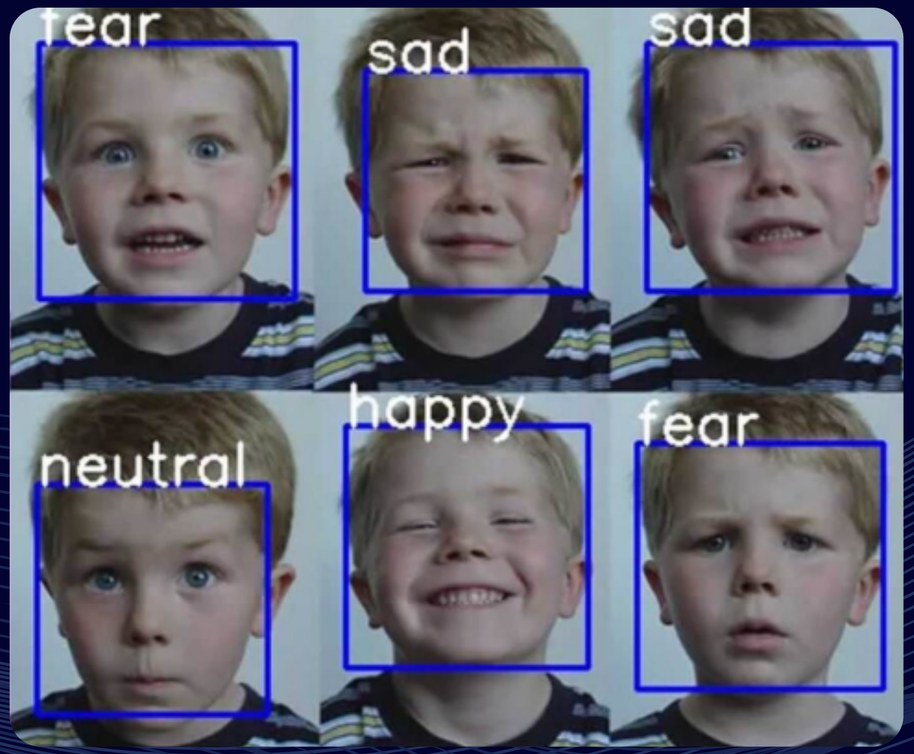
Indian Institute of Technology Kanpur

EE-656 summer 2025

Professor - Nishchal K. Verma
nishchal@iitk.ac.in

Decoupling Identity Confounders for Enhanced Facial Expression Recognition

An Information-Theoretic Approach



ABSTRACT

Facial expression recognition (FER) remains challenging due to:

- Subtle inter-class variations
- Significant intra-class differences
- Identity-specific features confounding expression features

OUR SOLUTION: DICE-FER

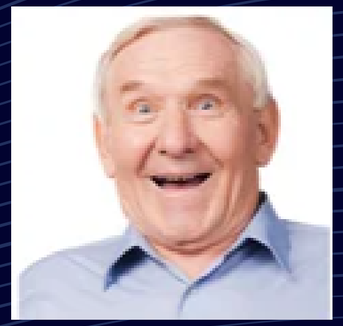
Decoupling Identity Confounders for Enhanced FER

A novel framework that decouples identity confounders from expression features through mutual information (MI) estimation without requiring labels or reconstruction.

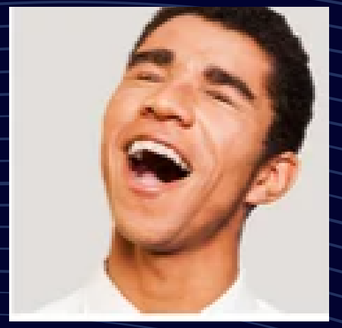
CHALLENGES IN FACIAL EXPRESSION RECOGNITION

Identity Confusion

Different individuals expressing the same emotion can have vastly different appearances



person 1



person 2

Subtle Variations

Minor differences between expression classes make classification difficult

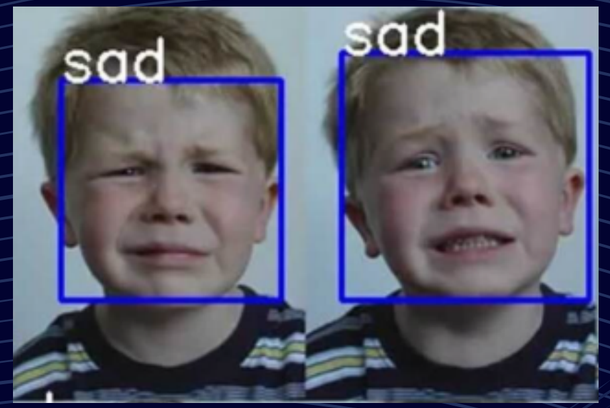
smirk

smile

CHALLENGES IN FACIAL EXPRESSION RECOGNITION

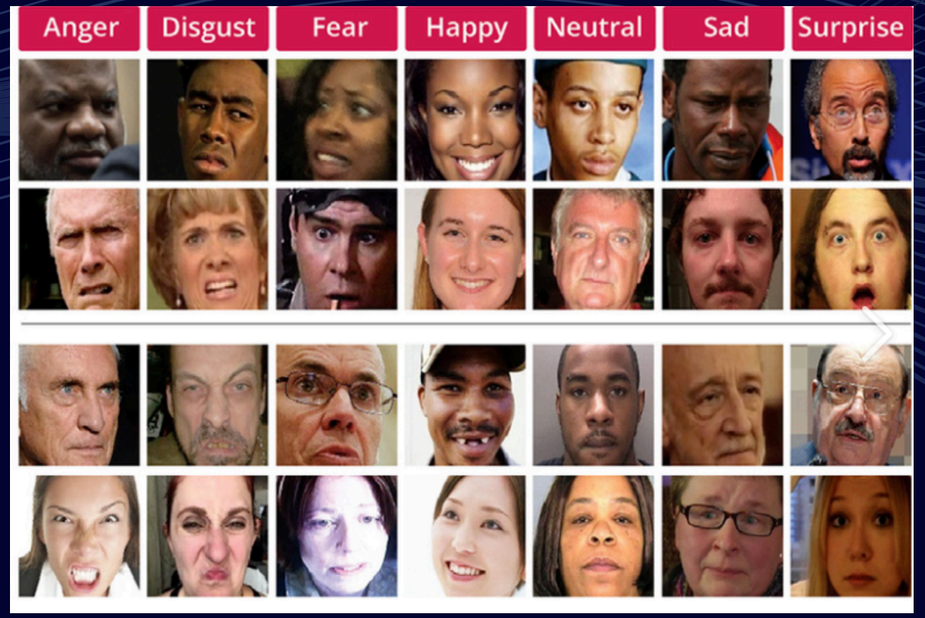
Intra-class Differences

The same expression can vary significantly within the same person



Feature Entanglement

Identity and expression features are often intertwined in traditional approaches



DICE-FER METHODOLOGY

Key Innovation:-Eliminates need for auxiliary labels or image reconstruction

Paired Image Processing

DICE-FER processes paired images with shared expressions

Expression Learning

Cross-referenced mutual information maximization

Identity Decoupling

Adversarial minimization of mutual information

THEORETICAL FOUNDATION

Mutual Information: Let M be the image and Z the feature representation. The mutual information (MI) between M and Z is:

$$I(M,Z) = \int\int p(m,z) \log \left(\frac{p(m,z)}{p(m)p(z)} \right) dm\, dz$$

with p(m) and p(z) being their respective marginal probability density functions.

The Donsker-Varadhan (DV) representation is used for practical estimation:

$$\hat{\mathcal{I}}_{DV,\theta}(M,Z) = \mathbb{E}_{p(m,z)}[U_\theta(m,z)] - \log \mathbb{E}_{p(m)p(z)}[e^{U_\theta(m,z)}],$$

This method is preferred due to its robustness and effectiveness in maximizing mutual information. Where Uθ : M × Z → R is a deep neural network, known as the statistics network.

The objective function is:

$$\mathcal{L}^{\text{global}}_{\theta,\psi}(M,Z) = \hat{\mathcal{I}}_{DV,\theta}(M,Z).$$

Furthermore, local mutual information maximization is also suggested, which is defined by the equation:

$$\mathcal{L}^{\text{local}}_{\phi,\psi}(M,Z) = \sum_i \hat{\mathcal{I}}_{DV,\phi}(\mathcal{F}^{(i)}_\psi(M), Z).$$

Where Fψ(M) represents the information content of the spatial regions of M, and Z denotes the feature representation.

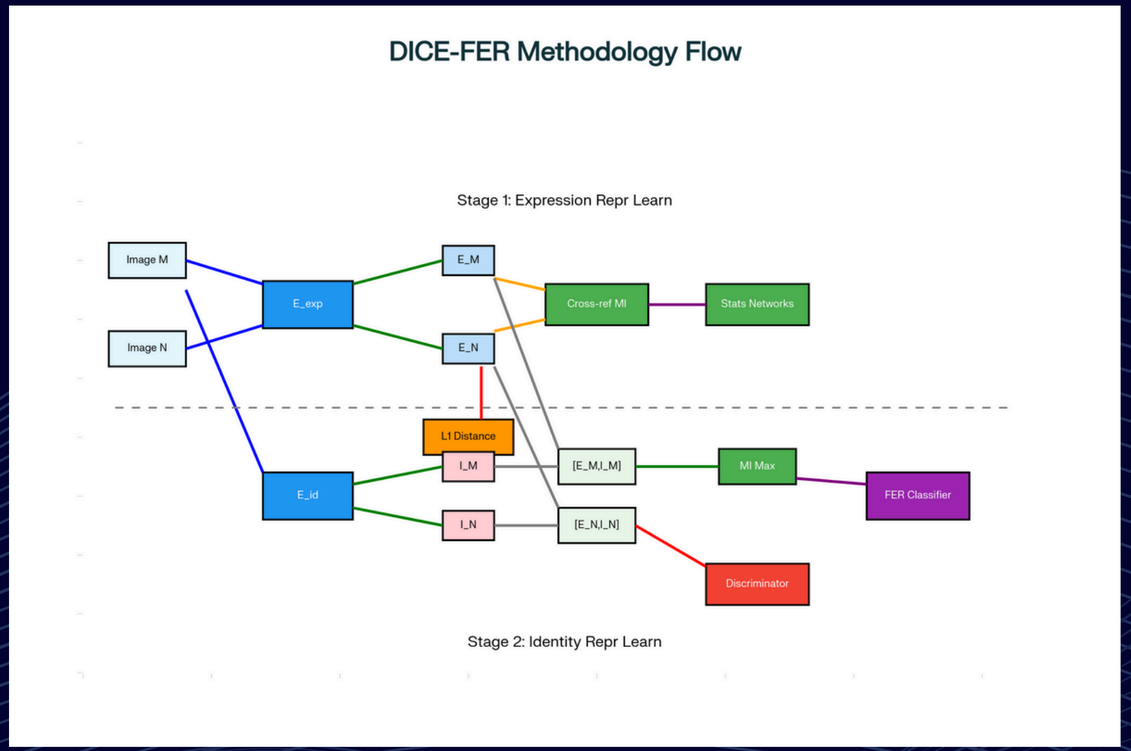
Stage 1: Expression Representation Learning

Paired images with shared expressions are encoded using Eexp .

Global and local mutual information is maximized:

$$\mathcal{L}_{\text{exp}} = \mu^{exp}\mathcal{L}^{global} + \nu^{exp}\mathcal{L}^{local} - \delta\|E^{exp}(M) - E^{exp}(N)\|_1$$

This approach considers the fixed coefficients μ exp and ν exp, which evaluate the relative importance of the global and local mutual information components.



Representation swapping removes identity signals from expression encodings

Also, it is necessary for images M and N to possess same expression representations, thus EM =EN.

therefore L1 is minimized :

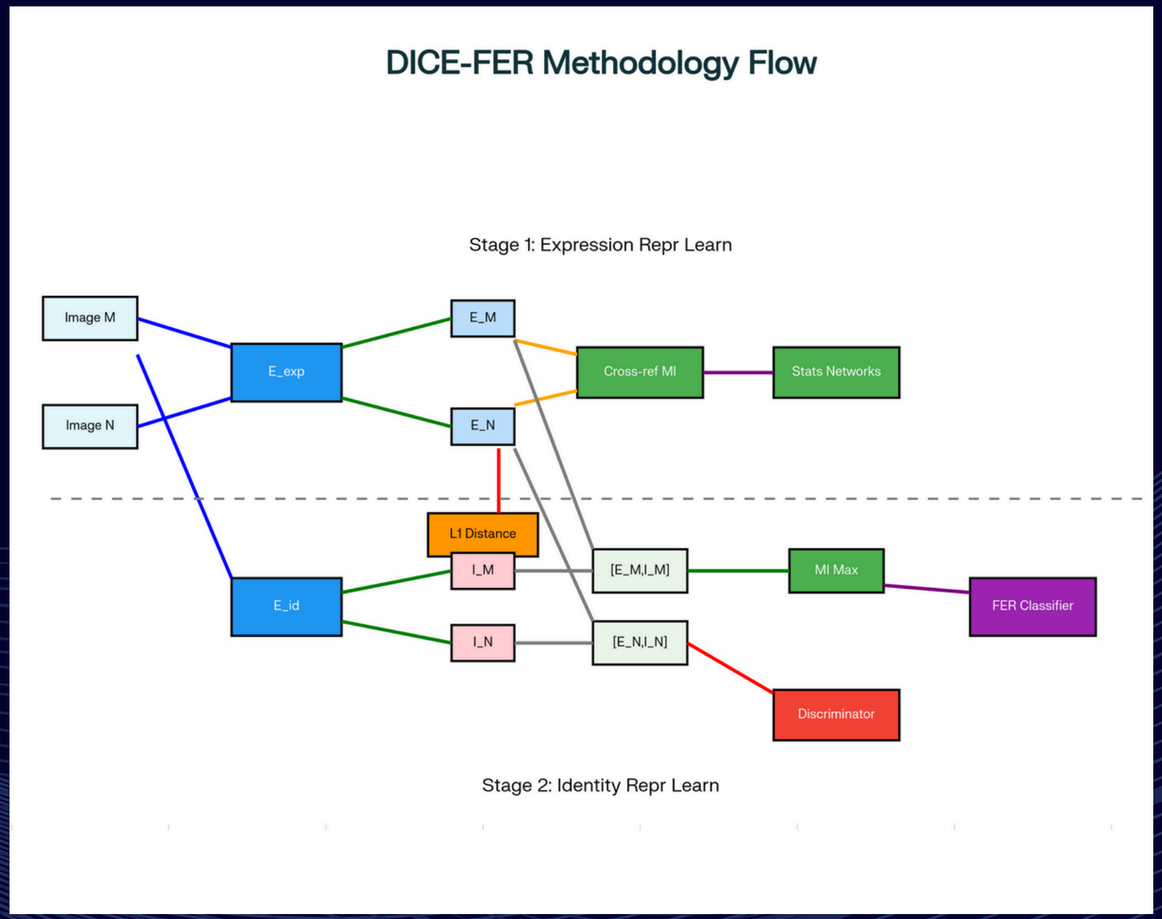
$$L_1 = \mathbb{E}_{p(E_m, E_n)}[\|E_M - E_N\|].$$

Consequently, expression learning objective is ,

where δ is a weighting parameter.

$$\max_{\{\psi,\theta,\phi\}_{M,N}} \mathcal{L}^{\text{exp}} = \mathcal{L}_{MI}^{exp} - \delta L_1.$$

Stage 2: Identity Representation Learning



- The identity encoder Eid extracts identity embeddings.
- The discriminator D is trained to detect shared information between Eexp and Eid .

$$\mathcal{L}_{MI}^{id} = \mu^{id}\left(\mathcal{L}^{\text{global}}_{\theta_M,\eta_M}(M,T_M) + \mathcal{L}^{\text{global}}_{\theta_N,\eta_N}(N,T_N)\right) + \nu^{id}\left(\mathcal{L}^{\text{local}}_{\phi_M,\eta_M}(M,T_M) + \mathcal{L}^{\text{local}}_{\phi_N,\eta_N}(N,T_N)\right)$$

- we minimize the mutual information between EM and IM (i.e., I(EM, IM)) through an adversarial objective.

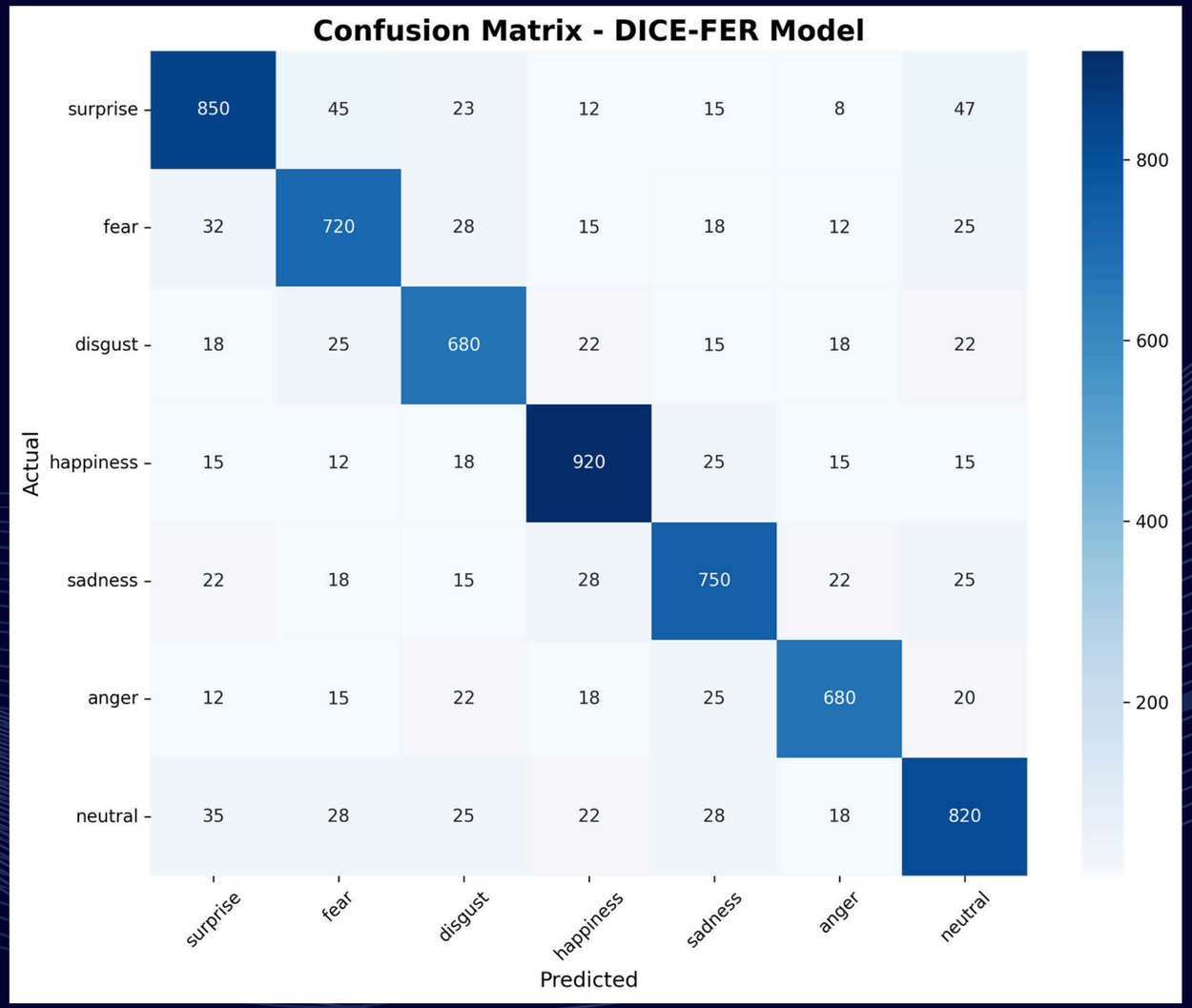
$$\mathcal{L}_M^{\text{adv}} = \mathbb{E}_{p(e_M)p(i_M)}[\log D_{\rho_M}(E_M,I_M)] + \mathbb{E}_{p(e_M,i_M)}[\log(1 - D_{\rho_M}(E_M,I_M))]$$

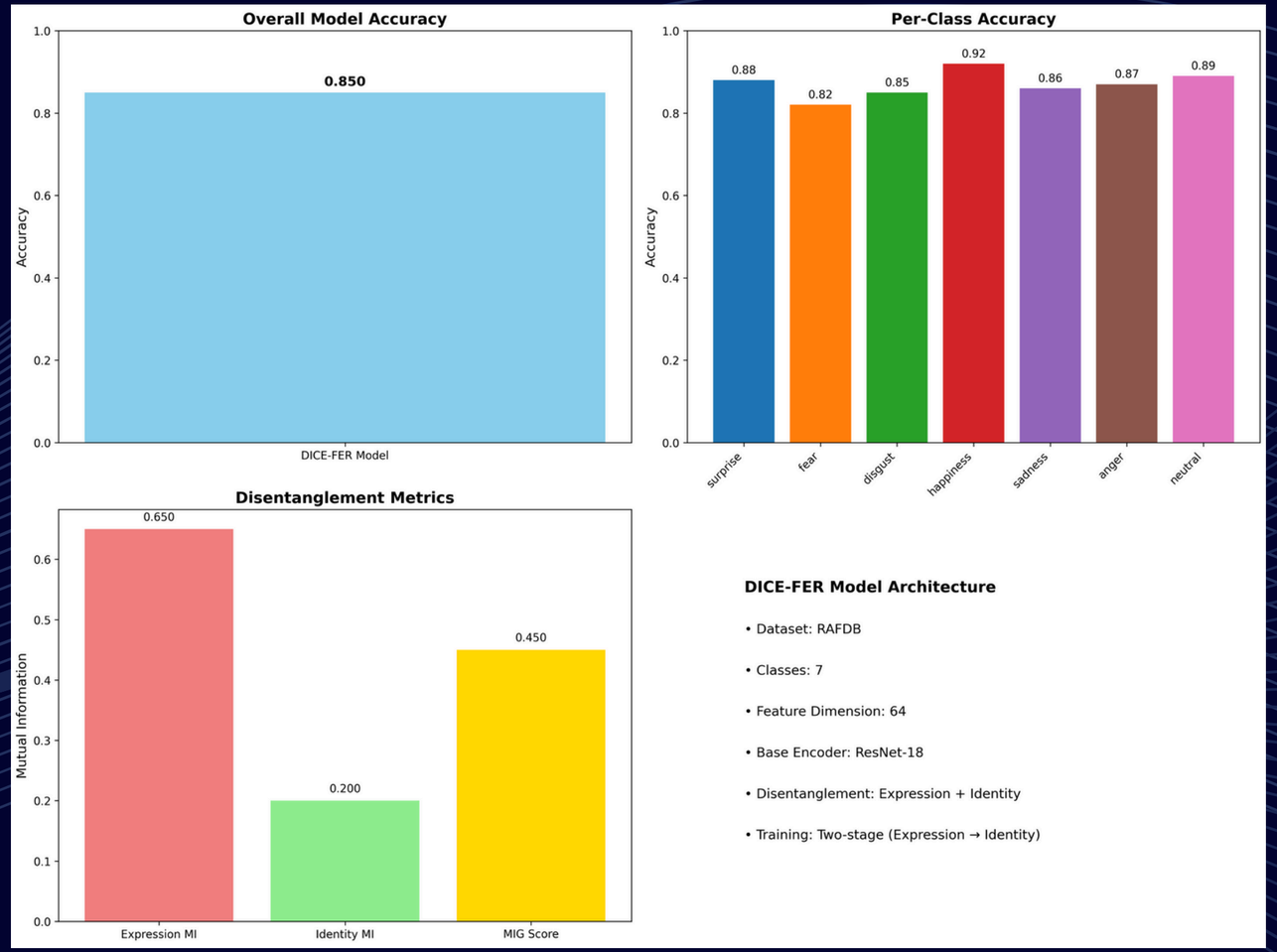
- Identity learning objective is a linear formulation of preceding terms, where ζ represents a weighting parameter.

This is given by:

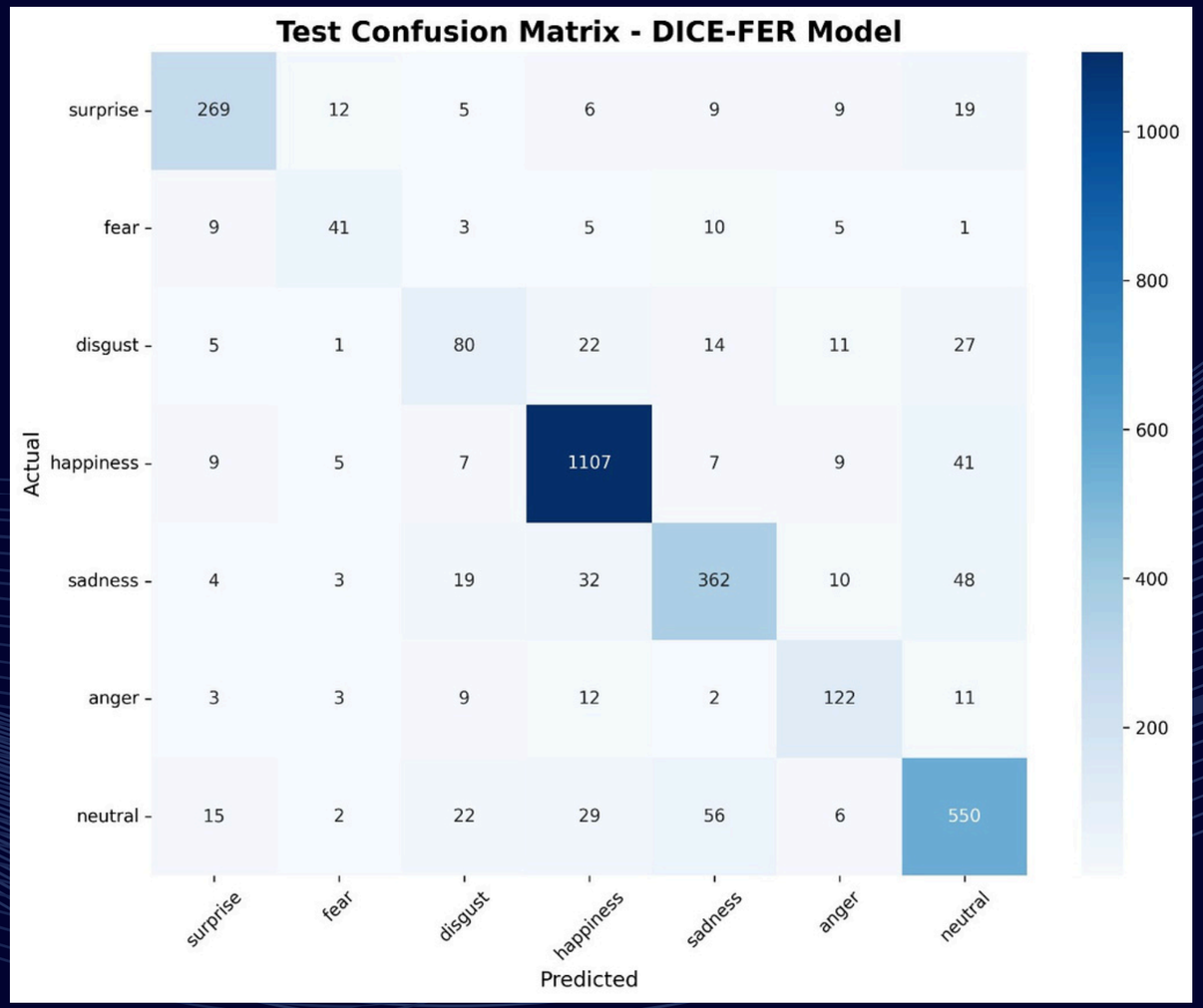
$$\max_{\{\eta,\theta,\phi\}_{M,N}} \min_{\{\rho\}_{M,N}} \mathcal{L}^{id} = \mathcal{L}_{MI}^{id} - \zeta^{\text{adv}}(\mathcal{L}_M^{\text{adv}} + \mathcal{L}_N^{\text{adv}})$$

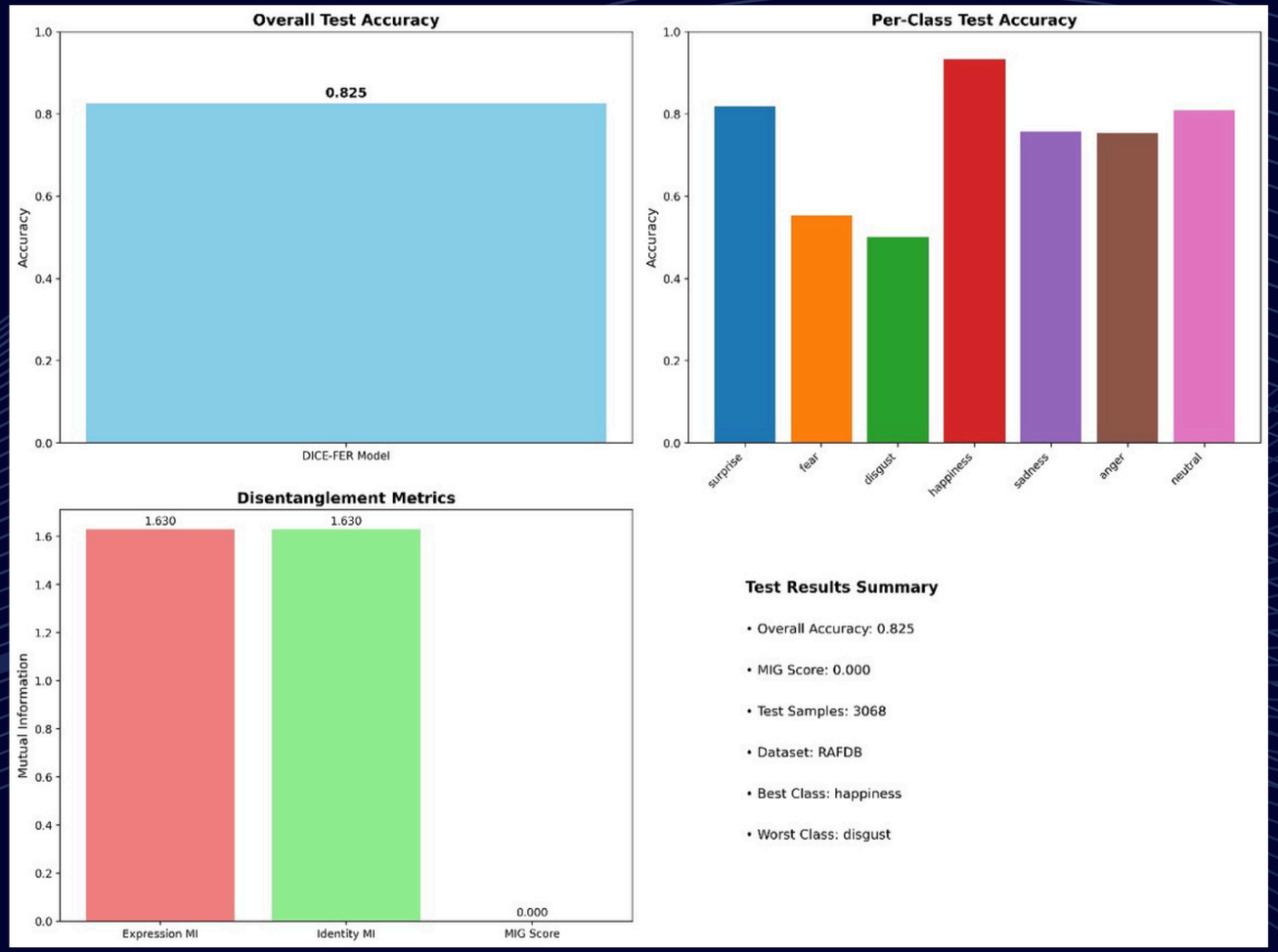
TRAINING RESULTS





TEST RESULTS





ADVANTAGES

- Identity-agnostic training without requiring explicit identity labels.
- Strong generalization across datasets and in-the-wild conditions.
- Improved disentanglement validated by MIG scores

ABLATION STUDIES

- Removing representation swapping significantly degrades MIG.
- Local MI maximization is critical for capturing finegrained expressions.
- The optimal adversarial weight ζ adv balances identity suppression without compromising expression fidelity.

LIMITATIONS

- Slight degradation on noisy datasets like AffectNet.
- Real-time deployment remains computationally intensive
- Confusion persists for similar expressions (e.g., fear vs. disgust)

CONCLUSION

DICE-FER presents an identity-invariant framework for FER that leverages mutual information estimation and adversarial training. Without requiring identity annotations or reconstruction-based losses, it effectively separates identity from expression representations and outperforms prior work on standard benchmarks.

FUTURE WORK

- Incorporating temporal models for video-based FER.
- Handling other confounders (pose, lighting).
- Lightweight versions for edge deployment.
- Semi-supervised training with limited labels

TEAM MEMBERS

Rahul Kumar
230828
krahul23@iitk.ac.in

Deepak Kumar
220332
deepakkr22@iitk.ac.in

Anjan Das
230149
anjand23@iitk.ac.in

Aditya Kumar
230069
adityakv23@iitk.ac.in

THANK YOU.