Do Emotion Recognition Models Generalize to Classrooms?

Robustness & Fairness Analysis in Real Learning Environments

Ashwin Tudur Sadashiva · Srigowri Mayasandra Prasanna · Joyce Horn Fonteles · Gautam Biswas

Department of Computer Science, Vanderbilt University, Nashville, TN, USA

CVPR JUNE 3-7, 2026 DENVER COLORADO

Overview

Emotion recognition powers affective computing and learning analytics — yet most models are trained on web/lab data that differs sharply from real classrooms.

We test three SOTA models on real classroom video for **robustness** and **fairness** under everyday conditions.

69 students · 3 environments · 3 models · 4 perturbations

Research Questions

- **RQ1** How sensitive are models to visual perturbations common in classrooms?
- **RQ2** Do prediction errors vary across Fitzpatrick skin tone groups?
- **RQ3** Does the learning environment influence prediction bias?
- **RQ4** Can lightweight mitigation reduce error and fairness gaps?

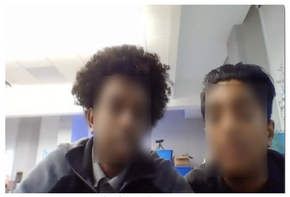
Dataset & Learning Environments

69 students (ages 11–18; 51% F / 49% M) across three IRB-approved classroom studies; no student appears in more than one dataset.

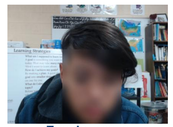
≈ 55,000 frames analyzed · 7,430 human-validated perturbation pairs (κ = 0.81)

Environment	Stud.	Group	Camera	Session
EcoJourneys (game-based)	25	3–4	Laptop cam	45m×2
GEM-STEP (embodied)	20	4–6	4 cameras	25m×6
C2STEM (collaborative)	24	2	Laptop cam	90m×4

GEM-STEP — Embodied



C2STEM



Eco Journeys

Methodology

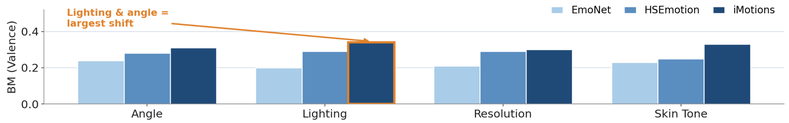
- **Models:** EmoNet (EmoFAN), HSEmotion (EfficientNet-B0), iMotions v10.1 — continuous valence–arousal.
- **Pipeline:** MTCNN face detection → crop → sparse sampling (1 frame/min), ≈ 55,000 frames.
- Paired perturbation — one factor at a time: **BM = $|\hat{y}_{ref} - \hat{y}_{pert}|$**
- **Metrics:** MAE (valence/arousal), ECE, fairness gap = MAE(VI) − MAE(I).
- **Validated pairs** (κ = 0.81): 1800 angle · 1680 lighting · 1550 resolution · 2400 skin tone.
- **Mitigation:** data augmentation + partial fine-tuning (frozen backbone, Adam, 10 epochs).

Baseline Performance

EmoNet transfers best; commercial iMotions has the highest error.

Model	MAE (Valence)	MAE (Arousal)
EmoNet	0.21	0.23
HSEmotion	0.24	0.26
iMotions (v10.1)	0.28	0.30

RQ1 — Robustness to Perturbations



All models shift under classroom conditions — iMotions is the most sensitive.

Effect Sizes (Cohen's d, avg. across models)

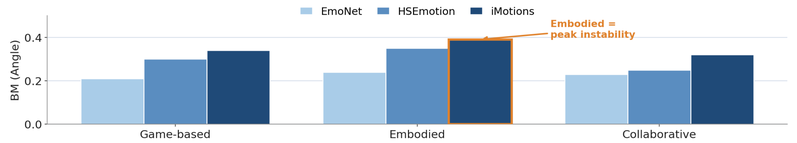
Perturbation	Avg Cohen's d	Effect	Significance
Camera angle	1.04	Large	$p < 0.001$
Lighting	1.07	Large	$p < 0.001$
Resolution	0.76	Medium	$p < 0.001$
Skin tone	0.92	Large	$p < 0.001$

RQ2 — Fairness Across Skin Tones (full results)

Group-wise MAE (valence/arousal) and ECE by Fitzpatrick Type — error rises I → VI for all models.

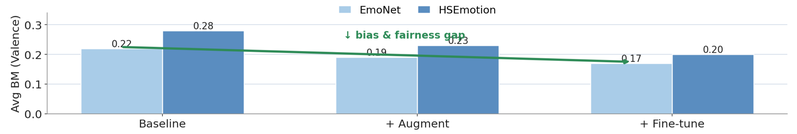
FT	EmoNet			HSEmotion			iMotions		
	Va	Ar	ECE	Va	Ar	ECE	Va	Ar	ECE
I	0.179	0.198	0.051	0.211	0.231	0.047	0.240	0.260	0.082
II	0.192	0.210	0.058	0.219	0.239	0.059	0.255	0.272	0.094
III	0.198	0.221	0.071	0.229	0.253	0.068	0.268	0.289	0.108
IV	0.211	0.229	0.080	0.242	0.257	0.079	0.281	0.300	0.120
V	0.220	0.240	0.092	0.250	0.271	0.091	0.295	0.318	0.137
VI	0.230	0.251	0.100	0.262	0.279	0.099	0.312	0.337	0.155
Gap	**0.051**	—	—	**0.051**	—	—	0.072	—	—

RQ3 — Learning Environment Effects



Embodied settings (movement + multi-camera) amplify instability vs. game-based webcams.

RQ4 — Mitigation Effects



Augmentation + fine-tuning cut bias magnitude and halve the fairness gap (0.06 → 0.03).

Key Takeaways & Conclusions

- ✓ Benchmark-validated emotion models do **not reliably generalize** to real classrooms.
- ✓ Camera angle, lighting, resolution & skin tone cause **systematic** valence–arousal shifts.
- ✓ Errors grow from Fitzpatrick I→VI; commercial iMotions is least robust and least fair.
- ✓ Embodied learning environments add extra sources of variation and instability.
- ✓ Lightweight augmentation + fine-tuning reduce error and fairness gaps — **but gaps remain.**
- ✓ Motivates **classroom-aware evaluation**, domain adaptation, and large K–12 datasets.

Highlights

69 learners · 3 classrooms

0.072 worst skin-tone gap

50% gap cut by fine-tuning

55K frames analyzed

Scan for the paper

Funding: U.S. Army CCDC Soldier Center #W912CG2220001 · NSF AI Institute #DRL-2112635

Contact: ashwin.tudur.sadashiva@vanderbilt.edu