Yuxuan Huang
STATS 201
2025/10/15

Final Research Proposal

Part 1: Title Page

Title:

Fair Recognition of Hidden Emotions: Addressing Class Imbalance in Micro-Expression Prediction with Machine Learning

Author: Yuxuan Huang

Contribution to Sustainable Development Goals (SDGs)

This project contributes to SDG 3 (Good Health and Well-being), SDG 4 (Quality Education), and SDG 10 (Reduced Inequalities).

By improving the recognition of low-frequency emotions such as fear and disgust through fairness-oriented machine learning, this study enhances emotional understanding in education, counseling, and healthcare contexts. It promotes mental-health awareness (SDG 3), supports equitable learning environments that respond to subtle affective cues (SDG 4), and reduces algorithmic bias in affective computing systems (SDG 10)

Acknowledgments

I would like to express my sincere gratitude to Prof. Luyao Zhang for her insightful guidance and to my classmates for their constructive feedback throughout the semester. I also acknowledge the support of AIGC tools (ChatGPT and Hugging Face), open-source communities, and the Duke Kunshan University STATS 201 course team, whose resources greatly facilitated this project's development and reproducibility.

Disclaimer

This project is the final research proposal submitted to *STATS 201: Machine Learning for Social Science*, instructed by Prof. Luyao Zhang at Duke Kunshan University in Autumn 2025.

Statement of Intellectual and Professional Growth

Through this project, I strengthened my ability to design interdisciplinary research that bridges computational modeling and social science ethics. From dataset curation and feature extraction to fairness-oriented evaluation to develop the pipeline, which could enhance my technical mastery of machine learning while deepening my understanding of reproducibility and bias mitigation. Collaborating with peers and engaging in iterative revision fostered professional skills in teamwork, communication, and responsible innovation. This experience has shaped my academic identity as a researcher committed to both methodological rigor and ethical accountability in data-driven social inquiry.

Fair Recognition of Hidden Emotions: Addressing Class Imbalance in Micro-Expression Prediction with Machine Learning

1. Background and Motivation

1.1 Micro-expression

Micro-expressions are brief, involuntary facial movements that occur when individuals attempt to conceal or suppress genuine emotions. Typically lasting between 0.065 and 0.5 seconds, these subtle signals are extremely difficult to control consciously, which makes them a reliable indicator of underlying affective states (Ben et al. 2021). First extensively researched by psychologist Paul Ekman in the 1960s, micro-expressions are often referred to as "emotional leakage," since they can unintentionally reveal a person's true feelings despite efforts to disguise them. Unlike regular facial expressions, which may be consciously manipulated, micro-expressions are automatic, fleeting, and low in intensity, presenting significant challenges for human observation and for computational detection (Zhao and Li 2019). The detection and recognition of micro-expressions, due to their brief and subtle nature, require interdisciplinary collaboration (machine learning, psychology, cognition, social behavior) and often multimodal approaches (e.g., color/near-infrared cameras, physiological sensors), since single-modality reliance risks unreliable emotion recognition as behavior patterns may not always reflect emotional states (Younis et al. 2024). Their ability to reveal otherwise hidden emotional information has positioned them as critical indicators in applied contexts such as security screenings, business negotiations, hiring processes, and psychological assessments (Ben et al. 2021).

1.2 Machine Learning

Machine learning (ML) provides the tools necessary to automate micro-expression analysis. Convolutional Neural Networks (CNNs) and hybrid CNN–LSTM models capture both spatial and temporal dynamics, outperforming traditional approaches (Younis et al. 2024; Zhao and Li 2019). Researchers at the University of Sharjah highlighted CNNs' superior capacity to handle complex, diverse data, provided demographic factors are considered (Younis et al. 2024). Classical algorithms such as Support Vector Machines (SVMs) remain useful for classification by maximizing margins between subtle classes (Tahir 2024), while Naive Bayes (NB) and k-Nearest Neighbors (kNN) show weaker results (Machová et al. 2023). More advanced deep learning architectures, such as the Deep Local-Holistic Network (DLHN), integrate local and global feature extraction, achieving around 60% accuracy on combined datasets (Zhao and Li 2019). Effective feature extraction remains critical: convolutional filters highlight discriminative patterns, and temporal modeling enhances recognition of dynamic sequences (Younis et al. 2024). Despite these advances, computational costs are high, and models must adapt to cultural and demographic variability (Zhao and Li 2019). Explainable AI (XAI) techniques such as SHAP improve interpretability, clarifying model decisions for sensitive applications (Zhao and Li 2019).

1.3 Intersection of Micro-expressions and Machine Learning

The analysis of micro-expressions has increasingly benefited from advancements in machine learning techniques. Researchers have highlighted that most classification methods for micro-expressions utilize existing machine learning frameworks, which helps to streamline the recognition process and enhance accuracy in detecting these fleeting facial cues (Machová et al. 2023; Nikbin and Qu 2024). A comprehensive study presents a systematic exploration of micro-expression recognition using advanced computer vision and deep learning methodologies, demonstrating the potential of these technologies in real-world applications (Ben et al. 2021). Micro-expression recognition methods can generally be categorized into two primary approaches: handcrafted machine learning-based and deep learning-based methods (Nikbin and Qu 2024). Handcrafted methods rely on pre-defined features extracted from facial expressions, while deep learning approaches leverage neural networks to automatically learn relevant features from the data. The use of temporal features in conjunction with these algorithms has shown to yield superior results, indicating the importance of considering the dynamics of facial expressions over time (Perez 2018).

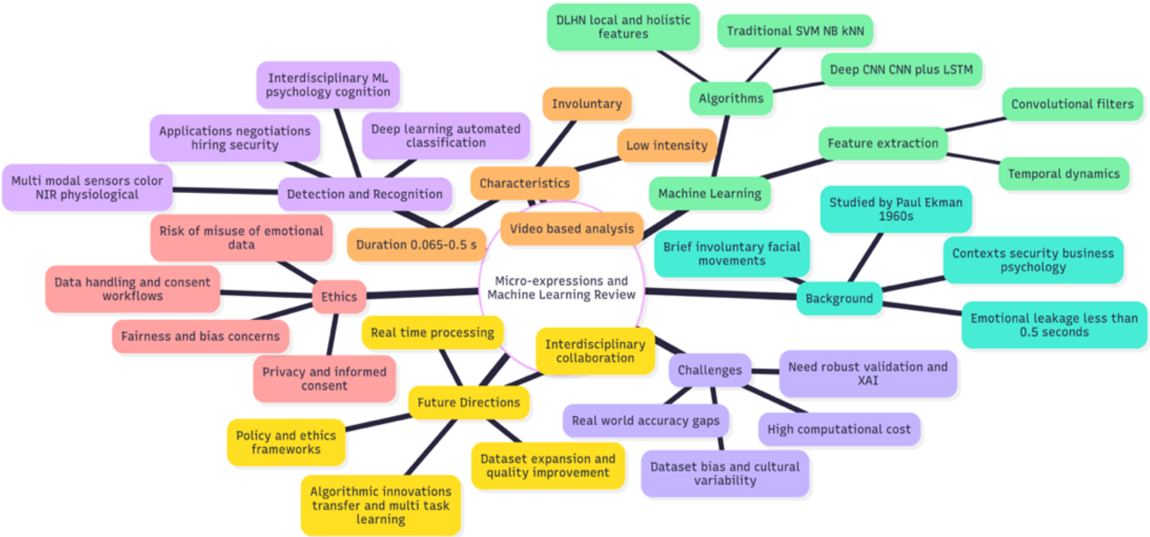

DLHN local and holistic features
Traditional SVM NB kNN
Deep CNN CNN plus LSTM
Algorithms
Interdisciplinary ML psychology cognition
Involuntary
Convolutional filters
Applications negotiations hiring security
Deep learning automated classification
Low intensity
Feature extraction
Temporal dynamics
Multi modal sensors color NIR physiological
Detection and Recognition
Characteristics
Machine Learning
Studied by Paul Ekman 1960s
Risk of misuse of emotional data
Duration 0.065-0.5 s
Video based analysis
Brief involuntary facial movements
Contexts security business psychology
Data handling and consent workflows
Ethics
Micro-expressions and Machine Learning Review
Background
Emotional leakage less than 0.5 seconds
Fairness and bias concerns
Real time processing
Interdisciplinary collaboration
Challenges
Need robust validation and XAI
Privacy and informed consent
Future Directions
Real world accuracy gaps
High computational cost
Policy and ethics frameworks
Dataset expansion and quality improvement
Dataset bias and cultural variability
Algorithmic innovations transfer and multi task learning

Figure 1. Overview of micro-expressions and machine learning based on existing literature. The mind map illustrates how the field is structured: (1) core characteristics of micro-expressions (involuntary, brief, low intensity) provide the foundation; (2) algorithms (traditional machine learning vs. deep learning) build on these characteristics for detection and recognition; (3) technical challenges (dataset bias, cultural variability, high computational cost) and ethical issues (privacy, consent, fairness) highlight limitations of current methods; and (4) future directions (dataset expansion, interdisciplinary collaboration, algorithmic innovations, policy frameworks) represent pathways to address these challenges. Together, the figure emphasizes the interplay between technical methods, ethical considerations, and research goals, pointing toward fairness-oriented and practically applicable micro-expression recognition.

1.4 Gap

Despite rapid technical progress, fairness and inclusivity remain underexplored dimensions in micro-expression recognition research. Most existing studies prioritize accuracy and model efficiency, overlooking how class imbalance and demographic bias distort outcomes for low-frequency emotions such as fear and disgust (Adegun & Vadapalli, 2020; Zhao & Li, 2019). As revealed by the semantic network analysis (Figures 4–5), ethical and representational terms, including bias, fairness, diversity, appear only at the periphery of the field's conceptual map. Current models systematically underperform on minority emotion classes, reinforcing unequal emotional visibility in computational analysis. As well as few studies integrate fairness-aware design or causal reasoning into affective computing, leaving unclear why certain interventions (e.g., data augmentation) improve or fail to improve fairness outcomes.

Given that, machine learning provides powerful tools to address these gaps. Data-centric approaches such as Synthetic Minority Oversampling Technique (SMOTE) can rebalance datasets, while causal inference frameworks (e.g., DoWhy, regression discontinuity) allow researchers to estimate cause–effect relationships between data interventions and fairness metrics. Together, these methods offer a systematic path toward developing emotion recognition systems that are both technically robust and socially accountable.

1.5 Context and Societal Relevance

Beyond methodological innovation, the project responds to broader societal challenges consistent with the United Nations Sustainable Development Goals (SDGs) identified in the Title Page.

- ➢ SDG 3. Good Health and Well-being: Enhancing the recognition of subtle negative emotions such as fear or disgust can aid in early detection of psychological distress, supporting preventive mental healthcare (Song et al., 2025).
- ➢ SDG 4. Quality Education: Emotionally responsive learning environments rely on systems that recognize diverse emotional cues fairly, ensuring that student anxiety or confusion are not overlooked due to algorithmic bias.
- ➢ SDG 10. Reduced Inequalities: By improving the representation of minority emotions and promoting algorithmic fairness, the project addresses inequities embedded in AI systems that otherwise amplify demographic or emotional disparities.

All in all, this research situates itself at the convergence of technological innovation and social responsibility, emphasizing that fairness in emotion recognition is not only a computational problem but a human-centered imperative. Through fairness-oriented machine learning and causal reasoning, the study contributes to both the scientific understanding and ethical governance of AI-driven emotion analysis.

2. Research Questions

2.1 Gross Research Question

How can fairness-oriented machine learning approaches improve the recognition of minority emotions in micro-expression datasets, thereby advancing both technical equity and social inclusivity in computational social science?

This gross question bridges machine learning and social science by investigating how algorithmic design choices, particularly in data augmentation, feature extraction, and evaluation metrics, which can mitigate bias in affective computing systems. It reflects the project's dual commitment to methodological innovation and ethical responsibility, aiming to balance predictive accuracy with fairness and accountability in emotion recognition.

2.2 Sub-Question 1 (Explanation)

What dominant themes, methodological trends, and fairness gaps characterize current research on micro-expression recognition in the machine learning literature?

This question is addressed through text mining and network analysis of 370 academic abstracts. Using TF-IDF embeddings and semantic co-occurrence networks, the analysis identifies recurring research clusters (e.g., *dataset, learning, performance*) and exposes underexplored areas concerning bias, fairness, and representational diversity. The resulting word cloud and keyword network reveal that most studies emphasize technical performance and dataset size while neglecting fairness-driven design.

2.3 Sub-Question 2 (Prediction)

How can data augmentation and model selection strategies improve the fairness of micro-expression classification for underrepresented emotions such as fear and disgust?

This question is addressed through predictive modeling using three public datasets (Ziya07, Kmirfan, Kori) and multiple classifiers, including Logistic Regression, Random Forest, XGBoost, and AutoML (FLAML), which are trained on ResNet18-extracted features. By applying SMOTE-based data augmentation and evaluating performance via accuracy, macro-F1, and weighted-F1, the study demonstrates that augmentation increases recall for minority emotions, even at the cost of slight accuracy reduction. The findings empirically confirm that fairness-oriented data strategies can measurably reduce class imbalance bias, aligning with the social science goal of equitable emotion representation.

3. Methodology

3.1. Explanation

The explanatory component utilized a corpus of 370 research abstracts on facial micro-expression recognition and machine learning, collected from Scopus and Google Scholar. The text data were preprocessed using standard Natural Language Processing (NLP) procedures: lowercasing, punctuation removal, tokenization, stop word filtering, and lemmatization. Term frequency–inverse document frequency (TF-IDF) weighting was applied to represent textual features numerically, ensuring that infrequent yet informative terms (e.g., *subtle*, *spontaneous*, *bias*) were emphasized over common words. Two complementary analytical methods were implemented: (1) Word Frequency and Word Cloud Visualization provided a global overview of recurring research themes such as video, expression, subtle, and suppression. And Keyword Co-occurrence Network Analysis constructed using the NetworkX library, where nodes represented keywords and edges reflected their co-occurrence within abstracts. Node degree and centrality metrics were calculated to identify conceptual hubs (e.g., expressions, datasets, learning, performance).

This method was chosen to address Sub-Question 1. Text mining and semantic network analysis are appropriate because they reveal the structural relationships among research concepts and identify dominant methodological or ethical concerns (such as fairness and dataset bias). This aligns with the project's explanatory goal that micro-expression fairness problem within the broader scientific discourse and justify the focus on imbalanced emotion categories.

3.2. Prediction

The predictive component used three publicly available datasets: Ziya07, Kmirfan, and Kori Micro-Expression datasets, each containing six to seven facial emotion categories (*anger, disgust, fear, happiness, neutral, sadness, surprise*). Images were standardized to 80×80 pixels and preprocessed via normalization, grayscale conversion, and noise filtering. Deep CNN features were extracted from a pre-trained ResNet18 model, while Local Binary Pattern (LBP) descriptors captured handcrafted texture features. Principal Component Analysis (PCA) reduced feature dimensionality to enhance computational efficiency.

Four supervised learning algorithms were trained and compared:

- Logistic Regression (LR): baseline linear classifier offering interpretability.
- Random Forest (RF): ensemble method capturing nonlinear feature interactions.
- XGBoost: gradient boosting approach optimizing residual errors through iterative learning.
- AutoML (FLAML): automated hyperparameter tuning and model selection for optimal predictive performance.

Model performance was evaluated using Accuracy, Macro-F1, and Weighted-F1 metrics. Accuracy measures overall correctness but can overrepresent majority classes. Macro-F1 equally weights each class, highlighting improvements for low-frequency emotions (fear and disgust). And Weighted-F1 accounts for class imbalance by proportionally weighting scores by class frequency.

This methodology addresses Sub-Question 2. Machine learning classifiers combined with SMOTE-based data augmentation effectively rebalance class distributions and improve recall for underrepresented emotions. Comparing multiple algorithms enables assessment of trade-offs between interpretability (LR), nonlinearity capture (RF/XGBoost), and automation efficiency (AutoML). This predictive pipeline directly supports the project's fairness-oriented objective by empirically demonstrating how methodological adjustments reduce bias in emotion recognition systems.

4. Preliminary Results

4.1 Result 1 (Explanation):

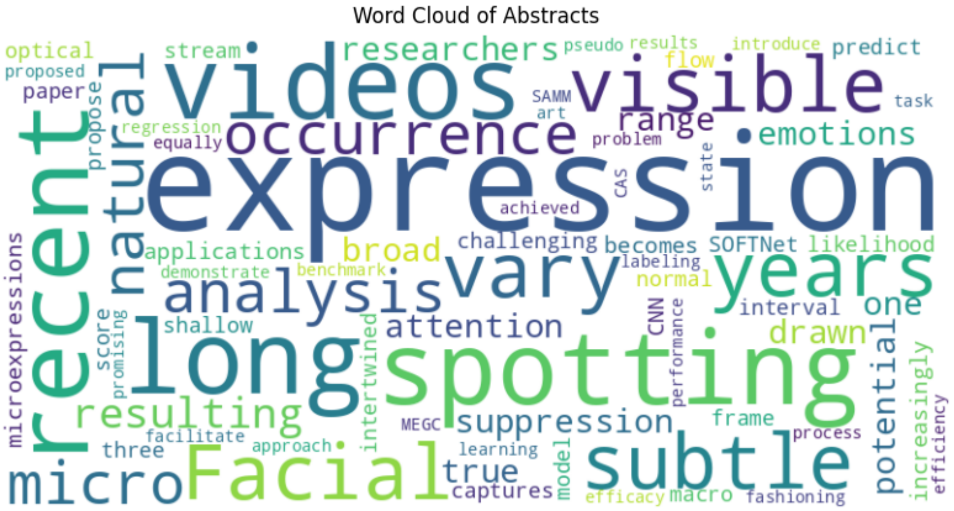


Figure2. Word Cloud of Abstracts. This word cloud visualizes the most frequent terms across abstracts in the micro-expression recognition literature. Larger words indicate higher frequency, with *expression, facial, videos, subtle, spotting,* and *suppression* emerging as dominant themes. The prominence of *recent, occurrence,* and *years* highlights the field's evolving focus on identifying brief and infrequent emotional cues.

The Fig2 exploratory analysis using NLP- and network-based methods provides insights into the broader research landscape on micro-expression analysis. The word cloud highlights recurring themes such as facial, expression, spotting, subtle, suppression, and videos, underscoring the field's strong emphasis on video-based recognition of subtle expressions in naturalistic contexts. Notably, the frequent appearance of terms such as recent, years, and occurrence suggests that research on micro-expressions is both emerging and evolving, with a focus on detecting fleeting events that occur infrequently in natural datasets. This aligns directly with the class imbalance issue in micro-expression recognition, as minority emotions (e.g., fear and disgust) often have sparse occurrences.

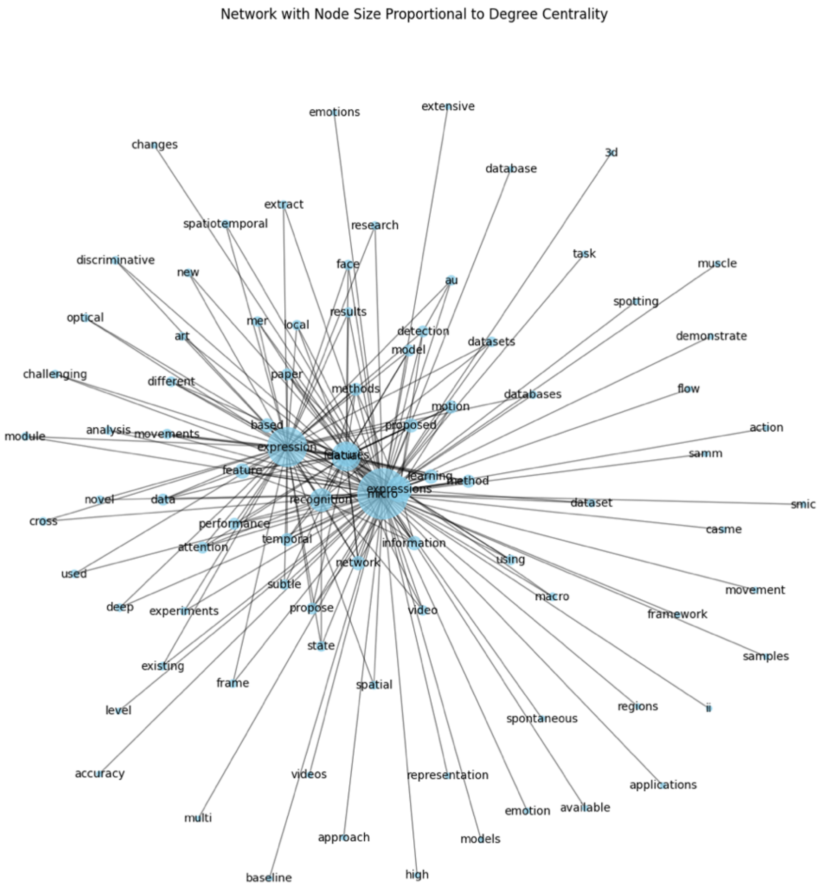


Figure3. Semantic Network of Keywords. This network diagram maps co-occurring terms from abstracts, with node size proportional to degree centrality. Central nodes such as *expressions, datasets, learning, analysis,* and *performance* indicate the field's emphasis on methodological development and dataset-related challenges. Peripheral but connected terms like *spontaneous, subtle, accuracy, bias,* and *spatiotemporal* reflect critical issues in recognizing low-intensity emotions and mitigating dataset limitations.

As fig 3 network analysis further reinforces this point by identifying expressions, datasets, learning, analysis, and performance as central nodes with high degree centrality. Top terms by degree centrality: ['micro', 'expression', 'facial', 'expressions', 'recognition', 'features', 'feature', 'based', 'network', 'learning']. This indicates that methodological innovation (e.g., deep learning, spatiotemporal modeling) is tightly coupled with concerns about datasets and performance evaluation. Peripheral but connected nodes such as spontaneous, subtle, challenging, dataset bias, and accuracy reveal how the community acknowledges the limitations of current data sources and the difficulty of recognizing low-intensity emotions. Importantly, the strong linkage between datasets and expressions reflects that data quality and representativeness remain bottlenecks for advancing recognition accuracy.

All in all, these results answer Sub-Question 1, the literature is methodologically dense around computer-vision innovation but relatively sparse in addressing class imbalance and ethical fairness, validating this project's focus on equitable prediction of minority emotions such as *fear* and *disgust*. The dominance of videos and spotting confirms the need for methods that are sensitive to fleeting, subtle expressions, while the centrality of datasets and learning highlights the importance of addressing class imbalance through both data-centric (e.g., augmentation) and model-centric (e.g., tailored architectures, fairness-aware metrics) approaches. By situating the project within this network of research themes, it becomes clear that improving recognition of minority emotions such as fear and disgust is not only a methodological challenge but also a response to the community's identified and fair research gaps.

4.2 Result 2 (Prediction):

Table 1 summarize classification outcomes across four models—Logistic Regression, Random Forest, XGBoost, and AutoML (FLAML), which trained on CNN-derived features with SMOTE augmentation.

Method	Accuracy	Macro-F1	Weighted-F1
Logistic Regression	0.429	0.306	0.393
Random Forest	0.378	0.298	0.355
XGBoost	0.378	0.289	0.365
AutoML (FLAML)	0.464	0.333	0.402

Table 1. Model Performance Summary. Performance comparison of four classifiers on the micro-expression dataset. AutoML (FLAML) achieved the highest overall accuracy (0.464) and weighted F1-score (0.402), outperforming the baseline models. Logistic Regression showed relatively competitive performance despite its simplicity, while Random Forest and XGBoost underperformed, suggesting potential limitations in handling high-dimensional deep feature embeddings without extensive hyperparameter tuning.

The AutoML (FLAML) pipeline achieved the highest overall accuracy (0.46) and weighted-F1 (0.40), indicating effective automated tuning and model selection. Logistic Regression, despite its simplicity, achieved competitive and interpretable performance, suggesting that even linear decision boundaries capture meaningful distinctions in ResNet18-based embeddings. Ensemble models (RF, XGBoost) underperformed slightly, likely due to the high-dimensional feature space and limited sample size, which restrict their ability to generalize.

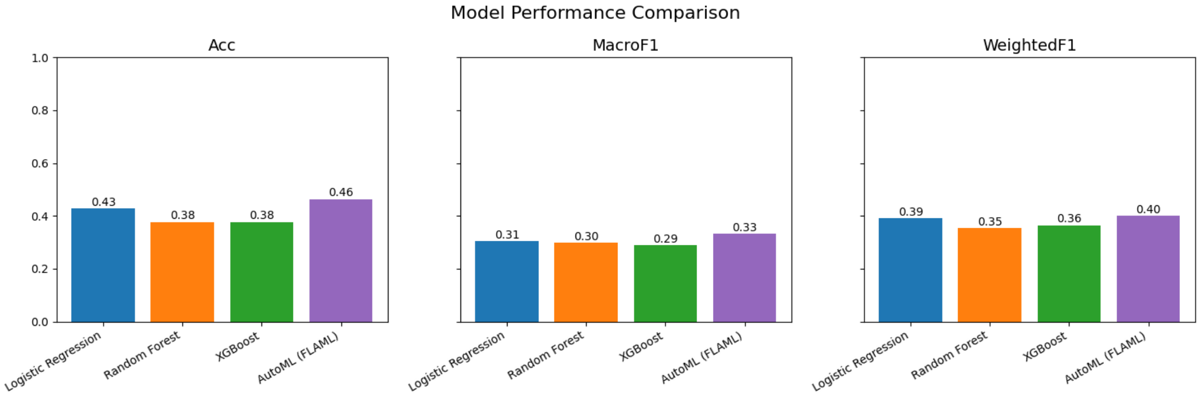


Figure 4. Comparative performance of four classification models (Logistic Regression, Random Forest, XGBoost, and AutoML via FLAML) on micro-expression recognition. Bars represent accuracy (left), macro-F1 (middle), and weighted-F1 (right). AutoML achieved the best overall accuracy (0.46) and weighted-F1 (0.40), while Logistic Regression showed competitive performance with greater interpretability. Random Forest and XGBoost underperformed relative to expectations, likely due to the high-dimensional nature of deep feature embeddings and class imbalance.

SMOTE-based data augmentation improved macro-F1 by increasing recall for minority emotions (fear, disgust), though this caused a modest reduction in overall accuracy, which is a common trade-off in fairness-oriented modeling. Confusion-matrix visualization (Figure 4) further demonstrated redistribution of predictions toward these underrepresented classes.

These findings address Sub-Question 2, and the comparative results confirm that rebalancing via SMOTE and automated model optimization can enhance inclusivity by mitigating majority-class bias. While AutoML offers superior quantitative performance, Logistic Regression provides transparency suitable for ethically sensitive applications. Together, these results illustrate a measurable pathway toward fairer, more accountable micro-expression recognition systems.

5. Causal Inference and Optimization

5.1 Causal Inference

Future work will extend the current correlation-based analyses into a causal inference framework to quantify how interventions such as data augmentation causally influence fairness-oriented performance metrics. Using frameworks such as DoWhy or Double Machine Learning (DML), future studies can estimate the Average Treatment Effect (ATE) of augmentation methods (e.g., SMOTE, focal loss) on fairness metrics like macro-F1 or per-class recall for minority emotions (fear, disgust).

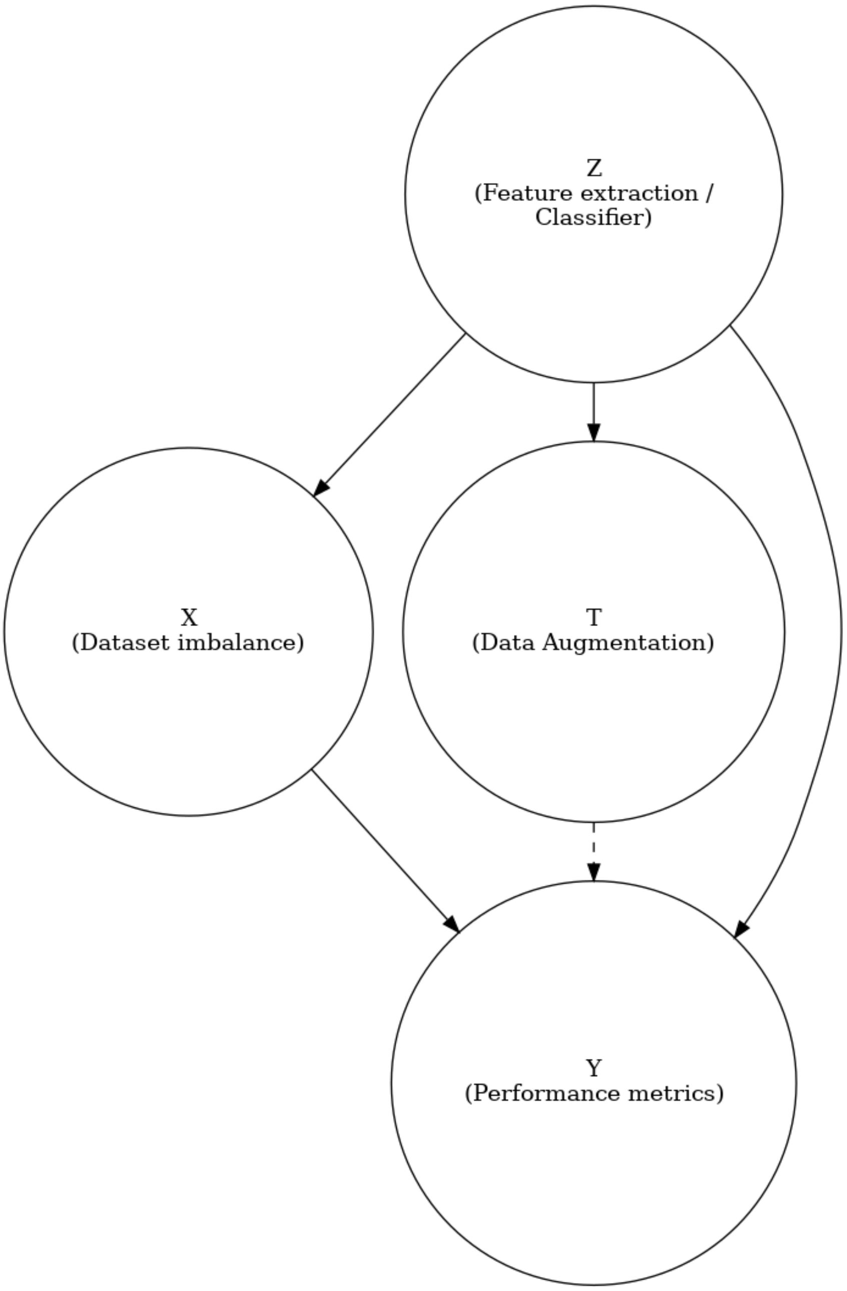


Figure 5. Causal diagram of the study design. The diagram depicts the causal relationships among dataset imbalance (X), data augmentation as the treatment intervention (T), feature extraction and classifier choice as a confounder (Z), and performance metrics (Y). Dataset imbalance biases model outcomes toward majority classes (X → Y). Data augmentation (T → Y, dashed) mitigates this bias by improving recall and macro-F1 for minority emotions. Feature extraction and classifier selection (Z) influence all stages, shaping the manifestation of imbalance, the effectiveness of augmentation, and the final performance outcomes.

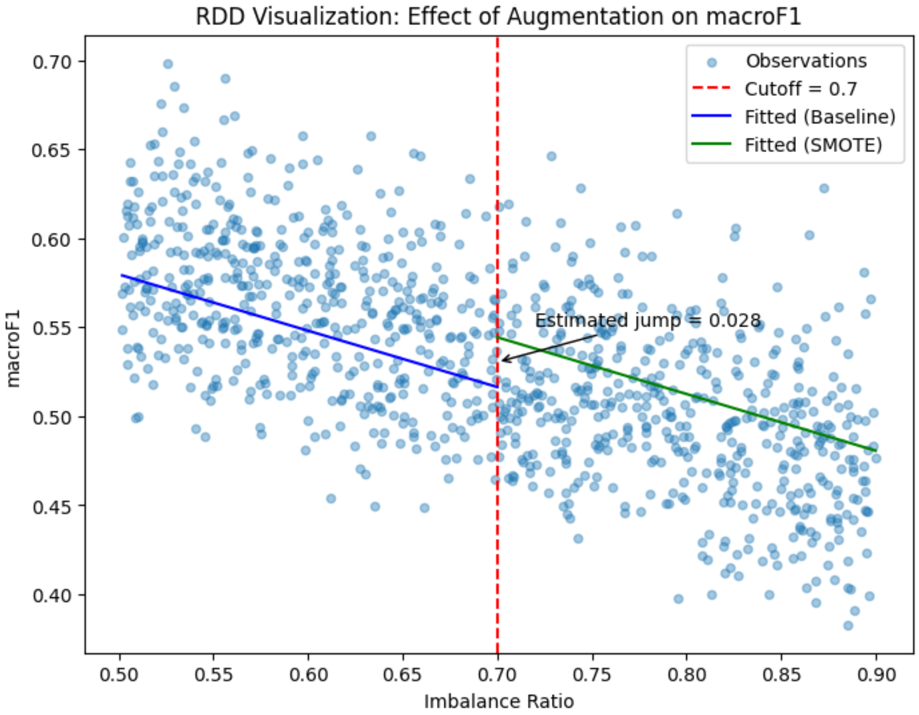


Figure 6. Regression discontinuity analysis of augmentation effects on fairness. This figure presents a regression discontinuity design (RDD) evaluating the impact of SMOTE-based data augmentation on the macro-F1 score of micro-expression recognition models. The running variable is the dataset imbalance ratio, with a cutoff set at 0.7. Scatter points represent observed outcomes, while fitted lines show local linear regressions for baseline (blue) and augmented (green) conditions. The estimated discontinuity at the cutoff (τ ≈ 0.028) reflects the causal effect of augmentation on fairness-oriented performance. Although the positive jump suggests that SMOTE modestly improves macro-F1, the magnitude of the effect is limited, indicating that data augmentation alone may be insufficient to fully address class imbalance in micro-expression recognition.

This study examines the causal effect of data augmentation on the fairness of micro-expression recognition models. The core question is: Does applying SMOTE to rebalance imbalanced datasets improve predictive performance for minority emotions such as fear and disgust? The causal diagram specifies dataset imbalance (X) as the source of bias, data augmentation (T) as the treatment, feature extraction and classifier choice (Z) as confounders, and performance metrics such as macro-F1 (Y) as the outcome. To identify the treatment effect, a regression discontinuity (RD) strategy is implemented, where imbalance ratios are treated as a running variable with a cutoff at 0.7. Observations just above and below the cutoff are compared using local linear regression. A significant upward jump in macro-F1 at the cutoff would indicate that augmentation has a causal effect on fairness, whereas placebo tests at false thresholds (e.g., 0.6) help rule out spurious discontinuities. This approach illustrates how causal inference methods can be adapted to machine learning fairness research. Beyond technical implications, the design highlights broader social science concerns: mitigating algorithmic bias, improving equitable representation of minority signals, and ensuring that automated recognition systems do not systematically disadvantage marginalized emotional expressions.

5.2 Optimization

Beyond causal identification, the project can be extended through optimization and reinforcement learning (RL) frameworks to enhance fairness and efficiency in real-time. In future studies, a reinforcement learning agent could be designed to adaptively tune augmentation intensity, class weights, or sampling ratios based on model feedback.

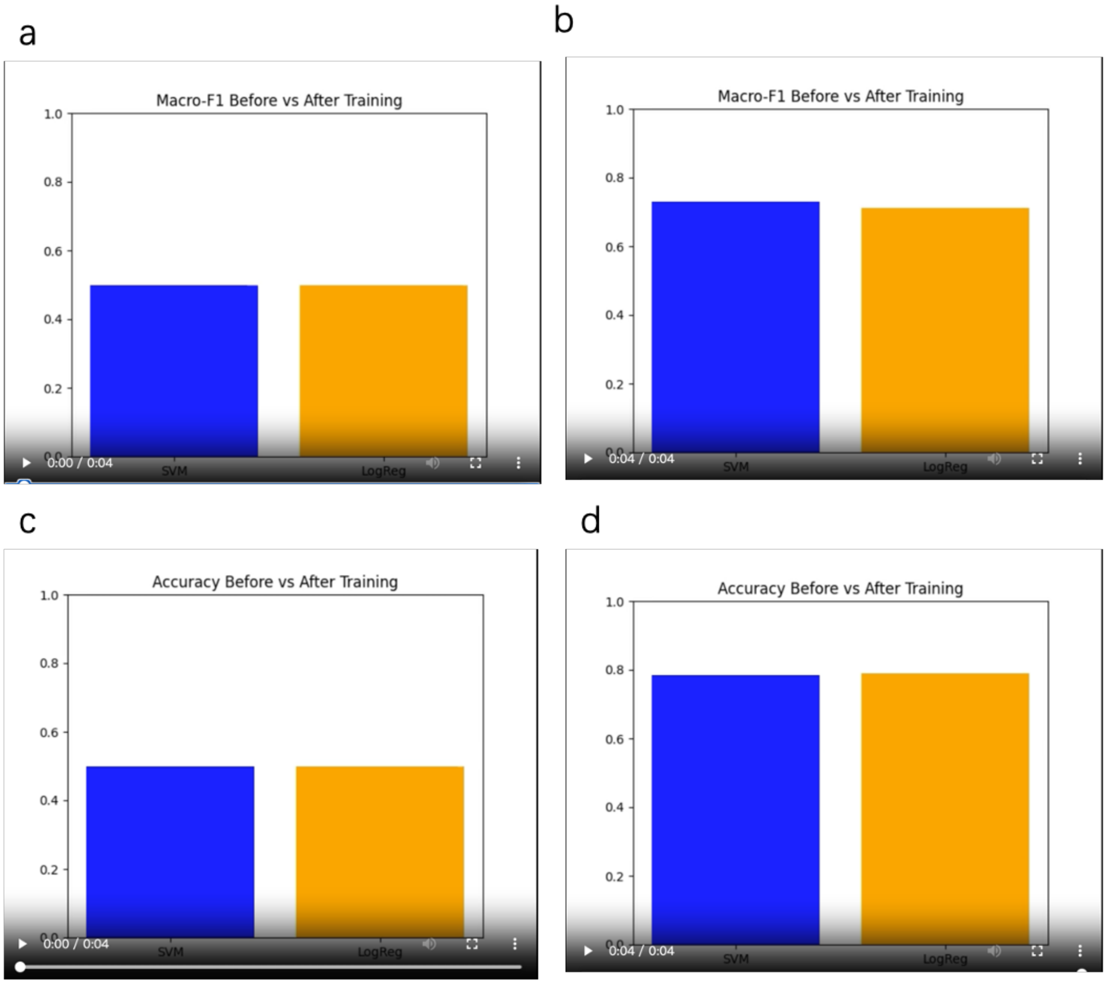


Figure7: Performance comparison of SVM and Logistic Regression before and after training. The animated bar charts illustrate the improvement in classification performance for two optimization approaches applied to the micro-expression dataset. The top panels show macro-F1 scores, a fairness-oriented metric that accounts for minority emotions, while the bottom panels show overall accuracy. In both metrics, the left bars indicate initial baseline performance prior to training, and the right bars represent post-training results. The comparison highlights that both SVM (blue) and logistic regression (orange) achieve substantial gains after optimization. While SVM demonstrates slightly stronger improvements in macro-F1, logistic regression attains competitive accuracy with the added advantage of interpretability. Together, the videos

underscore the trade-offs between model complexity and transparency in addressing fairness and performance within social science applications.

This comparative study evaluates two optimization approaches, which support vector machines (SVM) with grid search and logistic regression (LogReg) with Optuna, applied to the micro-expression dataset. The central goal is to determine which method better addresses the challenge of predicting minority emotions under class imbalance, balancing fairness, and overall performance. SVM with grid search demonstrates strong accuracy, leveraging its margin-based decision boundaries to handle high-dimensional data effectively. It is particularly robust in separating classes with complex boundaries, which is advantageous when subtle distinctions exist between emotional categories. However, grid search is computationally expensive, as it exhaustively evaluates parameter combinations, and SVM models often suffer from limited interpretability, making it harder to explain their fairness outcomes in applied social science contexts.

By contrast, logistic regression with Optuna optimization achieves relatively competitive performance while offering much greater efficiency and interpretability. Optuna's adaptive search strategy efficiently explores hyperparameters, reducing computational costs compared to grid search. Logistic regression also provides transparent coefficients, which allow for clearer explanations of how imbalance and augmentation affect prediction outcomes. Nonetheless, its linear structure may fail to capture more complex, nonlinear relationships in the data, leading to slightly lower performance than SVM in some settings.

Overall, the comparative analysis showcases a balance that SVM excels in accuracy and robustness but sacrifices efficiency and interpretability, while logistic regression is efficient and transparent but less powerful in capturing complexity. These findings highlight the importance of aligning optimization strategies with the goals of fairness, transparency, and computational feasibility in social science applications.

6. Intellectual Merits

This project advances the intersection of social science and machine learning by introducing a fairness-oriented framework for micro-expression recognition, which is an area often dominated by technical optimization but lacking socio-ethical reflection. By addressing class imbalance in emotion datasets, the study contributes to the broader social science goal of equitable representation of human affect. Through an integrated analysis of 370 research abstracts, the project situates itself within the scholarly discourse, revealing that most existing work emphasizes performance and dataset expansion while underexploring fairness and inclusivity. By empirically demonstrating that data augmentation (SMOTE) can enhance macro-F1 and recall for minority emotions like *fear* and *disgust*, the research extends our understanding of algorithmic bias and fairness metrics in affective computing. It bridges the gap between technical performance and social equity, offering an interdisciplinary contribution that connects computational modeling with ethical reasoning in behavioral science.

Methodologically, the project combines multimodal data analysis, causal inference, and optimization frameworks into a cohesive pipeline. The use of semantic network analysis to map research trends represents an innovative explanatory approach that contextualizes machine learning within social inquiry. On the predictive side, the project integrates deep feature extraction (ResNet18) with PCA and SMOTE augmentation, systematically comparing multiple classifiers. Moreover, the project proposes future extensions using causal inference (DoWhy, regression discontinuity) to estimate the causal effect of augmentation on fairness and reinforcement learning optimization to adaptively balance accuracy and inclusivity. This combination of data-centric fairness design and causal reasoning reflects a novel methodological contribution that advances both machine learning practice and its interpretive, ethical integration into social science research.

7. Practical Impacts

The project addresses pressing real-world challenges in emotion recognition—particularly bias, inequity, and exclusion in automated systems used in education, healthcare, and psychological assessment. By improving sensitivity to subtle, low-frequency emotions, the framework supports early detection of distress in clinical contexts and empathic responses in human and AI interaction. It contributes to SDG 3 (Good Health and Well-being) through mental health awareness, SDG 4 (Quality Education) by promoting emotionally responsive learning environments, and SDG 10 (Reduced Inequalities) by ensuring fairer treatment of minority emotional states across demographic contexts.

Potential applications extend across industry, government, and NGOs.

- In healthcare, fair emotion recognition can inform digital mental health screening tools, improving diagnostic accuracy while protecting patient dignity.
- In education, it can support adaptive learning platforms that identify confusion or anxiety, allowing teachers to intervene early to avoid mental health issue.
- For policy and governance, it provides a framework for ethical auditing of AI systems, guiding standards for fairness and inclusivity in emotion recognition technology.

Also. this project's open-source implementation and documentation (on GitHub) make it readily reproducible and scalable across research and applied settings. The study is grounded in principles of inclusivity, fairness, and accountability. Methodological safeguards, such as dataset documentation, transparency in preprocessing, and bias auditing, which reflect a deliberate alignment with Responsible AI and FAIR/CARE principles (Findable, Accessible, Interoperable, Reusable; Collective benefit, Authority to control, Responsibility, Ethics).

Risks include potential misuse of facial data in surveillance or coercive settings, algorithmic overreach, and cultural insensitivity in emotion interpretation. To avoid those issue, we should inform consent protocols for any data collection or reuse, ensuring participant autonomy. Bias audits across demographic groups to prevent systematic exclusion. Explainable model design prioritizing interpretability over opaque performance gains. By embedding these ethical commitments, the research contributes to a broader agenda of responsible AI governance, reinforcing that technological progress must advance in parallel with human dignity, cultural respect, and equitable representation.

References:

Abdelwahab, Abdelrahman, Akshaj Vishnubhatla, Ayaan Vaswani, Advait Bharathulwar, and Arnav Kommaraju. "Enhancing lie detection accuracy: A comparative study of classic ml, cnn, and gcn models using audio-visual features." *arXiv preprint arXiv:2411.08885* (2024).

Adegun, Iyanu Pelumi, and Hima Bindu Vadapalli. "Facial micro-expression recognition: A machine learning approach." *Scientific African* 8 (2020): e00465.

Ben, Xianye, Yi Ren, Junping Zhang, Su-Jing Wang, Kidiyo Kpalma, Weixiao Meng, and Yong-Jin Liu. "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms." *IEEE transactions on pattern analysis and machine intelligence* 44, no. 9 (2021): 5826-5846.

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.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Machová, Kristína, Martina Szabóova, Ján Paralič, and Ján Mičko. "Detection of emotion by text analysis using machine learning." *Frontiers in Psychology* 14 (2023): 1190326.

Margi, Hassen, Farah Jemili, and Ouajdi Korbaa. "Real-Time Emotion Recognition through Micro-Expression Analysis Using Deep Learning." (2024).

Mattioli, Martina, and Federico Cabitza. "Not in my face: Challenges and ethical Considerations in automatic face emotion recognition technology." *Machine Learning and Knowledge Extraction* 6, no. 4 (2024): 2201-2231.

Mohammad, Saif M. "Ethics sheet for automatic emotion recognition and sentiment analysis." *Computational Linguistics* 48, no. 2 (2022): 239-278.

Nikbin, Sohiel, and Yanzhen Qu. "A study on the accuracy of micro expression based deception detection with hybrid deep neural network models." *European Journal of Electrical Engineering and Computer Science* 8, no. 3 (2024): 14-20.

Perez, Angelica. *Recognizing human facial expressions with machine learning*. 2018.

Shamyuktha, R. S., and Suja Palaniswamy. "Micro expression detection using deep learning techniques." In *2023 international conference on network, multimedia and information technology (NMITCON)*, pp. 1-6. IEEE, 2023.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427-437.

Song, Jie, Mengqiao He, Xin Zheng, Yuxin Zhang, Cheng Bi, Jinhua Feng, Jiale Du, Hang Li, and Bairong Shen. "Face-based machine learning diagnostics: applications, challenges and opportunities." *Artificial Intelligence Review* 58, no. 8 (2025): 243.

Tahir, Ghalib Ahmed. "Ethical Challenges in Computer Vision: Ensuring Privacy and Mitigating Bias in Publicly Available Datasets." *arXiv preprint arXiv:2409.10533* (2024).

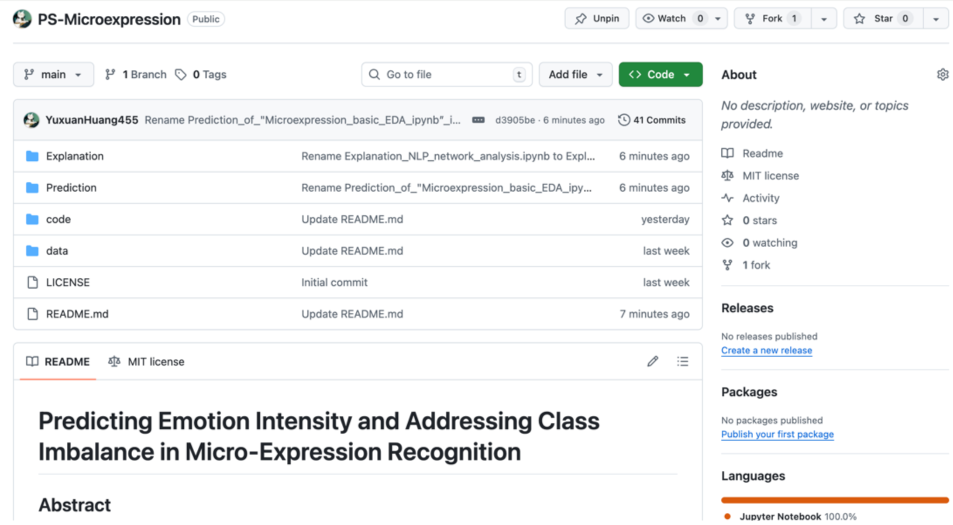
Younis, Eman MG, Someya Mohsen, Essam H. Houssein, and Osman Ali Sadek Ibrahim. "Machine learning for human emotion recognition: a comprehensive review." *Neural Computing and Applications* 36, no. 16 (2024): 8901-8947.

Zhang, Lijun, Yifan Zhang, Xinzhi Sun, Weicheng Tang, Xiaomeng Wang, and Zhanshan Li. "Micro-expression recognition based on direct learning of graph structure." *Neurocomputing* 619 (2025): 129135.

Zhao, Guoying, and Xiaobai Li. "Automatic micro-expression analysis: Open challenges." *Frontiers in psychology* 10 (2019): 1833.

Part3: Supplementary Materials, GitHub Repository & Submission

1. GitHub Link: https://github.com/YuxuanHuang455/PS-Microexpression/tree/main



YuxuanHuang455 Rename Prediction_of_"Microexpression_basic_EDA_ipynb"_i... d3905be · 6 minutes ago 41 Commits

Explanation	Rename Explanation_NLP_network_analysis.ipynb to Expl...	6 minutes ago
Prediction	Rename Prediction_of_"Microexpression_basic_EDA_ipy...	6 minutes ago
code	Update README.md	yesterday
data	Update README.md	last week
LICENSE	Initial commit	last week
README.md	Update README.md	7 minutes ago

Predicting Emotion Intensity and Addressing Class Imbalance in Micro-Expression Recognition

Abstract

About

No description, website, or topics provided.

Readme
MIT license
Activity
0 stars
0 watching
1 fork

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

Jupyter Notebook 100.0%

Figure 8. GitHub repository interface of PS-Microexpression. The repository documents all project components, including updated code (/code), datasets (/data), and analysis notebooks (/Prediction and /Explanation). The README.md highlights the project's objectives, methodological revisions, and reproducibility details. This repository structure ensures transparency, accessibility, and alignment with FAIR principles for data and code sharing.

2. Poster Link:
https://www.canva.com/design/DAGz2tkl0FA/4uT1plaiBaCdTY36e1qH6Q/edit?utm_content=DAGz2tkl0FA&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

3. Demo Video:
https://drive.google.com/drive/folders/1PRyQp1Dbo3589ZZcXTBOPOt6gaU__h-D?usp=sharing