DETECCIÓN Y BIOMARCADORES

Facial biomarkers of depression: from action units to WhatsApp-based detection

SentirIA Research Papers · 2026

Infraestructura de detección temprana en salud mental

Este documento es parte de la base científica de SentirIA,
plataforma de detección temprana y monitoreo continuo de deterioro en salud mental.
No constituye diagnóstico clínico. La evaluación es responsabilidad del profesional.

Facial biomarkers of depression: from action units to WhatsApp-based detection

Automated facial analysis can detect depression with 78–92% accuracy by tracking specific facial action units, gaze patterns, blink dynamics, and smile authenticity — but deploying this in a WhatsApp-based system like SentirIA requires navigating significant technical and ethical constraints. A growing body of research since 2013 has established that depression produces measurable changes in facial behavior, anchored in the Facial Action Coding System (FACS). These signals, when combined with vocal, textual, and behavioral biomarkers in multimodal pipelines, now approach clinical utility. Open-source tools like OpenFace 3.0 and MediaPipe make real-time extraction feasible even on mobile devices, though WhatsApp's compressed video introduces quality trade-offs that demand careful preprocessing and longitudinal aggregation. This report synthesizes the clinical evidence, available tools, and practical feasibility for building a depression detection pipeline from WhatsApp data.

The six action units that betray depression

The Facial Action Coding System, developed by Paul Ekman and Wallace Friesen in 1978, decomposes facial expressions into anatomically defined Action Units mapped to specific muscles. Research led by Jeffrey Cohn and Jeffrey Girard at Carnegie Mellon and the University of Pittsburgh has identified a core set of AUs whose activation patterns reliably distinguish depressed from non-depressed individuals.

AU1 (inner brow raise, frontalis pars medialis) ranks as the single most depression-associated AU across multiple studies. Parikh et al. (2024) found AU1 mean intensity was consistently higher in depressed participants over time in the EMPKINS dataset. A Japanese study confirmed AU1 presence was significantly elevated in subthreshold depression (BDI-II 11–20, $q < 0.05$ after FDR correction). AU4 (brow lowerer, corrugator supercilii and depressor supercilii) shows a distinctive pattern: EMG studies dating to Schwartz et al. (1976) and Greden et al. (1986) documented elevated *baseline* corrugator tension in depressed patients even at rest, while Rottenberg et al. (2005) found *reduced reactivity* of the same muscle to emotional stimuli. This paradox — high tonic tension but blunted phasic response — is a hallmark electromyographic signature of MDD.

AU12 (lip corner puller, zygomaticus major) produces the most robust single finding: smiling is consistently reduced in depression. Girard, Cohn, Mahoor et al. (2014) demonstrated a dose-dependent relationship in 33 adults tracked through a clinical trial — as depression severity increased, AU12 decreased, and when depression remitted, smiling recovered. AU14 (dimpler, buccinator) moves in the opposite direction, increasing with depression severity. This AU is associated with contempt, and its co-occurrence with residual smiles during severe depression produces what Girard et al. (2013) described as contempt-laced smiles — a hallmark of the Social

Risk Hypothesis proposed by Allen and Badcock (2003).

AU15 (lip corner depressor, depressor anguli oris) presents complex findings. Cross-sectionally, depressed patients show higher AU15 frequency than controls. But Girard et al. (2014) found that *within individuals*, AU15 paradoxically decreased when depression was most severe — consistent with Emotion Context Insensitivity, where even sadness expressions are dampened alongside happiness. AU17 (chin raiser, mentalis) co-activates with AU1 and AU4 in compound sadness-distress expressions and appears in disgust micro-expressions detected in depressed individuals. Additional depression-relevant AUs include AU10 (upper lip raiser, associated with disgust), AU7 (lid tightener), AU20 (lip stretcher), and AU26 (jaw drop), all found elevated in automated detection studies using LSTM networks.

Classification accuracy using AU features ranges from 79% (Cohn et al., 2009, automated) to 91.67% (LSTM with attention, 2023). Dibeklioğlu, Hammal, and Cohn (2018) achieved 78–85% using facial and head movement dynamics combined, while Tzirakis et al. (2019) reached 87.2% using a three-layer architecture for DASS prediction.

Flat affect, emotional inertia, and the dynamics of depressed faces

Depression alters not just *which* expressions occur but *how* they unfold temporally. The distinction between static expression snapshots and dynamic facial behavior is critical for automated detection.

Bylsma, Morris, and Rottenberg's 2008 meta-analysis established a medium effect size (d = 0.53) for reduced positive emotional reactivity in MDD and a smaller effect (d = 0.25) for reduced negative reactivity. This asymmetry is explained by Rottenberg's Emotion Context Insensitivity (ECI) theory: depression systematically dampens reactivity to *all* emotional stimuli, not just positive ones, but positive expressions are affected more severely. Gehricke and Shapiro (2000) confirmed via EMG that depressed patients show reduced facial muscle activity during both happy *and* sad imagery — evidence of psychomotor retardation affecting the face. This retardation, present in 50–75% of MDD patients, produces the clinical observation of flat or masked affect.

The temporal dynamics of facial movement carry stronger diagnostic signal than static features. Dibeklioğlu et al. (2018) found that facial movement dynamics (onset velocity, offset velocity, duration, amplitude) were more predictive of depression severity than head movement or vocal prosody. The SFTNet study (2023) analyzing 156 participants found that depressed subjects produced expressions with average duration under 20 frames (~667ms) and showed less expression variety, achieving 87.3% detection accuracy from micro-expression analysis alone. Depressed individuals also exhibit higher emotional inertia — their affects are "stickier," showing stronger autocorrelation and less responsiveness to environmental changes.

Research on concealed depression using micro-expressions (lasting 40–500ms) reveals that involuntary negative facial expressions leak through attempted neutral displays. Chen and Luo (2023) found strong correlations between underlying depression and negative micro-expressions, with the eyebrow and lip regions being most informative. However, micro-expression detection from standard 30fps video is marginal (1–6 frames per expression), making macro-level expressivity metrics more practical for real-world deployment.

Where depressed eyes look — and don't look

Gaze aversion in depression is one of the most consistently replicated nonverbal findings in psychiatry, with evidence spanning four decades from clinical observation to modern eye-tracking.

Depressed individuals show reduced fixation on eye regions of faces, particularly during sustained attention. A 2024 BMC Psychiatry study with 93 participants found this effect was specific to late/sustained attention rather than initial orienting — depressed individuals orient normally but fail to maintain gaze on socially relevant stimuli. Armstrong and Olatunji's meta-analysis confirmed reduced gaze maintenance on positive stimuli and increased maintenance on negative stimuli, while Lazarov et al. (2018) found both high-depressive students and clinically diagnosed MDD patients dwelled longer on sad faces compared to happy faces, whereas non-depressed individuals showed a strong happy-face preference. Huang and Zhu's 2022 meta-analysis quantified these effects: large effect size (SMD > 0.8) for fixation count on positive stimuli and medium effect sizes for fixation duration on both positive and negative stimuli.

Automated gaze-based depression detection has produced promising results. A multimodal classifier using POV glasses achieved 84.7% accuracy (sensitivity 90.9%, specificity 78%, AUC 0.89) using gaze, facial expressions, and head movement. Zhang et al. reached 80.1% accuracy using random forest classification of pupil size, gaze position, and gaze duration. An AI meta-analysis in npj Digital Medicine (2025) found that multimodal methods incorporating gaze achieved a pooled AUC of 0.95, substantially outperforming unimodal approaches (0.84–0.92).

The relationship between blink rate and depression is mediated by dopamine. Karson's foundational 1983 study established that spontaneous eye blink rate (normal baseline: 15–24 blinks/minute) correlates with dopaminergic activity, validated by Taylor et al. (1999) who demonstrated strong correlation between blink rate and caudate dopamine concentration. In depression, findings are nuanced: Ebert et al. (1996) found basal blink rate was *not different* in non-retarded depression but that sleep deprivation (which has antidepressant effects) increased blink rate proportionally to mood improvement, implicating dopamine activation. The connection between reduced dopaminergic transmission, psychomotor retardation, and blink rate suggests blink metrics may be most informative for melancholic depression subtypes with prominent motor slowing.

Why not all smiles are created equal

The distinction between Duchenne smiles (genuine, involving orbicularis oculi pars lateralis producing AU6 + zygomaticus major producing AU12) and non-Duchenne smiles (social, AU12 only) has been central to depression research since Ekman estimated that only ~10% of people can voluntarily produce AU6.

Girard et al.'s landmark longitudinal studies (2013, 2014) demonstrated that during high depression severity, participants smiled less overall (reduced AU12), and smiles that did occur were more likely contaminated by AU14 (contempt) — producing non-genuine, ambivalent expressions. As treatment progressed and symptoms decreased, both smile frequency and Duchenne smile proportion increased. Reed et al. confirmed that negative-affect smiles (smiles co-occurring with negative emotion AUs) were more common in currently depressed participants. Automated detection of this distinction achieves 87% accuracy using Gabor filters with linear SVM on controlled datasets, while human untrained observers perform at near-chance levels (53–66%).

However, Girard et al.'s own later work (2019/2021) challenged the simplicity of the Duchenne marker through the "artifact hypothesis": AU6 may partly result from mechanical coupling with intense AU12 rather than representing an independent positive emotion signal. When controlling for AU12 intensity, AU6's predictive power for genuine positive emotion was substantially reduced. This suggests that depression detection should not rely solely on the AU6 presence/absence binary but should incorporate smile intensity, duration, onset/offset dynamics, and co-occurring action units. Smile duration may be more diagnostic than the Duchenne marker per se — training observers on mouth movement duration improved authenticity discrimination more than training on AU6 detection.

Open-source tools for building a facial depression detector

Six tools merit serious consideration, each with distinct strengths for depression-relevant facial analysis.

OpenFace 2.0/3.0 (Baltrušaitis, CMU) remains the research gold standard. Version 2.0 detects 18 AUs with both presence and intensity outputs, achieving mean AUC of 0.81 on posed data and F1 ~50% on in-the-wild DISFA data. The newly released OpenFace 3.0 (2025) substantially improves performance: F1 of 60% on DISFA and 62% on BP4D, gaze angular error of 2.56° (beating prior SOTA), and 0.60 accuracy on AffectNet emotion recognition — all in a unified 29.4M-parameter model running at 26 FPS on CPU. Critical limitation: non-commercial license with commercial rates of $10,000–18,000/year.

MediaPipe Face Landmarker (Google) offers the strongest mobile deployment story. Its 478 3D landmarks and 52 blendshape scores (loosely mapping to FACS AUs) run at ~165+ FPS on GPU with full on-device processing. Under Apache 2.0 licensing with first-class Android/iOS/web support, it is the clear choice for privacy-sensitive mobile applications. The blendshape-to-AU mapping is approximate and lacks formal FACS validation, but scores like browDownLeft/Right, mouthFrownLeft/Right, and cheekSquintLeft/Right provide depression-relevant signal.

LibreFace (USC/ICT, WACV 2024) achieves the highest AU intensity accuracy — Pearson correlation 0.63 on DISFA, 7% higher than OpenFace 2.0, running 2x faster. It detects 12 AUs plus 8 emotions including contempt, using a MAE-pretrained ResNet-18. py-feat (MIT license) provides the most complete research pipeline: 20 AUs, modular backends, built-in statistical analysis, and integration of MediaPipe blendshapes, though it's too slow for real-time use. DeepFace (Serengil, MIT license) is easiest to integrate but provides only emotion-level classification at ~57% accuracy on FER2013 — below the 65% human baseline and unsuitable for clinical use without model replacement.

Tool	Depression AUs	Best AU Metric	Real-Time	Mobile	License
OpenFace 3.0	AU1,4,6,12,14,15,17 ✓	F1 62% (BP4D)	26 FPS CPU	No	Research only
MediaPipe	Via 52 blendshapes	Not benchmarked	165+ FPS	Yes	Apache 2.0
LibreFace	12 AUs + contempt	PCC 0.63 (DISFA)	Moderate	No	Non-commercial
py-feat	20 AUs	Comparable to OF2	Offline only	No	MIT
DeepFace	Emotions only	57% (FER2013)	Variable	Limited	MIT

Can WhatsApp video messages reveal depression?

WhatsApp video messages encode at 720p resolution (1280×720), H.264 codec, 30fps, within a 16MB file limit — specifications that constrain but do not preclude facial biomarker extraction.

At 720p, face regions in selfie-style video messages typically span 200–400+ pixels, well above the 64×64 pixel critical threshold for landmark detection. Research shows 4x downsampling from 256×256 to 64×64 does not significantly affect recognition accuracy. Face detection in MPEG-compressed video achieves 85–92% detection rates even operating on DCT coefficients. The primary challenges are not resolution or compression but rather lighting variability in user-generated content, head pose variation, and the 30fps frame rate being marginal for micro-expression capture (yielding only 1–6 frames per micro-expression at 40–200ms duration).

The 60-second duration cap on WhatsApp video notes is a meaningful limitation. Multi-timescale

feature extraction research argues that complete recordings capturing minute-level dependencies outperform short clips. However, the DEPAC corpus study found speech samples under one minute provide sufficient signal when content is appropriate, and depression-relevant macro-features (overall expressivity, AU frequency distributions, smile rates) can be extracted from short clips. The key architectural insight is that longitudinal aggregation across multiple messages over weeks compensates for per-message brevity — tracking within-subject trends rather than making single-clip diagnoses.

Profile photo analysis offers a complementary low-frequency signal. Reece and Danforth's landmark 2017 study of 43,950 Instagram photos demonstrated that depressed users' photos were bluer, grayer, and darker, with a preference for black-and-white filters, achieving 70% detection accuracy — outperforming general practitioners' 42% diagnostic rate. Depressed users posted more photos with faces but fewer faces per photo. While no published research specifically examines WhatsApp profile photos, the same computational features (hue, saturation, brightness, face presence, expression) are extractable, and longitudinal tracking of change frequency could provide behavioral signal aligned with social withdrawal patterns.

A preprocessing pipeline for WhatsApp video should include: (1) face detection via MTCNN or RetinaFace, (2) landmark-based alignment to canonical position, (3) quality filtering to discard occluded/blurred/extreme-pose frames, (4) AU extraction via OpenFace 3.0 or MediaPipe blendshapes, (5) per-subject normalization to establish individual baselines, and (6) temporal aggregation of statistical summaries across each message and across messages over time.

Building SentirIA: multimodal fusion and the ethical minefield

A WhatsApp-based depression detection system should treat facial biomarkers as one modality within a multimodal pipeline — and likely not the primary one. Text-based models frequently outperform or match multimodal systems because depression symptoms are "directly articulated in language" (Nature, 2024). Sadeghi et al. achieved MAE of 2.85 on PHQ-8 using LLM-enhanced text features, while multimodal voice-text fusion reached F1 of 96.66% in classification. The AVEC challenge benchmarks show consistent improvement from multimodal approaches, with recent systems achieving RMSE below 4.0 for PHQ-8 regression.

The optimal fusion architecture for WhatsApp is late fusion or attention-based fusion, since modalities arrive asynchronously (a text message Monday, a voice note Wednesday, a video message Friday) and any modality may be absent on a given day. An LSTM with attention mechanism using intermediate feature fusion achieved 91.67% accuracy for binary depression classification. The system should weight modalities by reliability and availability:

- Text messages (primary): always available, richest depression signal via sentiment, vocabulary shifts, and response patterns

- Voice messages (secondary): well-validated vocal biomarkers including reduced pitch variability, slower speech rate, longer pauses, and elevated jitter/shimmer
- Video messages (supplementary): facial AU distributions, gaze patterns, and expressivity metrics, degraded by compression but valuable longitudinally
- Behavioral metadata (passive): response latency, message frequency, active hours, typing dynamics — BiAffect research showed phone movement during typing correlates with anhedonia ($\beta = -0.12$, $p < 0.001$)
- Profile photos (low-frequency): color, brightness, face presence, and change frequency

Ethical and regulatory challenges are substantial. False positives risk harmful labeling and unnecessary clinical intervention, while false negatives create liability. Mental health data falls under GDPR Article 9 protections and HIPAA requirements. A depression screening tool would likely qualify as a Class II medical device under FDA SaMD classification (requiring 510(k) clearance) and Class IIa under EU MDR Rule 11. Only 7 of 84 potential AI mental health clinical decision support tools met full inclusion criteria in a recent product review, reflecting an immature regulatory landscape. WhatsApp's end-to-end encryption means analysis must occur on-device or with explicit user consent for data sharing — making MediaPipe's on-device inference capability particularly valuable.

Conclusion

The evidence base for facial biomarkers of depression is robust: AU1 and AU4 elevation, AU12 reduction, AU14 increase, and dampened overall expressivity form a reliable signature detectable by automated systems at 78–92% accuracy in controlled settings. The transition from laboratory to WhatsApp introduces degradation from compression, lighting, and duration constraints that likely reduce accuracy to an estimated 65–80% for facial modalities alone. The practical path forward for SentirIA is a multimodal late-fusion architecture where text serves as the primary signal, voice as secondary, facial features as supplementary longitudinal markers, and behavioral metadata as a passive monitoring layer. The most actionable technical choice is MediaPipe for on-device facial feature extraction (mobile-first, Apache 2.0, privacy-preserving) with custom blendshape-to-AU mapping validated against OpenFace 3.0 outputs. The most consequential non-technical choice is whether to pursue medical device classification — which determines everything from required clinical validation sample sizes to permissible deployment contexts and the fundamental question of whether the system provides information, screening, or diagnosis.