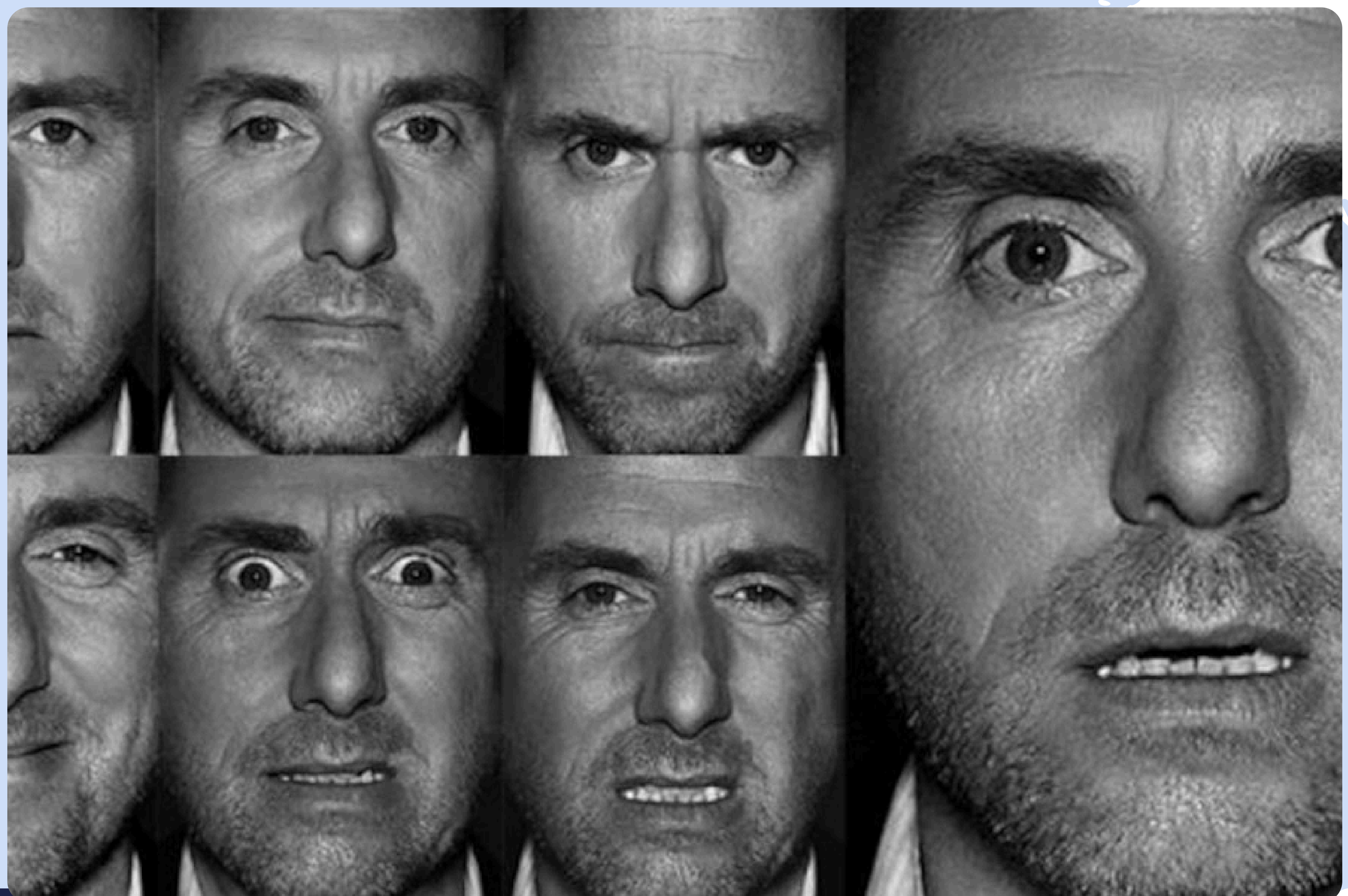
DUKE KUNSHAN

Fair Recognition of Hidden Emotions:

Addressing Class Imbalance in Micro-Expression Prediction with Machine Learning



Introduction

- Micro-expressions are brief, involuntary facial cues that reveal hidden emotions, but their subtlety makes detection difficult.
- Machine learning methods such as CNNs and SVMs enhance recognition but struggle with dataset bias and class imbalance.
- Minority emotions (e.g., fear, disgust) are underrepresented, leading to poor recall and fairness concerns in prediction.

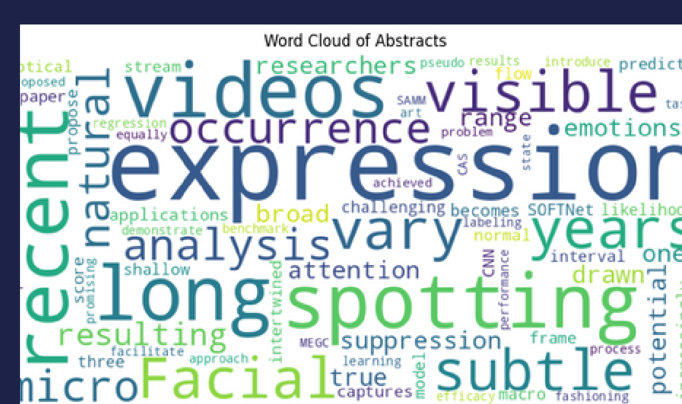


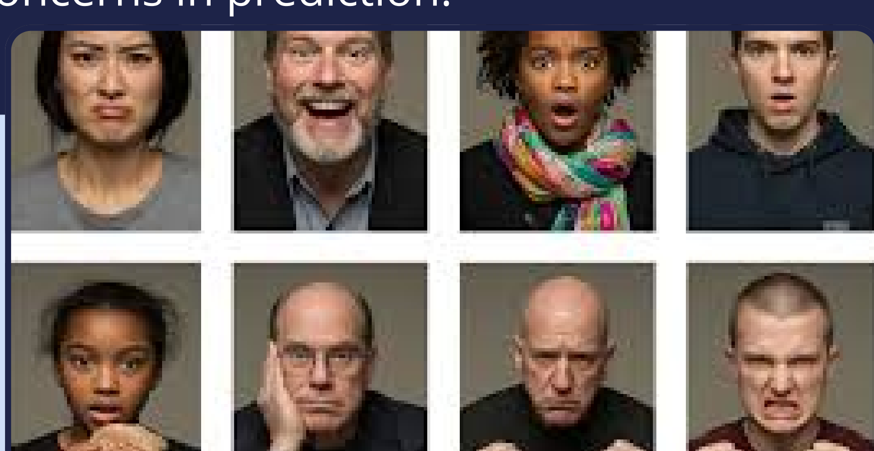
Figure1. Word Cloud of Abstracts.

Figure2. Project linked to SDGs.

Research Objectives

- To improve recognition of underrepresented emotions by applying data-centric (SMOTE, augmentation) and feature-centric (embedding-based) strategies.
- To evaluate trade-offs between accuracy, fairness, and interpretability across models such as Logistic Regression, SVM, and AutoML.
- To advance equitable and transparent micro-expression recognition for practical use in mental health, education, and use AI ethics.

Figure3. Semantic Network of Keywords.



Research Question

How can machine learning methods address class imbalance to improve the fair recognition of low-frequency emotions such as fear and disgust in micro-expression datasets?

Method

3 public micro-expression datasets (Ziya07~2,000 samples, Kmirfans~1,500 samples, Kori~1,800 samples)

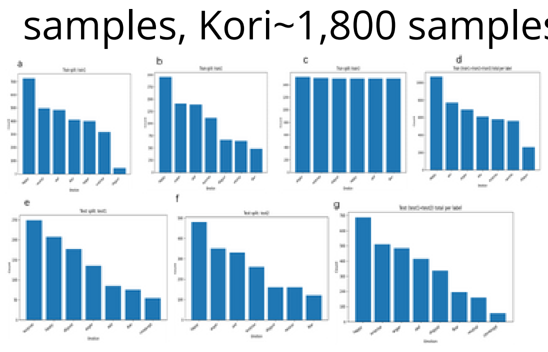


Figure4. Class distributions across different training and testing splits.

7 emotion categories (anger, disgust, fear, happy, neutral, surprise, and contempt) .

Feature Extraction

ResNet18 used as CNN backbone for deep features; Local Binary Patterns (LBP) for handcrafted texture features.

Dimensionality Reduction

Principal Component Analysis (PCA) applied to reduce redundancy and computational cost.

Classification

Support Vector Machine (SVM, RBF kernel) and Logistic Regression used to evaluate feature sets.

Data Augmentation

Synthetic Minority Oversampling Technique (SMOTE) applied to generate synthetic samples for minority classes, improving recall and macro-F1.

Result & Discussion

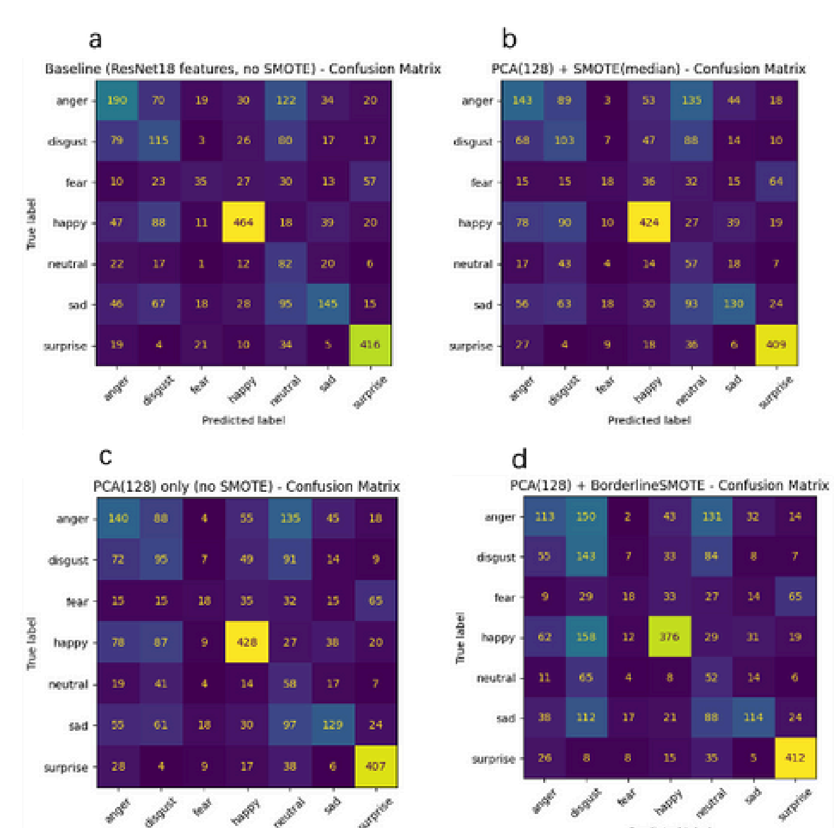
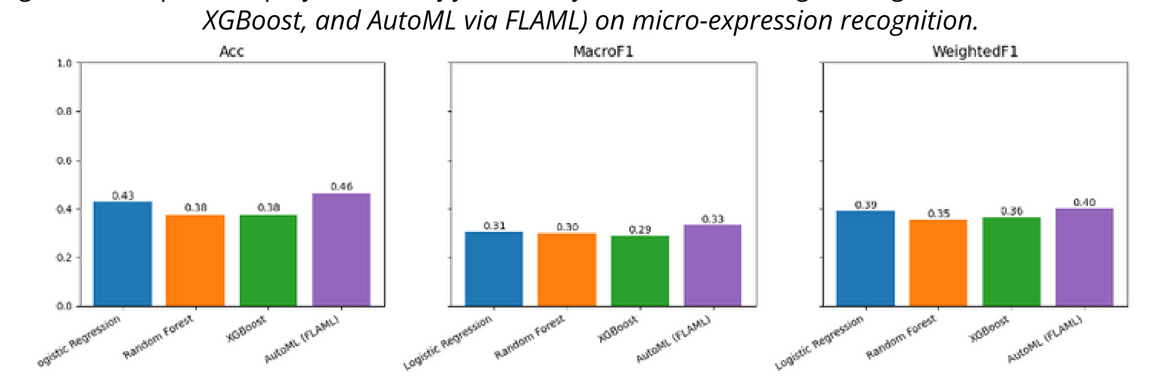


Figure5. Confusion matrices of different classification settings.

Figure 6. Comparative performance of four classification models (Logistic Regression, Random Forest, XGBoost, and AutoML via FLAML) on micro-expression recognition.



Baseline vs. Augmentation

ResNet18 baseline mainly recognized frequent emotions (happy, neutral).

SMOTE boosted recall and Macro-F1 for fear and disgust, despite reduced accuracy.

Model Comparison

AutoML achieved highest accuracy (0.46) and Weighted-F1 (0.40).

Logistic Regression was simple yet competitive; Random Forest/XGBoost struggled with deep embeddings.

Challenges Identified

- Micro-expression datasets remain small, imbalanced, and lack demographic diversity, limiting generalizability.
- Ethical risks such as privacy concerns, unclear consent, and potential misuse in surveillance undermine public trust.

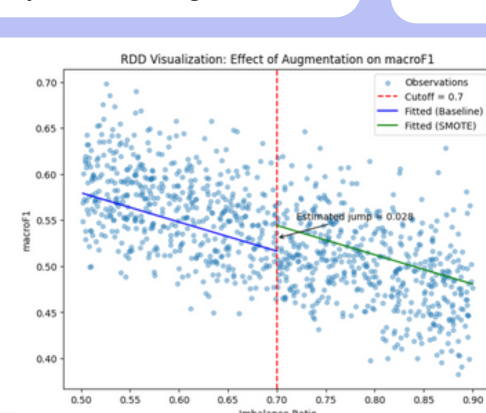


Figure 7. Regression discontinuity analysis of augmentation effects on fairness.

Recommendations

- Expand datasets with diverse, ethically collected samples and apply fairness-aware augmentation methods.
- Promote transparency and accountability by adopting interpretable models, bias auditing, and clear consent protocols.

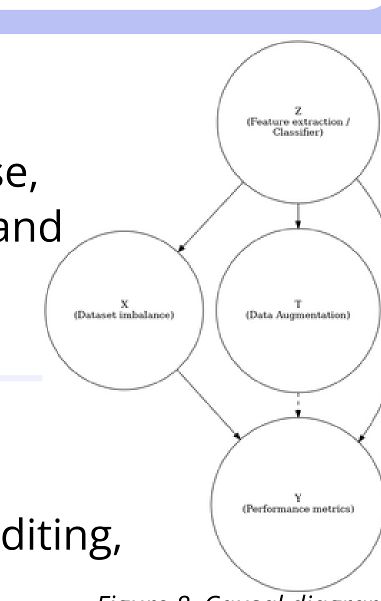


Figure 8. Causal diagram of the study design.

Conclusion

- SMOTE and PCA improved recall and Macro-F1 for minority emotions like fear and disgust, even though overall accuracy decreased slightly.
- Better recognition of subtle emotions can support earlier detection of distress in mental health and foster more inclusive responses in education and counseling.

References

- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
- Ben, Xianye, Yi Ren, Junping Zhang, Su-Jing Wang, Kidiyo Kpalma, Weixiao Meng, and Yong-Jin Liu. (2021). Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. IEEE transactions on pattern analysis and machine intelligence, 44(9), 5826-5846.

Additional references on deep learning (He et al., 2016; Zhao & Li, 2019), fairness and ethics (Mohammad, 2022; Mattioli & Cabitza, 2024), and recent advances in micro-expression recognition (Younis et al., 2024; Zhang et al., 2025) are cited in the full manuscript.

Researcher Yuxuan Huang **instruction professor** Prof. Luyao Zhang **Email address** yh368@duke.edu

Timeline Aug–Oct 2025 | STATS 201