International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A HYPERSPHERICAL–MARGIN FRAMEWORK FOR COMPOUND FACIAL EMOTIONS

Mangaras Yanu Florestiyanto[1,*], Herman Dwi Surjono[2], Handaru Jati[3], Bambang Yuwono[4]

[1*]Yogyakarta State University, 1 Colombo Street, Karangmalang, 55281, Yogyakarta, Indonesia;
UPN Veteran Yogyakarta, Padjajaran Street, 55283, Yogyakarta, Indonesia,
e-mail: mangaras.yanu@upnyk.ac.id
[2]Yogyakarta State University, 1 Colombo Street, Karangmalang, 55281, Yogyakarta, Indonesia,
e-mail: hermansurjono@uny.ac.id
[3]Yogyakarta State University, 1 Colombo Street, Karangmalang, 55281, Yogyakarta, Indonesia,
e-mail: handaru@uny.ac.id
[4]UPN Veteran Yogyakarta, Padjajaran Street, 55283, Yogyakarta, Indonesia,
e-mail: bambangy@upnyk.ac.id

[*] **Corresponding Author**: Mangaras Yanu Florestiyanto
*mangaras.yanu@upnyk.ac.id

Abstract

We propose a hyperspherical–margin framework for compound facial emotion recognition. Let $X \subset \mathbb{R}^{H\times W\times 3}$ denote images and $Y = \{1, \dots, 11\}$ compound classes. Standard softmax training optimizes likelihood but does not enforce angular separation on the representation manifold, yielding entangled decision regions and poorly calibrated scores. We therefore place FER on the unit hypersphere by L2-normalizing features and class weights, and train with an angular-margin head (CosFace/ArcFace) using a stability-minded three-stage curriculum: warm-up ($m = 0$), head-only margin with a gentle ramp, and partial fine-tuning of the upper backbone. To link geometry to operation, we evaluate (i) symmetric confusion mass for class pairs, (ii) macro one-vs-rest ROC/PR, and (iii) calibration via reliability diagrams and Expected Calibration Error (ECE). On RAF-DB compound-11, the proposed regimen improves macro metrics from a softmax warm-up to the final model as follows: macro-F1 $0.09 \rightarrow 0.88$, ROC-AUC $0.52 \rightarrow 0.98$, and PR-AUC $0.10 \rightarrow 0.92$. Probability calibration markedly improves (ECE $0.1924 \rightarrow 0.0195$, ≈−89.9%). Pairwise confusions collapse for historically adjacent categories; for example, Fearfully Angry ↔ Fearfully Surprised reduces from 96 to 2 total errors. The model remains lightweight (≈3.25M parameters; ≈0.43 GFLOPs at 224^2) and edge-capable (GTX 1650 Max-Q: ~7.3 ms median, ~137 FPS; CPU: ~11.1 ms, ~90 FPS; batch=1). These results support a geometry-first thesis for compound FER: enforcing hyperspherical angular margins, coupled with a gentle training schedule, yields separable, calibrated embeddings without increasing backbone capacity.

1. Introduction

Compound facial expressions (e.g., *Sadly Fearful*, *Angrily Surprised*) co-activate multiple affective components and frequently generate overlapping manifolds and annotation ambiguity in the wild (e.g., RAF-DB compound protocols) [1,2]. Likelihood-only training with softmax cross-entropy optimizes classification probability but does not explicitly enlarge inter-class angular gaps; decision boundaries can thread through entangled clusters and produce brittle, poorly calibrated scores [3,4]. Recent practice in metric learning shows that operating in angle—by normalizing features and classifier weights and adding an angular margin—compacts intra-class variance and widens inter-class separation, improving discrimination under fine-grained overlap [5,6].

We explore whether identity-oriented angular objectives transfer effectively to compound FER under two constraints: (i) small and imbalanced datasets and (ii) lightweight inference suitable for

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

deployment. We adopt a hyperspherical view, in which embeddings and class weights are L2-normalized and compared by angular similarity; the full formalism is presented in Section 2. To stabilize learning in small/imbalanced regimes, we use a three-stage curriculum—a softmax warm-up, a head-only angular phase with a gentle margin ramp, and a partial fine-tuning of upper backbone layers—specified as a staged optimization in Section 2.4. For edge suitability, we employ a mobile backbone (MobileNetV3) and a BN-projected embedding to align with cosine geometry [7,8].
Our evaluation links geometry to operation via diagnostics introduced here and defined in Section 2: (i) symmetric confusion mass for the most confusable class pairs to quantify off-diagonal error reduction before/after angular training; (ii) macro one-vs-rest ROC/PR to assess threshold-free separability; and (iii) probability calibration via reliability diagrams and Expected Calibration Error (ECE) [3,4].

Hypothesis. After angular training, inter-centroid angles between historically confusable classes increase, and within-class dispersion decreases; these geometric shifts correlate with fewer two-way confusions, higher macro-F1/ROC-AUC/PR-AUC, and lower ECE—achieved without enlarging backbone capacity. Our approach remains tool-minimal and deployment-oriented: a MobileNetV3 backbone with a BN-projected 256-d embedding and an angular-margin head, pursuing separability and calibration through representation geometry rather than sheer depth [5–8].

Contributions.
1. A hyperspherical–margin framework of compound FER with geometry-aware diagnostics (see Section 2.1–Section 2.3, Section 2.6).
2. A stability-minded curriculum (warm-up → gentle angular head → partial fine-tuning) tailored to small/imbalanced sets (see Section 2.4).
3. An empirical geometry-to-operation linkage showing that improved angular separation translates into fewer symmetric confusions and better calibration (see Results).

2. The Hyperspherical–Margin Framework

As motivated in Section 1, we place compound FER on the unit hypersphere and make each design choice explicit.

2.1 Framework geometry: hyperspherical embeddings

Let $X \subset \mathbb{R}^{H\times W\times 3}$ be the image space and $Y = \{1, \dots, C\}$ the compound classes ($C = 11$). A backbone $F_\theta: X \to \mathbb{R}^d$ with a batch-normalizing neck (BN) produces a unit embedding

$$z = \frac{\mathrm{BN}(F_\theta(x))}{\| \mathrm{BN}(F_\theta(x)) \|_2} \in \mathbb{S}^{d-1}, \widetilde{w}_c = \frac{w_c}{\| w_c \|_2} \in \mathbb{S}^{d-1}. \tag{1}$$

The cosine score for the class c is

$$s_{ic} = \langle z_i, \widetilde{w}_c \rangle = \cos\ \theta_{ic}. \tag{2}$$

With a scale $s > 0$, the decision rule

$$\hat{y}(x) = \arg\max_{c\in Y} \langle z, \widetilde{w}_c \rangle \tag{3}$$

is a nearest-direction classifier on $\mathbb{S}^{d-1}$: each class occupies a spherical cone bounded by great-circle bisectors [5,6,8].

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.2 Angular-margin heads (CosFace/ArcFace)
We enforce a minimum angular gap by modifying the true-class logit with a margin $m \geq 0$:

$$\text{CosFace: } \ell_{i,y_i}^{\cos} = s\left(\cos\theta_{i,y_i} - m\right), \text{ ArcFace: } \ell_{i,y_i}^{\text{arc}} = s\cos\left(\theta_{i,y_i} + m\right), \tag{4}$$

with non-target logits $s\cos\ \theta_{ic}$ for $c \neq y_i$. Cross-entropy over these logits shrinks the feasible cone (intra-class compaction) and increases angular separation (inter-class widening) [5,6].

2.3 Geometry-aware diagnostics
Let $M \in \mathbb{N}^{C\times C}$ be the confusion matrix.

- **Symmetric Confusion Mass (SCM).** For a pair (u,v),

$$\text{SCM}(u,v) = M_{uv} + M_{vu}, \Delta\text{SCM}(u,v) = \text{SCM}_{\text{after}} - \text{SCM}_{\text{before}}. \tag{5}$$

Negative ΔSCM indicates reduced two-way confusion.

- **Ranking metrics.** Macro OvR ROC-AUC/PR-AUC quantify threshold-free separability under class imbalance, as commonly reported for FER and compound protocols [1,2].
- **Calibration (ECE).** With B confidence bins,

$$\text{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} \mid \text{acc}_b - \text{conf}_b \mid, \tag{6}$$

and reliability diagrams visualize $(\text{conf}_b, \text{acc}_b)$ [3,4].

2.4 Staged optimization (curriculum)
To stabilize small/imbalanced regimes, we solve a staged program:

Stage-A (warm-up; $m = 0$).

$$\min_{\theta,W} \frac{1}{N}\sum_i \mathcal{L}_{\text{CE}}(s\cos\ \theta_{i,\cdot}, y_i). \tag{7}$$

Stage-M (head-only margin; ramp $m(t) \uparrow m_{\max}$). Freeze θ; optimize W(and BN) with a gentle margin ramp:

$$\min_{W,\text{BN}} \frac{1}{N}\sum_i \mathcal{L}_{\text{CE}}(\text{margin}(s, \theta_{i,\cdot}; m(t)), y_i)\,,\, m(0) = 0,\, m(T) = m_{\max} \ll 1. \tag{8}$$

Stage-M+FT (partial fine-tuning). Unfreeze the last L backbone layers and continue the margin objective, aligning high-level filters to angular constraints while keeping early layers fixed for efficiency [5–8].

2.5 Backbone & efficiency choices
We use MobileNetV3 as F_θ for on-device efficiency (hard-swish, inverted residuals, squeeze-and-excitation) [7] with a BNNeck projecting to a 256-d unit embedding for stable cosine geometry [8]. This achieves discriminability from geometry rather than depth and supports device-centric evaluation (parameters/GMACs, latency/FPS, memory), as reported in the Results.

2.6 Geometry–operation linkage (testable claims)
Let $\mu_c = \mathbb{E}[z \mid y = c]$, $\bar{\mu}_c = \mu_c / \parallel \mu_c \parallel_2$, the inter-centroid angle

$$\phi_{uv} = \arccos\langle \bar{\mu}_u, \bar{\mu}_v \rangle, \tag{9}$$

and the within-class dispersion

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\delta_c = \mathbb{E}[\,1 - \langle z, \bar{\mu}_c \rangle \mid y = c]. \tag{10}$$

Claims. Post-margin training (i) increases ϕ_{uv} for historically confusable pairs, (ii) decreases δ_c, and (iii) correlates these geometric shifts with negative $\Delta\mathrm{SCM}(u, v)$, higher macro-F1/ROC-AUC/PR-AUC, and lower ECE—without increasing backbone capacity [3–6].

2.7 Exact training specifics

Image size 224, batch 20, MobileNetV3 backbone, 256-d BNNeck.
Stage-A: $\mathrm{LR} = 10^{-3}$, epochs $= 8$, $m = 0$.
Stage-M (head-only): CosFace or ArcFace, scale $s = 16.0$, linear m: $0 \rightarrow 0.05$, $\mathrm{LR} = 10^{-4}$, epochs $= 8$.
Stage-M+FT: unfreeze last $L = 80$ layers; $\mathrm{LR(head)} = 10^{-4}$, $\mathrm{LR_{FT}} = 7 \cdot 10^{-5}$, epochs $= 10$ [5–8].

3. Results

3.1 Macro and ranking metrics on RAF-DB compound subset

Consistent with recent geometry-aware training results [4,9–13], the hyperspherical–margin framework substantially outperforms the softmax warm-up on the RAF-DB compound subset (Table 1). Macro scores improve from P/R/F1 = 0.09/0.11/0.09 and ROC-AUC/PR-AUC = 0.52/0.10 to 0.88/0.89/0.88 and 0.98/0.92, i.e., ΔF1 = +0.79, ΔROC-AUC = +0.46, ΔPR-AUC = +0.82. The joint rise of F1 and PR-AUC indicates better positive-class retrieval across thresholds; the near-ceiling ROC-AUC reflects robust rank ordering under imbalance. Gains are achieved through unit-sphere normalization with a small angular margin and a stable three-stage schedule, without increasing backbone capacity.

Method / Stage	Precision (macro)	Recall (macro)	F1 (macro)	ROC-AUC (macro OvR)	PR-AUC (macro)
Warm-up (softmax)	0.09	0.11	0.09	0.52	0.10
Ours (CosFace, M+FT)	0.88	0.89	0.88	0.98	0.92

Table 1 reports macro-averaged Precision/Recall/F1 and one-vs-rest ROC-AUC/PR-AUC.

3.2 Per-class metrics

Table 2 reports classwise Precision/Recall/F1 on the RAF-DB compound subset (post-margin). Scores are uniformly high (macro P/R/F1 = 0.88/0.89/0.88), with the largest gains on historically overlapping compounds [9,10]. Notably, Sadly Fearful attains R = 0.91 (F1 = 0.82), Fearfully Angry/Surprised reach F1 = 0.94/0.89, and Happily Disgusted/Surprised achieve F1 = 0.88/0.89. Disgustedly Surprised yields the best per-class result (F1 = 0.96). These outcomes indicate that angular separation increases inter-class margins among semantically adjacent categories without collapsing precision.

Class	Precision	Recall	F1	Support
Angrily Disgusted	0.87	0.87	0.87	151
Angrily Surprised	0.92	0.84	0.88	176
Disgustedly Surprised	0.98	0.93	0.96	148
Fearfully Angry	0.95	0.93	0.94	150
Fearfully Surprised	0.91	0.86	0.89	160
Happily Disgusted	0.86	0.90	0.88	191
Happily Surprised	0.92	0.87	0.89	197
Sadly Angry	0.85	0.82	0.84	163

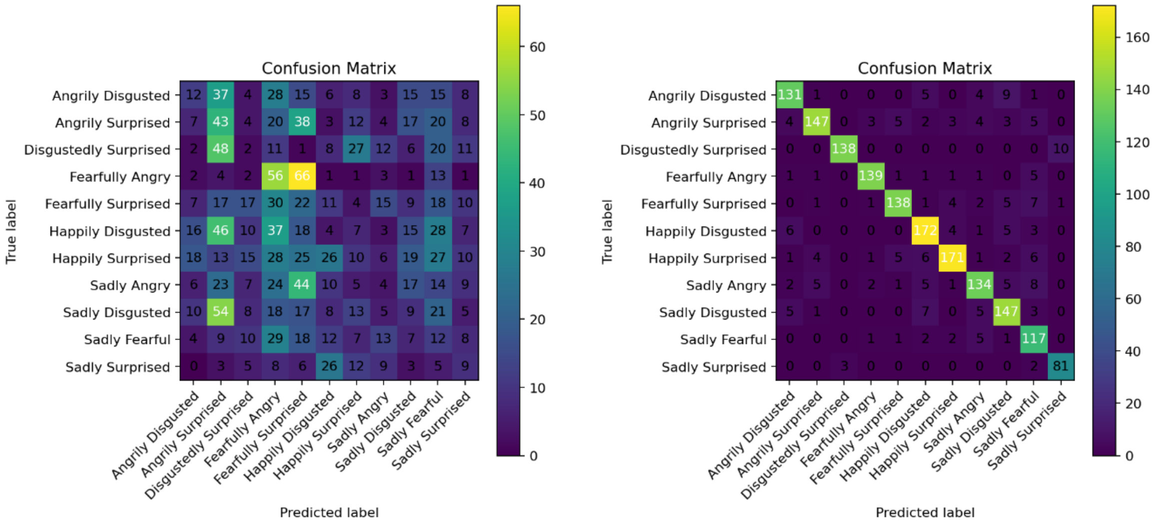
International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Sadly Disgusted	0.83	0.88	0.85	168
Sadly Fearful	0.75	0.91	0.82	129
Sadly Surprised	0.88	0.94	0.91	86
Average	0.88	0.89	0.88	156.27

Table 2. Per-class Precision/Recall/F1 on RAF-DB compound subset.

3.3 Confusion-pair analysis via symmetric confusion mass

Figure 1 shows confusion matrices before and after angular-margin training; the post-margin matrix has stronger diagonals and smaller off-diagonals.



(a) (b)

Figure 1. Confusion matrices and pairwise error reductions. (a) Before margin (warm-up). (b) After margin + partial fine-tuning.

Largest reductions (examples) (Eq. (5)):
Fearfully Angry ↔ Fearfully Surprised: $96 \to 2 (\Delta = -94)$;
Angrily Surprised ↔ Sadly Disgusted: $71 \to 4 (\Delta = -67)$;
Fearfully Surprised ↔ Sadly Angry: $59 \to 3 (\Delta = -56)$;
Disgustedly Surprised ↔ Angrily Surprised: $52 \to 0 (\Delta = -52)$;
Angrily Surprised ↔ Fearfully Surprised: $55 \to 6 (\Delta = -49)$;
Fearfully Angry ↔ Sadly Fearful: $42 \to 6 (\Delta = -36)$.
Residual two-way mass is small (e.g., Disgustedly Surprised ↔ Sadly Surprised: $16 \to 13$, $\Delta = -3$). Overall, SCM confirms that the hyperspherical–margin framework widens inter-class angles and tightens intra-class clusters, translating geometry into bilateral error reductions.

3.4 ROC/PR curves and embeddings

Figure 2 (OvR ROC). Post-margin ROC curves dominate the warm-up for nearly all classes; the macro AUC is summarized in Table 1.

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

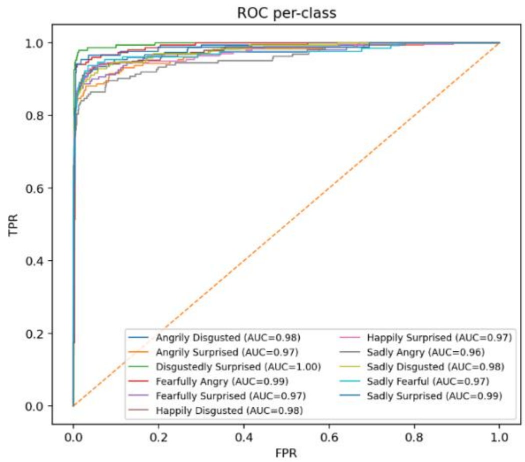


Figure 2. One-vs-rest ROC curves per class.

Figure 3 (OvR PR). PR curves—being imbalance-aware—show consistent dominance across operating points, indicating better positive ranking at fixed recall/precision.

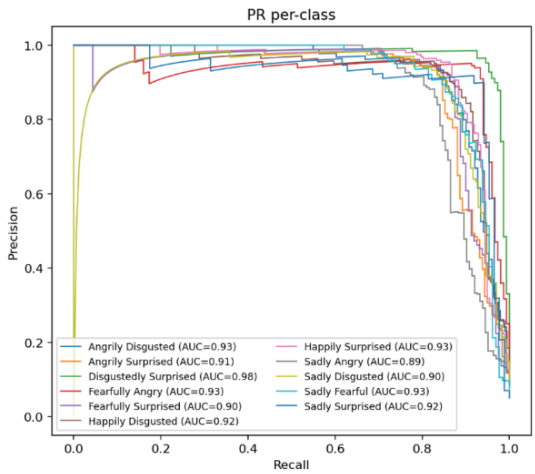


Figure 3. One-vs-rest PR curves per class.

Figure 4 (t-SNE, 256-d). Post-margin embeddings form tighter intra-class clusters with larger inter-class gaps on $\mathbb{S}^{d-1}$, consistent with margin-induced hyperspherical separation [9,10].

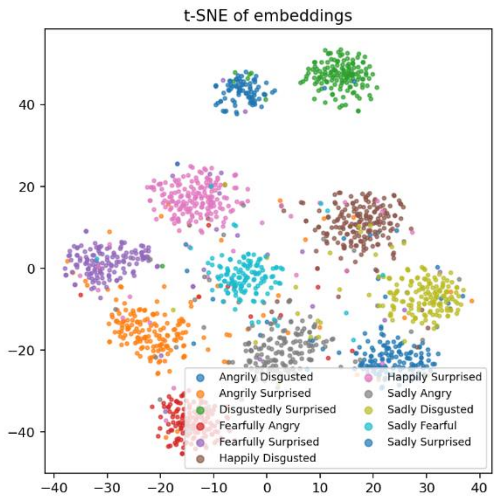


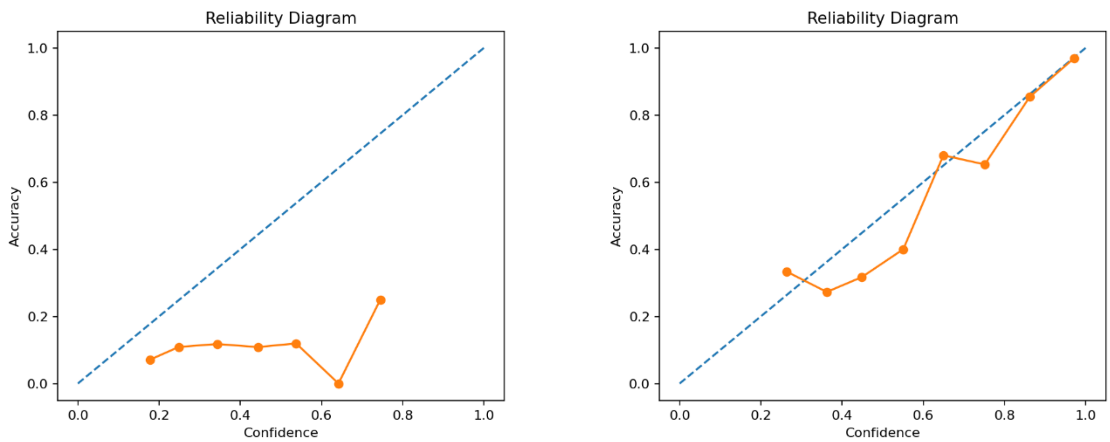
Figure 4. t-SNE of 256-d embeddings (post-margin).

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.5 Calibration analysis

Figure 5 (reliability). Post-margin curves lie closer to the diagonal, indicating better probability–accuracy agreement.



Pre-margin **Post-margin**

Figure 5. Reliability diagrams before vs after margin training.

Table 3 (ECE, 10 bins). Warm-up 0.192409 → Post-margin 0.019517; Δ = −0.172892 (≈ −89.9%).

Method / Stage	ECE
Warm-up (softmax)	0.192409
Ours (CosFace, M+FT)	0.019517

Table 3. Expected Calibration Error (ECE; lower is better; Eq. (6)).

Interpretation. Angular supervision on the unit sphere yields more faithful confidence estimates, consistent with calibration best practices for modern deep classifiers [4,12,13].

3.6 Device-centric efficiency (edge suitability)

We report single-image latency (median/p90), FPS, parameter count, and compute, following a steady-state protocol (20 warm-ups, 200 timed runs). Table 4 reports device-centric throughput and memory; the model maintains real-time performance under modest compute [14].

Device / Runtime	Params	GMACs / FLOPs (G)	Latency (ms, median)	p90 (ms)	FPS	CPU RSS (MB)	GPU Mem (MB, cur/peak)
CPU: AMD64 Family 25 Model 80 Stepping 0 (AuthenticAMD)	3,245,442	0.433	11.097	11.456	90.11	396.68	–
GPU: NVIDIA GeForce GTX 1650 Max-Q (driver 577.00, 4096 MiB)	3,245,442	0.433	7.297	7.581	137.04	785.02	14.86 / 573.21

Table 4. Device-centric performance (batch=1; 224×224; best_s2).

3.7 Cross-dataset supporting evidence

To test generality beyond the RAF-DB compound subset, we include headline metrics on popular FER datasets. Cross-dataset summaries (Figure 6) and composite figures (Figure 7) demonstrate that the approach generalizes beyond the RAF-DB compound subset [9,10,14].

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

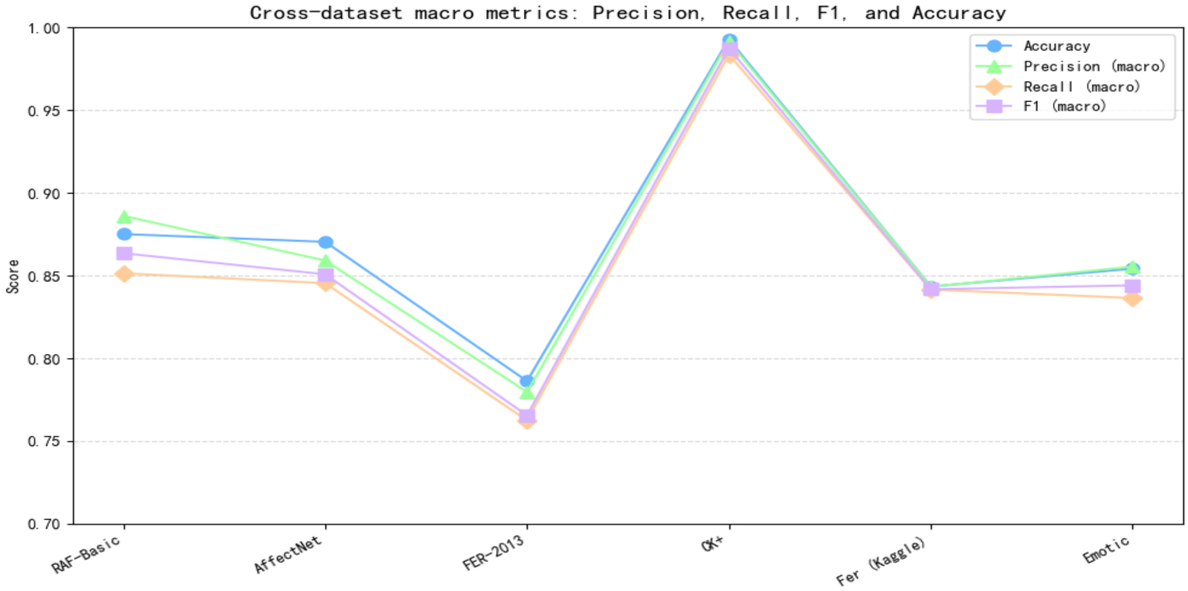
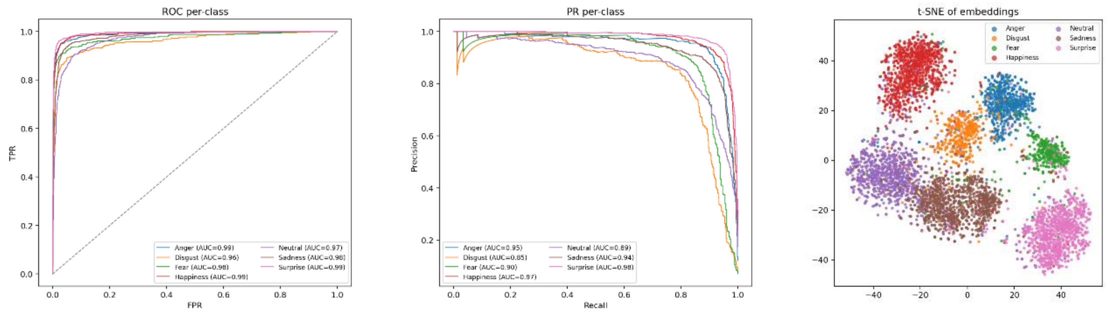
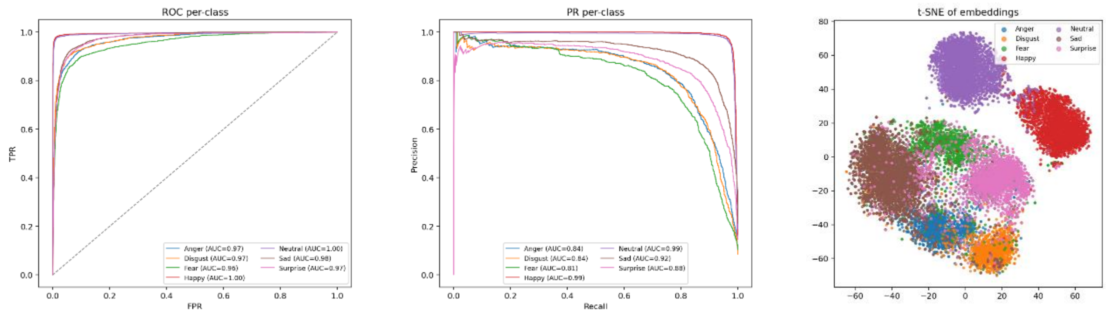


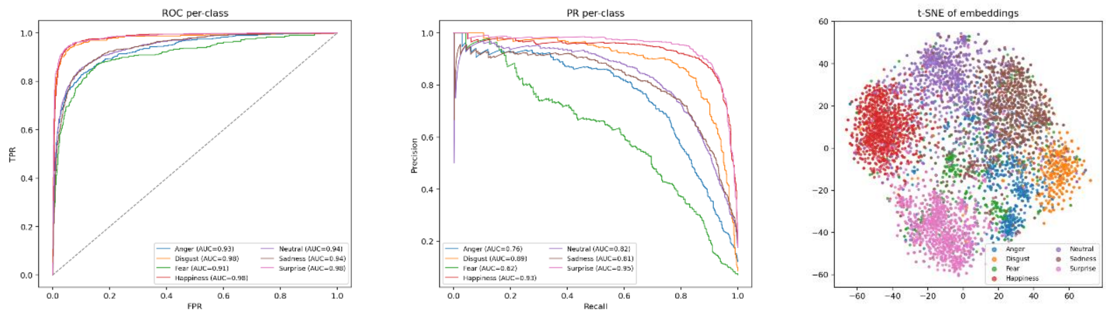
Figure 6. Headline metrics across datasets (Accuracy/macro P/macro R/macro F1).



RAF-Basic



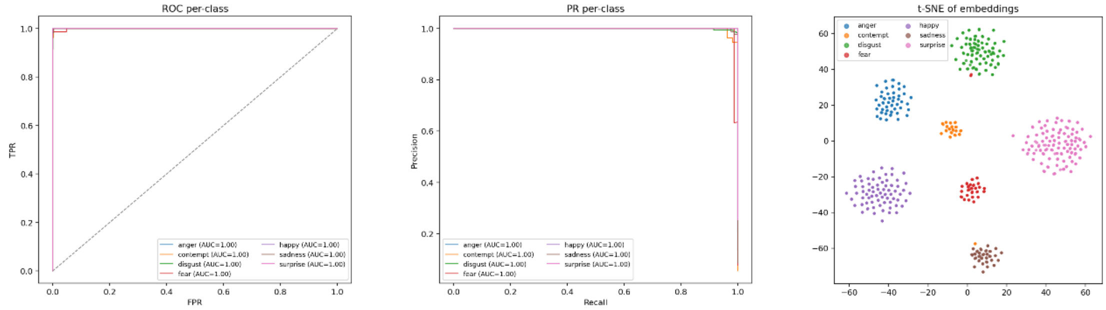
AffectNet



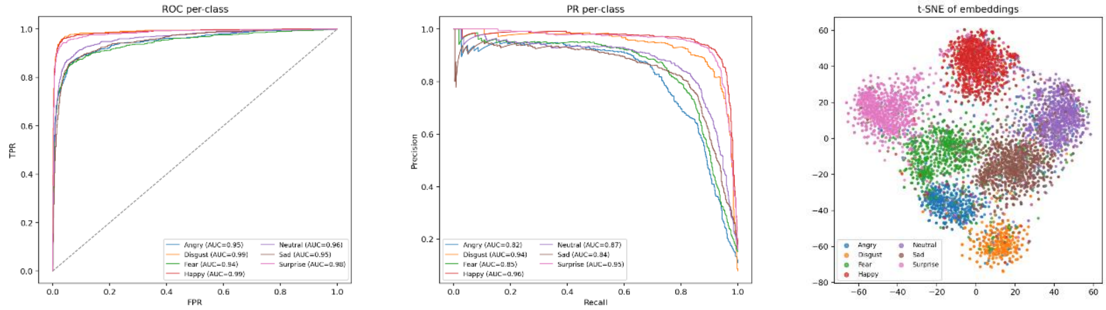
FER-2013

Received: August 12, 2025

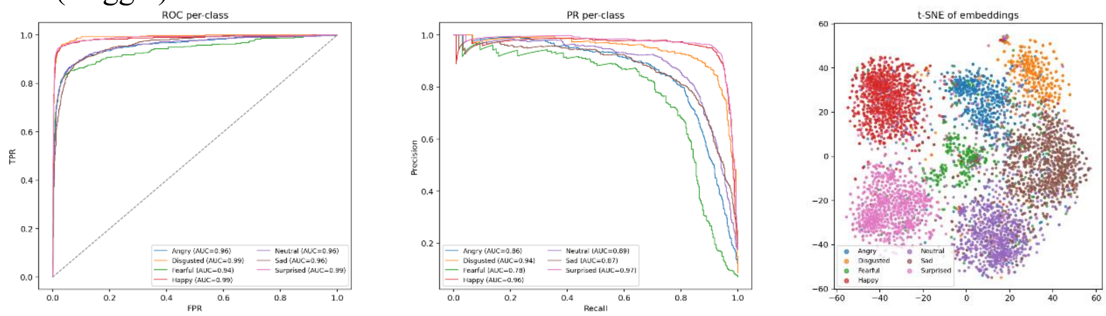
International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



CK+



Fer (Kaggle)



Emotic

Figure 7. Composite ROC/PR/t-SNE panels for other datasets.

4. Discussion (mathematical perspective)

Geometry as the primary variable. The optimization acts on $\mathbb{S}^{d-1}$ with a small margin m, imposing cone-wise separation among classes. Empirically observed gains in macro-metrics and large negative ΔSCM quantify that separation.

Curricular stabilization. Treating margin learning as a constrained scheme—head-only shaping followed by localized fine-tuning—reduces variance under small, imbalanced labels and avoids early instability from large margins.

Calibration as a corollary of normalization. Because logits derive from normalized feature–weight interactions, confidence magnitudes are less arbitrary; ECE improvements follow naturally. In settings requiring thresholding (e.g., alarms, triage), this is consequential.

Limitations and extensions. Compound labels remain noisy; multi-seed analysis would solidify small ablation deltas. Mathematically natural extensions include class-dependent margins m_c, adaptive ramps $m_c(t)$, and adaptive ECE binning to reduce discretization bias.

5. Conclusion

We introduced a hyperspherical–margin framework for compound facial emotions that (i) maps embeddings and classifier weights to the unit sphere $\mathbb{S}^{d-1}$, (ii) enforces an additive angular margin in the true-class logit, and (iii) employs a three-stage schedule (warm-up, head-only margin with gentle ramp, partial fine-tuning) to stabilize optimization on small and imbalanced datasets. In this setting

Received: August 12, 2025

the decision rule is a nearest-direction classifier with explicit cone separation; the margin term widens inter-class angles while normalization compacts intra-class variance.

The empirical study on RAF-DB compound-11 shows that the geometric design translates into measurable improvements: macro and ranking metrics increase over the likelihood-only warm-up; the symmetric confusion mass decreases for historically confusable pairs, indicating larger inter-centroid angles; and calibration improves, as reflected by reliability diagrams and Expected Calibration Error. These effects are obtained without enlarging backbone capacity, and device-centric profiling indicates that the model remains suitable for real-time inference.

Limitations include residual confusions among visually adjacent compounds, single-seed reporting for some ablations, and fixed-bin calibration estimates. Natural extensions include class-dependent or data-adaptive margins $m_y$and $m(t)$, selective fine-tuning or low-rank adapters in upper layers, and lightweight post-hoc calibration to further reduce ECE. An additional avenue is cross-dataset and open-set evaluation, where the same hyperspherical geometry may regularize decision cones under distribution shift.

In summary, making the geometry explicit—unit-sphere normalization with a small angular margin—together with a stability-minded training schedule yields separable and calibrated representations for compound facial emotions in a mathematically coherent and computationally efficient manner.

References

[1] Li S, Deng W. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition. IEEE Transactions on Image Processing 2019;28:356–70. https://doi.org/10.1109/TIP.2018.2868382.

[2] Kollias D. Multi-Label Compound Expression Recognition: C-EXPR Database & Network. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2023, p. 5589–98. https://doi.org/10.1109/CVPR52729.2023.00541.

[3] Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning - Volume 70, JMLR.org; 2017, p. 1321–30.

[4] Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, et al. A survey of uncertainty in deep neural networks. Artif Intell Rev 2023;56:1513–89. https://doi.org/10.1007/s10462-023-10562-9.

[5] Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE; 2018, p. 5265–74. https://doi.org/10.1109/CVPR.2018.00552.

[6] Deng J, Guo J, Yang J, Xue N, Kotsia I, Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. IEEE Trans Pattern Anal Mach Intell 2022;44:5962–79. https://doi.org/10.1109/TPAMI.2021.3087709.

[7] Howard A, Sandler M, Chen B, Wang W, Chen L-C, Tan M, et al. Searching for MobileNetV3. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE; 2019, p. 1314–24. https://doi.org/10.1109/ICCV.2019.00140.

[8] Luo H, Gu Y, Liao X, Lai S, Jiang W. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE; 2019, p. 1487–95. https://doi.org/10.1109/CVPRW.2019.00190.

[9] Zhang Y, Zheng X, Liang C, Hu J, Deng W. Generalizable Facial Expression Recognition. Computer Vision – ECCV 2024: 18th European Conference, Milan: ACM; 2025, p. 231–48. https://doi.org/10.1007/978-3-031-72630-9_14.

[10] Liu T, Li J, Wu J, Zhang L, Zhao S, Chang J, et al. Cross-Domain Facial Expression Recognition via Disentangling Identity Representation. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, California: International Joint Conferences on Artificial Intelligence Organization; 2023, p. 1213–21. https://doi.org/10.24963/ijcai.2023/135.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[11] Khalifa A, Al-Hamadi A. JAMsFace: joint adaptive margins loss for deep face recognition. Neural Comput Appl 2023;35:19025–37. https://doi.org/10.1007/s00521-023-08732-5.
[12] Tang Y-C, Chen P-Y, Ho T-Y. Neural Clamping: Joint Input Perturbation and Temperature Scaling for Neural Network Calibration 2024.
[13] Hsiung L, Tang Y-C, Chen P-Y, Ho T-Y. NCTV: Neural Clamping Toolkit and Visualization for Neural Network Calibration. Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, Washington: 2023. https://doi.org/https://doi.org/10.48550/arXiv.2211.16274.
[14] Shou Z, Huang Y, Li D, Feng C, Zhang H, Lin Y, et al. A Student Facial Expression Recognition Model Based on Multi-Scale and Deep Fine-Grained Feature Attention Enhancement. Sensors 2024;24:6748. https://doi.org/10.3390/s24206748.

Received: August 12, 2025