Jazmin.IA

WIML WORKSHOP · NEURIPS 2026 · SYDNEY

Jazmín.IA

An embedded computer vision system for analyzing behavioral indicators and their relationship with user performance

Yuliana Ituyan Angulo
Mechatronics Engineering · Universidad Autónoma de Occidente · Cali, Colombia



MOTIVATION

Performance metrics tell you how much you did, not how you felt

The face speaks

Frowning, frequent blinking and yawning are visible signs of load and fatigue, measurable without touching anyone.

The gap

Most facial analysis runs in the cloud, classifies emotions, and is rarely linked to real task performance.

Our approach

Describe facial Action Units instead of emotions, on a low-cost board, with no image ever leaving it.

SYSTEM OVERVIEW

Every frame is analyzed inside an Arduino UNO Q

1 · Camera
USB webcam, ≈ 5.8 fps

2 · Landmarks
MediaPipe Face Mesh, 468 points

3 · Features
Geometry + 13-AU model every 0.5 s

4 · Baseline
10 s personal neutral calibration

5 · Signals
Tension, tiredness or well-being

6 · Feedback
LED matrix, screen, report

468
face landmarks per frame

13 AUs
from our own model

2.91 MB
int8 model, –90% size

0
images stored, numbers only

HARDWARE AND SOFTWARE

One board, two processors, three programs

MPU · Qualcomm QRB2210 · Debian · Python 3.11

Main program

Camera, MediaPipe, TensorFlow Lite, signal engine, CSV logging and a FastAPI web interface.

MPU · Python ≥ 3.13

LED bridge

A tiny local HTTP server that turns each pattern into an RPC call to the microcontroller.

MCU · STM32U585

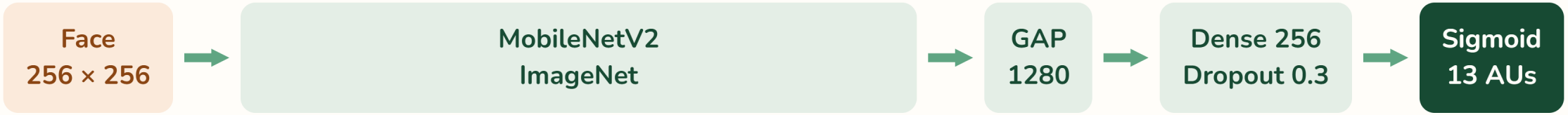
Matrix sketch

Draws “!”, a heart or a smiley on the built-in 8 × 13 LED matrix.

Why two Python processes? MediaPipe supports up to Python 3.12, while Arduino’s Bridge library needs 3.13. They talk over localhost; if the matrix fails, analysis keeps running and reconnects every 3 s.

ACTION UNIT MODEL

One face image in, 13 AU probabilities out



A lightweight take on ADMME

- 40,584 CASME II + SAMM frames, AUs pseudo-labeled with Py-Feat
- Weighted binary cross-entropy for rare AUs
- Frozen backbone 15 epochs, then full fine-tuning 10 epochs
- Per-AU thresholds chosen on validation, never on test

0.872
macro AUC-ROC

0.706
macro F1

2.91 MB
int8, from 30.06 MB

0.5 s
per inference on board

THE KEY FINDING

Good test metrics did not survive a live webcam

0%

AU12 (smile) activation during 47 s of smiling

73–100%

AU07 active with a neutral face, for one participant

99%

of the time in alert, with fixed thresholds

Lab micro-expressions (no speech, little positive affect, high-speed cameras) are far from a person working in front of a webcam. **Decision: geometry leads, the AU model supports.**

GEOMETRY AND PERSONAL CALIBRATION

Each person is compared with their own neutral face

Blink

EAR below a personal threshold for 0.03–0.6 s

Yawn

MAR ≥ 1.8 × neutral for ≥ 2 s, no smile or frown

Frown

Brows ≥ 7% lower than an adaptive reference

Smile

Mouth ≥ 8% wider than during calibration

Tension alert

AU tension index ≥ 1.20 over 20 s **and** a geometric frown in ≥ 15% of the last 6 s, **or** a sustained frown in ≥ 40%. A smile cancels it unless the frown reaches 50%.

Stable by design

All windows use the clock, not frame counts. Hysteresis cut state changes from 20 to 8 in a 388 s session, without losing any real transition.

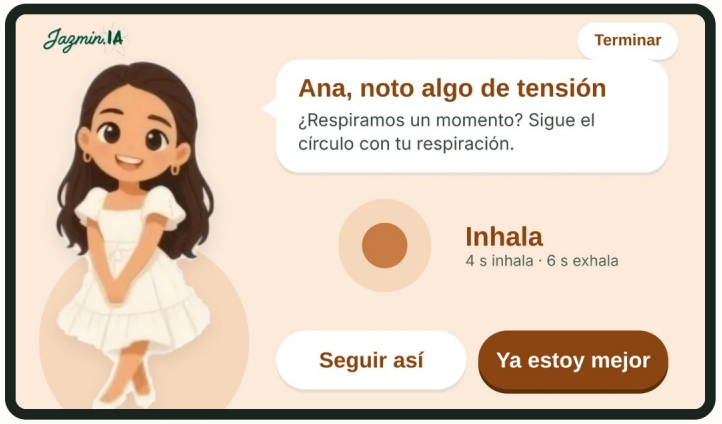
USER EXPERIENCE

Jazmín reacts to the face, never shows it



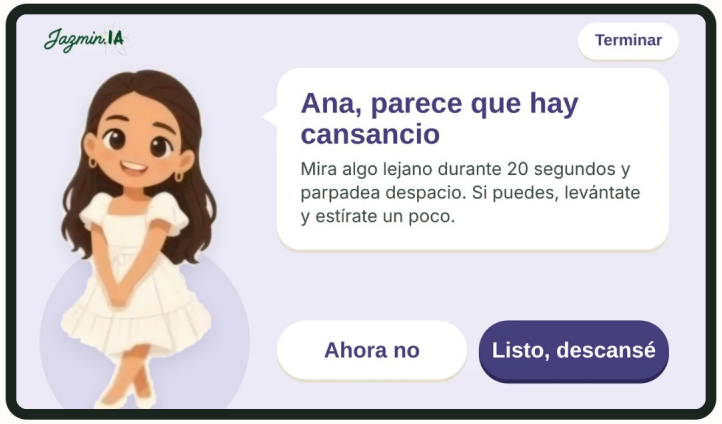
Calibrating

10 s of relaxed face to learn your neutral. No alerts yet.



Tension → breathe

Guided breathing: 4 s in, 6 s out. The matrix shows “!”.



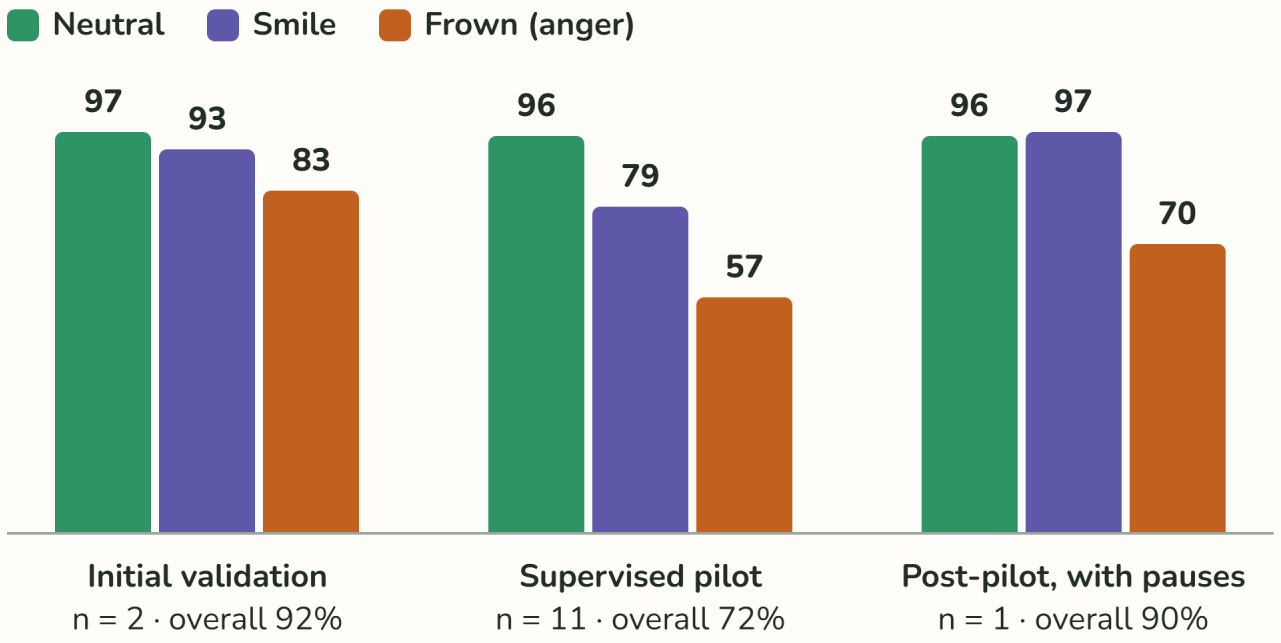
Tiredness → rest

High blink rate plus yawns: look away, blink slowly, stretch.

When all is well, the LED matrix alternates a heart and a smiley. A soft sound plays only if an alert lasts 20 s or more. Interface shown in Spanish, as deployed.

RESULTS · GUIDED EXPRESSIONS, FRAME-LEVEL ACCURACY

Neutral is reliable; mild frowns are the weak spot



18% → 92%

from the first labeled test (18% on neutral) to overall accuracy on new validation sessions, after iterative tuning with session replay

0.5%

time in alert during 30 min of natural use (teaching), 0 false yawns

LIMITATIONS AND FAIRNESS

What still fails, and who it may fail more

Known failure modes

- Mild frowns (brows down < 7%): 53% vs. 70% for strong ones
- Squinting smiles can read as a frown
- AU labels come from Py-Feat, not human FACS coders
- Acted expressions, small and young sample

Skin tone and lighting

Mean pilot accuracy: **86%** light skin (n = 3), **70%** medium (n = 7), **38%** for the only very dark-skinned participant, who was also under harsh overhead light.

Descriptive only, it cannot isolate a cause. The system now measures face lighting on every frame, and next tests will report results by Monk Skin Tone.

NEXT STEPS

From expressions to performance

- **Task sessions:** arithmetic and a 3-min reaction test (PVT-B) to relate facial indicators to response time, accuracy and errors
- **One-week case study** of real remote work with the device
- **Broader, fairer evaluation:** more participants, Monk Skin Tone, ethics review and in-app consent
- **Better calibration:** guided smile and frown, split brow height from brow distance



Thank you! · Yuliana Ituyan Angulo · JazminIA Project Page Link