Demo: Tracing a DSM-5 Depression Assessment from Score to Formula on Edge Hardware

Tibor Haller, Bruno Rodrigues
Sensing Group SG, School of Computer Science SCS,
University of St. Gallen HSG, Switzerland
E-mail: {tibor.haller, bruno.rodrigues}@unisg.ch

***Abstract*—This demonstration allows attendees to interrogate every step of an explainable on-device mental-health assessment. The *Contextual Multimodal Linkage* (CML) Framework maps speech and network biomarkers to the nine DSM-5 indicators of a depressive episode through five closed-form layers, with situational context re-weighting each biomarker contribution. On a commodity mini-PC, attendees run an assessment, unfold any indicator to the formula and weight behind it, and switch the context model to watch the same evidence re-weighted. Because the pipeline is closed-form rather than learned, every score traces to the exact formula and weights that produced it, which a black-box model does not expose, and every run is deterministic and re-runnable. The audience inspects the decision procedure rather than a post-hoc rationalization on synthetic data with no detection accuracy claimed. Demo video: https://youtu.be/g5d3ZWxAQRM**

***Index Terms*—Explainable AI, edge intelligence, mobile health, multimodal fusion, DSM-5**

I. Motivation and System

Continuous in-home monitoring promises earlier awareness of depressive episodes than episodic clinical visits, and passive mobile and behavioral sensing has shown that daily-life signals correlate with depressive symptom severity [1], [2]. Yet dominant deep-learning pipelines are opaque to the clinicians who would act on them, while regulation (FDA Good Machine Learning Practice [3], EU AI Act Articles 13–14) now demands traceable, reproducible decisions. The companion paper submitted to this workshop [4] introduces the *CML Framework*: an on-device pipeline that links raw measurements to a DSM-5-aligned [5] assessment through five inspectable layers (Measurement → Metric → Biomarker → Indicator → Assessment), unifying prior acoustic [6], [7] and home-router network [8] linkage to DSM-5 indicators.

Two design choices carry the explainability: a modality-agnostic late-fusion boundary that preserves per-modality traceability, and a closed-form context-as-modifier [9] that re-weights each biomarker's fuzzy membership [10] by the prevailing situation (Fig. 1). Nothing is learned end-to-end, so the pipeline is interpretable by design rather than by post-hoc attribution [11]–[13]: the explanation is not reconstructed after the fact, it *is* the computation, a deliberate trade of the accuracy end-to-end learning can bring for per-step auditability. The mappings and weights are configured by the author from the clinical literature and are illustrative, so *CML Framework* is a research prototype to inspect a decision procedure.

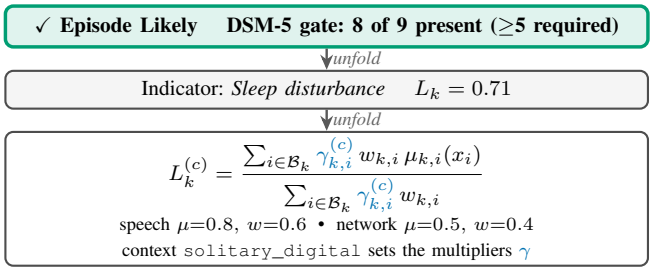

Fig. 1. The path attendees unfold on the device: the episode result opens to an indicator, and the indicator to the closed-form operator that produced it, with every biomarker's membership μ and weight w. Switching the context changes the multipliers γ (blue), re-weighting the same evidence live.

The demo makes that property tangible on a commodity mini-PC (Intel N150, 16 GB) of the kind that an in-home deployment would use, with no cloud dependency and no internet uplink, and the network biomarkers are sensed locally, so no data leaves the device.

II. Demonstration

The demonstration runs the full proof-of-concept [14], pipeline plus dashboard, on the mini-PC with a laptop as display. All results are computed on synthetic workloads engineered to exercise the pipeline. A guided 90-second walk-through (Fig. 1) runs as follows.

1) **Run an assessment.** The attendee selects a mock user and triggers the analysis over a seeded 14-day workload (2,688 biomarker and 336 context-marker records across nine indicators).
2) **Read the result, then unfold it.** The results view (Fig. 2) states whether an episode is likely, with the DSM-5 indicator gate (*e.g.,* 8 of 9 present, ≥5 required), and any indicator opens to its formula, intermediate values, and per-biomarker weight (Fig. 1).
3) **Change the context.** On identical input, the attendee swaps the active weight profile, *e.g.,* from *solitary_digital* to a social-gaming profile (speech 1.5×, network 0.5×), and the same evidence, re-weighted by the multipliers γ, changes the indicator scores live.

Ethics. The demonstration uses synthetic data only and is a research prototype, not a diagnostic tool. Every indicator, including suicidality, is shown as an illustrative, non-diagnostic score.

References

[1] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," *J. Med. Internet Res.*, vol. 17, no. 7, p. e175, 2015.

[2] R. Wang *et al.*, "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing (UbiComp)*, 2014, pp. 3–14.

[3] U.S. FDA, Health Canada, and UK MHRA, "Good machine learning practice for medical device development: Guiding principles," 2021.

[4] T. Haller and B. Rodrigues, "Explainable edge AI for context-aware multimodal health monitoring in IoT households," submitted to *AIMEH'26 (WiMob 2026 Workshops)*, 2026.

[5] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Association Publishing, 2013.

[6] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Commun.*, vol. 71, pp. 10–49, 2015.

[7] J. Länzlinger, K. O. E. Müller, B. Stiller, and B. Rodrigues, "Towards interpretable depression detection: Linking acoustic features to DSM-5 indicators," in *IEEE Int. Conf. Pervasive Computing and Communications (PerCom), Work-in-Progress*, Pisa, Italy, 2026.

[8] S. Nef and B. Rodrigues, "CareNet: Linking home-router network traffic to DSM-5 depressive behavior indicators," 2025, arXiv:2511.12772.

[9] A. K. Dey, "Understanding and using context," *Pers. Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, 2001.

[10] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[11] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[12] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *The Lancet Digital Health*, vol. 3, no. 11, pp. e745–e750, 2021.

[13] P. W. Koh, T. Nguyen, Y. S. Tang, S. Mussmann, E. Pierson, B. Kim, and P. Liang, "Concept bottleneck models," in *Proc. 37th Int. Conf. Machine Learning (ICML)*, 2020, pp. 5338–5348.

[14] T. Haller, "CML proof-of-concept repository," 2026. [Online]. Available: https://github.com/972C8/cml-depression-poc

Appendix
Demo Setup and Requirements

This appendix describes the on-site setup and the attendee experience. It is provided for review and is not intended for the proceedings.

A. Equipment (all brought by the authors)

- Blackview MP60 mini-PC (Intel N150, 16 GB, Windows 11 Pro) running the full CML pipeline and dashboard server.
- One laptop as display and input for the browser-based dashboard, connected to the mini-PC over a direct local link.
- Optional external monitor if the venue can provide one (HDMI) so bystanders can follow the interaction, though it is not required.

B. Venue requirements

- One table (approx. 1 m) and two power sockets.
- No network connectivity required: the system is fully self-contained on the edge device, which is itself part of the demonstrated claim.

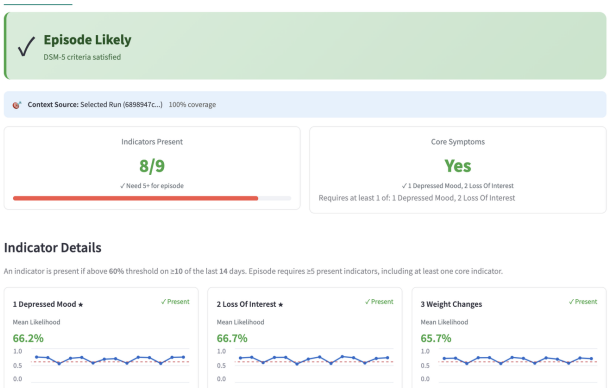

Fig. 2. Results view attendees read first: the episode verdict, the DSM-5 gate (8 of 9 indicators present, 5 required, at least one core symptom), and per-indicator likelihood over the observation window.

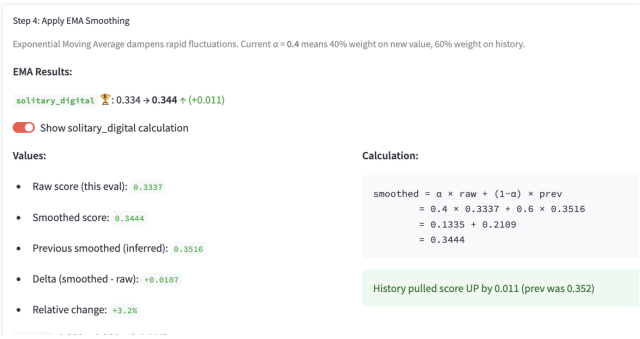

Fig. 3. Pipeline-transparency view from the live dashboard: any computation step unfolds to its formula, inputs, and result. Shown is the EMA-smoothing step of context evaluation, where the displayed formula and the executed calculation are the same object.

C. Attendee experience

The demo runs as a guided free exploration. Attendees drive the dashboard themselves through the walkthrough of Section II, in any order, on synthetic multimodal workloads produced by the built-in mock-data generator (no personal data is shown or collected). The transparency view (Fig. 3) is the centerpiece: reviewers of explainable-AI claims can pick any indicator score and verify, step by step, that the displayed explanation and the executed computation are the same object. A full 14-day run completes in about 23 minutes on a Raspberry Pi 4 and about 2 minutes on a laptop in the companion paper's evaluation. The demo's Intel N150 mini-PC is laptop-class, so full runs complete in minutes. For the booth interaction we keep a completed run loaded for inspection, trigger live context evaluations (seconds) on demand, and launch a full run when asked. The Experiment view additionally lets attendees adjust pipeline parameters (weights, thresholds) and re-run, and the Data view shows the raw ingested records behind every score.