

Beyond self-report: a systematic review of wearable and multimodal digital biomarkers for detecting psychological, affective, and fatigue states in athletes

Nilma Zola¹, Ifdil Ifdil^{*)1}, M. Fahli Zatrahadi²

¹Center for Educational Neuroscience, Trauma, and Human Behavior, Universitas Negeri Padang

²Universitas Islam Negeri Sultan Syarif Kasim, Indonesia

*) Correspondence regarding this article should be addressed to: Author address e-mail: ifdil@fip.unp.ac.id

Abstract: Subjective, retrospection-based questionnaires of self-reports are not sufficient sources for early signs of psychological and fatigue conditions in athletes which opens the possibility of digital biomarkers. The review is on the basis of the PRISMA 2020 framework of the 799 records initially found through Scopus, of which 2 duplicates were excluded resulting in 797, among whom, 44 were considered for the inclusion of 16 and the rest were the exclusion of abstract titles. Papers in English describing sensor-based detection of psychological, affective, or fatigue states in sport or exercise were selected. Electrocardiography was the main modality used alongside heartbeat rate variability with the most studies focusing on the detection of fatigue and stress states, and there was an overall increase in the use of deep and machine-learning classifiers with some studies having classification accuracy beyond 90%. Multimodal fusion gave better results compared to the single-signal method. The limitations to the actual implementation of these systems were mainly the use of small sample sizes, weak external validation, inconsistent labelling of the reference signal. This review presents modalities and stresses the importance of standardized protocols, ecological validity, and transparent reporting both for researchers and practitioners.

Keywords: Digital biomarkers; Wearable sensors; Athlete monitoring; Psychological state detection; Machine learning

Article History: Received on 02/01/2025; Revised on 27/01/2025; Accepted on 28/02/2025; Published Online: 10/03/2025.



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2025 by author.

INTRODUCTION

Elite and recreational sports have both turned to ongoing physiological monitoring as a means of ensuring that athletes perform the best they can, recover adequately, and remain in good mental health. This change has been helped by the widespread use of wearable gadgets. The availability of inexpensive hardware, smartphones, and machine-learning techniques have made it possible to assess both physical and psychological functioning of humans beyond the confines of the laboratory (Yin et al., 2025; Nahavandi et al., 2022). In both healthcare and sports, these tools provide objective, frequent observations that can supplement or even rival traditional methods of assessment (Jacobsen

et al., 2020). The psychological component of athlete's functioning that is related to stress, fatigue, mood, and emotion, on the other hand, remains a fairly difficult thing to track compared to others and in this respect, the idea of a digital biomarker has particularly taken roots (Sameh et al., 2025).

Generally, sports psychologists and athletes have relied on self-reported questionnaire measures which are retrospective in nature, require a lot of mental and manual work, and also prone to recall and social-desirability bias. Recent studies have revealed that digitally sensed parameters are not only different from what athletes declare about themselves but the two sets of information may be in conflict also, highlighting the limitations of the subjective monitoring method (Rao et al., 2023). States such as competitive anxiety, mental exhaustion, and emotional arousal fluctuate very quickly thus when athletes respond to episodic questionnaires the information gathered reflects mainly their stable characteristics while what they do moment by moment (their transient characteristics) that determine mainly their performance and well being is missing. This situation has prompted researchers to look for more objective markers that are continuously available and could be derived from both wearable and other types of sensors.

The use of wearable smart devices and artificial intelligence as tools to infer health and mental states is a growing trend in this field (Khan et al., Surveys of mobile-health and affective-computing systems have shown a rapid increase in emotion-detection, stress, and mood recognition pipelines (Maithri et al., Bhatt et al., 2022), while digital-phenotype frameworks have transitioned from being used primarily in patient populations to those involved in everyday activities (Jung et al., 2025; Lipschitz et al., 2025). For example, physiological signals, namely heart rate variability, photoplethysmography, and electrodermal activity have been demonstrated as correlates of negative emotional and stress states (Tsai et al., 2025; Booth, 2022) and have become a basis for the transformation of raw sensor data into psychobiologically significant measures.

The methodological progression has also led to the advancement of this transformation. Now, multilayered networks can capture intricate, non-linear, and even latent relations of different types of signals with various psychological latent states and, sport scientists report more and more use of real-time analytic systems (Vec et al., 2024). In addition, there are digital biomarkers for major depression, mood instability, and psychological distress which have been proven to be accurate by clinicians and have been applied to other domains too (Rykov et al., Książek et al., 2024), and personalisation of emotional models can be useful for the generalisation of such models across individuals (Kargarandehkordi et al., Berkemeier et al., 2025). In summary, these advancements indicate that most of the analytical elements for psychological monitoring of specific athletes are present. They, however, come from a variety of disciplines.

Still, there are quite a few gaps in the research literature devoted specifically to athletes. Work that is centered around sports mainly emphasizes factors like biomechanical load and injury whereas psychological aspects hardly come into picture. Also, reviews in sport of wearables hardly separate emotional or mental effects (Woll et al., 2025). Thus, the first missing element identified here is a comprehensive chart detailing the various methods of sensing that can be employed to detect the corresponding psychological and fatigue states in sporting populations. Without such a chart, researchers are unable to easily determine which physiological signal is best suitable for which state, and they have hardly any clue regarding which modality-state pairings are redundant to each other or still a mystery.

A second missing factor is methodology. Many experiments draw on very small and rather similar participant groups, simulated stressors introduced in a lab, and inconsistent "ground truth" labels all of which compromise reliability and generality (Tump et al., 2022), (Park et al., 2022). The distinction is it a high heart rate due to sports or a high level of stress is often omitted even though it is crucial to valid results in physically active people (Perez-Pozuelo et al., 2020). These limitations

are rarely assessed in a coherent manner which leaves professionals with no idea regarding the reliability of current digital-biomarker methods for athlete monitoring (Debard et al., 2020).

In view of the fact that technology changes very fast and the athlete's health is really at stake here, a comprehensive analysis would definitely be both convenient and needed. Putting various pieces of evidence together will enable everyone to know where exactly you are, where there is still some doubt, and where you should put all your focus next. Based on this logic, the article is going to apply the PRISMA 2020 protocol (Page et al., 2021) to study the scholarly literature relating to wearable and multimodal digital biomarkers for psychological, emotional and fatigue states in athletes and individuals who lead an active lifestyle. The three major points of inquiry that lead the review are the following:

RQ1: Which digital-biomarker modalities and wearable sensing systems have been used to detect psychological, affective, and fatigue-related states in athletes and physically active individuals?. **RQ2:** Which psychological and affective states are most frequently targeted, and what analytical or machine-learning approaches are applied to detect them?. **RQ3:** What is the reported validity, performance, and practical readiness of these multimodal digital-biomarker approaches, and what limitations constrain their deployment?

METHOD

Research Design and Framework

A systematic literature review (SLR) was adopted because it offers a transparent, replicable, and auditable procedure for aggregating dispersed evidence and mitigating selection bias (Tranfield et al., 2003; Liberati et al., 2009). Reporting followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021; Moher et al., 2009), which structures the identification, screening, eligibility, and inclusion phases and improves the completeness of systematic reporting.

Search Strategy

Search concepts were derived using the PICO framework to ensure systematic identification of key terms. Population (P) comprised athletes and physically active individuals; Interest (I) comprised wearable or multimodal digital-biomarker sensing; Comparison (C) was not applicable, as the review is descriptive rather than comparative; and Outcome (O) comprised the detection of psychological, affective, or fatigue states. Each element was expanded into a cluster of synonyms and combined with Boolean operators. The conceptual search string is presented below.

("wearable*" OR "digital biomarker*" OR "physiological sensor*" OR "multimodal sensing")
AND ("athlete*" OR "sport*" OR "exercise" OR "physical activity")
AND ("psychological state*" OR "affect*" OR "emotion*" OR "stress" OR "fatigue" OR "mood")

Truncation (asterisk) captured morphological variants (e.g., athlete/athletes; emotion/emotional), and field codes restricted matching to title, abstract, and keywords (TITLE-ABS-KEY). Limiters were applied for document type, language, and publication window, as detailed in Section 2.4. The string was piloted and refined iteratively; an initial version returned an unmanageable volume of tangential engineering records, prompting tighter coupling of the population and outcome clusters.

Database and Information Sources

Scopus served as the primary and authoritative bibliographic source owing to its broad coverage of engineering, computer-science, health, and sport-science literature. The dataset analysed in this review is the Scopus metadata export used for screening and synthesis. The search and export were

executed as a single consolidated retrieval; where fields such as author affiliation and country were not present in the export, the corresponding attributes are reported as not available in the source file rather than inferred.

Eligibility Criteria

Eligibility criteria are summarised in Table 1. Studies were included when they were English-language journal articles or reviews that addressed sensor-based detection of psychological, affective, or fatigue states in athletes or physically active populations and were available in full text.

Table 1. Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
Language	English	Non-English
Document type	Article, Review	Conference paper, book, book chapter, editorial, note
Publication period	2020–2025	Before 2020
Focus	Detection of psychological, affective, or fatigue states via digital biomarkers	No psychological/affective/fatigue outcome
Population	Athletes or physically active individuals	Non-human or purely clinical, non-exercise samples
Accessibility	Full text available	Abstract only / inaccessible

Study Selection Process

Selection proceeded in three stages. First, duplicate records were identified and removed. Second, titles and abstracts were screened against the eligibility criteria, excluding records whose population, outcome, or study type fell outside scope. Third, the remaining reports were assessed in full for eligibility, and studies satisfying all criteria were retained for synthesis. The numerical outcome of each stage is reported in Section 4 and Figure 1.

Quality Assessment

Methodological appraisal was conceived independently of the PICO-based search formulation. The Mixed Methods Appraisal Tool (MMAT), version 2018 (Hong et al., 2018), was selected for its suitability across the heterogeneous designs anticipated—quantitative descriptive, quantitative non-randomised, review, and mixed methods. For each study, the design category was first classified, after which an overall methodological-confidence rating (High, Medium, or Low) was assigned. Consistent with MMAT developer guidance, no study was excluded on the basis of appraisal alone. Because the evidence base for this review is the Scopus metadata export, confidence ratings reflect study design, reported sampling, and validation as described in the source records; full item-level MMAT scoring against complete texts is identified as a complementary step. A design-and-confidence overview is provided in Table 2.

Table 2. Study design and methodological-confidence overview of the included studies.

Author(s), Year	Study design	Methodological confidence
Jiang & He, 2023	Quantitative descriptive	Medium
Brunyé et al., 2024	Narrative review	Medium
Nashrullah et al., 2025	Systematic review	High
Mu et al., 2024	Quantitative (algorithmic)	Medium
Zhou & Wang, 2024	Quantitative descriptive	Medium
Kakhi et al., 2025	Systematic review	High
Raggi et al., 2025	Topical review / guidelines	Medium
Seçkin et al., 2023	Comprehensive review	High
Spetz et al., 2025	Quantitative observational	High
Dai et al., 2025	Critical review	Medium
Dai, 2022	Quantitative (modelling)	Medium
Stojchevska et al., 2022	Quantitative (comparative ML)	Medium
de Zambotti et al., 2020	State-of-the-art review	High
Luo et al., 2020	Quantitative pilot	Medium
Kodithuwakku Arachchige et al., 2022	Review	Medium
Li, 2022	Quantitative descriptive	Medium

Data Extraction Procedure

For each included study the following were extracted: authors, year, source journal, document type, biomarker modality, target psychological or fatigue state, analytical approach, and principal finding. Author affiliation and country were not present in the source export and are therefore not tabulated.

Bibliometric Context

To situate the included studies within wider research activity, the annual volume of the retrieved corpus was tabulated to characterise growth in the field (Figure 2). This descriptive bibliometric step contextualises, but does not substitute for, the qualitative synthesis.

Data Analysis and Synthesis

Findings were integrated using thematic synthesis (Thomas & Harden, 2008). Studies were coded by biomarker modality and by target state; codes were then organised into descriptive themes aligned with the three research questions, and analytical themes were developed to interpret agreements, tensions, and gaps across the corpus.

Reporting and Documentation

The review adheres to the PRISMA 2020 checklist (Page et al., 2021). The flow of records across identification, screening, eligibility, and inclusion is documented in Figure 1, and all reported counts are mutually consistent across the Abstract, Methods, and Results. Prisma 2020 flow diagram.

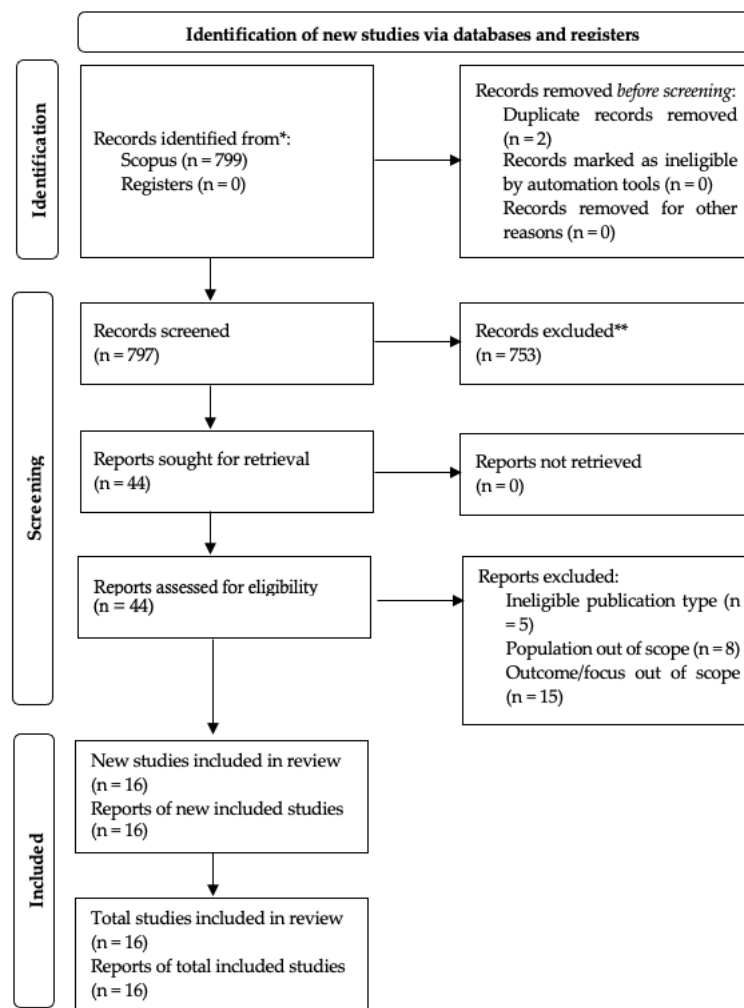


Figure 1. PRISMA 2020 flow diagram of the study-selection process (adapted from Page et al., 2021).

RESULTS AND DISCUSSION

Study Selection Results

The Scopus search identified 799 records. After 2 duplicates were removed, 797 records underwent title and abstract screening, of which 753 were excluded as outside the population, outcome, or document-type scope. The remaining 44 reports were assessed in full, and 28 were excluded: 5 of ineligible publication type, 8 with populations out of scope, and 15 whose outcome or focus lay outside psychological, affective, or fatigue-state detection. Sixteen studies met all criteria and were retained for synthesis (Figure 1).

The retrieved corpus grew steadily across the review window, rising from 59 records in 2020 to 272 in 2025, indicating rapidly intensifying research interest in wearable and digital-biomarker methods relevant to physical and mental monitoring (Figure 2).

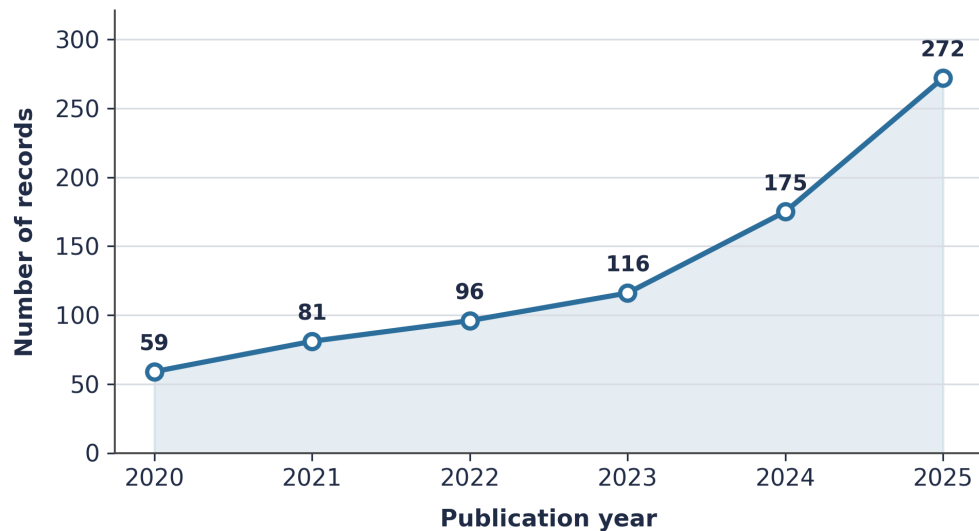


Figure 2. Annual publication trend of the retrieved corpus (n = 799).

Descriptive Characteristics

The 16 included studies span 2020–2025 and combine primary experimental work with systematic and narrative reviews. Table 3 summarises each study; Table 4 classifies them by biomarker modality, target state, and analytical approach. Figures 3 and 4 quantify modality and target-state coverage across the corpus.

Table 3. Summary of included studies.

Title	Author(s) / Year	Journal	Study type	Key findings
EEG Signal Analysis based on machine learning in psychological balance detection of athletes	Jiang & He (2023)	Evolutionary Intelligence	Article	An LSTM–CNN model classifies athletes' pre-competition psychological balance from EEG signals by fusing time-domain and space-domain features.
Movement Sensing Opportunities for Monitoring Dynamic Cognitive States	Brunyé et al. (2024)	Sensors	Review	Wearable IMUs and markerless optical motion tracking can classify transient cognitive states such as stress and workload from overt movement across sport and occupational domains.
A Systematic Review of Heart Rate-based Fatigue Prediction Using Machine Learning and Deep Learning Techniques	Nashrullah et al. (2025)	International Journal of Computing and Digital Systems	Article	Heart-rate-based ML and DL models can predict fatigue; deep-learning approaches generally improve accuracy, and wearables enable field-level deployment.
Intelligent fatigue detection based on hierarchical multi-scale ECG representations and HRV measures	Mu et al. (2024)	Biomedical Signal Processing and Control	Article	A hierarchical transformer integrating multi-scale ECG and HRV features improves exercise-induced fatigue detection from wearable ECG.
Athlete's facial emotion recognition	Zhou & Wang (2024)	International Journal of	Article	Fusing multiple physiological signals recognises athletes' facial

Title	Author(s) / Year	Journal	Study type	Key findings
method based on multi physiological information fusion		Reasoning-based Intelligent Systems		emotions with an F1 score above 0.97 and low latency.
Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities	Kakhi et al. (2025)	Computers in Biology and Medicine	Review	Integrating wearables with AI supports real-time fatigue detection; the review maps signal-processing pipelines, persistent challenges, and future opportunities.
Inducing and Assessing Acute Mental Stress in Controlled Conditions: Topical Review and Guidelines for Effective Experimental Protocols	Raggi et al. (2025)	IEEE Access	Review	Reliable stress-induction protocols and physiological plus AI pipelines are required to reduce measurement bias in acute mental-stress detection.
Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities	Seçkin et al. (2023)	Applied Sciences (Switzerland)	Review	Wearable sensors support performance monitoring, injury prevention and rehabilitation in athletes, though further optimisation of their use is needed.
Validating Subjective Ratings with Wearable Data for a Nuanced Understanding of Load-Recovery Status in Elite Endurance Athletes	Spetz et al. (2025)	Sports Medicine - Open	Article	Wearable-derived metrics associate with subjective mood and readiness; combining metrics yields a more nuanced load-recovery picture in elite endurance athletes.
Cost-effective and portable electrochemical sensing of sweat constituents via printed electrodes	Dai et al. (2025)	International Journal of Electrochemical Science	Review	Portable printed electrochemical sensors non-invasively track exercise-induced sweat biomarkers linked to muscle fatigue, hydration and stress.
Endurance Monitoring Method for Rock Climbers Based on Multisensor FDA Model	Dai (2022)	Mathematical Problems in Engineering	Article	A multisensor FDA model enables systematic, quantitative endurance monitoring of rock climbers during training.
Assessing the added value of context during stress detection from wearable data	Stojchevska et al. (2022)	BMC Medical Informatics and Decision Making	Article	Adding contextual information (e.g., physical activity) improves stress detection from wearable physiological data compared with physiology alone.
Sensors Capabilities, Performance, and	de Zambotti et al. (2020)	Sleep Medicine Clinics	Review	Consumer sleep wearables can capture sleep and physiology at

Title	Author(s) / Year	Journal	Study type	Key findings
Use of Consumer Sleep Technology				scale with relevance to mood and recovery, but validation against gold standards remains variable.
Assessment of Fatigue Using Wearable Sensors: A Pilot Study	Luo et al. (2020)	Digital Biomarkers	Article	Wearable multisensor data combined with machine learning can derive digital fatigue biomarkers, demonstrated as feasible in a pilot cohort.
The use of wearable devices in cognitive fatigue: current trends and future intentions	Kodithuwakku Arachchige et al. (2022)	Theoretical Issues in Ergonomics Science	Review	Wearables show promise for early detection of cognitive fatigue through physiological alterations, but standardisation of measures is still required.
Athletes' State Monitoring under Data Mining and Random Forest	Li (2022)	Journal of Sensors	Article	EEG spectral and entropy features with a random-forest classifier distinguish archers' competitive states with accuracy above 90%, outperforming an SVM baseline.

Table 4. Classification of included studies by biomarker modality, target state, and analytical approach.

Author(s) / Year	Biomarker modality	Target state	Analytical approach	Key outcome
Jiang & He (2023)	EEG	Pre-competition psychological balance	LSTM-CNN learning (time + space domain)	An LSTM-CNN model classifies athletes' pre-competition psychological balance from EEG signals by fusing time-domain and space-domain features.
Brunyé et al. (2024)	Movement IMU / markerless optical	Stress, mental workload, cognitive states	Narrative review of movement-sensing methods	Wearable IMUs and markerless optical motion tracking can classify transient cognitive states such as stress and workload from overt movement across sport and occupational domains.
Nashrullah et al. (2025)	Heart rate / HRV (plus EEG, EMG, facial)	Fatigue	Systematic review of ML vs. DL	Heart-rate-based ML and DL models can predict fatigue; deep-learning approaches generally improve accuracy, and wearables enable field-level deployment.
Mu et al. (2024)	ECG / HRV	Exercise-induced fatigue	Hierarchical multi-scale transformer	A hierarchical transformer integrating multi-scale ECG and HRV features improves exercise-

Author(s) / Year	Biomarker modality	Target state	Analytical approach	Key outcome
				induced fatigue detection from wearable ECG.
Zhou & Wang (2024)	Multi-physiological (ECG, respiration, pulse, EDA) + facial	Emotion	Multi-signal fusion with least-squares SVM	Fusing multiple physiological signals recognises athletes' facial emotions with an F1 score above 0.97 and low latency.
Kakhi et al. (2025)	Wearable ECG/EMG and multimodal	Fatigue	PRISMA systematic review of wearables + AI	Integrating wearables with AI supports real-time fatigue detection; the review maps signal-processing pipelines, persistent challenges, and future opportunities.
Raggi et al. (2025)	Physiological signals + AI	Acute mental stress	Topical review and protocol guidelines	Reliable stress-induction protocols and physiological plus AI pipelines are required to reduce measurement bias in acute mental-stress detection.
Seçkin et al. (2023)	Body-worn wearable sensors	Performance and monitoring (incl. stress)	Comprehensive narrative review	Wearable sensors support performance monitoring, injury prevention and rehabilitation in athletes, though further optimisation of their use is needed.
Spetz et al. (2025)	Wearables (sleep, continuous glucose, exercise) + POMS	Mood and load–recovery status	One-year field validation, 20 elite athletes	Wearable-derived metrics associate with subjective mood and readiness; combining metrics yields a more nuanced load–recovery picture in elite endurance athletes.
Dai et al. (2025)	Electrochemical sweat sensors (printed electrodes)	Fatigue, hydration, stress (via sweat biomarkers)	Critical review of printed-electrode platforms	Portable printed electrochemical sensors non-invasively track exercise-induced sweat biomarkers linked to muscle fatigue, hydration and stress.
Dai (2022)	Multisensor system	Endurance / physical state	Functional data analysis (FDA) model	A multisensor FDA model enables systematic, quantitative endurance monitoring of rock climbers during training.

Author(s) / Year	Biomarker modality	Target state	Analytical approach	Key outcome
Stojchevska et al. (2022)	Wearable physiological + contextual data	Stress	ML with vs. without context features	Adding contextual information (e.g., physical activity) improves stress detection from wearable physiological data compared with physiology alone.
de Zambotti et al. (2020)	Consumer multisensor wearables	Sleep, recovery, mood and wellbeing	State-of-the-art review	Consumer sleep wearables can capture sleep and physiology at scale with relevance to mood and recovery, but validation against gold standards remains variable.
Luo et al. (2020)	Multisensor wearable	Fatigue	RNN imputation + supervised/unsupervised ML	Wearable multisensor data combined with machine learning can derive digital fatigue biomarkers, demonstrated as feasible in a pilot cohort.
Kodithuwakku Arachchige et al. (2022)	Wearable physiological sensors	Cognitive fatigue	Review of trends and future directions	Wearables show promise for early detection of cognitive fatigue through physiological alterations, but standardisation of measures is still required.
Li (2022)	EEG	Pre-competition competitive state (incl. central fatigue)	Random-forest classification	EEG spectral and entropy features with a random-forest classifier distinguish archers' competitive states with accuracy above 90%, outperforming an SVM baseline.

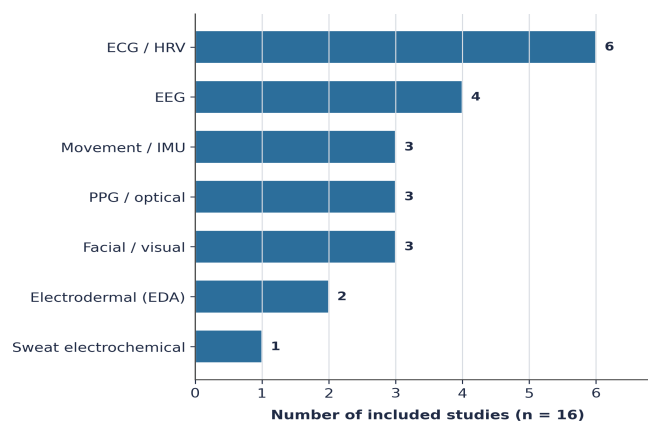


Figure 3. Digital-biomarker modality coverage across the included studies (studies may address multiple modalities).

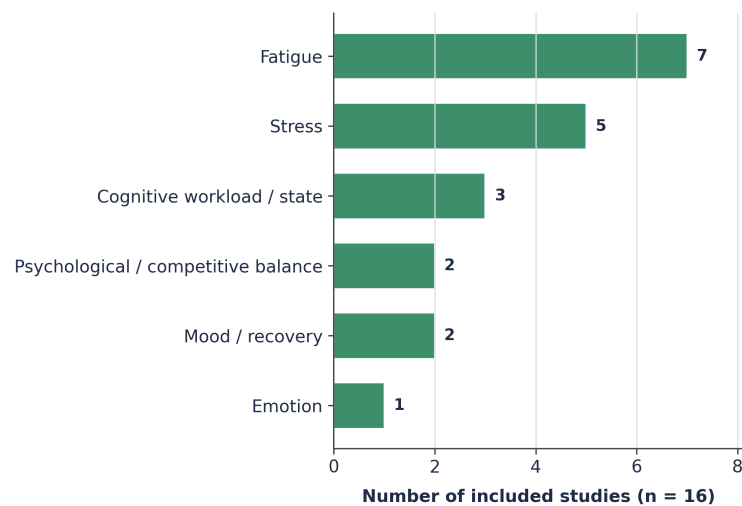


Figure 4. Target psychological, affective, and fatigue states across the included studies (studies may address multiple states).

Thematic Synthesis

RQ1 - Modalities and sensing systems

The included studies employ a heterogeneous set of sensing modalities, but cardiac signals predominate. Electrocardiography and heart-rate variability underpin the largest share of the corpus, anchoring fatigue and stress detection in both algorithmic and applied work (Mu et al., 2024), (Nashrullah et al., 2025). Wearable ECG has been paired with hierarchical, multi-scale representations to improve exercise-induced fatigue detection (Mu et al., 2024), while heart-rate-centred reviews confirm its status as the default physiological channel for fatigue prediction (Nashrullah et al., 2025). Electroencephalography forms a second cluster, applied to competitive and pre-competition psychological states in athletes through deep networks and ensemble classifiers (Jiang & He, 2023), (Li, 2022).

Beyond cardiac and cortical signals, the corpus reveals a clear movement toward multimodal fusion. Facial and visual channels are combined with electrocardiography, respiration, pulse, and electrodermal activity for emotion recognition (Zhou & Wang, 2024), and movement or inertial sensing is used to infer cognitive states from overt behaviour (Brunyé et al., 2024). Emerging platforms extend the sensing envelope further-electrochemical sweat sensors track exercise-induced biomarkers of fatigue and stress (Dai et al., 2025), multisensor systems model endurance in field sport (Dai, 2022), and consumer wearables aggregate sleep, cardiac, and activity streams for recovery-oriented monitoring (de Zambotti et al., 2020), (Spetz et al., 2025). Comprehensive sport-wearable reviews confirm that body-worn sensing now spans the full performance-recovery continuum (Seçkin et al., 2023).

RQ2 - Target states and analytical approaches

Fatigue is the most frequently targeted construct, encompassing physical, cognitive, and central fatigue and appearing across primary studies and reviews alike (Mu et al., 2024), (Kakhi et al., 2025), (Luo et al., 2020), (Kodithuwakku Arachchige et al., 2022). Stress forms the second major cluster, addressed through physiological pipelines and, critically, through work that incorporates context to disambiguate exercise-driven from emotion-driven arousal (Stojchevska et al., 2022), (Raggi et al., 2025). Emotion, mood, recovery, and pre-competition psychological balance complete the state landscape, with mood and readiness validated against subjective instruments in elite athletes (Spetz et al., 2025) and psychological balance and competitive state inferred from cortical activity (Jiang & He, 2023), (Li, 2022) (Figure 4).

Analytically, the field is converging on machine and deep learning. Deep architectures-recurrent, convolutional, and transformer-based-dominate recent primary studies and consistently improve accuracy over classical baselines (Mu et al., 2024), (Jiang & He, 2023), and ensemble methods such as random forests achieve high classification accuracy for athlete state monitoring (Li, 2022). Fusion of multiple physiological channels repeatedly outperforms single-signal models, exemplified by facial-emotion recognition reaching F1 scores above 0.97 (Zhou & Wang, 2024). Reviews of wearable-plus-AI fatigue detection map the signal-processing pipelines that make such performance possible while cautioning that reported gains are often obtained on small, controlled datasets (Kakhi et al., 2025), (Nashrullah et al., 2025).

Comparative and Critical Analysis

Methodologically, two trajectories are visible. Primary studies tend to be small, single-cohort, and laboratory- or session-bound, prioritising algorithmic novelty over generalisation (Jiang & He, 2023), (Zhou & Wang, 2024), (Dai, 2022), (Li, 2022). Reviews, by contrast, emphasise breadth and translational potential but inherit the heterogeneity and inconsistent reporting of the underlying primary work (Kakhi et al., 2025), (Seçkin et al., 2023), (Dai et al., 2025). A recurring and consequential weakness is the treatment of context: because elevated heart rate or sweating can reflect either exertion or emotion, models that ignore activity context risk conflating physiological and psychological signals-precisely the confound that context-aware stress detection was designed to address (Stojchevska et al., 2022). Studies that validate sensed metrics against established subjective instruments, such as mood profiling in elite endurance athletes, represent the more rigorous end of the spectrum and point toward the validation standard the field should adopt (Spetz et al., 2025).

Discussion

Interpreted together, the evidence indicates that wearable and multimodal digital biomarkers can detect a range of psychological, affective, and fatigue states in athletes, but that the maturity of detection varies sharply by state and modality. Cardiac-based fatigue detection is comparatively well developed, whereas emotion and mood detection, though promising, rest on fewer and smaller studies (Zhou & Wang, 2024), (Spetz et al., 2025). Theoretically, the corpus supports a shift from episodic self-report toward continuous, physiologically grounded inference, echoing the digital-phenotyping paradigm now established in adjacent mental-health research (Jung et al., 2025), (Sameh et al., 2025). This reframes athlete psychological monitoring as a signal-processing and machine-learning problem as much as a psychometric one.

Practically, these findings have direct implications for coaches, sport scientists, and clinicians. Multimodal fusion and context-awareness should be treated as design requirements rather than enhancements, because single-signal, context-blind systems are prone to misclassification in active populations (Stojchevska et al., 2022), (Booth et al., 2022). Recovery- and readiness-oriented monitoring that anchors sensor data to validated subjective instruments offers an immediately actionable model for practice (Spetz et al., 2025), (Perez-Pozuelo et al., 2020). At the same time, the broader digital-biomarker literature shows that validation against clinical or criterion measures is essential before such tools inform consequential decisions (Rykov et al., 2021), (Jacobsen et al., 2020).

Situated against prior reviews, this synthesis both confirms and extends earlier observations. Reviews of affective computing and mobile-health sensing report the same convergence on deep learning and the same validation deficits (Maithri et al., 2022), (Bhatt et al., 2022), and sport-specific real-time AI reviews likewise flag small samples and limited external testing (Vec et al., 2024). Contradictions nonetheless persist: some studies report near-ceiling accuracy for state classification (Zhou & Wang, 2024), (Li, 2022), whereas field-validation work reveals only moderate associations between sensed and self-reported states (Spetz et al., 2025), (Rao et al., 2023). The most plausible

explanation is differing evaluation conditions-controlled, within-subject protocols inflate apparent performance relative to longitudinal, ecologically valid designs.

Three research gaps follow. First, modality-state coverage is uneven: emotion and mood detection in athletes are markedly under-studied relative to fatigue (Zhou & Wang, 2024). Second, ecological validity is limited, with few longitudinal, in-competition datasets and sparse external validation (Park et al., 2022), (Tump et al., 2022). Third, ground-truth labelling is inconsistent, undermining cross-study comparison and meta-analytic aggregation (Rao et al., 2023). This review has its own limitations: it draws on a single database (Scopus), which, although comprehensive, may omit records indexed elsewhere; author affiliation and country were absent from the export, precluding geographic analysis; and appraisal was based on source metadata rather than full item-level MMAT scoring, as noted in Section 2.6.

The future agenda is correspondingly concrete. Priority one is the development of standardised, openly reported protocols for state induction, labelling, and evaluation, so that reported accuracies become comparable across studies (Raggi et al., 2025). Priority two is longitudinal, multimodal, context-aware data collection in real training and competition, moving beyond laboratory sessions (Brunyé et al., 2024), (Perez-Pozuelo et al., 2020). Priority three is transparent model reporting-including external validation and calibration-to close the gap between headline accuracy and field performance (Kakhi et al., 2025), (Vec et al., 2024). In direct answer to the research questions: cardiac (ECG/HRV) and cortical (EEG) modalities, increasingly fused with facial, movement, and electrochemical signals, dominate the field (RQ1); fatigue and stress are the most frequently targeted states, detected principally through machine- and deep-learning classifiers (RQ2); and while reported performance is often high, genuine validity and deployment readiness remain constrained by small samples, weak external validation, and inconsistent ground truth (RQ3).

CONCLUSIONS

This systematic review synthesised 16 studies on wearable and multimodal digital biomarkers for detecting psychological, affective, and fatigue states in athletes and physically active individuals. In answer to the research questions, cardiac and cortical modalities-increasingly fused with facial, movement, and electrochemical signals-dominate the field, and fatigue and stress are the states most often targeted, detected chiefly through machine- and deep-learning models whose reported accuracy is frequently high but unevenly validated. The review contributes a consolidated modality-and-state map that clarifies where evidence is strong and where it is thin. For practice, it argues that multimodal fusion, context-awareness, and validation against established instruments are prerequisites for trustworthy athlete monitoring. Its main limitations are reliance on a single database and metadata-based appraisal. Future research should prioritise standardised protocols, ecologically valid longitudinal data, and transparent model reporting to move digital biomarkers from promising prototypes toward dependable tools for athlete welfare and performance.

REFERENCES

- Berkemeier, L., Kamphuis, W., Brouwer, A.-M., de Vries, H., Schadd, M., van Baardewijk, J. U., Oldenhuis, H., Verdaasdonk, R., & van Gemert-Pijnen, L. (2025). Measuring Affective State: Subject-Dependent and -Independent Prediction Based on Longitudinal Multimodal Sensing. *IEEE Transactions on Affective Computing*, 16(2), 827–843. <https://doi.org/10.1109/TAFFC.2024.3474098>

- Bhatt, P., Liu, J., Gong, Y., Wang, J., & Guo, Y. (2022). Emerging Artificial Intelligence-Empowered mHealth: Scoping Review. *JMIR mHealth and uHealth*, 10(6), Article e35053. <https://doi.org/10.2196/35053>
- Booth, B. M., Vrzakova, H., Mattingly, S. M., Martinez, G. J., Faust, L., & D'Mello, S. K. (2022). Toward Robust Stress Prediction in the Age of Wearables: Modeling Perceived Stress in a Longitudinal Study With Information Workers. *IEEE Transactions on Affective Computing*, 13(4), 2201–2217. <https://doi.org/10.1109/TAFFC.2022.3188006>
- Brunyé, T. T., McIntyre, J., Hughes, G. I., & Miller, E. L. (2024). Movement Sensing Opportunities for Monitoring Dynamic Cognitive States. *Sensors*, 24(23), Article 7530. <https://doi.org/10.3390/s24237530>
- Dai, J., Lu, J., & Zhou, D. (2025). Cost-effective and portable electrochemical sensing of sweat constituents via printed electrodes. *International Journal of Electrochemical Science*, 20(11), Article 101164. <https://doi.org/10.1016/j.ijoes.2025.101164>
- Dai, Y. (2022). Endurance Monitoring Method for Rock Climbers Based on Multisensor FDA Model. *Mathematical Problems in Engineering*, 2022, Article 2683399. <https://doi.org/10.1155/2022/2683399>
- Debard, G., De Witte, N., Sels, R., Mertens, M., Van Daele, T., & Bonroy, B. (2020). Making wearable technology available for mental healthcare through an online platform with stress detection algorithms: The CareWear project. *Journal of Sensors*, 2020, Article 8846077. <https://doi.org/10.1155/2020/8846077>
- de Zambotti, M., Cellini, N., Menghini, L., Sarlo, M., & Baker, F. C. (2020). Sensors Capabilities, Performance, and Use of Consumer Sleep Technology. *Sleep Medicine Clinics*, 15(1), 1–30. <https://doi.org/10.1016/j.jsmc.2019.11.003>
- Hoang, H., Ha, H., Nguyen, H., Watton, P., & Ngo, L. (2025). Advancing mental health diagnostics: a review on the role of smartphones, wearable devices, and artificial intelligence in depression and anxiety detection. *Biomedical Engineering Letters*, 15(6), 1003–1012. <https://doi.org/10.1007/s13534-025-00505-4>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). Mixed Methods Appraisal Tool (MMAT), version 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.
- Jacobsen, M., Dembek, T. A., Kobbe, G., Gaidzik, P. W., & Heinemann, L. (2020). Noninvasive Continuous Monitoring of Vital Signs With Wearables: Fit for Medical Use?. *Journal of Diabetes Science and Technology*, 15(1), 34–43. <https://doi.org/10.1177/1932296820904947>
- Jiang, Z., Van Zoest, V., Deng, W., Ngai, E. C. H., & Liu, J. (2023). Leveraging Machine Learning for Disease Diagnoses Based on Wearable Devices: A Survey. *IEEE Internet of Things Journal*, 10(24), 21959–21981. <https://doi.org/10.1109/JIOT.2023.3313158>
- Jiang, P., & He, H. (2023). EEG Signal Analysis based on machine learning in psychological balance detection of athletes. *Evolutionary Intelligence*, 16(5), 1533–1538. <https://doi.org/10.1007/s12065-022-00770-7>
- Jung, H. W., Kim, D. Y., Lee, I., Kim, O., Lee, S., Lee, S., Chung, U. S., Kim, J.-H., Kim, S., Kim, J. W., Shin, A. L., & Lee, J. J. (2025). Key Features of Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, 27, Article e77331. <https://doi.org/10.2196/77331>
- Kakhi, K., Jagatheesaperumal, S. K., Khosravi, A., Alizadehsani, R., & Acharya, U. R. (2025). Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities. *Computers in Biology and Medicine*, 195, Article 110461. <https://doi.org/10.1016/j.compbiomed.2025.110461>

- Kargarandehkordi, A., Kaisti, M., & Washington, P. (2024). Personalization of Affective Models Using Classical Machine Learning: A Feasibility Study. *Applied Sciences (Switzerland)*, 14(4), Article 1337. <https://doi.org/10.3390/app14041337>
- Khan, N., Deshpande, N., Singh, P., & Kadam, P. (2025). A review of AI-based health prediction using apple watch and fitbit data. *Discover Applied Sciences*, 7(10), Article 1217. <https://doi.org/10.1007/s42452-025-07783-8>
- Kodithuwakku Arachchige, S. N. K., Burch V, R. F., Chander, H., Turner, A. J., & Knight, A. C. (2022). The use of wearable devices in cognitive fatigue: current trends and future intentions. *Theoretical Issues in Ergonomics Science*, 23(3), 374–386. <https://doi.org/10.1080/1463922X.2021.1965670>
- Książek, K., Masarczyk, W., Głomb, P., Romaszewski, M., Stokłosa, I., Ścisło, P., Dębski, P., Pudło, R., Buza, K., Gorczyca, P., & Piegza, M. (2024). Assessment of symptom severity in psychotic disorder patients based on heart rate variability and accelerometer mobility data. *Computers in Biology and Medicine*, 176, Article 108544. <https://doi.org/10.1016/j.compbiomed.2024.108544>
- Li, X. (2022). Athletes' State Monitoring under Data Mining and Random Forest. *Journal of Sensors*, 2022, Article 1966786. <https://doi.org/10.1155/2022/1966786>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7), e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Lipschitz, J. M., Lin, S., Saghafian, S., Pike, C. K., & Burdick, K. E. (2025). Digital phenotyping in bipolar disorder: Using longitudinal Fitbit data and personalized machine learning to predict mood symptomatology. *Acta Psychiatrica Scandinavica*, 151(3), 434–447. <https://doi.org/10.1111/acps.13765>
- Luo, H., Lee, P.-A., Clay, I., Jaggi, M., & De Luca, V. (2020). Assessment of Fatigue Using Wearable Sensors: A Pilot Study. *Digital Biomarkers*, 4, 59–72. <https://doi.org/10.1159/000512166>
- Maithri, M., Raghavendra, U., Gudigar, A., Samanth, J., Prabal Datta, Barua, Murugappan, M., Chakole, Y., & Acharya, U. R. (2022). Automated emotion recognition: Current trends and future perspectives. *Computer Methods and Programs in Biomedicine*, 215, Article 106646. <https://doi.org/10.1016/j.cmpb.2022.106646>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mu, S., Liao, S., Tao, K., & Shen, Y. (2024). Intelligent fatigue detection based on hierarchical multi-scale ECG representations and HRV measures. *Biomedical Signal Processing and Control*, 92, Article 106127. <https://doi.org/10.1016/j.bspc.2024.106127>
- Nahavandi, D., Alizadehsani, R., Khosravi, A., & Acharya, U. R. (2022). Application of artificial intelligence in wearable devices: Opportunities and challenges. *Computer Methods and Programs in Biomedicine*, 213, Article 106541. <https://doi.org/10.1016/j.cmpb.2021.106541>
- Nashrullah, M. A., Puspitaningayu, P., Tjahyaningtjias, H. P. A., Funabiki, N., Alan, M. B. F., Suartana, I. M., Rusdiawan, A., Kusuma, D. A., Fathir, L. W., Wiyono, A., & Indriyanti, A. D. (2025). A Systematic Review of Heart Rate-based Fatigue Prediction Using Machine Learning and Deep Learning Techniques. *International Journal of Computing and Digital Systems*, 18(1). <https://doi.org/10.12785/ijcds/1571100750>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher,

- D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Park, Y., Go, T.-H., Hong, S. H., Kim, S. H., Han, J. H., Kang, Y., & Kang, D. R. (2022). Digital Biomarkers in Living Labs for Vulnerable and Susceptible Individuals: An Integrative Literature Review. *Yonsei Medical Journal*, 63, S43–S55. <https://doi.org/10.3349/ymj.2022.63.S43>
- Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J. M., Taheri, S., Guan, Y., & Fernandez-Luque, L. (2020). The future of sleep health: a data-driven revolution in sleep science and medicine. *npj Digital Medicine*, 3(1), Article 42. <https://doi.org/10.1038/s41746-020-0244-4>
- Raggi, M., Moore, L. J., & Mesin, L. (2025). Inducing and Assessing Acute Mental Stress in Controlled Conditions: Topical Review and Guidelines for Effective Experimental Protocols. *IEEE Access*, 13, 10022–10042. <https://doi.org/10.1109/ACCESS.2025.3528518>
- Rao, C., Di Lascio, E., Demanase, D., Marshall, N., Sopala, M., & De Luca, V. (2023). Association of digital measures and self-reported fatigue: a remote observational study in healthy participants and participants with chronic inflammatory rheumatic disease. *Frontiers in Digital Health*, 5, Article 1099456. <https://doi.org/10.3389/fdgth.2023.1099456>
- Rykov, Y., Thach, T.-Q., Bojic, I., Christopoulos, G., & Car, J. (2021). Digital biomarkers for depression screening with wearable devices: Cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, 9(10), Article e24872. <https://doi.org/10.2196/24872>
- Sameh, A., Rostami, M., Oussalah, M., Korpelainen, R., & Farrahi, V. (2025). Digital phenotypes and digital biomarkers for health and diseases: a systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones. *Artificial Intelligence Review*, 58(2), Article 66. <https://doi.org/10.1007/s10462-024-11009-5>
- Seçkin, A.Ç., Ateş, B., & Seçkin, M. (2023). Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities. *Applied Sciences (Switzerland)*, 13(18), Article 10399. <https://doi.org/10.3390/app131810399>
- Spetz, L., Rogestedt, J., Nilsson, R., Mattsson, C. M., & Larsen, F. J. (2025). Validating Subjective Ratings with Wearable Data for a Nuanced Understanding of Load-Recovery Status in Elite Endurance Athletes. *Sports Medicine - Open*, 11(1), Article 154. <https://doi.org/10.1186/s40798-025-00958-y>
- Stojchevska, M., Steenwinckel, B., Van Der Donckt, J., De Brouwer, M., Goris, A., De Turck, F., Van Hoecke, S., & Ongenaes, F. (2022). Assessing the added value of context during stress detection from wearable data. *BMC Medical Informatics and Decision Making*, 22(1), Article 268. <https://doi.org/10.1186/s12911-022-02010-5>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. <https://doi.org/10.1186/1471-2288-8-45>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Tsai, Y.-Y., Chen, Y.-J., Lin, Y.-F., Hsiao, F.-C., Hsu, C.-H., & Liao, L.-D. (2025). Photoplethysmography-based HRV analysis and machine learning for real-time stress quantification in mental health applications. *APL Bioengineering*, 9(2), Article 026103. <https://doi.org/10.1063/5.0256590>
- Tump, D., Narayan, N., Verbiest, V., Hermsen, S., Goris, A., Chiu, C.-D., & Van Stiphout, R. (2022). Stressors and Distressors in Working From Home Based on Context and Physiology From

-
- Self-Reports and Smartwatch Measurements: International Observational Study Trial. *JMIR Formative Research*, 6(11), Article e38562. <https://doi.org/10.2196/38562>
- Vec, V., Tomažič, S., Kos, A., & Umek, A. (2024). Trends in real-time artificial intelligence methods in sports: a systematic review. *Journal of Big Data*, 11(1), Article 148. <https://doi.org/10.1186/s40537-024-01026-0>
- Woll, S., Birkenmaier, D., Biri, G., Nissen, R., Lutz, L., Schroth, M., Ebner-Priemer, U. W., & Giurgiu, M. (2025). Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review. *JMIR mHealth and uHealth*, 13, Article e59660. <https://doi.org/10.2196/59660>
- Yin, F., Chen, J., Xue, H., Kang, K., Lu, C., Chen, X., & Li, Y. (2025). Integration of wearable electronics and heart rate variability for human physical and mental well-being assessment. *Journal of Semiconductors*, 46(1), Article 011603. <https://doi.org/10.1088/1674-4926/24080026>
- Zhou, X., & Wang, J. (2024). Athlete's facial emotion recognition method based on multi physiological information fusion. *International Journal of Reasoning-based Intelligent Systems*, 16(2), 107–117. <https://doi.org/10.1504/IJRIS.2024.138626>