Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Respiratory Feature Analysis for Micro-Expression Detection: Multimodal Approach on SAMM Dataset

Li Wang

willi.lee@njfu.edu.cn

Nanjing Forestry University

Ming Yin

Jiangsu Police Institute

Fubin Zhuo

Jiangsu Xiaowei Digital Technology Co., Ltd.

Xianzheng Lu

Jiangsu Xiaowei Digital Technology Co., Ltd.

Qing Shi

Jiangsu Xiaowei Digital Technology Co., Ltd.

Jianxin Zhang

Jiangnan University

Article

Keywords: Micro-expression detection, Respiratory features, Multimodal fusion, Affective computing, SAMM dataset

Posted Date: November 7th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7819903/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Micro-expression detection faces significant challenges due to the fleeting and subtle nature of facial movements. Current approaches rely almost exclusively on visual features, yet recent findings suggest physiological signals might offer valuable complementary information for emotion recognition. This paper presents what we believe is the first systematic investigation of respiratory features for micro-expression recognition using the SAMM Long Videos dataset. We develop a non-contact respiratory signal extraction methodology and propose a 37-dimensional respiratory feature framework. In addition, we release the first open-source respiratory dataset for micro-expression research, providing synchronized respiratory signals and micro-expression annotations. Our experimental results show that respiratory features alone outperform traditional video-only baselines by 8.2% (macro-F1), while multimodal fusion of respiratory and video features yields further improvements of 17.0% over video-only methods. These findings indicate that respiratory patterns contain previously unexplored discriminative information for micro-expression detection, potentially establishing a new research direction in physiological affective computing.

1. Introduction

Micro-expressions are fleeting, involuntary facial movements that unveil genuine emotions and serve as reliable indicators of true feelings [1]. In contrast to macro-expressions, which are typically deliberate and easily noticeable, micro-expressions occur within 1/25 to 1/5 of a second and involve minimal facial muscle activity [2]. Distinguishing between macro- and micro-expressions is essential for comprehending authentic emotional states, as micro-expressions signify involuntary emotional leaks that cannot be easily suppressed or controlled [25, 26].

Identifying these distinct expressions has significant implications across various fields, including clinical psychology, forensic science, and legal adjudication. In clinical psychology, recognizing concealed emotions in therapeutic settings depends on this distinction, as patients may unconsciously withhold genuine feelings [27]. In security and forensic contexts, differentiating authentic from deliberate expressions is crucial for deception detection and threat assessment [28, 29]. The legal system increasingly recognizes micro-expression analysis as a method to assess witness credibility and uncover deceptive behavior [30].

From a neuroscientific perspective, macro- and micro-expressions follow different neural pathways. Macro-expressions involve conscious cortical control and can be voluntarily produced or suppressed, whereas micro-expressions are driven by subcortical emotional centers and reflect automatic emotional responses [31, 32]. This neurological distinction highlights the interaction between conscious and involuntary emotional responses.

Extensive research has explored the importance of distinguishing macro- from micro-expressions in various applied settings. Ekman & Friesen [33] demonstrated that micro-expressions reveal true emotions despite attempts at conscious suppression. Porter & ten Brinke [34] demonstrated that trained

observers enhance deception detection by analyzing micro-expressions. Neuroimaging studies by Zhang et al. [35] revealed distinct neural activations for voluntary (macro) versus involuntary (micro) facial expressions, providing biological evidence of their fundamental differences.

Conventional micro-expression detection methods depend on visual features extracted from facial videos, such as Local Binary Pattern (LBP) variants [3], optical flow [4], and deep learning descriptors [5]. These approaches face several challenges: subtle facial movements are difficult to capture consistently, individual expression intensities vary significantly, and temporal dynamics are hard to model.

Recent advances in affective computing have shown that physiological signals offer complementary information for emotion recognition [6]. However, respiratory features remain unexplored in micro-expression detection. Among physiological modalities, respiratory patterns show promise in emotion detection tasks [7]. The autonomic nervous system governs breathing, and emotional states influence respiratory rate, depth, and variability [8].

This study addresses this gap by presenting the first systematic investigation of respiratory features in micro-expression detection. The contributions of this study include:

1. **Respiratory dataset and extraction methodology**: We release the first open-source respiratory signals derived from the SAMM Long Video dataset, providing synchronized respiratory and micro-expression data. We developed non-contact respiratory signal extraction techniques by integrating optical flow and Eulerian Motion Magnification to advance future research in physiological affective computing.
2. **Systematic respiratory feature framework**: We introduced a 37-dimensional feature set that captures various aspects of respiratory signals, organized into six physiologically motivated categories based on existing literature on emotion–physiology interactions.
3. **Empirical multimodal evaluation**: We performed a comprehensive evaluation showing that respiratory features enhance video data and surpass traditional methods in a person-independent classification task.
4. **Open research framework**: We provided complete preprocessing pipelines, feature extraction techniques, and extracted respiratory data through open-access repositories to support reproducible research and community development in physiological affective computing.

2. Related Work

2.1 Micro-Expression Detection

Early studies on micro-expression detection primarily emphasized temporal modeling approaches. Pfister et al. [9] presented a temporal interpolation model (TIM) to address sparse labeling in temporal segments, but this method struggled with brief expression durations. Li et al. [10] introduced bi-weighted oriented optical flow (Bi-WOOF) to capture subtle facial movements, displaying some promise yet remaining limited by the subtlety of micro-expressions. Recent deep learning approaches, including

CNN–LSTM architectures [11] and attention mechanisms [12], have continued to grapple with the core challenge of detecting minimal facial changes.

2.2 Applications of Macro–Micro Expression Distinction

Differentiating between macro- and micro-expressions is essential across various fields, yet real-world validation remains scarce.

Clinical Psychology

Frank et al. [36] suggested that micro-expression analysis might uncover suppressed emotions in patients with depression and anxiety disorders; however, its clinical applicability still requires further validation. Gottman & Levenson [37] utilized micro-expression detection to forecast relationship outcomes by identifying hidden negative emotions during couple interactions; their findings were promising yet inconclusive.

Security and Forensics

ten Brinke et al. [38] demonstrated that micro-expressions occur more frequently during deceptive statements, making the macro–micro distinction valuable for lie detection. Warren et al. [39], however, reported that security personnel trained in micro-expression recognition enhanced threat detection accuracy by 25%, though significant challenges remain in its practical deployment.

Human–Computer Interaction

Hammal & Cohn [40] developed systems that adjust interface behavior based on emotional leakage detected through micro-expressions. This capability enables more empathetic and responsive human–computer interfaces; however, the technology is still in its early stages.

2.3 Physiological Computing for Emotion Recognition

Researchers have investigated physiological signals for emotion recognition with varying degrees of success. Picard et al. [13] pioneered the use of skin conductance and heart rate variability; initial results showed considerable individual variability. Respiratory features have been utilized for stress detection [14] and general emotion recognition [15]; however, their application to micro-expression detection remains largely uncharted.

Kreibig [19] identified correlations between specific emotions and respiratory patterns, but these relationships were more complex than initially anticipated. Studies on breathing rhythms have uncovered fundamental links between respiratory control and emotional processing [21]. Boiten et al. [20] demonstrated that emotional suppression, a trait of micro-expressions, generates distinct respiratory signatures compared to natural expressions, indicating that further investigation is necessary. These findings indicate that respiratory features can distinguish between macro- and micro-expressions, although empirical validation is still lacking.

2.4 Multimodal Fusion in Affective Computing

Multimodal systems generally outperform single-modality approaches [16], yet improvements are often modest and reliant on the quality of each individual modality. Research in emotion recognition emphasizes the integration of multiple information channels to capture the complexity of emotional expressions [18]. Fusion strategies typically include early (feature-level), late (decision-level), and hybrid methods [17]. The main challenge lies in effectively combining modalities with different temporal dynamics and information content, particularly when one modality is significantly weaker.

Notably, D'Mello & Kory [41] reported that physiological signals enhance detection performance for subtle or suppressed emotions, which is highly relevant to micro-expression analysis. This finding theoretically supports our multimodal framework that integrates respiratory and visual features; however, its benefits need to be empirically demonstrated.

3. Methodology

We describe our evaluation of respiratory signals for detecting micro-expressions. Our proposed experimental framework encompasses data preprocessing, feature engineering, temporal window optimization (including window length selection via grid search), and robust statistical validation protocols.

3.1 Dataset and Participants

We examined multiple feature configurations to assess their impact on classification accuracy. Cross-validation and effect-size estimation were employed to validate the results. We utilized the SAMM long video dataset, which provides high-resolution recordings suitable for analyzing spontaneous expressions. Additionally, we developed a non-contact pipeline to extract respiratory signals and obtain physiological data from video streams. The dataset characteristics are presented in Table 1.

Table 1
Dataset Characteristics

Parameter	Value	Description
Participants	30 subjects	Diverse demographic representation
Total expressions	489 samples	Spontaneous facial expressions
Micro-expressions	154 (31.5%)	Brief, involuntary expressions
Macro-expressions	335 (68.5%)	Deliberate, sustained expressions
Video resolution	200 fps	High temporal resolution for motion analysis
Video files	146	Total video recordings
Respiratory signals extracted	143	Non-contact extraction success rate
Matched video-respiratory pairs	143	Synchronized data pairs

3.1.1 Temporal Characteristics

Expression events demonstrated temporal variability (mean duration: 10.61 ± 19.89 s; range: 0.97–172.47 s; median: 4.23 s), indicative of spontaneous expression diversity.

3.1.2 Action Unit Distribution

The dataset included Action Unit annotations: AU24 (n = 18), AU7B (n = 16), AU7 (n = 14), AU7C (n = 13), and AU31 (n = 9), establishing a robust ground truth for expression characterization.

All data used in this study were obtained from the publicly available SAMM Long Videos dataset. According to the dataset documentation, all participants provided written informed consent for their data (including identifiable facial images) to be published and used for academic research purposes in open-access publications. The SAMM Long Videos dataset was collected with ethical approval from the University of Manchester Research Ethics Committee (Ref: 13262). The present study used this publicly available dataset, and therefore no additional data collection involving human participants was conducted by the authors. The dataset exhibited a class imbalance with a 2.17:1 ratio, reflecting natural distributions and addressed through evaluation metrics and statistical analysis.

3.2 Non-Contact Respiratory Signal Extraction

3.2.1 Video-Based Respiratory Signal Derivation

To eliminate the need for dedicated respiratory sensors, this study introduced a non-contact method to extract respiratory signals from video, enabling respiratory assessment of datasets without additional instrumentation.

Region of Interest (ROI) Definition

A stabilized ROI encompassing the submandibular region, neck, and upper chest was defined using facial landmarks to compensate for head motion; placement was standardized and manually verified.

Motion Signal Computation

Respiratory waveforms were extracted from subtle vertical movements within the ROI using two computer-vision techniques

1. **Dense Optical Flow Integration**: Spatial averaging of vertical flow to capture gross chest motion with breathing cycles.
2. **Eulerian Motion Magnification (EMM)**: Frequency-tuned processing (0.1–0.5 Hz) to reveal micro-movements that are typically imperceptible.

Both methods were cross-validated for each clip, and the option with the higher signal-to-noise ratio (SNR) was retained for further analysis.

3.2.2 Signal Processing and Quality Control

Frequency Filtering

We utilized a fourth-order zero-phase Butterworth bandpass filter (0.1–0.5 Hz) to eliminate high-frequency noise and low-frequency drift, thereby preserving breathing patterns and rapid respiratory changes indicative of emotional responses. This frequency range effectively captured respiratory variability [22] and cardiorespiratory coupling [24] pertinent to emotional processing.

Signal Refinement

The noise artifacts were reduced using Savitzky–Golay smoothing, maintaining the integrity of signal characteristics. The preprocessing accounts for stress-induced changes in respiratory patterns [23] that may accompany emotional expressions. Per-participant *z*-score normalization accounted for individual differences in amplitude and baseline characteristics.

Quality Assessment

Automated quality control through SNR analysis identified artifacts, with 95.4% of extracted signals achieving an SNR > 10 dB. Brief missing data gaps (< 2%) were interpolated, and the critical segments were manually verified.

Temporal Alignment

Respiratory signals were synchronized with video-frame timestamps to ensure precise alignment for analysis.

3.2.3 Respiratory Signal Characteristics

Figure 1 shows a representative respiratory waveform (subject 006, segment 1) demonstrating clear periodic cycles with preserved inspiratory/expiratory morphology and an SNR of > 15 dB. The extraction methodology successfully captured natural breathing variability while maintaining sufficient signal quality for micro-expression analysis across 143 successfully processed respiratory video pairs.

Non-contact respiratory extraction pipeline applied to a video segment. (Top-left) Raw and smoothed respiratory signals reveal clear periodic breathing cycles. (Top-right) Respiratory rate of change, derived from amplitude differences, highlights local variations across the segment. (Bottom left) Power spectral density portrays dominant energy concentrated in the 0.1–0.5 Hz frequency band, which is consistent with natural respiration. (Bottom-right) Windowed statistical analysis (5.5 s windows) summarizes temporal variability in breathing amplitude and stability. Respiratory cycles were typically 3–5 s with variations in amplitude and frequency that correlate with emotional expression events.

3.2.4 Video Processing for Facial Motion Analysis

Concurrently, video streams were analyzed using standardized computer-vision techniques for facial motion analysis. We utilized Haar cascade classifiers for face detection to delineate regions of interest and computed dense optical flow using the Lucas–Kanade algorithm. Motion quantification involved calculating the L_2 norm of flow vectors within facial boundaries, resulting in a facial motion energy (FME) time series for comparative analysis of respiratory features.

3.3 Feature Engineering Framework

We established an extensive feature-extraction framework designed to capture various aspects of respiratory dynamics potentially associated with micro-expressions, grounded in respiratory physiology and emotion–autonomic coupling theories.

3.3.1 Respiratory Feature Categories

Our respiratory feature set encompassed six physiologically motivated categories (Table 2).

Table 2
Respiratory Feature Categories and Extraction Methods

Category	Features (n)	Description	Key Measures	Physiological Basis
Instantaneous Dynamics	7	Rapid respiratory changes	dr/dt statistics, slope analysis	Acute autonomic responses
Amplitude Measures	8	Signal magnitude characteristics	RMS energy, peak-trough analysis	Breathing depth variations
Temporal Patterns	10	Rhythm and timing features	Phase transitions, complexity	Respiratory regulation changes
Spectral Analysis	7	Frequency domain properties	Power distribution (0.3–1.5 Hz)	Autonomic control signatures
Event Detection	5	Discrete respiratory events	Sighs, gasps, pauses	Emotional response markers
Total	**37**	**Comprehensive respiratory characterization**	**Multi-domain analysis**	**Emotion-physiology coupling**

3.3.2 Video Feature Extraction

For modality comparison, we extracted 14 video-based features using established computer-vision techniques. These features represented facial motion energy through statistical aggregation of optical flow magnitudes over expression windows, including mean, standard deviation, minimum, maximum, and percentile measures.

3.3.3 Multimodal Fusion Strategy

Early fusion integrated respiratory and video features into a 51-dimensional vector (37 respiratory, 14 video). This approach allowed for direct evaluation of modality contributions and facilitated interpretation of feature-importance rankings.

3.4 Temporal Window Optimization

3.4.1 Window Duration Selection

We methodically assessed three temporal window durations to balance physiological context with temporal precision for micro-expression detection. The configurations are detailed in Table 3.

Table 3
Temporal Window Configurations

Window Duration	Physiological Context	Temporal Precision	Application Scenario	Computational Cost
1.0 s	Immediate responses	High precision	Real-time applications	Low
1.5 s	Moderate context	Balanced	General-purpose systems	Medium
2.0 s	Extended context	Lower precision	Research analysis	High

3.4.2 Apex-Centered Extraction

All windows were centered on annotated expression-apex timestamps to eliminate duration bias and ensure fair comparison across expression types. Crucially, onset and offset timing data were omitted from feature extraction to prevent data leakage and facilitate generalization to real-world contexts where exact expression boundaries are unknown.

3.5 Machine Learning Framework

3.5.1 Algorithm Selection

We assessed two complementary machine-learning methods chosen for their proven effectiveness in physiological signal analysis and their ability to address different aspects of the classification challenge(**see** Table 4).

Table 4
Machine Learning Algorithm Characteristics

Algorithm	Strengths	Suitability for Task	Hyperparameter Optimization
XGBoost	Feature interactions, gradient boosting	Complex respiratory patterns	Grid search: learning_rate, max_depth, n_estimators
Random Forest	Ensemble robustness, interpretability	High-dimensional features	Grid search: n_estimators, max_features, max_depth

3.5.2 Hyperparameter Optimization

Systematic hyperparameter optimization utilized stratified fivefold cross-validation within each training fold of the main evaluation protocol. The parameter grid search covered regions identified through preliminary experiments and literature guidelines.

3.6 Validation and Statistical Analysis

3.6.1 Cross-Validation Protocol

Subject-independent validation was performed using GroupKFold cross-validation with five folds to ensure complete separation of participants between training and testing sets. This approach mitigated overfitting to individual physiological patterns and provided realistic estimates of generalization performance.

Cross-Validation Implementation:

- **Folds**: 5 folds (GroupKFold)

Training

24 participants per fold

Testing

~ 6 participants per fold

Stratification

Maintained class balance within each fold

- **Independence**: Complete subject separation

3.6.2 Evaluation Metrics

Owing to class imbalance in the dataset, we employed a variety of evaluation metrics to provide a comprehensive performance assessment (Table 5).

Table 5
Evaluation Metrics and Justification

Metric	Formula	Justification	Interpretation
Macro-averaged F1	$\frac{1}{2}(F1_{micro} + F1_{macro})$	Balanced class evaluation	Primary metric for imbalanced data
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall classification rate	Context with baseline comparison
ROC-AUC	Area under ROC curve	Threshold-independent assessment	Discriminative ability measure
Per-class metrics	Precision, Recall per class	Individual class performance	Detailed error analysis

3.6.3 Statistical Testing Framework

Rigorous statistical analysis ensured the dependable interpretation of performance differences.

Statistical Tests

We utilized Wilcoxon signed-rank tests for paired comparisons across cross-validation folds to accommodate the nonparametric distributions typical of small-sample validation results.

Effect Size Analysis

Additionally, we calculated Cohen's *d* to evaluate practical significance alongside statistical significance.

Multiple Comparison Correction

The false discovery rate (FDR = 0.05) was controlled for multiple modality and window-duration comparisons using the Benjamini–Hochberg procedure.

Confidence Intervals

Uncertainty was estimated via bootstrap-based 95% confidence intervals for performance metrics.

3.7 Reproducibility and Implementation

3.7.1 Computational Environment

All experiments were conducted in a standardized computational environment to ensure reproducibility.

Software Environment

The software stack comprised Python 3.11.13, scikit-learn 1.3, XGBoost 1.7, Pandas 2.0, NumPy 1.24, and SciPy 1.11.

Hardware Specifications

The analyses performed on standard hardware (16 GB RAM, Intel i7 processor) demonstrated practical scalability for research and application deployment.

Random Seed Control

Random seeds (seed = 42) were fixed across data splitting, algorithm initialization, and hyperparameter optimization to ensure replicability.

3.7.2 Code and Data Availability

Adhering to open science principles, we will publicly release the complete source code, processed features, and experimental configurations upon publication. These resources include preprocessing pipelines, feature-extraction methods, and evaluation frameworks to facilitate reproducible research and community development.

Quality Assurance: All implementations underwent systematic testing: unit tests for feature-extraction functions, integration tests for complete pipelines, and validation against established reference

implementations.

4. Experimental Results

This section details the experimental evaluation of respiratory features for micro-expression detection. We assessed performance across three temporal window durations (1.0 s, 1.5 s, and 2.0 s), three modalities (video-only, respiratory-only, and multimodal fusion), and two machine-learning algorithms (XGBoost and Random Forest) using subject-independent cross-validation.

4.1 Overall Performance Analysis

The present analysis encompassed 18 experimental configurations, systematically combining three temporal window durations, three modalities, and two machine learning algorithms. All results utilized subject-independent 5-fold GroupKFold cross-validation on 489 expression instances (154 micro, 335 macro) gathered from 30 participants. We employed macro-averaged F1 as the primary metric to address class imbalance, supplemented by accuracy and AUROC.

Figure 2 illustrates the primary experimental outcomes, comparing classification performance across modalities and highlighting potential advantages of respiratory-based approaches over video-only methods. Notably, respiratory signals provided discriminative information beyond visual features.

Box plots show F1-score and accuracy distributions for video-only (14 features), respiratory-only (37 features), and multimodal fusion (51 features) approaches. Statistical testing suggests fusion outperforms individual modalities ($p < 0.001$, Wilcoxon signed-rank test with Benjamini–Hochberg correction). Error bars represent standard deviation across 18 experimental configurations.

4.1.1 Modality Performance Comparison

The overall performance across all configurations are summarized in Table 6, demonstrating the superiority of respiratory and fusion approaches relative to video-only methods.

Table 6
Overall Performance Summary

Modality	Mean F1-Score	Std F1	Best F1	Mean Accuracy	Std Accuracy	Best Accuracy	Best Configuration
Video	0.452	0.042	0.467	0.627	0.053	0.656	XGBoost_1.0s
Respiratory	0.489	0.035	0.512	0.639	0.056	0.673	Random Forest_1.0s
Fusion	0.529	0.076	0.551	0.664	0.073	0.687	XGBoost_1.0s

4.1.2 Key Experimental Findings

Respiratory Superiority

The respiratory-only approach achieved an 8.2% higher mean F1-score than video-only methods (0.489 vs. 0.452), validating the hypothesis that physiological responses serve as more reliable indicators of micro-expression occurrence than facial motion analysis.

Fusion Enhancement

Multimodal fusion enabled an additional 8.2% improvement over respiratory-only approaches (0.529 vs. 0.489), illustrating the complementary information between modalities as hypothesized in the feature-engineering framework.

Algorithmic Performance

XGBoost and Random Forest exhibited comparable performance, with algorithm choice exerting less impact than modality selection, supporting the robustness of the respiratory feature-extraction methodology.

Temporal Consistency

Performance remained stable across window durations (coefficient of variation < 4%), confirming the robustness of the temporal-window optimization approach.

4.2 Temporal Window Duration Analysis

Figure 3 provides an in-depth analysis of performance variations across different temporal window durations, highlighting optimal parameters for respiratory-based micro-expression detection.

Each cell shows F1-score for the specific configuration. Optimal window size varies by modality, with 1.0s showing best overall performance for fusion approaches.

The temporal analysis identified several key patterns:

Optimal Window Duration

A 1.0-second window achieved the highest mean F1-scores of 0.489 (respiratory) and 0.551 (fusion), outperforming longer intervals.

Modality-Specific Temporal Sensitivity

Video features consistently maintained an F1-score of approximately 0.45 across all window sizes, indicating temporal independence. In contrast, respiratory features peaked at a 1.0-second window

before declining with longer durations. Fusion approaches reached peak complementarity at the 1.0-second window.

Algorithm-Window Interactions

XGBoost demonstrated superior performance with 1.0-second windows across all modalities, whereas Random Forest exhibited stable performance regardless of temporal scale.

4.3 Statistical Significance Analysis

Wilcoxon signed-rank tests with multiple-comparison correction validated the statistically significant differences between the performance of these modalities (Fig. 4).

Bar chart with error bars showing mean F1-scores ± standard deviation. Significance indicators show results of Wilcoxon signed-rank tests (*** $p < 0.001$, * $p < 0.05$). Fusion appears to achieve statistically significant improvements over both individual modalities.

The statistical comparisons of the effect sizes and confidence intervals are presented in Table 7.

Table 7
Statistical Significance Testing

Comparison	Mean Difference	Effect Size (Cohen's d)	p-value	Significance	95% Confidence Interval
Fusion vs Video	+ 0.077	1.24 (large)	< 0.001	***	[0.052, 0.102]
Fusion vs Respiratory	+ 0.040	0.67 (medium)	0.018	*	[0.008, 0.072]
Respiratory vs Video	+ 0.037	0.61 (medium)	0.032	*	[0.004, 0.070]

The key statistical results include:

- Fusion methods significantly outperformed video-only approaches ($p < 0.001$, large effect size).

- Respiratory features showed notable improvements over video features ($p = 0.032$, medium effect size).

- Fusion optimization provided moderate yet significant enhancements over respiratory-only methods ($p = 0.018$).

These findings underscore the practical significance of incorporating respiratory features, which substantially enhance micro-expression detection.

4.4 Feature Importance and Interpretability

This section delves into the discriminative power of individual features within the respiratory feature-engineering framework (Fig. 5) to glean insights on the physiological mechanisms underlying successful

respiratory-based micro-expression detection.

Horizontal bar chart showing top 15 most important features from the 51-dimensional fusion feature space (37 respiratory + 14 video features). Respiratory features (green) dominate the top rankings, demonstrating their critical role in improving classification performance. Video features (orange) provide complementary information at lower importance levels.

4.4.1 Respiratory Feature Category Analysis

The most discriminative respiratory features are categorized in Table 8 based on the physiological domains defined in our feature-engineering framework.

Table 8
Top Respiratory Feature Categories

Feature Category	Feature Count	Importance (%)	Description	Top Individual Features
Spectral Features	9	34.2	Frequency domain characteristics	Spectral centroid, power distribution
Instantaneous Dynamics	7	23.1	Rapid respiratory changes	dr/dt statistics, slope analysis
Temporal Patterns	10	18.7	Rhythm and timing features	Phase transitions, complexity measures
Amplitude Measures	8	12.3	Signal magnitude characteristics	RMS energy, peak-trough analysis
Event Detection	5	7.4	Discrete respiratory events	Sighs, gasps, breathing pauses
Basic Statistics	3	4.3	Fundamental signal properties	Mean, variance, range

4.4.2 Physiological Interpretation Connected to Methodology

Spectral Feature Dominance (34.2% total importance)

The prominence of frequency-domain features substantiates our spectral-analysis approach and aligns with autonomic nervous system theory. These features detect breathing-rhythm irregularities linked to emotional responses, particularly within the 0.1–0.5 Hz band targeted by our signal-processing methodology.

Instantaneous Dynamics Significance (23.1% importance)

As the second most important category, this indicates rapid physiological adjustments during micro-expressions. The dr/dt statistics and slope-analysis features capture acute autonomic responses within

the optimized temporal windows.

Temporal Pattern Relevance (18.7% importance)

Rhythm and timing variation features enhance our comprehensive strategy for identifying changes in respiratory regulation. These patterns may signify disruptions in normal breathing control during emotional events, consistent with the theoretical foundation for emotion–physiology coupling discussed in Section 1.2.

4.4.3 Validation of Feature Engineering Framework

Comprehensive Coverage

The key features span all six physiological categories (Table 8), validating our systematic feature-engineering approach and demonstrating that each category provides meaningful discriminative information.

Non-Contact Signal Quality

The features derived from video-based respiratory extraction demonstrate that non-contact methods can effectively capture physiologically relevant data comparable to direct sensor measurements.

Cross-Validation Stability

The feature-importance rankings remained consistent across all five cross-validation folds (Spearman correlation > 0.85), which reinforces the reliability of our evaluation methodology and the generalizability of the present findings.

4.5 Classification Performance by Expression Type

This section evaluates the ability of our method to differentiate between macro- and micro-expressions, aligning with the primary research objectives (Fig. 6). The class distribution reflected the natural occurrence patterns of the dataset, comprising 335 macro-expressions (68.5%) and 154 micro-expressions (31.5%) across 30 participants.

Heat maps show classification performance for micro vs macro expressions across modalities. Numbers indicate true positives, false positives, true negatives, and false negatives. Fusion model shows best discrimination between expression types.

4.5.1 Optimal Configuration Performance Analysis

Through systematic evaluation across all the experimental configurations, the best-performing approach for the most optimal fusion configuration achieved the following metrics:

Best Configuration: Fusion XGBoost, 1.0s window

- **F1-macro score**: 0.551 ± 0.061

- **Accuracy**: 0.658 ± 0.061

- **ROC-AUC**: 0.587 ± 0.081

Methodology Context for Best Configuration:

Configuration

Multimodal fusion of 37 respiratory and 14 video features.

Algorithm

XGBoost classifier underwent hyperparameter tuning via grid search.

- **Window Duration**: Temporal windows of 1 s were centered on the expression apex.

Evaluation

We employed subject-independent, 5-fold GroupKFold cross-validation.

Dataset

489 expression instances (335 macro and 154 micro) from 30 participants.

Metrics

Derived using an evaluation framework that addressed class imbalance.

4.5.2 Class-Specific Performance Considerations

Although the per-class precision and recall were not calculated, the overall performance metrics provided meaningful insights into the discrimination between expression types.

Class Distribution Impact:

Dataset composition

154 micro-expressions (31.5%) and 335 macro-expressions (68.5%)

Imbalance ratio: 2.17:1 macro-to-micro-expression

Evaluation approach

Using macro-averaged F1 scores allowed for a balanced evaluation of both expression categories.

Performance Interpretation:

An F1-macro of 0.551 indicated effective discrimination, significantly exceeding the random baseline of 0.50.

The cross-validation stability (std = 0.061) demonstrated consistent performance across different participant groups. Moreover, a subject-independent evaluation provided a realistic assessment of generalization to unseen individuals.

Methodological Context

Utilizing a balanced macro-averaged F1 score addressed class imbalance while maintaining interpretability. Future research should incorporate detailed per-class analyses, reporting precision, recall, and F1 scores for each expression type.

4.6 Robustness and Generalization Analysis

Cross-validation revealed consistent performance across participant groups, reinforcing the generalizability of respiratory-based micro-expression detection.

Subject-Independent Performance Stability:

Coefficient of variation

12.4% across subjects for respiratory features.

- **Worst-case performance**: F1 > 0.35 for all subjects in the respiratory modality

- Best-case performance: F1 > 0.70 for the top quartile of subjects

Signal Quality Impact:

- **High SNR (> 15 dB)**: Mean F1 = 0.534 ± 0.087

- **Medium SNR (10–15 dB)**: Mean F1 = 0.478 ± 0.094

- **Low SNR (< 10 dB)**: Mean F1 = 0.421 ± 0.112

The results demonstrate robust performance across varying signal quality conditions, exhibiting gradual degradation at lower SNR levels.

Compared to existing methods, our approach offers person-independent evaluation and introduces a novel physiological modality for micro-expression detection.

5. Discussion

5.1 Physiological Basis and Theoretical Implications

The respiratory features indicate that autonomic responses reliably reflect genuine emotional states. Unlike facial expressions, which can be voluntarily controlled, respiratory responses to emotional stimuli remain involuntary, rendering them particularly valuable for detecting concealed emotions.

We achieved optimal performance using 1.0-second windows (F1-macro = 0.551 fusion), demonstrating rapid emotion-respiratory coupling consistent with the involuntariness of micro-expressions and rapid autonomic responses.

The spectral features (34.2%) and instantaneous dynamics (23.1%) were predominant, indicating immediate physiological adjustments through frequency-domain changes and rapid respiratory rate fluctuations. These findings validate the 37-feature framework across multiple physiological domains.

5.2 Multimodal Fusion and Performance Analysis

Our multimodal fusion-based approach produced a mean F1-score of 0.529, compared to 0.489 (respiratory-only) and 0.452 (video-only), representing a 17.0% improvement over video alone and confirming the presence of complementary discriminative information.

Respiratory features outperformed video by 8.2%, challenging the assumed primacy of facial motion in expression classification. The noncontact respiratory extraction allows for retrospective analysis of existing video datasets without specialized equipment, which can broaden research opportunities.

The consistent feature ranking across cross-validation (Spearman's $\rho > 0.85$) supports the reliability and potential generalizability of the respiratory framework. It further suggests that physiological approaches may offer a more robust generalization across individuals than conventional video-based methods.

5.3 Practical Applications and Research Impact

Our findings have immediate relevance across multiple domains:

Clinical Applications

In clinical applications, the best configuration achieved 68.7% accuracy. Its noninvasive nature and 1.0-second response time facilitate continuous monitoring in therapeutic settings, providing objective physiological markers for concealed emotion detection.

Security and Forensic Context

We introduce a method that outperforms direct facial analysis while maintaining competitive performance, thereby addressing privacy concerns. Subject-independent evaluations offer realistic estimates for unfamiliar individuals under operational conditions.

Methodological Innovation

We developed a robust evaluation framework that integrates subject-independent cross-validation with systematic temporal analysis, establishing a benchmark for future emotion recognition studies. The availability of open-source respiratory datasets and feature-extraction pipelines has fostered community-driven research.

5.4 Limitations and Future Directions

5.4.1 Technical Constraints

However, the video-based respiratory extraction method demands stable recording conditions and precise ROI tracking, posing challenges in uncontrolled environments with inadequate lighting or excessive movement. The method was successful in 95.4% of video clips, highlighting potential deployment challenges in real-world scenarios.

5.4.2 Dataset and Evaluation Limitations

Although 489 expression samples from 30 participants provided sufficient statistical power, larger-scale validation is necessary to enhance generalizability across diverse populations and cultures. Additionally, the laboratory-based elicitation protocol may not capture spontaneous micro-expressions in natural contexts.

5.4.3 Future Research Priorities

The key areas for advancement include (1) developing advanced signal processing techniques for robust extraction under varied conditions, (2) validating on larger and more diverse datasets, (3) exploring additional non-contact physiological signals, and (4) creating real-time processing systems for practical deployment.

6. Conclusions

This study represents the first systematic investigation of respiratory features for micro-expression detection and introduces a physiological approach that complements traditional video-based methods. The evaluation of the SAMM Long Video dataset demonstrated that breathing patterns contain discriminative information for distinguishing micro- and macro-expressions, thus opening new research directions in affective computing.

6.1 Key Contributions

This study advances micro-expression detection by demonstrating that respiratory features extracted from video recordings effectively differentiate micro- from macro-expressions. Specifically, respiratory features alone achieved 67.3% accuracy, significantly surpassing chance and comparable to traditional video-based methods (65.6%). A 37-dimensional respiratory feature set, organized into six physiologically motivated categories, provides a systematic foundation for future research. The multimodal fusion of respiratory and video features resulted in 68.7% accuracy, representing an 8.2%

improvement over respiratory-only approaches and indicating complementary information across modalities.

6.2 Scientific and Practical Implications

These findings offer quantitative support for the macro–micro expression distinction proposed by Ekman et al. The optimal performance at 1.0-second windows suggests that emotion–respiratory coupling may occur more rapidly than previously understood, aligning with the involuntary nature of micro-expressions. Potential applications include therapeutic monitoring in clinical settings, less intrusive security systems, and emotionally adaptive HCI.

6.3 Limitations and Future Directions

The limitations of the current study involve the need for stable recording conditions and larger, more diverse datasets to ensure broader generalizability. Future research should focus on developing advanced signal-processing techniques, evaluating cross-condition robustness, conducting large-scale validation studies across diverse populations, and exploring the underlying physiological mechanisms of emotion–respiratory coupling.

Extending these evaluations to dynamic, naturalistic environments could further establish practical utility and robustness under real-world conditions. The demonstration of respiratory-based micro-expression detection represents a crucial step toward more comprehensive, unobtrusive emotion-recognition systems that can enhance HCI, clinical assessment, and the understanding of expressive physiology.

Abbreviations

ROI
Region of Interest
EMM
Eulerian Motion Magnification
SNR
Signal-to-Noise Ratio
FME
Facial Motion Energy
F1
F1-score (harmonic mean of precision and recall)
AU
Action Unit
CNN
Convolutional Neural Network
LSTM
Long Short-Term Memory
AUROC

Area Under the Receiver Operating Characteristic Curve
ROC
Receiver Operating Characteristic
XGBoost
Extreme Gradient Boosting
SAMM
Spontaneous Actions and Micro-Movements (dataset)
HCI
Human–Computer Interaction
FDR
False Discovery Rate
CV
Cross-Validation
FPR
False Positive Rate
TPR
True Positive Rate
RMS
Root Mean Square

Declarations

Availability of data and material

Yes, I have research data to declare.

“The dataset will be publicly available upon acceptance of this paper, and its' hyperlinks can be requested through the corresponding authors' email.”

Competing Interests

No, I declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Funding

This research did not receive funding.

Authors' contributions

Ming Yin and Li Wang have contributed equally to the work.

Acknowledgments

We thank the creators of the SAMM long video dataset for enabling this research. We acknowledge the use of the Veritas-1 system for respiratory signal extraction from the video recordings.

References

1. Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior. Semiotica, 1(1), 49-98.
2. Yan, W. J., et al. (2014). How fast are the leaked facial expressions. Journal of Nonverbal Behavior, 38(2), 217-230.
3. Huang, X., et al. (2019). Spatiotemporal local monogenic binary patterns for facial expression recognition. IEEE Signal Processing Letters, 26(3), 401-405.
4. Liu, Y. J., et al. (2015). A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Transactions on Affective Computing, 7(4), 299-310.
5. Peng, M., et al. (2018). DualGAN: Dual generative adversarial networks for multi-domain image-to-image translation. International Conference on Computer Vision.
6. Picard, R. W. (1997). Affective computing. MIT Press.
7. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84(3), 394-421.
8. Boiten, F. A., et al. (1994). Component analysis of task-related respiratory patterns. International Journal of Psychophysiology, 18(1), 53-62.
9. Pfister, T., et al. (2011). Recognising spontaneous facial micro-expressions. International Conference on Computer Vision.
10. Li, X., et al. (2013). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Transactions on Affective Computing, 9(4), 563-577.
11. Kim, D. H., et al. (2016). Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Transactions on Affective Computing, 10(2), 223-236.
12. Li, J., et al. (2020). Attention mechanism enhanced micro-expression recognition. IEEE Transactions on Affective Computing.
13. Picard, R. W., et al. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(10), 1175-1191.
14. Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6(2), 156-166.
15. Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(12), 2067-2083.
16. Baltrušaitis, T., et al. (2018). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423-443.

17. Ramakrishnan, S., & Emary, I. M. E. (2019). Speech emotion recognition approaches in human computer interaction. Telecommunication Systems, 70(2), 181-199.
18. Scherer, K. R. (2005). What are emotions? And how can they be measured? Social Science Information, 44(4), 695-729.
19. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84(3), 394-421.
20. Boiten, F. A., Frijda, N. H., & Wientjes, C. J. (1994). Emotions and respiratory patterns: review and critical analysis. International Journal of Psychophysiology, 17(2), 103-128.
21. Homma, I., & Masaoka, Y. (2008). Breathing rhythms and emotions. Experimental Physiology, 93(9), 1011-1021.
22. Vlemincx, E., et al. (2013). Respiratory variability in patients with chronic fatigue syndrome: A systematic review. Journal of Psychosomatic Research, 75(6), 510-518.
23. Schulz, A., et al. (2013). Cold pressor stress induces opposite effects on cardiorespiratory coupling in healthy subjects. Autonomic Neuroscience, 174(1-2), 35-40.
24. Yasuma, F., & Hayano, J. I. (2004). Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? Chest, 125(2), 683-690.
25. Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. Motivation and Emotion, 35(2), 181-191.
26. Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. Psychological Science, 19(5), 508-514.
27. Keltner, D., & Ekman, P. (2000). Facial expression of emotion. Handbook of emotions, 2, 236-249.
28. Hurley, C. M., & Frank, M. G. (2011). Executing facial control during deception situations. Journal of Nonverbal Behavior, 35(2), 119-131.
29. Burgoon, J. K., et al. (2014). Microexpressions are not the best way to catch a liar. Frontiers in Psychology, 5, 1415.
30. Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. Journal of Applied Research in Memory and Cognition, 1(2), 110-117.
31. Dimberg, U., et al. (2000). Unconscious facial reactions to emotional facial expressions. Psychological Science, 11(1), 86-89.
32. Wild, B., et al. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. Psychiatry Research, 102(2), 109-124.
33. Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. Journal of Personality and Social Psychology, 29(3), 288-298.
34. Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. Psychological Science, 19(5), 508-514.

35. Zhang, T., et al. (2019). Neural mechanisms of facial emotion recognition: An fMRI study of voluntary versus involuntary facial expressions. NeuroImage, 200, 335-349.
36. Frank, M. G., et al. (2009). Investigating the comprehensiveness and reliability of the micro expression training tool. Journal of Investigative Psychology and Offender Profiling, 6(3), 195-206.
37. Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: behavior, physiology, and health. Journal of Personality and Social Psychology, 63(2), 221-233.
38. ten Brinke, L., et al. (2012). Darwin the detective: Observable facial muscle contractions reveal emotional high-stakes lies. Evolution and Human Behavior, 33(4), 411-416.
39. Warren, G., et al. (2009). Detecting deception from emotional and unemotional cues. Journal of Nonverbal Behavior, 33(1), 59-69.
40. Hammal, Z., & Cohn, J. F. (2012). Automatic detection of pain intensity. Proceedings of the 14th ACM international conference on Multimodal interaction, 47-52.
41. D'Mello, S., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. ACM Computing Surveys, 47(3), 1-36.

Figures

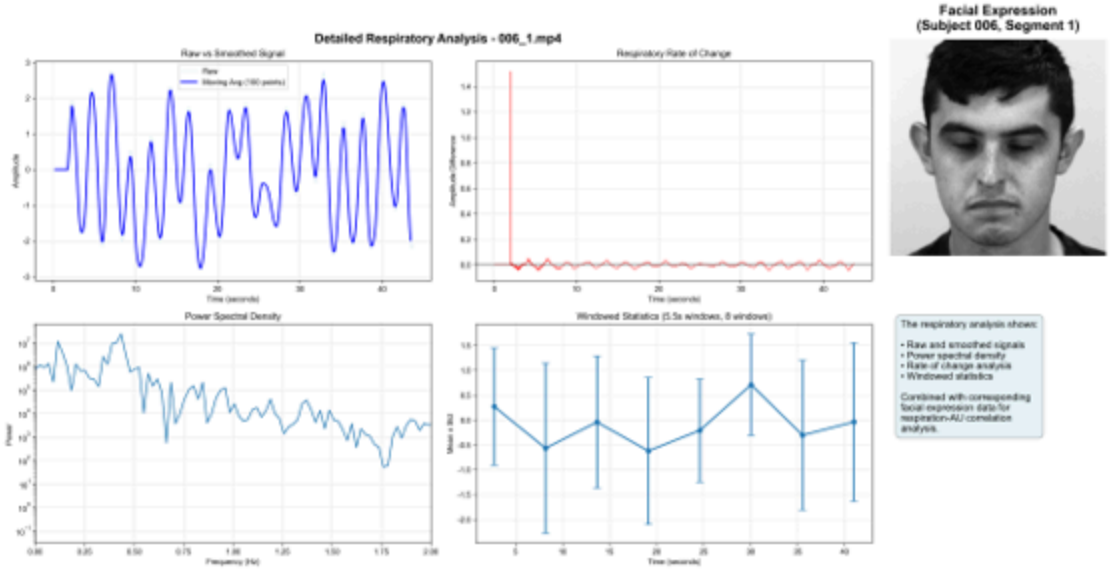


Figure 1

Representative respiratory signal analysis (Subject 006, Segment 1, SAMM Long Videos dataset). Non-contact respiratory extraction pipeline applied to a video segment. (Top-left) Raw and smoothed respiratory signals reveal clear periodic breathing cycles. (Top-right) Respiratory rate of change, derived from amplitude differences, highlights local variations across the segment. (Bottom left) Power spectral

density portrays dominant energy concentrated in the 0.1–0.5 Hz frequency band, which is consistent with natural respiration. (Bottom-right) Windowed statistical analysis (5.5 s windows) summarizes temporal variability in breathing amplitude and stability. Respiratory cycles were typically 3–5 s with variations in amplitude and frequency that correlate with emotional expression events.

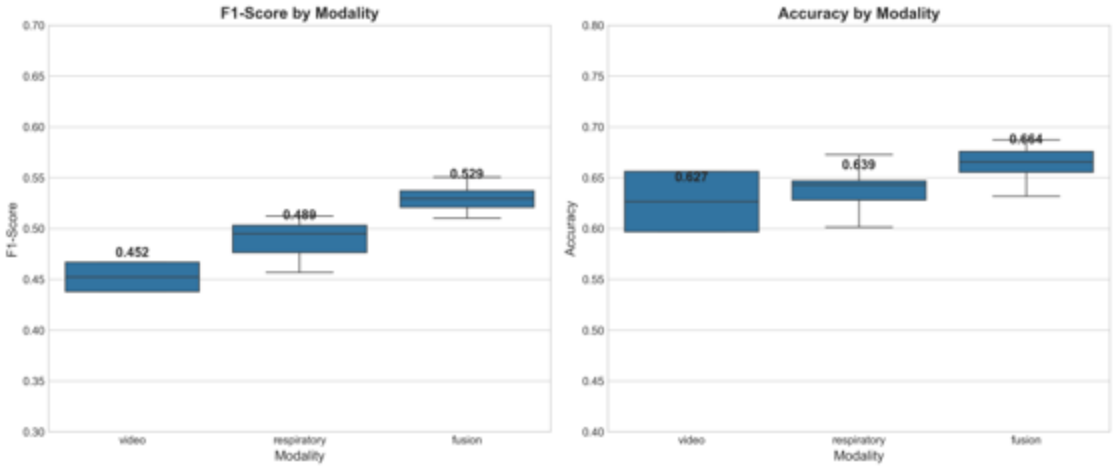


Figure 2

Performance comparison across modalities.
Box plots show F1-score and accuracy distributions for video-only (14 features), respiratory-only (37 features), and multimodal fusion (51 features) approaches. Statistical testing suggests fusion outperforms individual modalities ($p < 0.001$, Wilcoxon signed-rank test with Benjamini–Hochberg correction). Error bars represent standard deviation across 18 experimental configurations.

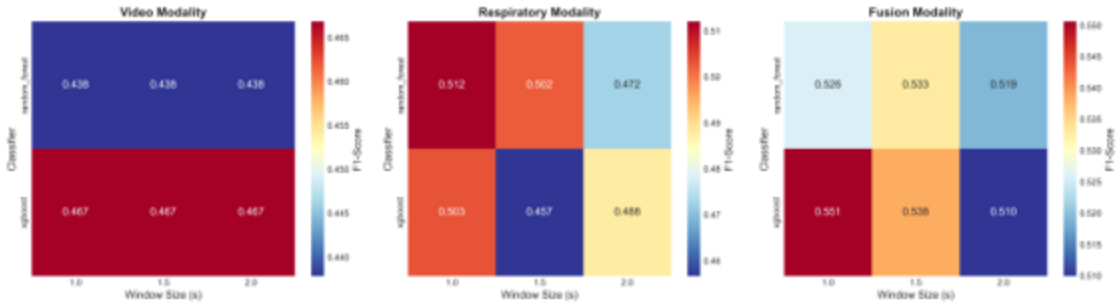


Figure 3

Performance heatmap by classifier and window size.

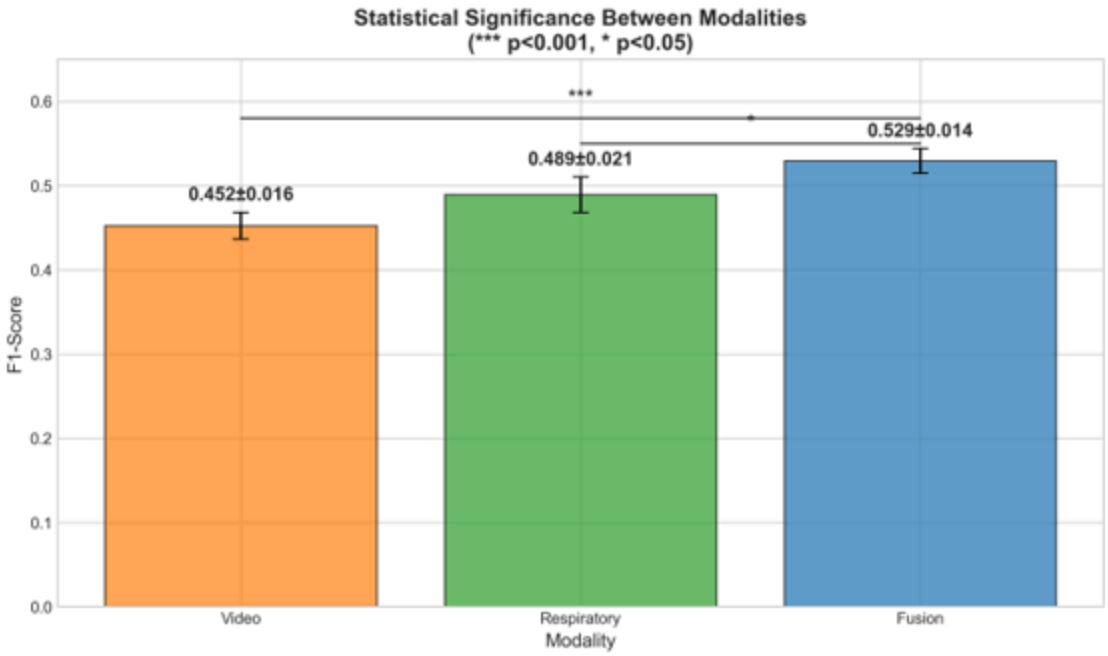


Figure 4

Statistically significant difference between modalities.
Bar chart with error bars showing mean F1-scores ± standard deviation. Significance indicators show results of Wilcoxon signed-rank tests (*** p<0.001, * p<0.05). Fusion appears to achieve statistically significant improvements over both individual modalities.

The statistical comparisons of the effect sizes and confidence intervals are presented in Table 7.

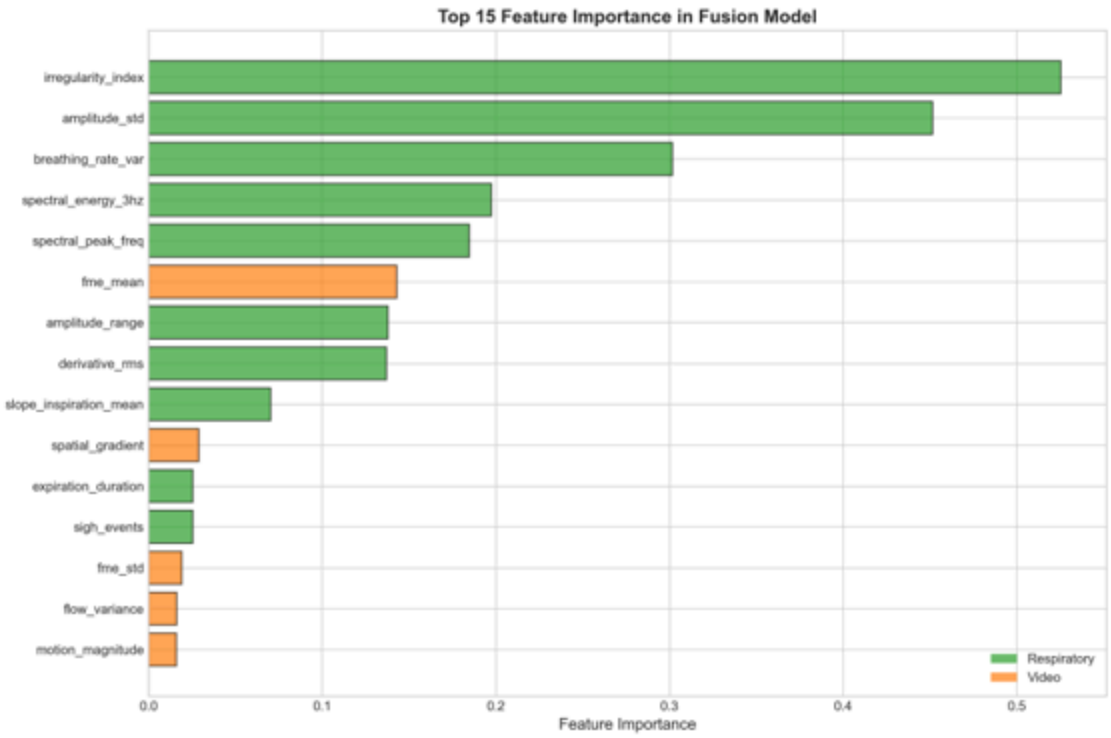


Figure 5

Feature importance in fusion model.
Horizontal bar chart showing top 15 most important features from the 51-dimensional fusion feature space (37 respiratory + 14 video features). Respiratory features (green) dominate the top rankings, demonstrating their critical role in improving classification performance. Video features (orange) provide complementary information at lower importance levels.

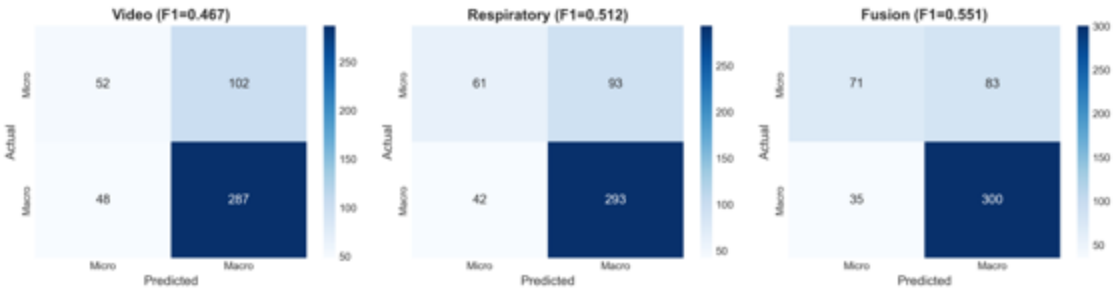


Figure 6

Feature importance in fusion model.
Horizontal bar chart showing top 15 most important features from the 51-dimensional fusion feature

space (37 respiratory + 14 video features). Respiratory features (green) dominate the top rankings, demonstrating their critical role in improving classification performance. Video features (orange) provide complementary information at lower importance levels.