Evidence-grounded multimodal AI for sport and health monitoring

Tsedeke Temesgen Habe, Ph.D.[1] | Annika Hangleiter, M.Sc.[2] | Keijo Haataja, Ph.D.[1] | Prof. Pekka Toivanen[1]

[1] Computational Intelligence Research Group, School of Computing, University of Eastern Finland, Kuopio, Finland | [2] Wearable Technologies AG, Herrsching a. Ammersee, Germany

COMMIT RC Workshop | 25 August 2026 | Tietoteknia, Kuopio Campus, Finland

POSTER 016

WHAT THIS POSTER SHOWS

Synchronized video and wearable muscle signals are aligned on one clock, classified only as running or skiing, and displayed with traceable evidence. Missing, low-confidence or unsupported outputs are withheld for human review.

01 Evidence, supported task, and internal result

Privacy-safe synthetic live-video illustration. No participant, identity or project recording is shown.

Question and label contract

Thirteen heterogeneous source packages were reconstructed without inventing video-sensor correspondence. Video is the master clock; EMG and vendor summaries enter only where interval masks confirm measured overlap. The only supervised sport label is **running versus skiing**.

Evidence used

Temporal intersection produced **77 aligned clips** from **16 video-bearing sessions**. Ten source packages act as subject surrogates because identity linkage is incomplete; two additional sessions contain sensor-only data. Unavailable signals remain unavailable.

77	0.876	0.935	15 / 16
CLIPS TESTED ONCE	CLIP MACRO F1	SESSION MACRO F1	SESSIONS CORRECT

Interpretation

Five subject-surrogate-grouped folds produced one held-out prediction per aligned clip. Macro F1 gives running and skiing equal weight. The small, folder-labelled archive demonstrates internal feasibility only; it does not establish population performance or validate fatigue, technique, injury or safety claims.

02 Models used - and why there is more than one

VIDEO: R3D-18

An **18-layer 3D residual network**, pretrained on Kinetics-400, encodes motion in every available camera. Shared camera embeddings are pooled with an availability mask. This route measures what video alone contributes.

SENSORS: TCN

Separate **mask-aware temporal convolution encoders** represent raw EMG and vendor-summary sequences. Inputs use non-overlapping **20-second windows** and training-fold robust normalization. This route measures what wearable signals contribute.

MISSING-DATA FUSION

Learned gates combine only available video, EMG and summary embeddings. Availability masks and **0.15 modality dropout** stop the model from assuming every stream exists. Routing weights select information; they are not causal importance.

OTHER ANALYSES

Three-cluster K-means creates latent states A-C for navigation, not named phases. A separate CPET extension compares regularized linear, tree-ensemble, boosting and support-vector regressors. Muscle load and balance remain deterministic calculations.

WHY MULTIPLE MODELS? Motion, sensor sequences, missing-modality fusion, exploratory navigation and oxygen-uptake prediction have different data structures and targets. Their outputs answer different questions and must not be treated as interchangeable.

03 Training, testing, and artifact-driven display

1 TRAIN

Use activity- and session-balanced sampling. Fit robust normalization on training packages only. Train R3D-18, TCN and fusion for at most 50 epochs with AdamW; sensor windows do not overlap.

2 VALIDATE

Choose the checkpoint by validation session macro F1, with sample macro F1 as tie-break. Fit temperature scaling and the accept-or-review confidence threshold on validation outputs only.

3 TEST ONCE

Use five stratified subject-surrogate-grouped folds. Every eligible package is held out once, giving one out-of-fold prediction for each of 77 clips. Test data affect no preprocessing or selection.

4 DISPLAY ARTIFACTS

The interface reads the checkpoint, preprocessing statistics, training configuration, prediction row, output contract, model card and file hashes. It displays evidence roles; it does not invent missing results.

The visible demo result is therefore traceable to an exact saved artifact set and to a prediction produced while that clip was held out.

04 Demo: one time cursor and an explicit evidence role

The synchronized cursor connects the video interval, activity probability and muscle calculations. Each visible item is labelled as **measured**, **learned**, **calculated**, **rule-based** or **unavailable**; unsupported outputs are withheld.

Enlarged prototype interface, preserved without structural changes. The left side shows time-aligned activity evidence; the right side separates deterministic muscle measures from learned outputs.

RESPONSIBLE-USE BOUNDARY

SUPPORTED NOW

Route a reviewed session to the running or skiing workflow and open the synchronized evidence interval that influenced the result.

RESEARCH ONLY

Latent states and future phase or event candidates support navigation only. Reviewed temporal labels and further validation are required.

NOT SUPPORTED

Do not diagnose fatigue, pain or injury, assess exercise safety, or make a safe-to-continue decision. Prospective validation is still required.

RESEARCH PROTOTYPE - NOT CLINICALLY VALIDATED - HUMAN REVIEW IS MANDATORY

ACKNOWLEDGMENT

H2TRAIN project is supported by the Chips Joint Undertaking (GA 101140052) and its members, including top-up funding by the National Funding Agency for Finland, the Innovation Funding Agency Business Finland.