EquiME: Equitable Micro-Expression Dataset for Cross-Demographic Emotion Recognition

Pei-Sze Tan
tan.peisze@monash.edu
Monash University
Malaysia

Sailaja Rajanala
sailaja.rajanala@monash.edu
Monash University
Malaysia

Yee-Fan Tan
tan.yeefan@monash.edu
Monash University
Malaysia

Raphael C.W Phan
raphael.phan@monash.edu
Monash University
Malaysia

Huey-Fang Ong
ong.hueyfang@monash.edu
Monash University
Malaysia

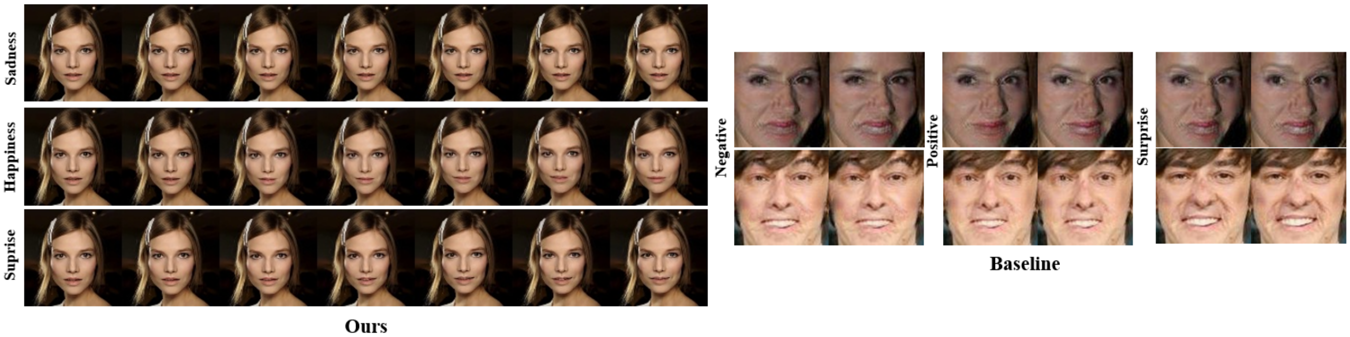


Figure 1: Teaser of EquiME Datasets VS MiE-X (Baseline).

Abstract

Micro-expression (ME) recognition is a challenging task due to the subtle and transient nature of these facial movements. Existing real-world ME datasets are limited in scale, diversity, and emotional breadth, hindering the development of robust recognition models. In this work, we introduce **EquiME**, a large-scale synthetic dataset for micro-expression analysis, generated using the image-to-video model. By leveraging a structured causal modeling approach, we employ Facial Action Units (AUs) as intermediate representations that drive the generation of realistic ME sequences. Our dataset achieves significant demographic diversity (55.6% women, 44.4% men; representation across six racial groups) while simultaneously capturing five distinct emotion categories—a balance rarely achieved in real-world data collection where recruiting diverse participants capable of expressing all target emotions remains challenging. Experimental results show that training on EquiME, followed by cross-validation evaluation, it shows consistency of performance across different model architectures. This paper presents a streamlined pipeline for generating synthetic micro-expression datasets, designed to be accessible to users without a computer science background. Project page: https://kirito-blade.github.io/me-vlm/

CCS Concepts

• **Computing methodologies** → **Machine learning approaches**; **Computer vision**; • **Human-centered computing** → *Human computer interaction (HCI)*; • **Applied computing** → *Psychology*.

Keywords

Micro-Expression, Text-to-video model, Action Units, Dataset Generation

ACM Reference Format:
Pei-Sze Tan, Sailaja Rajanala, Yee-Fan Tan, Raphael C.W Phan, and Huey-Fang Ong. 2025. EquiME: Equitable Micro-Expression Dataset for Cross-Demographic Emotion Recognition. In *Proceedings of ACM Multimedia 2025 (ACMMM '25).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Micro-expression (ME) has emerged as a critical area of research with applications spanning from security and deception detection to healthcare and human-computer interaction [16]. Despite its importance, progress in this field has been significantly hindered by the severe limitations of existing datasets, which suffer from extremely small sample sizes (typically <300 samples), limited demographic diversity, and restricted emotional breadth [27]. These constraints stand in stark contrast to the millions of samples commonly available for general face recognition tasks, creating

ACMMM '25, Dublin, Ireland

ACM ISBN 978-1-4503-XXXX-X/25/10
https://doi.org/XXXXXXX.XXXXXXX

Characteristics	**Real Human ME Datasets**	**MiE-X Dataset**	**EquiME (Ours)**
Data Format	Video Sequences (Spontaneous Micro-Expressions)	Static Images (Apex & Onset frames)	Video Sequences (Temporal Data)
Dataset Scale			
Subjects	~40	5,000	15,000
Samples	~500	45,000	75,000 (15,000 × 5 class)
Emotion Classes	3-7 Classes • Happiness • Sadness • Surprise • Disgust • Anger • Fear • Contempt (some datasets)	3 Classes • Positive • Negative • Surprise	5 Classes • Happiness • Sadness • Surprise • Disgust • Anger
Technical Specifications			
Resolution	640 × 480 pixels (varies)	128 × 128 pixels	256 × 256 pixels
Customization	No	No	Yes
Multimodal Training Support	No	No	Yes
Attributes Label (Gender, Race, Age)	No	No	Yes
Note: EquiME offers comprehensive temporal and emotional diversity compared to MiE-X and real-human datasets			

Table 1: Comparison between real-human datasets (e.g CAS(ME)^2, SAMM and etc), MiE-X (synthetic), and EquiME (ours)

a substantial barrier to developing robust and generalizable MER systems.

The lack of **large-scale, diverse micro-expression datasets** presents a fundamental challenge for developing effective recognition models. Current benchmark datasets like CASME^3, SAMM, and SMIC not only contain limited samples but also exhibit significant demographic homogeneity, with most participants belonging to similar age groups, ethnicities, and cultural backgrounds [3, 14, 15]. Even in the current literature, there are methods for ME image-driven generation for synthetic datasets [30], as well as approaches that use wild images to synthesize large datasets [17]. However, current methods face limitations such as over-reliance on existing micro-expression datasets and the generation of low-resolution or visually unconvincing images. This homogeneity introduces inherent biases into trained models, limiting their ability to generalise across diverse populations. Furthermore, the dominant validation methodology in MER research—Leave-One-Subject-Out (LOSO) cross-validation—while providing some insights into cross-subject generalization, fails to adequately address whether models learn meaningful micro-expression features rather than subject-specific characteristics across demographically similar individuals [2, 9].

Advantages of Synthetic Data is that EquiME addresses several ethical concerns commonly associated with traditional micro-expression datasets. First, it ensures privacy protection by avoiding the need to record real individuals in emotionally vulnerable states. Second, it sidesteps consent issues related to inducing or capturing genuine emotional reactions. Third, it allows for more deliberate control over demographic representation, enabling more balanced datasets across gender, age, and ethnicity groups—something that is difficult to achieve in real-world data collection.

Demographic diversity in facial expression datasets is not merely a matter of representation but fundamentally impacts the fairness, accuracy, and applicability of resulting technologies. Research has demonstrated that facial recognition and emotion analysis systems trained on homogeneous datasets perform inconsistently across different demographic groups, perpetuating biases and potentially leading to harmful applications [4, 28]. For micro-expressions—subtle, involuntary facial movements that occur within fractions of a second, these biases can be even more pronounced due to potential cultural and phenotypic variations in expression. The creation of demographically balanced datasets is therefore essential not only for advancing technical performance but also for ensuring that MER technologies serve diverse populations equitably.

Our work directly addresses these critical gaps by introducing a large-scale, demographically diverse synthetic dataset for micro-expression analysis. By carefully balancing gender representation (55.6% women, 44.4% men), including participants from six racial groups, and spanning multiple age ranges while simultaneously capturing five distinct emotion categories, we provide a resource that overcomes the limitations of existing datasets. This balanced approach enables the development of more robust models that can generalize across demographic boundaries, representing a significant step toward more equitable affective computing systems.

Our main contributions in this paper are threefold:

(1) We introduce **EquiME**, a novel large-scale synthetic dataset for micro-expression analysis that achieves unprecedented demographic diversity while maintaining comprehensive emotional coverage;
(2) We develop a structured causal modeling approach using Facial Action Units (AUs) as intermediate representations to generate realistic micro-expression sequences across diverse identities;

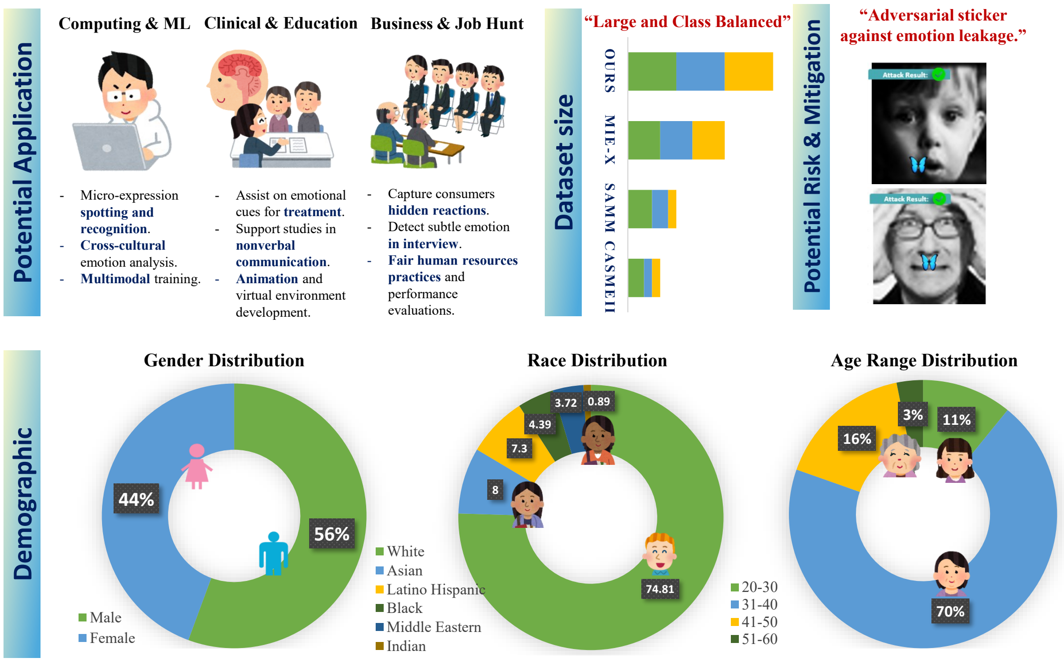


Figure 2: EquiME supports research in micro-expression recognition [11, 19, 23, 26], temporal spotting [22, 24, 31–33], and cross-cultural analysis [5, 23]. It aids mental health and HCI systems [1], offers educational value [7], and benefits industry in sentiment analysis, security, and digital avatar creation [17, 20]. Demographic distribution of our emotion recognition dataset showing (a) gender distribution (Woman: 55.6%, Man: 44.4%), (b) race distribution (White: 74.8%, Asian: 8.9%, Latino/Hispanic: 7.3%, Black: 4.4%, Middle Eastern: 3.7%, Indian: 0.9%), and (c) age distribution (20-30: 55.4%, 31-40: 34.7%, 41-50: 8.2%, 51-60: 1.6%, Under 20: 0.05%, Over 60: 0.06%). Our dataset achieves greater demographic diversity than many existing emotion recognition benchmarks while capturing five distinct emotion categories simultaneously.

(3) We demonstrate significant improvements in cross-dataset performance by leveraging our diverse dataset, achieving superior generalization compared to models trained on existing limited datasets.

As illustrated in Figure 2, the proposed dataset and methodology advance not only the technical state of the art in micro-expression recognition but also address critical ethical concerns regarding fairness and inclusivity in affective computing technologies, providing a foundation for developing more equitable systems that perform consistently across diverse populations.

1.1 Comparison with Existing Synthetic Dataset

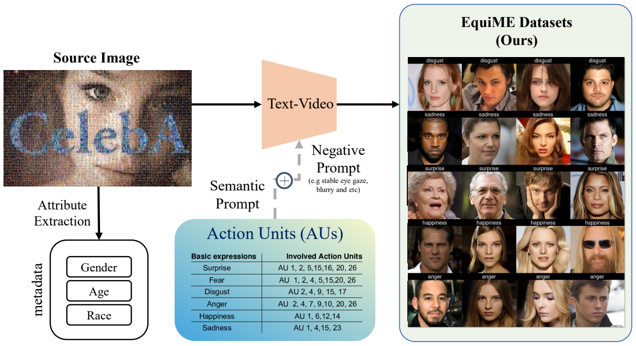
Table 1 presents a comparative overview of EquiME against both real-human micro-expression datasets and the synthetic MiE-X dataset. Unlike real-world participants dataset, which consist of spontaneous video sequences collected from a limited number of subjects, EquiME provides large-scale, controllable, and richly annotated synthetic video sequences that simulate realistic micro-expression dynamics. While MiE-X provides a larger volume of data compared to traditional datasets, it is limited to static images, specifically onset and apex frames, and lacks the temporal continuity needed to spot and analyze expression transitions. This limitation is illustrated in Figure 1 and compared with our dataset's instances.

EquiME addresses this gap by offering full video sequences that span the onset, apex, and offset phases of expression, thereby enabling more nuanced temporal analysis. Additionally, EquiME expands the range of emotion classes to five (Happiness, Sadness, Surprise, Disgust, and Anger), improving the dataset's granularity and realism compared to the three broad categories in existing datasets.

From a technical perspective, EquiME delivers higher-resolution imagery than MiE-X, and uniquely supports dataset customization. This includes the ability to vary facial attributes, environmental conditions, and expression intensity, which is not possible in either CAS(ME)^2, SAMM, or MiE-X. Furthermore, EquiME is the first to support multimodal training configurations, enabling future research to combine visual, semantic, and possibly physiological cues. These advances make EquiME a comprehensive and flexible resource for advancing the field of micro-expression research.

2 Micro-Expression Video Generation Protocol with Text-to-Video Model

To generate a large-scale synthetic micro-expression dataset, we utilize the CelebA-HQ dataset [13], which consists of high-quality static facial images sourced from diverse celebrity photographs. This dataset provides a wide range of facial structures, lighting conditions, and demographics, ensuring diversity in generated micro-expression sequences. No preprocessing is applied to the images before they are used for video synthesis.



Basic expressions	Involved Action Units
Surprise	AU 1, 2, 5,15,16, 20, 26
Fear	AU 1, 2, 4, 5,15,20, 26
Disgust	AU 2, 4, 9, 15, 17
Anger	AU 2, 4, 7, 9,10, 20, 26
Happiness	AU 1, 6,12,14
Sadness	AU 1, 4,15, 23

Figure 3: Pipeline for generating synthetic micro-expression datasets using a text-to-video model guided by source images.

2.1 Generation Pipeline

We introduce a structured pipeline for generating micro-expression videos using the LTX-Video model [10], an advanced image-to-video synthesis framework as shown in Figure 3. The model takes a single static image as input and generates a short video sequence that simulates micro-expressions based on predefined emotion prompts.

2.1.1 *Prompt Engineering for Emotion Control*

To guide the model in producing realistic micro-expressions, we construct detailed prompts based on the Facial Action Coding System (FACS)[6]. Each prompt specifies the combination of Action Units (AUs) that correspond to a particular emotion category. The considered emotion classes include happiness, sadness, surprise, fear, anger, disgust, and contempt. Each of these emotions is defined by a specific set of facial muscle movements, such as cheek raising for happiness, inner brow raising and lip corner depression for sadness, or jaw dropping for surprise. By carefully crafting these prompts, we ensure that the synthesized videos exhibit subtle and accurate micro-expressions.

2.1.2 *Negative Prompt Constraints*

To maintain high-quality and realistic micro-expression synthesis, a negative prompt is applied to the LTX-Video generation process. This prompt explicitly discourages unwanted artifacts, including motion blur, camera shake, exaggerated expressions, open mouths, teeth visibility, distorted facial features, unrealistic skin textures, etc. Additionally, visual distractions such as visible clothing, accessories, background details, multiple faces, non-frontal perspectives, and low resolution are minimized. By enforcing these constraints, the generated videos remain focused on subtle micro-expressions without introducing unnatural distortions.

2.2 Structural Causal Model for Micro-Expression Generation

Our earlier research revealed that biases in micro-expression recognition models, particularly those related to AUs, gender, and dataset composition, can significantly influence both model accuracy and fairness [23]. Among these factors, AU-related bias proved to be especially impactful, highlighting the importance of how facial muscle activations shape emotion perception.

This insight prompted a shift in our approach. Rather than treating AUs as passive features, we began to view them as active causal variables that drive emotional expressions. If AUs can directly influence how emotions are expressed and perceived, then controlling them provides a principled way to guide the generation of micro-expressions. Based on this reasoning, we introduce a Structural Causal Model (SCM) that formalizes the relationships between AUs, emotions, and the video synthesis process.

2.2.1 *SCM Definition*

The Structural Causal Model consists of three primary variables: Z, Y, and X. The variable Z represents the Action Units (AU coefficients) as a vector that encodes facial muscle movements. The variable Y denotes the emotion category, which is treated as a discrete variable corresponding to the perceived emotional state. Finally, X refers to the source video, which is the generated micro-expression video sequence.

The causal relationships among these variables defined as:

$$Y = f_Z(Z) + \epsilon_Y \tag{1}$$

In this expression, $f_Z(\cdot)$ is a deterministic function that maps the configuration of AUs to an emotion label. The term ϵ_Y represents independent noise that captures variability in emotion perception not explained by the AUs.

The generation of the synthetic micro-expression video is modeled as:

$$X = G(Z, Y) + \epsilon_X \tag{2}$$

Here, $G(\cdot)$ is a generative function that synthesizes the video based on the given AU vector and emotion label. The term ϵ_X accounts for stochastic variations that may arise during the video synthesis process.

By modeling AUs as causal factors driving emotion, this formulation allows precise control over the generation of micro-expressions. As a result, the synthesized videos remain consistent with both the intended emotional state and the corresponding facial muscle activations.

2.3 Dataset Generation and Structure

Static facial images are processed through the LTX-Video model using emotion-specific prompts and negative constraints to generate 161-frame videos at 256 × 256 resolution, capturing micro-expression temporal dynamics through multiple inference steps. The resulting dataset contains synthetic micro-expression videos with emotion labels corresponding to generation prompts and demographic metadata (age, gender, ethnicity) enabling balanced subset creation through stratified sampling. Videos are organized by emotion type and demographic attributes, provided in MP4 format for micro-expression recognition research.

3 Introducing the EquiME Dataset

We officially present **EquiME**, a large-scale synthetic micro-expression dataset generated using the LTX-Video model [10]. EquiME addresses the scarcity of high-quality, diverse, and well-annotated micro-expression datasets, providing a robust benchmark for facial expression recognition research.

Our dataset provides a valuable contribution to emotion recognition research by addressing a significant limitation in existing datasets: the difficulty of simultaneously capturing demographic diversity while representing multiple emotion categories. While our data exhibits reasonable gender balance (55.6% women, 44.4% men), we acknowledge limitations in racial representation (74.8% white) and age distribution (90.1% between 20-40 years). These imbalances reflect the persistent challenges in real-world data collection where recruiting participants across diverse demographic backgrounds who can authentically express all five emotion categories remains difficult. Previous datasets typically sacrifice either demographic diversity or emotional breadth, focusing on limited emotion categories or homogeneous participant groups. Our contribution bridges this gap by providing a dataset that maintains representation across genders while capturing a more inclusive racial distribution than many benchmark datasets, all while consistently representing five distinct emotion categories. This balanced approach enables more robust emotion recognition models that can generalize across demographic groups, an essential step toward creating fair and unbiased affective computing systems.

License The EquiME dataset is released under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). This allows free use for academic and research purposes while requiring attribution and prohibiting commercial use without explicit permission.

Experiment Configurations All experiments reported in this paper can be replicated using the provided configuration files in our GitHub repository. Each repository is carefully documented and modularised to facilitate ease of use, reproducibility, and adaptation for future research.

4 Baseline Micro-Expression Recognition Datasets and Models

To assess the effectiveness of our synthetic dataset, we tested on 2 real-world datasets and trained with both real and synthetic datasets for cross-database validation.

The datasets used in this experiment include SAMM [3], CASMEII [29], and MiE-X [17]. **SAMM** is a high-resolution, temporally precise micro-expression dataset designed for objective facial muscle movement analysis, with detailed FACS annotations and diverse subject demographics. **CASMEII** is a well-established spontaneous micro-expression dataset collected under controlled conditions, offering high frame-rate video (200 fps) and extensive emotion labeling, making it widely used for temporal dynamics studies. **MiE-X** is a recent synthetic micro-expression dataset generated with generative models, incorporating diverse facial attributes and expression intensities to support scalable model training and bias analysis.

We implement four baseline micro-expression recognition models with varying architectural complexity. **ST-CNN [21]** is a specialised 3D CNN tailored for micro-expression analysis with all frames included for model training. While **ResNet3D** is a widely used classifier, a deep 3D CNN with residual blocks to mitigate vanishing gradients and enhance temporal feature learning. Also, **MobileNet-based hybrid** is a lightweight model analysing only the middle frame of each sequence using a pre-trained MobileNetV2. **Simple3DCNN** is a compact 3D CNN using 16 filters and a 64-unit dense layer, designed for rapid training and deployment in low-resource environments. All models are trained using categorical cross-entropy loss and the Adam optimizer (learning rate 10^{-4}), with early stopping. Performance is evaluated through cross-dataset validation to test generalization under diverse conditions.

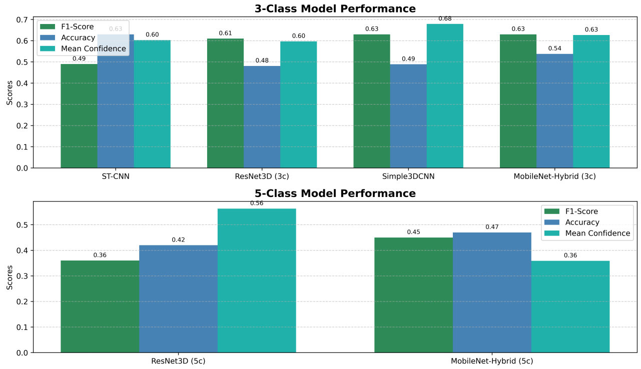


Figure 4: Evaluation results of 3 and 5 class classification by training our dataset, EquiME, with four baselines on the SAMM dataset. Comparison of models with different class setups in terms of weighted F1-score, accuracy, and mean confidence.

4.1 Quantitative Results on Different Datasets

Our experimental evaluation highlights several important findings regarding the performance and reliability of micro-expression recognition. In Table 4, our proposed dataset performs competitively in both 3-class and 5-class classification tasks, with the MobileNet-Hybrid architecture showing particular strength in fine-grained emotion recognition. Although performance declines slightly in the more challenging 5-class setup, our method maintains a balanced performance across all metrics.

Cross-dataset evaluations in Table 5 reveal strong generalisation capabilities. Unlike baseline models, which often achieve high scores under optimal conditions but degrade significantly under domain shifts, our method delivers stable performance across different source–target dataset combinations. This robustness is especially evident in evaluations across the SAMM and CASMEII datasets.

A key strength of our approach lies in its consistency. While some baselines may reach higher peak scores, they often exhibit large performance fluctuations that undermine reliability. In contrast, our method consistently delivers strong results across multiple architectures and evaluation settings.

4.2 Quality Evaluation on Generated Video

A comprehensive comparison of video quality metrics is presented across both real (SAMM, SMIC) and synthetic (MiE-X, Ours) micro-expression datasets. The metrics evaluated include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Total Variation (TV), Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), and CLIP-IQA [12]. These metrics jointly assess various aspects of visual fidelity, structural similarity, perceptual quality, and semantic alignment with natural images.

Table 2 reports the performance of our synthetic dataset in comparison to existing real and synthetic datasets using a suite of video quality metrics. Our method demonstrates strong overall performance, particularly in terms of perceptual and semantic quality. With a PSNR of 41.03 and an SSIM of 0.9684, our dataset achieves

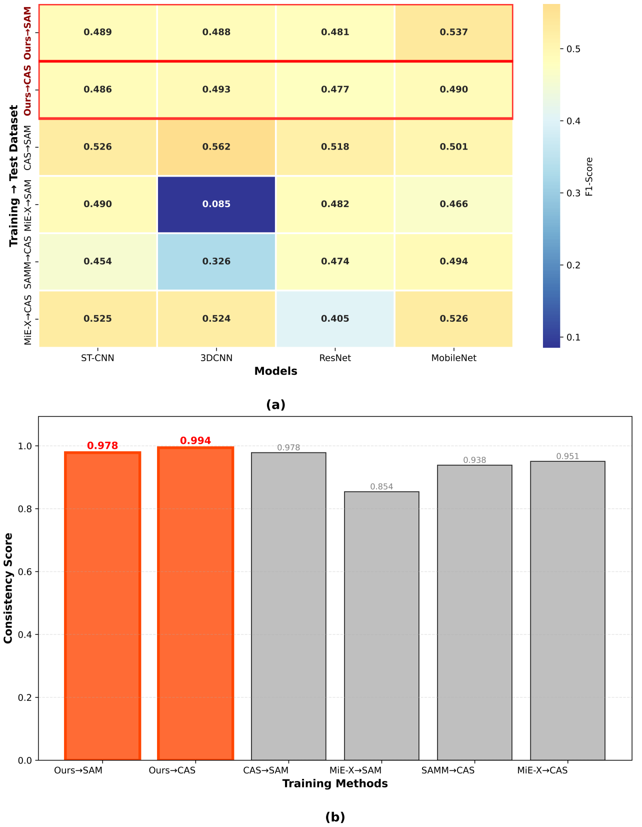


Figure 5: Comparison with other training datasets. (a) F1-score performance matrix and (b) consistency analysis across training methods. Our methods (starred/orange) show reliable performance across different model architectures.

Dataset	PSNR	SSIM	TV	BRISQUE	CLIP-IQA
Real Datasets					
SAMM	39.15	0.9167	26.40	7.96	0.48
SMIC	39.94	0.9460	15.22	34.47	0.25
Synthetic Datasets					
MiE-X	42.49	0.9787	6.96	27.28	0.61
Ours	**41.03**	**0.9684**	**13.80**	**14.66**	**0.79**

Table 2: Comparison of video quality metrics between real and synthetic datasets. We evaluate PSNR, SSIM, Total Variation (TV), BRISQUE, and CLIP-IQA to assess visual fidelity. Our results are based on a 6-second video synthesized via a text-to-video generation model. For PSNR, SSIM, TV, and BRISQUE, lower deviation from the real dataset values indicates higher realism and quality. CLIP-IQA, however, is an exception; since our video is directly generated from text prompts, it is expected to score higher on CLIP-IQA due to better semantic alignment with the input text.

high fidelity and structural similarity, closely approximating the quality of natural images while maintaining temporal coherence. Although MiE-X records slightly higher values in these two metrics, it does so at the cost of perceptual quality, as indicated by its higher BRISQUE score of 27.28. In contrast, our dataset achieves a significantly lower BRISQUE score of 14.66, suggesting that the generated frames are more perceptually natural and less prone to artificial distortions. Our method achieves a CLIP-IQA score of 0.79, the highest among all evaluated datasets. This indicates strong alignment with real-world imagery in both visual and semantic dimensions, consistent with our approach of using text prompts to generate videos. While the TV of 13.80 is marginally higher than that of MiE-X, it remains substantially lower than in real datasets, preserving temporal smoothness without sacrificing detail. Collectively, these results confirm that our generation approach strikes a more effective balance between pixel-level quality, structural integrity, perceptual realism, and semantic consistency, making it a superior choice for synthetic micro-expression data generation.

5 Ethical Considerations and Privacy

The development and deployment of micro-expression datasets and recognition systems raise important ethical questions, particularly in relation to privacy, consent, and the potential for misuse. Given the sensitive nature of micro-expressions, which often reveal involuntary emotional cues, careful consideration must be given to how such data is collected, modelled, and applied.

Therefore, we acknowledge several ethical risks associated with micro-expression recognition technology. These include possible **surveillance applications**, where systems could be used to monitor emotional states without consent; **emotional privacy violations**, where individuals' subtle expressions might be analyzed against their will; and **misinterpretation risks**, as emotion recognition systems may not account for cultural or individual variability. To mitigate these concerns, we enforce restrictions on commercial use through licensing agreements, provide clear ethical guidelines for appropriate applications, promote transparency in the deployment of systems built from EquiME, and support broader public discourse on emotional privacy norms.

Adversarial Attack and Privacy Defence : To further protect against unintended emotion leakage, especially in real-world applications, we consider the threat of adversarial attacks. Industry applications can apply adversarial defence techniques such as Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) [8, 18] or methods developed in our prior work [25] as shown in Figure 2. In addition, physical countermeasures such as adversarial stickers can be deployed to obscure or neutralise micro-expression cues in practical scenarios, providing users with additional privacy protection.

6 Conclusion

EquiME offers a novel and comprehensive solution to the challenges of micro-expression research by providing a large-scale, demographically balanced synthetic dataset with precise emotional and temporal annotations. Its use of Facial Action Units and causal modeling enables the generation of realistic micro-expression sequences that enhance recognition model training. By addressing limitations of existing datasets in diversity and scale, EquiME paves the way for more inclusive and accurate micro-expression analysis, with an accessible pipeline that supports broad adoption in both research and applied settings.

References

[1] Rochelle Ackerley, Jean-Marc Aimonetti, and Edith Ribot-Ciscar. Emotions alter muscle proprioceptive coding of movements in humans. *Scientific reports*, 7(1): 8465, 2017.

[2] George I Austin, Itsik Pe'er, and Tal Korem. Distributional bias compromises leave-one-out cross-validation. *ArXiv*, pages arXiv–2406, 2025.

[3] Adrian K Davison, Cliff Lansley, Nicholas Costen, Kevin Tan, and Moi Hoon Yap. Samm: A spontaneous micro-facial movement dataset. *IEEE transactions on affective computing*, 9(1):116–129, 2016.

[4] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. Metrics for dataset demographic bias: A case study on facial expression recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5209–5226, 2024.

[5] Paul Ekman. Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review*, 169222(1):45–60, 1973.

[6] Paul Ekman and Erika L. Rosenberg. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, 2005.

[7] Jennifer Endres and Anita Laidlaw. Micro-expression recognition training in medical students: a pilot study. *BMC medical education*, 9(1):1–6, 2009.

[8] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *stat*, 1050:20, 2015.

[9] Quentin F Gronau and Eric-Jan Wagenmakers. Limitations of bayesian leave-one-out cross-validation for model selection. *Computational brain & behavior*, 2 (1):1–11, 2019.

[10] Yoav HaCohen, Nisan Chiprut, Benny Brazowski, Daniel Shalem, Dudu Moshe, Eitan Richardson, Eran Levin, Guy Shiran, Nir Zabari, Ori Gordon, Poriya Panet, Sapir Weissbuch, Victor Kulikov, Yaki Bitterman, Zeev Melumian, and Ofir Bibi. Ltx-video: Realtime video latent diffusion. *arXiv preprint arXiv:2501.00103*, 2024.

[11] Xingxun Jiang, Yuan Zong, Wenming Zheng, Jiateng Liu, and Mengting Wei. Seeking salient facial regions for cross-database micro-expression recognition. arXiv, 2021.

[12] Sergey Kastryulin, Jamil Zakirov, Denis Prokopenko, and Dmitry V. Dylov. Pytorch image quality: Metrics for image quality assessment, 2022.

[13] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[14] Jingting Li, Zizhao Dong, Shaoyuan Lu, Su-Jing Wang, Wen-Jing Yan, Yinhuan Ma, Ye Liu, Changbing Huang, and Xiaolan Fu. Cas (me) 3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

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

[16] Yante Li, Jinsheng Wei, Yang Liu, Janne Kauttonen, and Guoying Zhao. Deep learning for micro-expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(4):2028–2046, 2022.

[17] Yuchi Liu, Zhongdao Wang, Tom Gedeon, and Liang Zheng. How to synthesize a large-scale and trainable micro-expression dataset? In *ECCV*, 2022.

[18] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.

[19] Walied Merghani and Moi Hoon Yap. Adaptive mask for region-based facial micro-expression recognition. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 765–770, 2020.

[20] Albert Pumarola, Antonio Agudo, Aleix M. Martinez, Alberto Sanfeliu, and Francesc Moreno-Noguer. Ganimation: Anatomically-aware facial animation from a single image. In *Computer Vision – ECCV 2018*, 2018.

[21] Sai Prasanna Teja Reddy, Surya Teja Karri, Shiv Ram Dubey, and Snehasis Mukherjee. Spontaneous facial micro-expression recognition using 3d spatiotemporal convolutional neural networks, 2019.

[22] Pei-Sze Tan, Sailaja Rajanala, Arghya Pal, Raphaël C.-W. Phan, and Huey-Fang Ong. Unbiased decision-making framework in long-video macro & micro-expression spotting. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 84–89, 2023.

[23] Pei-Sze Tan, Sailaja Rajanala, Arghya Pal, Shu-Min Leong, Raphaël C-W Phan, and Huey Fang Ong. Causally uncovering bias in video micro-expression recognition. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5790–5794. IEEE, 2024.

[24] Pei-Sze Tan, Sailaja Rajanala, Arghya Pal, Raphaël C-W Phan, and Huey-Fang Ong. Causal-ex: Causal graph-based micro and macro expression spotting. *arXiv preprint arXiv:2503.09098*, 2025.

[25] Pei-Sze Tan, Sailaja Rajanala, Yee-Fan Tan, Arghya Pal, Chun-Ling Tan, Raphaël C.-W. Phan, and Huey-Fang Ong. Post-hoc adversarial stickers against micro-expression leakage. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[26] Monu Verma, Priyanka Lubal, Santosh Kumar Vipparthi, and Mohamed Abdel-Mottaleb. Rnas-mer: A refined neural architecture search with hybrid spatiotemporal operations for micro-expression recognition. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4759–4768, 2023.

[27] Hong-Xia Xie, Ling Lo, Hong-Han Shuai, and Wen-Huang Cheng. An overview of facial micro-expression analysis: Data, methodology and challenge. *IEEE Transactions on Affective Computing*, 14(3):1857–1875, 2022.

[28] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. Investigating bias and fairness in facial expression recognition. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, pages 506–523. Springer, 2020.

[29] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. 9:1–8, 2014.

[30] Chuin Hong Yap, Ryan Cunningham, Adrian K Davison, and Moi Hoon Yap. Synthesising facial macro-and micro-expressions using reference guided style transfer. *Journal of Imaging*, 7(8):142, 2021.

[31] Chuin Hong Yap, Moi Hoon Yap, Adrian Davison, Connah Kendrick, Jingting Li, Su-Jing Wang, and Ryan Cunningham. 3d-cnn for facial micro- and macro-expression spotting on long video sequences using temporal oriented reference frame. New York, NY, USA, 2022. Association for Computing Machinery.

[32] Shukang Yin, Shiwei Wu, Tong Xu, Shifeng Liu, Sirui Zhao, and Enhong Chen. Au-aware graph convolutional network for macro-and micro-expression spotting. *arXiv preprint arXiv:2303.09114*, 2023.

[33] Yuan Zhao, Xin Tong, Zichong Zhu, Jianda Sheng, Lei Dai, Lingling Xu, Xuehai Xia, Yu Jiang, and Jiao Li. Rethinking optical flow methods for micro-expression spotting. New York, NY, USA, 2022. Association for Computing Machinery.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009