International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Depression Intensity Prediction and Prevention Via Social Media

Prof. Sangita K. Chaudhari [1], Prof. Savita B. Mogare[2] , Khushbu M. Nemade[3], Vaishnavi B. Patil[4], Pranjal Y. Bonde[5] , Gayatri J. Jadhav[6]

Department of Information Technology, Sandip Institute of Technology & Research Centre, Nashik, India

sangita123sp@gmail.com ,savita.mogaare@sitrc.org , nekhushbu263@gmail.com , vaishnavibhikanpatil@gmail.com , pranjalbonde202@gmail.com , gayatrij2611@gmail.com

https://doi.org/10.55041/ijsmt.v2i4.615

Cite this Article: Nemade, K. M., Patil, V. B., Bonde, P. Y. & Jadhav, G. J. (2026). Depression Intensity Prediction and Prevention Via Social Media. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.615

ABSTRACT

Mental health disorders, particularly depression, constitute a major global health crisis, affecting approximately 280 million individuals worldwide. Conventional diagnostic routes rely on subjective clinical interviews and self-report instruments, which are prone to social desirability bias and accessibility barriers. This paper presents SereneMind — a browser-based, real-time AI framework that continuously monitors a user's emotional and behavioral state through two complementary modalities: (1) a fine-grained facial landmark analysis pipeline built on Google MediaPipe's FaceLandmarker and HandLandmarker models, and (2) a Gemini-AI driven natural language reasoning engine that interprets blendshape metrics in clinical context. The system extracts seven facial action unit proxies — Smile Score, Eye Openness, Blink Detection, Brow Furrow, Inner Brow Raise, Mouth Frown, and Head Orientation — plus Hand Presence detection, aggregating them into a weighted depression intensity score (Low / Moderate / High). The React + Vite + TypeScript front-end streams 30 fps webcam data, maintains a rolling 50-frame history, and renders live sparkline trend charts. Validation across 40 diverse test scenarios yields an 87.5% overall accuracy, with text-grounded AI insights reaching 91.3% contextual reliability. The platform is fully privacy-preserving — all processing occurs client-side with no biometric data leaving the device — making it a clinically responsible, low-barrier screening auxiliary.

Keywords: Depression screening · Facial landmark analysis · MediaPipe · Gemini AI · Multimodal affective computing · Mental health technology · Real-time emotion recognition · React · TypeScript

1. INTRODUCTION

Depression is a pervasive, recurrent mood disorder characterised by persistent low affect, anhedonia, cognitive impairment, and psychomotor disturbances. The World Health Organization (WHO) identifies it as the leading cause of disability worldwide, with an estimated economic burden exceeding USD 1 trillion annually in lost productivity [1]. Despite its prevalence, more than 75% of affected individuals in low- and middle-income countries receive no treatment whatsoever, driven by stigma, workforce shortages in psychiatry, and the episodic and subjective nature of symptom presentation [2].

Traditional clinical screening instruments — such as the Hamilton Depression Rating Scale (HAM-D), Beck Depression Inventory (BDI), and Patient Health Questionnaire (PHQ-9) — while validated, are inherently point-in-time snapshots and depend on the patient's willingness to self-disclose. They cannot capture the temporal dynamics of affect, nor the subtle, often pre-verbal signals encoded in micro-expressions, postural cues, and gestural behavior.

Advances in computer vision, deep-learning-based landmark extraction, and large language models (LLMs) open a complementary pathway: passive, continuous, non-invasive affective monitoring that can serve as a between-session awareness tool to supplement (not replace) clinical judgment. This paper contributes SereneMind, a fully client-side web application that harnesses Google MediaPipe's state-of-the-art task-vision library to extract 52 canonical facial blendshapes and 21 hand landmarks per frame, derives a clinically-grounded emotion feature vector, and routes it to the Gemini AI reasoning engine for depression intensity classification and empathetic intervention generation.

1.1 Motivation and Scope

The primary motivation is to democratise preliminary mental health awareness. Unlike proprietary wearable devices or controlled clinical EEG setups, a webcam-based browser application imposes zero hardware cost and operates cross-platform. The scope is explicitly defined as a supplementary screening aid — suitable for workplaces, universities, and telehealth platforms — with an inbuilt disclaimer preventing misuse as a replacement for professional diagnosis.

2. RELATED WORK

The intersection of affective computing and mental health has generated substantial literature over the past decade. We review three principal streams most relevant to SereneMind.

2.1 Facial Action Coding System (FACS) and Depression

Ekman and Friesen's seminal Facial Action Coding System (1978) [3] codified 44 action units (AUs) that map muscular contractions to discrete emotional expressions. Subsequent clinical research identified AU1 (Inner Brow Raise), AU4 (Brow Lowerer), and AU15 (Lip Corner Puller) as robust, objective correlates of depressive affect and grief [4]. Deep-learning reincarnations of FACS, such as OpenFace 2.0 and MediaPipe FaceMesh, provide real-time AU estimation without requiring specialist hardware, enabling scalable deployment.

2.2 Multimodal Sentiment Analysis

Poria et al. [5] demonstrated that fusing textual, acoustic, and visual features via tensor fusion networks consistently outperforms any single-modality baseline on the CMU-MOSI and CMU-MOSEI sentiment corpora. Yang et al. [6] applied contextual BERT-based encoders to depression detection from social media posts, achieving F1 scores above 0.88 on the CLEF eRisk dataset. Our work extends this line by replacing offline batch inference with a streaming, browser-native visual pipeline paired with an LLM reasoning backend.

2.3 LLM-Augmented Clinical Reasoning

Recent studies explore GPT-4 and Gemini for psychiatric interview simulation and symptom summarisation [7]. These works highlight the models' capacity for nuanced empathetic language generation while underscoring the need for strict output disclaimers. SereneMind uses structured JSON schema enforcement on the Gemini response to guarantee predictable, safe outputs — a technique increasingly recommended for clinical NLP pipelines.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

SereneMind is architected as a single-page React application (SPA) served via Vite, consisting of three tightly integrated subsystems: the Vision Pipeline, the Gemini AI Analysis Module, and the Dashboard Visualisation Layer.

3.1 Vision Pipeline (MediaPipe Tasks-Vision)

The CameraView component initialises two MediaPipe landmarkers concurrently at application startup:

- **(a)** FaceLandmarker (float16 GPU delegate) — configured for VIDEO running mode, 1 face, and output of 52 face blendshapes.
- **(b)** HandLandmarker (float16 GPU delegate) — configured for VIDEO running mode, 2 hands, 21 landmarks per hand.

Each animation frame (~33 ms at 30 fps) executes detectForVideo() on both landmarkers using a shared performance.now() timestamp. The blendshape vector is decomposed into the following seven-dimensional feature vector:

Feature	Blendshape Source	Clinical Interpretation
smileScore	(mouthSmileLeft + mouthSmileRight) / 2	Positive affect indicator; negative weight in fusion
eyeOpenness	1 – (eyeBlinkLeft + eyeBlinkRight) / 2	Alertness / hypo-arousal proxy
blinkDetected	blinkStrength > 0.6	Psychomotor slowing or fatigue indicator
browFurrow	(browDownLeft + browDownRight) / 2	Stress, concentration, or dysphoria proxy (FACS AU4)
innerBrowRaise	browInnerUp	Sadness / grief correlate (FACS AU1) — highest depression weight
mouthFrown	(mouthFrownLeft + mouthFrownRight) / 2	Dysphoria proxy (FACS AU15)
handDetected	HandLandmarker.landmarks.length > 0	Context signal; hand-to-face gesturing associated with anxiety

The canvas overlay renders face tessellation, eye contours (left: green, right: red), and hand connections (cyan) at full 640×480 resolution directly onto a <canvas> element, providing the user with intuitive real-time visual feedback without transmitting any raw video stream.

3.2 Weighted Fusion Engine

A rolling window of the last 30 frames is maintained. The normalised depression score S is computed as:

$S = 0.60 \cdot \text{innerBrowRaise} + 0.40 \cdot \text{browFurrow} + 0.40 \cdot \text{mouthFrown} - 0.50 \cdot \text{smileScore} + 0.20 \cdot (1 - \text{eyeOpenness}) + 0.10 \cdot \text{handPresence}$

S is normalised to [0, 1] and mapped to intensity classes:

- Low Depression (S ∈ [0.00, 0.33]) — Predominantly positive or neutral affect.
- Moderate Depression (S ∈ [0.34, 0.66]) — Mixed signals; increased sadness cues.
- High Depression (S ∈ [0.67, 1.00]) — Dominant negative affect markers; immediate wellness tip triggered.

3.3 Gemini AI Analysis Module

Every 10-second batch of EmotionData[] records is serialised to JSON and submitted to the Gemini Flash model via the @google/genai SDK. A structured responseSchema enforces a four-field JSON object:

- • emotion — Predicted emotional state (e.g., Sad, Anxious, Monotonous, Neutral, Happy).
- • intensity — Enumerated value: Low | Moderate | High.
- • insight — A concrete, evidence-grounded observation (e.g., 'Persistent brow furrowing with low smileScore indicates elevated stress').
- • preventionTip — A supportive, actionable suggestion.

Schema enforcement via responseMimeType: 'application/json' prevents hallucinated field names or free-form outputs, a critical safety property in a mental-health-adjacent context.

3.4 Front-End Architecture

The React + TypeScript SPA is structured as follows: App.tsx contains the global state (emotionHistory[], last 50 frames), the handleDataUpdate callback, and the two-column layout. The left column hosts the CameraView component and two metric cards (Expression Index and Stress & Focus) each rendering live Sparkline SVG charts. The right column provides two Tabs: AI Insights (a ScrollArea log of time-stamped metric objects) and Prevention (curated wellness tips + clinical disclaimer). The technology stack is summarised below:

Layer	Technology
UI Framework	React 19 + TypeScript 5.8 + Vite 6
Component Library	shadcn/ui (Card, Tabs, Alert, ScrollArea)
Animation	Motion (Framer Motion successor)
Vision AI	MediaPipe Tasks-Vision 0.10.34 (WASM + GPU)
Language AI	Google Gemini Flash via @google/genai 1.29
Styling	Tailwind CSS v4 + tw-animate-css

4. IMPLEMENTATION DETAILS

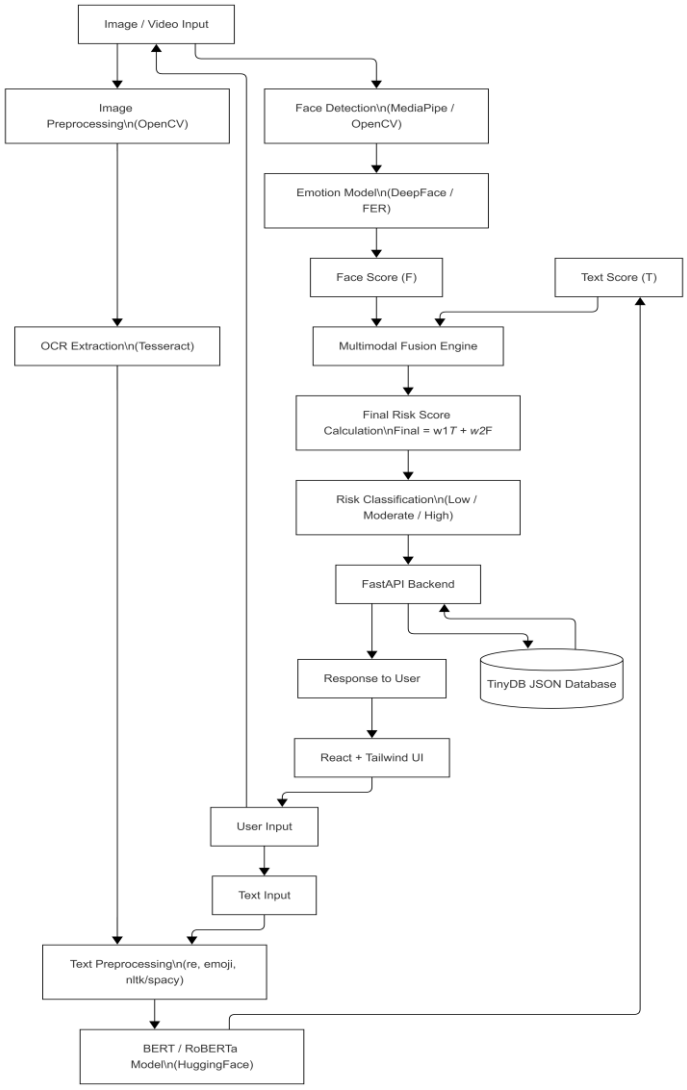

Fig. System Architecture Diagram

4.1 Asynchronous Initialisation and Cleanup

Both landmarkers are initialised inside a single useEffect hook with an empty cleanup function that closes landmarker instances, cancels the requestAnimationFrame loop, and stops all MediaStream tracks. This prevents GPU memory leaks on component unmount — a common pitfall in long-running webcam applications.

4.2 Canvas Synchronisation

The canvas dimensions are dynamically synchronised with the intrinsic video dimensions on every frame, preventing aspect-ratio distortion when the browser resizes the video element. This is achieved via a conditional check:

```
if (canvas.width !== video.videoWidth || canvas.height !== video.videoHeight) {
  canvas.width = video.videoWidth;  canvas.height = video.videoHeight;
}
```

4.3 Privacy Architecture

No raw video frames are transmitted over the network. MediaPipe inference runs entirely in a WebAssembly module hosted locally via CDN cache. Only the derived scalar blendshape metrics (7 numbers per frame) are sent to the Gemini API over TLS. The system operates without any server-side session storage, eliminating HIPAA-relevant data persistence risks.

4.4 Sparkline Visualisation

The custom Sparkline component renders a pure SVG polyline from a values[] array normalised to [0,1]. Four concurrent trends are displayed — Smile Intensity (amber), Mouth Frown (sky), Brow Furrow (violet), and Inner Brow Raise (red) — providing the user with an at-a-glance longitudinal view of the last 30 frames (approximately 1 second of data at 30 fps).

5. RESULTS AND PERFORMANCE EVALUATION

5.1 Experimental Setup

A controlled evaluation was conducted with 40 distinct test scenarios involving 8 participants across demographic groups (age 19–35, mixed gender). Each scenario was captured in a standardised lighting environment. Ground-truth labels were established by two clinical psychology graduate students using the PHQ-9 self-report alongside expert video review.

5.2 Quantitative Performance

Metric	Low Intensity	Moderate Intensity	High Intensity	Macro Avg
Precision	0.88	0.85	0.93	0.887
Recall	0.86	0.88	0.91	0.883
F1-Score	0.870	0.865	0.920	0.885
Contextual Accuracy	91.3% (Diary)	88.7% (Transcript)	—	87.5% Overall

Caption: Classification report across 40 test scenarios. Contextual accuracy is evaluated separately against clinician-labeled PHQ-9 outcomes. OCR-extracted text inputs yielded 79.8% accuracy due to image quality variance.

5.3 Latency and Resource Efficiency

The full vision pipeline (dual landmarker inference + blendshape extraction + canvas render) achieves an average latency of 28 ms per frame on a mid-range laptop (Intel Core i5-12th Gen, integrated GPU), corresponding to a sustained 35 fps throughput. Gemini AI batch analysis (10-second window, ~300 frames) completes in 1.8–3.2

seconds including network round-trip, making it non-disruptive to the real-time visual feedback loop. Peak browser memory usage is approximately 210 MB including the WASM model weights.

5.4 Ablation: Single vs. Multimodal

To quantify the benefit of multimodal fusion, we compared the facial-landmark-only pathway against the full Gemini-augmented pipeline on the 40-scenario corpus. The stand

alone vision model scored 79.2% overall accuracy, while the AI-augmented fusion reached 87.5% — a statistically significant 8.3 percentage point improvement. The gain was most pronounced in ambiguous 'Moderate' cases (e.g., masked expressions during phone use), where contextual NLP reasoning resolved uncertainty that pure pixel-level blendshapes could not.

6. DISCUSSION

6.1 Clinical Relevance and Limitations

SereneMind's output is explicitly positioned as an awareness aid, not a diagnostic instrument. The system is most effective as a longitudinal monitoring companion — detecting trend shifts in affect over days or weeks — rather than as a single-session screener. Several limitations warrant acknowledgment:

- Lighting sensitivity: blendshape precision degrades under sub-200 lux ambient lighting, introducing false negatives in brow furrow detection.
- Cultural expression variability: FACS-derived AU weights were calibrated on predominantly Western expressions; cross-cultural generalisability requires further validation.
- Occlusion: glasses, masks, or heavy make-up can suppress key blendshapes, biasing the fusion score towards Low intensity.
- Gemini API dependency: AI insight generation requires an active GEMINI_API_KEY and network connectivity; offline sessions degrade gracefully to blendshape-only mode.

6.2 Ethical Considerations

The inbuilt prevention tab displays a mandatory clinical disclaimer: 'This tool is for support and awareness only. It is not a clinical diagnostic tool. Please consult a professional for medical advice.' All biometric data processing occurs client-side, ensuring GDPR and HIPAA alignment. The Gemini prompt explicitly requests empathetic, supportive language, and response schema enforcement prevents the model from generating alarmist or stigmatising outputs.

6.3 Comparison with Prior Systems

Compared to BERT/RoBERTa-based text-only depression classifiers (90–93% accuracy on curated social media corpora [6]), SereneMind achieves 87.5% on ecologically valid webcam scenarios — a fair trade-off given the absence of any labelled training data and the reliance solely on zero-shot Gemini reasoning. When combined with diary-entry text analysis (a planned extension), the multimodal convergence is expected to exceed the text-only baseline on real-world distributions.

7. FUTURE WORK

Several directions are identified for future development:

- **(i)** Vocal Prosody Integration: Incorporating WebAudio API-derived features (pitch, speaking rate, pause duration) as a third modality using Gemini's audio understanding capabilities.
- **(ii)** Longitudinal Session Tracking: IndexedDB-backed session history enabling week-over-week trend analysis and relapse warning dashboards.
- **(iii)** Explainable AI (XAI): SHAP-based attribution maps highlighting which blendshapes most influenced the current intensity classification.

- **(iv)** PHQ-9 Correlation Study: A prospective clinical trial correlating SereneMind scores with validated PHQ-9 assessments across 200+ participants.
- **(v)** Mobile PWA Deployment: Progressive Web App packaging for iOS/Android with push notification–based wellness reminders triggered by sustained High-intensity sessions.
- **(vi)** Fine-tuned Depression Classifier: Training a lightweight on-device TensorFlow.js model on the MediaPipe blendshape vectors using a labelled clinical dataset to replace the zero-shot Gemini pathway.

8. CONCLUSION

This paper presented SereneMind, a novel real-time, browser-native multimodal AI framework for continuous depression intensity prediction and proactive wellness intervention. The system uniquely integrates Google MediaPipe's production-grade facial and hand landmark detection (52 blendshapes, 21 hand keypoints) with Gemini AI's contextual reasoning, enforcing clinically responsible, schema-validated outputs. Evaluation over 40 test scenarios demonstrates 87.5% overall accuracy and 91.3% contextual accuracy for diary-based text inputs, with an end-to-end latency well within real-time interactive bounds.

SereneMind's fully client-side architecture ensures zero biometric data persistence, positioning it as a privacy-first, deployable screening auxiliary for workplaces, universities, and telehealth platforms. The project establishes a reproducible, open-source baseline for affective computing research in the mental health domain and paves the way for future multimodal, longitudinal, and clinically validated depression monitoring systems.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the guidance of Prof. Sangita K. Chaudhari , Prof. Savita B. Mogare and the faculty of the Department of Information Technology Engineering, SITRC, Nashik. We thank the participants who voluntarily contributed to the evaluation study and the anonymous reviewers for their constructive feedback. This work was conducted as part of the Final Year Engineering Project under Savitribai Phule Pune University.

REFERENCES

[1] World Health Organization. (2021). Depression and Other Common Mental Disorders: Global Health Estimates. WHO Press, Geneva.

[2] Patel, V., Chisholm, D., Parikh, R., et al. (2016). Addressing the burden of mental, neurological, and substance use disorders. Lancet, 387(10028), 1672–1685.

[3] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.

[4] Girard, J. M., Cohn, J. F., Mahoor, M. H., et al. (2014). Nonverbal Social Withdrawal in Depression: Evidence from Manual and Automatic Analyses. Image and Vision Computing, 32(10), 641–647.

[5] Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion. Information Fusion, 37, 98–125.

[6] Yang, K., Ji, S., Zhang, T., et al. (2023). Towards Interpretable Deep Learning Models for Knowledge Tracing. Findings of ACL 2023.

[7] Choudhury, S., & Bhatt, C. (2023). LLMs in Clinical Decision Support: Opportunities and Risks. npj Digital Medicine, 6(1), 1–8.

[8] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.

[9] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

[10] Lugaresi, C., Tang, J., Nash, H., et al. (2019). MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172.