2], Ehsan Adeli[3] and Guoying Zhao[1,*]

[1]CMVS, University of Oulu, Finland
[2]GLAM, Imperial College London, United Kingdom
[3]Stanford University, USA

Abstract

This paper presents a summary of the 3rd Challenge on Human Behavior Analysis for Emotion Understanding (MiGA) 2025, held in conjunction with IJCAI 2025 in Guangzhou, China. As a continuation of the MiGA workshop series inaugurated in 2023 and 2024, this year's edition advances the field from recognition to emotion understanding, marking an important step toward cognitively grounded emotion understanding through subtle body behaviors. The 2025 challenge consists of three independent tracks: (1) multi-modal micro-gesture classification from pre-segmented clips, (2) online micro-gesture recognition in continuous sequences, and the **newly opened** track (3) gesture-based emotion reasoning for predicting hidden affective states from real-world interviews. Two large-scale datasets—iMiGUE and SMG—were released with additional emotion annotations. The competition attracted **over 80 registered entrants** from 17 research institutions worldwide and was hosted on the Kaggle platform. Results indicate continuous improvements over previous years: the top team achieved 73.1% accuracy in Track 1 and 0.275 F1 score in Track 2, while the new emotion recognition track achieved an average accuracy of above 0.60. Analyses reveal growing trends in hybrid Transformer–Mamba architectures, CLIP-guided semantic alignment, and large language model-based multi-modal fusion for emotion reasoning. MiGA 2025 highlights the shift in affective computing research from perceptual recognition to inferential understanding—bridging human behavior, cognition, and artificial intelligence. The series continues to serve as a global benchmark for emotion understanding through human behavior analysis.

Keywords

Affective Computing, Micro-Gesture Analysis, Multimodal Emotion Understanding, Emotion Reasoning, Human Behavior Analysis, Gesture Recognition

1. Introduction

Understanding subtle human behaviors is a long-standing challenge in affective computing and human-centered AI [1, 2, 3, 4, 5, 6]. Among various non-verbal cues, micro-gestures—brief, involuntary, and low-amplitude movements of hands, face, or upper body—play a crucial role in revealing suppressed or implicit emotional states. Unlike conventional actions or expressive gestures, micro-gestures are sparse in time, localized in space, and highly context-dependent, making them difficult to annotate and recognize automatically [7, 8, 9].

The Micro-Gesture Analysis for Hidden Emotion Understanding (MiGA) challenge series was established to systematically advance this research direction, continuously hosted on the IJCAI conference every year since 2023 [10]. In 2023 and 2024, MiGA organized the first [1] and second challenges [2], attracting more than 100 participants in total. The challenges are on two public MG datasets (iMiGUE

MiGA@IJCAI25: International IJCAI Workshop on 3rd Human Behavior Analysis for Emotion Understanding, August 29, 2025, Guangzhou, China.
*Corresponding author.
chen.haoyu@oulu.fi (H. Chen); bjoern.schuller@imperial.ac.uk (B. W. Schuller); eadeli@stanford.edu (E. Adeli); guoying.zhao@oulu.fi (G. Zhao)
0000-0003-3267-2664 (H. Chen); 0000-0002-6478-8699 (B. W. Schuller); 0000-0002-0579-7763 (E. Adeli); 0000-0003-3694-206X (G. Zhao)

[1]https://cv-ac.github.io/MiGA2023/
[2]https://cv-ac.github.io/MiGA2/

and SMG) [8, 7, 9] with the tasks of MG classification and online recognition. With more research interests gained in the recognition of micro-gestures [11, 12, 13, 14, 15], we chose to continuously host the MiGA competition this year. Building upon previous editions, MiGA-IJCAI 2025 [3] further expands the scope by integrating micro-gesture classification, online detection, and behavior-based emotion understanding.

This article provides a comprehensive overview of the MiGA-IJCAI 2025 workshop and challenge, summarizing the tasks, dominant methodological trends, representative solutions, and emerging research directions revealed through participating works.

2. Competition Tracks and Datasets

MiGA 2025 consists of multiple tracks addressing complementary aspects of micro-gesture understanding: **1. Micro-Gesture Classification**: fine-grained recognition of short gesture clips into predefined categories on the iMiGUE dataset; **2. Online Micro-Gesture Recognition**: temporal localization and classification of micro-gestures in long, untrimmed videos on the SMG dataset; and a new introduced track **3. Behavior-based Emotion Understanding**: inferring emotional states from non-verbal behavioral cues on the iMiGUE dataset, where the goal is to predict wether professional tennis players win or lose their match based on their post-match press interview.

All tracks are developed based on the iMiGUE dataset [8] and the SMG dataset [9], both of which consist of identity-anonymized interview recordings of professional tennis players. These datasets provide fine-grained annotations of micro-gesture categories together with emotion-related labels. Notably, they pose substantial challenges for affective behavior modeling, such as pronounced class imbalance, high semantic similarity across categories, and extremely subtle motion patterns, making them well aligned with real-world affective analysis scenarios.

3. Competition Tracks and Datasets

MiGA 2025 consists of multiple tracks addressing complementary aspects of micro-gesture understanding: **1. Micro-Gesture Classification**: fine-grained recognition of short gesture clips into predefined categories on the iMiGUE dataset; **2. Online Micro-Gesture Recognition**: temporal localization and classification of micro-gestures in long, untrimmed videos on the SMG dataset; and a new introduced track **3. Behavior-based Emotion Understanding**: inferring emotional states from non-verbal behavioral cues on the iMiGUE dataset, where the goal is to predict wether professional tennis players win or lose their match based on their post-match press interview.

All tracks are developed based on the iMiGUE dataset [8] and the SMG dataset [9], both of which consist of identity-anonymized interview recordings of professional tennis players. These datasets provide fine-grained annotations of micro-gesture categories together with emotion-related labels. Notably, they pose substantial challenges for affective behavior modeling, such as pronounced class imbalance, high semantic similarity across categories, and extremely subtle motion patterns, making them well aligned with real-world affective analysis scenarios.

4. Competition Itinerary

This section encompasses both the competition schedule and relevant participant details.

[3] https://cv-ac.github.io/MiGA2025/

4.1. Competition agenda

The challenge was managed using the Kaggle competition platform [4] [5] [6]. The schedule of the competition was as follows:
Mar. 30th, 2025. Call for Challenge online. Registration starts.
Apr. 05, 2025. Release of training data, development toolkit, and sample codes.
May 11, 2025. Release of testing data.
May 25, 2025. Final test submission deadline. Registration ends.
May 27, 2025. Release of challenge results.
Jun. 07, 2025. Paper submission deadline.
Jun. 10, 2025. Notification to authors.
Jun. 15, 2025. Camera-ready deadline.
Aug. 29, 2025. MiGA IJCAI 2025 Workshop, Guangzhou, China.

4.2. Participants

The competition has been conducted using Kaggle, a commonly recognized challenge open-source platform. We created a different competition for each track, having separate information and leaderboards (see the Kaggle platform links). A total of 82 entrants have been registered on the Kaggle platform, 35 for track 1, 17 for track 2, and 30 for track 3.

All these users were able to access the data for the developing stage and submit their predictions for this stage. For the final evaluation stage, team registration was mandatory, and a total of 38 participants were successfully registered: 16 for track 1, and 6 for track 2, and 16 for track 3. During the challenge period, a total of 428 submissions were made, with 152 for track 1, 55 for track 2, and 221 for track 3.

5. Challenge Results and Methods

In this section, we present the winning methods on all three tasks. For all three tracks, we verify their validation of the methods by asking the top three teams to submit their source code and predictions for the test sets. Below, we introduce the ranking of each track.

5.1. Track 1: Multi-modal MG classification

Table 1 summarizes the methods of the top three teams on the test set of track 1. The top three solutions in the MiGA 2025 Track 1 micro-gesture recognition challenge represent three distinct yet complementary technical paradigms. The first-place method, MM-Gesture [16], follows a large-scale multimodal fusion strategy, integrating skeletal information (joint and limb) with multiple visual modalities, including RGB, Taylor-series temporal videos, optical flow, and depth. It employs PoseConv3D [17] and Video Swin Transformer backbones for modality-specific spatiotemporal modeling, enhanced by transfer learning on the MA-52 dataset [11] and a weighted ensemble scheme to fully exploit cross-modal complementarity. The second-place approach [18] focuses on efficient single-modality representation learning, proposing a Global-Aware Importance Estimation (GAIE) module within a Video Vision Transformer to mitigate background redundancy by estimating token-level global importance and adaptively aggregating less informative background tokens into salient foreground regions, achieving strong performance using RGB input alone. In contrast, the third-place method [19] adopts a skeleton-centric spatiotemporal modeling pipeline based on PoseC3D, where 2D skeletal sequences are encoded as 3D heatmaps and enhanced through topology-aware joint connectivity, structure-preserving temporal alignment, and auxiliary semantic embedding supervision, emphasizing the role of structural priors and temporal coherence in fine-grained micro-gesture recognition.

[4] https://www.kaggle.com/competitions/the-3rd-mi-ga-ijcai-challenge-track-1/
[5] https://www.kaggle.com/competitions/the-3rd-mi-ga-ijcai-challenge-track-2/
[6] https://www.kaggle.com/competitions/the-3rd-mi-ga-ijcai-challenge-track-3/

Table 1
Summary of the Top-3 Methods in MiGA 2025 Track 1 (Micro-Gesture Classification on the iMiGUE dataset).

Team	Rank	Modality	Backbone	Accuracy
HFUT-VUT (MM-Gesture [16])	1	Skeleton + RGB + Flow + Depth	PoseConv3D + Video Swin	73.2%
awuniverse (GAIE-ViT) [18]	2	RGB only	Video ViT	68.7%
Lonelysheep (PoseC3D J&L) [19]	3	Skeleton (Joint + Limb)	PoseC3D	67.0%

Table 2
Top-2 Methods of multi-modal MG online recognition results in MiGA 2025 on the SMG dataset.

Team	Rank	Modality	Backbone / Detector	F1 score
HFUT-VUT[20]	1	RGB	VideoMAEv2-g + DyFADet (STA)	0.3803
Chutian Meng [21]	2	RGB	VideoMAE-g + DyFADet	0.3153

Table 3
Top-3 Methods of multi-modal MG-based emotion recognition results in MiGA 2025 on the iMiGUE dataset.

Team	Rank	Modality	Backbone / Key Technique	Accuracy
backpacker [24]	1	RGB + Pose	YOLO11x+DINOv2 + VLM Pseudo-Label (Gemini)	0.6923
ISPCAST [25]	2	RGB (Global+Face)	MViTv2-S + SwinFace + InterFusion	0.6346
Haozhe Bu et al. [26]	3	RGB (Full+Face)	ResNet34 + Emotion Priors + Class Balancing	0.6346

5.2. Track 2: Multi-modal MG online recognition

Table 3 summarizes the methods of the top two teams ranked on the test set of track 2. The MiGA 2025 Track 2 (Online Micro-Gesture Recognition) Top-2 solutions both follow a strong two-stage paradigm built on pretrained video representations and temporal action detection, but differ in their key enhancements, as shown in Table 3. Both of the methods surpass the baseline method [22, 23] (0.203) by a large margin. The 1st-place method (HFUT-VUT) [20] is based on the DyFADet framework with VideoMAEv2-g features, and introduces two major contributions: (i) a category-frequency adaptive data augmentation strategy to address the severe class imbalance of spontaneous micro-gestures, and (ii) a multi-scale spatial–temporal attention module inserted into the detection head to improve boundary localization and fine-grained gesture discrimination. This design achieved the best performance with an F1 score of 38.03, ranking first on the leaderboard. The 2nd-place method (Chutian Meng et al.) [21] also adopts a two-stage pipeline, using the VideoMAE family of self-supervised video transformers to extract expressive spatiotemporal RGB features, followed by the query-based DyFADet temporal detector for dynamic duration modeling and precise online localization in long untrimmed videos. Their approach highlights the effectiveness of combining powerful pretrained visual backbones with query-based temporal detection, ultimately securing second place in Track 2.

5.3. Track 3: MG-based emotion recognition

Table 3 summarizes the methods of the top three teams ranked on track 3, MG-based emotion recognition. Track 3 (Behavior-Based Hidden Emotion Understanding) in MiGA 2025 attracted three highly competitive Top-3 solutions, each leveraging distinct strategies for inferring win/loss outcomes from post-match interview videos under subtle emotion concealment and severe class imbalance. The 1st-place method ("backpacker") [24] proposes a powerful weak-to-strong multimodal training paradigm: YOLO11x + DINOv2 are used to extract dense portrait-level visual sequences, while Gemini 2.5 Pro with Chain-of-Thought + Reflection prompting generates pseudo-labels and reasoning texts as weak supervision. These pseudo-labeled samples are merged back into training, combined with OpenPose keypoint streams and ultra-long Transformers for holistic temporal modeling, achieving the best accuracy of 69.23%. The 2nd-place solution (ISPCAST) [25] adopts a dual-stream transformer fusion

framework that explicitly separates contextual and facial dynamics: MViTv2-S encodes global video cues, SwinFace extracts face-specific embeddings, and an InterFusion gated fusion module iteratively integrates both modalities, leading to robust multimodal emotion reasoning and securing second place with 63.46%. The 3rd-place method (Haozhe Bu et al.) [26] emphasizes robustness under imbalance through emotional priors and ensemble balancing: a DeepFace-based emotion prior module provides semantic tendency scores, while dual-channel ResNet34 models jointly encode full-body behaviors and facial expressions. Severe skew is mitigated via intra-class partitioning of the majority class, and final predictions are produced by majority voting across multiple complementary classifiers, yielding 63.46% accuracy and ranking third despite using only RGB data. Note that both the second and third schemes obtained the same prediction accuracy: 63.46%, the 2nd-place solution uses fewer entries on the Kaggle platform, thus has a higher ranking.

6. Discussion

This paper has described the main characteristics of the MiGA 2025 Challenge hosted at IJCAI 2025, Guangzhou, China, which included tracks on (i) Multi-modal MG classification, (ii) Multi-modal MG online recognition and (iii) MG-based emotion recognition. Two large datasets (the SMG and iMiGUE datasets) were introduced and made available on the competition platform with corresponding toolkits to the participants for a fair comparison of the performance results.

Analyzing the methods introduced by the above participants, overall, MiGA 2025 demonstrates both strong progress and clear limitations across the three tracks. Track 1–2 confirm that modern video foundation backbones (e.g., VideoMAE) coupled with temporal detection frameworks (DyFADet) form an effective paradigm for spotting and online recognizing micro-gestures in long videos, yet the best Track 2 F1 remains only 0.38, highlighting persistent bottlenecks in subtle boundary localization, extreme class imbalance, sparse event frequency, and limited multimodal exploitation. Track 3 achieves higher top accuracy (0.69), but success relies heavily on holistic behavioral reasoning and external priors such as VLM-based pseudo-labeling, while the sharp drop to 0.63 for lower-ranked methods indicates weak generalization and the intrinsic difficulty of inferring concealed emotions from interviews alone.

Aside from those winning schemes proposed for the MiGA competition 2025, some other interesting research related to MG is also included in the MiGA 2025 workshop. For instance, Zhou et al. [27] introduce an Unreal Engine 5 based synthetic data generator to augment scarce wearable motion datasets and improve cross-person human behaviour recognition performance. Wang et al. [28] propose GraphAU-Pain, a graph neural network that explicitly models facial Action Units and their relations for interpretable pain intensity estimation, achieving strong gains on UNBC. Patapati et al. further present CLIP-MG [29], a pose-guided semantic attention adaptation of CLIP that fuses skeleton and RGB cues for micro-gesture recognition on iMiGUE.

Future work will focus on scaling micro-gesture emotion understanding via self-supervised and weakly supervised learning, enabling models to exploit vast unannotated behavioral data while reducing reliance on costly manual labels. Another promising direction is to move beyond isolated recognition toward cognitively grounded, temporally coherent behavior modeling that integrates long-term context, causal reasoning, and multimodal priors for robust real-world emotion inference [30].

Acknowledgments

We wish to acknowledge all the 80 entrants who participated in MiGA 2025. We especially thank Associate Professor Xiaobai Li for assisting in organizing the event, and Yueyi Yang and Fang Kang for the technical support. We also wish to thank the platforms Kaggle for providing free sources for us to organize the challenges of MiGA 2024 and 2025.

This MiGA challenge and workshop was supported by the Academy of Finland for Academy Professor project EmotionAI (grants 336116, 345122), the University of Oulu & The Academy of Finland Profi 7 Hybrid Intelligence (grant 352788), and Research Fellow project (grant 371019), as well as the Ministry

of Education and Culture of Finland for AI forum project. We also wish to acknowledge the CSC – IT Center for Science, Finland, for computational resources.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] G. Zhao, Y. Li, Q. Xu, From emotion ai to cognitive ai, International Journal of Network Dynamics and Intelligence (2022) 65–72.

[2] Z. Lian, H. Chen, L. Chen, H. Sun, L. Sun, Y. Ren, Z. Cheng, B. Liu, R. Liu, X. Peng, et al., Affectgpt: A new dataset, model, and benchmark for emotion understanding with multimodal large language models, ICML (2025).

[3] S. Järvelä, G. Zhao, A. Nguyen, H. Chen, Hybrid intelligence: Human-ai coevolution and learning (2025).

[4] H. Chen, Human emotion understanding via human behaviors and go beyond, in: Proceedings of the 3rd International Workshop on Multimodal and Responsible Affective Computing, 2025, pp. 1–1.

[5] Z. Lian, H. Sun, L. Sun, H. Chen, L. Chen, H. Gu, Z. Wen, S. Chen, S. Zhang, H. Yao, et al., Ov-mer: Towards open-vocabulary multimodal emotion recognition, ICML (2025).

[6] Z. Lian, L. Sun, L. Chen, H. Chen, Z. Cheng, F. Zhang, Z. Jia, Z. Ma, F. Ma, X. Peng, et al., Emoprefer: Can large language models understand human emotion preferences?, ICLR (2026).

[7] H. Chen, X. Liu, X. Li, H. Shi, G. Zhao, Analyze spontaneous gestures for emotional stress state recognition: A micro-gesture dataset and analysis with deep learning, in: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), IEEE, 2019, pp. 1–8.

[8] X. Liu, H. Shi, H. Chen, Z. Yu, X. Li, G. Zhao, imigue: An identity-free video dataset for micro-gesture understanding and emotion analysis, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 10631–10642.

[9] H. Chen, H. Shi, X. Liu, X. Li, G. Zhao, Smg: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis, International Journal of Computer Vision 131 (2023) 1346–1366.

[10] H. Chen, B. W. Schuller, E. Adeli, G. Zhao, The 2nd challenge on micro-gesture analysis for hidden emotion understanding (miga) 2024: Dataset and results, in: MiGA 2024: Proceedings of IJCAI 2024 Workshop&Challenge on Micro-gesture Analysis for Hidden Emotion Understanding (MiGA 2024) co-located with 33rd International Joint Conference on Artificial Intelligence (IJCAI 2024), 2024.

[11] D. Guo, K. Li, B. Hu, Y. Zhang, M. Wang, Benchmarking micro-action recognition: Dataset, method, and application, IEEE Transactions on Circuits and Systems for Video Technology (2024).

[12] D. Guo, X. Li, K. Li, H. Chen, J. Hu, G. Zhao, Y. Yang, M. Wang, Mac 2024: Micro-action analysis grand challenge, in: Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 11304–11305.

[13] A. Shah, H. Chen, G. Zhao, Representation learning for topology-adaptive micro-gesture recognition and analysis, in: IJCAI-MIGA Workshop & Challenge on Micro-gesture Analysis for Hidden Emotion Understanding (MiGA) July 21, 2023 Macao, China, Redaktion Sun SITE, 2023.

[14] A. Shah, H. Chen, G. Zhao, Naive data augmentation might be toxic: Data-prior guided self-supervised representation learning for micro-gesture recognition, in: 2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG), IEEE, 2024, pp. 1–9.

[15] Z. Xia, H. Huang, H. Chen, X. Feng, G. Zhao, Hybrid-supervised hypergraph-enhanced transformer for micro-gesture based emotion recognition, IEEE Transactions on Affective Computing (2025).
[16] J. Gu, F. Wang, K. Li, Y. Wei, Z. Wu, D. Guo, Mm-gesture: towards precise micro-gesture recognition through multimodal fusion, MiGA@ IJCAI, 2025 (2025).
[17] H. Duan, Y. Zhao, K. Chen, D. Lin, B. Dai, Revisiting skeleton-based action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 2969–2978.
[18] X. Hu, C. Pu, Y. Li, K. Xie, M. Qiguang, Enhancing micro-gesture classification via global-aware importance estimation in vision transformer, MiGA@ IJCAI, 2025 (2025).
[19] H. Xu, L. Cheng, Y. Wang, S. Tang, Z. Zhong, Towards fine-grained emotion understanding via skeleton-based micro-gesture recognition, MiGA@ IJCAI, 2025 (2025).
[20] P. Liu, K. Li, F. Wang, Y. Wei, J. She, D. Guo, Online micro-gesture recognition using data augmentation and spatial-temporal attention, MiGA@ IJCAI, 2025 (2025).
[21] C. Meng, F. Ma, C. Zhang, J. Miao, Y. Yang, Y. Zhuang, Online micro-gesture recognition in long videos via spatiotemporal feature encoding and query-based temporal detection, MiGA@ IJCAI, 2025 (2025).
[22] H. Chen, X. Liu, J. Shi, G. Zhao, Temporal hierarchical dictionary guided decoding for online gesture segmentation and recognition, IEEE Transactions on Image Processing 29 (2020) 9689–9702.
[23] H. Chen, X. Liu, G. Zhao, Temporal hierarchical dictionary with hmm for fast gesture recognition, in: 2018 24th international conference on pattern recognition (ICPR), IEEE, 2018, pp. 3378–3383.
[24] T. backpacker, Weak-to-strong: Vlm-based pseudo-label annotation for an alternating self-supervised, weakly-supervised, and semi-supervised training strategy, MiGA@ IJCAI, 2025 (2025).
[25] A. Martirosyan, S. Tigranyan, M. Razzhivina, A. Aslanyan, N. Salikhova, I. Makarov, A. Savchenko, A. Avetisyan, Multi-track multimodal learning on imigue: Micro-gesture and emotion recognition, MiGA@ IJCAI, 2025 (2025).
[26] H. Bu, et al., Behavior-based tennis match outcome prediction via fusion of class balancing strategy and emotional priors, MiGA@ IJCAI, 2025 (2025).
[27] Z. Xingyu, M. Keisuke, T. Kento, Unreal engine-based data augmentation to improve real-world human activity recognition with wearable devices, MiGA@ IJCAI, 2025 (2025).
[28] W. Zhiyu, L. Yang, G. Hatice, Graphau-pain: Graph-based action unit representation for pain intensity estimation, MiGA@ IJCAI, 2025 (2025).
[29] P. Santosh, S. Trisanth, A. Amith, Clip-mg: Guiding semantic attention with skeletal pose features and rgb data for micro-gesture recognition on the imigue dataset, MiGA@ IJCAI, 2025 (2025).
[30] Y. Yang, H. Liu, F. Kang, M. Zhang, Z. Lian, H. Tang, H. Chen, Saynext-bench: Why do llms struggle with next-utterance prediction?, arXiv preprint arXiv:2602.00327 (2026).