Vol. 6, No. 2, 2025, pp. 181-195
DOI: https://doi.org/10.29210/993500

ELORA

Contents lists available at Elora Center

Lentera Negeri

Journal homepage: https://lentera.eloracenter.org/lentera

AI, computer vision, and sensor-based assessment of psychological states in tennis: a systematic review

Andre Igoresky[*)], Ifdil Ifdil, M. Arie Desman
Universitas Negeri Padang, Indonesia

Article Info

Article history:
Received Nov 22th, 2025
Revised Nov 24th, 2025
Accepted Des 30th, 2025

Keyword:
Digital phenotyping,
Psychological states,
Tennis players,
Wearable sensors,
Artificial intelligence

ABSTRACT

This review examined how wearable sensors, computer vision, and artificial intelligence (AI) measure or infer psychological states in tennis and racket sports. Outcomes included competitive stress, anxiety, mental fatigue, cognitive or attentional states, recovery, and flow or "zone" experiences. Computer-vision studies were treated as behavioural proxies unless a psychological target was reported. Reporting followed PRISMA 2020. A Scopus search identified 253 records from 2015–2025; 43 texts were assessed, and 10 studies met the criteria. Eligible studies involved racket-sport athletes, psychological/affective constructs or behavioural proxies, and digital sensing or computational modalities. Approaches included physiological, EEG, inertial, computer-vision, and multimodal methods. Findings indicate overall feasibility, but performance was not comparable because constructs, labels, samples, settings, and validation procedures differed. No external-validation dataset was identified. Cohen's kappa could not be reconstructed, and FICO was an internal appraisal rubric. Priorities include longitudinal field validation, standardized labeling, psychometric anchoring, explainable models, and privacy-preserving data governance.

Corresponding Author:
Andre Igoresky,
andre.igo88@fik.unp.ac.id

Introduction

Elite and developmental sport today sit inside a data-intensive ecosystem. Physiological, biomechanical, and behavioural signals are captured with ever-finer granularity, and that's changing how performance gets understood and managed. Across most disciplines, the rise of digital sensing and analytical tools has nudged athlete monitoring away from episodic laboratory checks toward continuous, field-based measurement (Wang, 2025; Nan, 2022). Tennis is a clear case in point. The sport is globally popular, physically demanding, and psychologically loaded; players have to regulate emotion, hold attention, and make split-second perceptual-motor decisions while the scoreline shifts beneath them. Small wonder, then, that the convergence of wearable biosensors, computer vision, and artificial intelligence (AI) has drawn intense scholarly attention. It promises to make the previously invisible psychological side of performance measurable and actionable (Havlucu et al., 2022; Li & Wu, 2023).

What actually motivates this review is a specific phenomenon: the objective quantification of psychological states during real tennis practice and competition. That includes affect, anxiety, arousal, cognitive load, mental fatigue, and the elusive optimal-performance or 'zone' experience. Traditional psychological assessment in tennis is another story. It has leaned almost entirely on retrospective self-report instruments, usually administered before or after play, which are prone to recall bias and blind to within-point dynamics (Fernandez-Fernandez et al., 2015). Digital phenotyping offers a more compelling route. The approach is defined as the moment-by-moment quantification of the individual-level human phenotype using data from personal digital devices, and it infers latent psychological states from continuous biosignal and behavioural streams (Havlucu et al., 2022; Wang, 2025).

A fair amount of research has already probed the physiological correlates of psychological load in tennis. Take studies of psychophysiological stress. They have combined heart rate, salivary cortisol, and perceived-exertion measures to tell competition apart from training demands (Fernandez-Fernandez et al., 2015; Gescheit et al., 2015). Heart-rate variability (HRV), for its part, has emerged as a sensitive marker of autonomic balance and recovery in junior and elite players alike (Villafaina et al., 2023; Cusano et al., 2025). Electroencephalographic (EEG) investigations then add a neural layer, characterising the signatures of stress, expertise, and mental fatigue in racket-sport athletes (Tsai et al., 2022; Habay et al., 2021; Liu & Wang, 2018).

Recent technological and methodological advances have only accelerated the field. Low-cost inertial measurement units (IMUs), smartwatches, and biosensor wristbands now stream multichannel data that machine-learning models can map onto psychological labels (Havlucu et al., 2022; Tabrizi et al., 2021). Computer vision and deep learning add another layer. Pipelines ranging from attention-enhanced recurrent networks to convolutional architectures now recognise strokes, actions, and movement quality automatically, from both video and sensor arrays (Gao & Ju, 2024; Ye et al., 2024; Pan et al., 2022). So too do functional near-infrared spectroscopy and event-related potential paradigms, which are beginning to reveal the cognitive and social-affective neuroscience behind on-court interaction (Zhang et al., 2025; Wang et al., 2022; Wei et al., 2025).

For all that momentum, the first major gap is the fragmentation of evidence across modalities and constructs. Wearable, EEG, vision, and biochemical approaches have mostly grown up in parallel silos, each one aimed at a narrow slice of the psychological spectrum. There is little integrative synthesis to show which sensing modality best indexes which psychological state (Wang, 2025; Ding et al., 2024). The upshot is that practitioners have no consolidated evidence base to turn to when choosing technology for monitoring affect, fatigue, or cognition in tennis.

A second gap is more theoretical and methodological in nature. Plenty of studies infer psychological states from proxy signals without ever validating those inferences against established psychological constructs. Models are also frequently trained on small, homogeneous samples, with limited external validation (Havlucu et al., 2022; Tsai et al., 2022). Ecological validity, label reliability, and model interpretability all remain underexplored, which raises a nagging question: do laboratory-derived classifiers actually generalise to the noisy realities of the tennis court (Beck et al., 2025; Kaye et al., 2025)?

That makes the justification for this systematic review both timely and pressing. Commercial wearables and AI analytics are spreading fast through professional and developmental tennis, so evidence-based guidance is needed to ensure psychological inferences stay valid, reliable, and ethically deployed (Wang, 2025; Nan, 2022). A rigorous synthesis can clarify what has been achieved so far, expose the methodological weaknesses, and chart a coherent agenda for computational sport psychology in tennis.

Objective and review scope. This systematic literature review was designed to synthesize evidence on the validity and practical utility of digital signals for psychological-state inference in tennis and related racket sports. Here, digital phenotyping is operationalized as the use of continuous or repeated digital sensing, physiological measurement, or computational analysis to quantify a psychological construct directly or through an explicitly justified behavioural or physiological proxy. The review was structured around a Scopus search using the TITLE-ABS-KEY Boolean query reported below, with eligibility limited during screening to English-language, peer-reviewed publications from 2015–2025. The objective was to synthesize the sensing modalities, psychological constructs, analytical models, validation procedures, and translation barriers reported in this literature.

To distinguish direct psychological phenotyping from adjacent sports analytics, studies were treated as direct evidence when the target was stress, anxiety, fatigue, cognition/attention, recovery-related state, or flow/zone; studies that measured only stroke or movement recognition were retained only as contextual behavioural-proxy evidence.

This review takes on those gaps with three research questions. The first one examines the landscape of measurement. Answering it yields a modality-by-construct map, something that would help consolidate a badly fragmented literature.

The second question interrogates analytical performance and rigour. It contributes a critical appraisal of how well machine-learning and deep-learning models infer psychological states, and of how trustworthy those inferences actually are.

The third question turns to application and translation. Here the contribution is an evidence-informed agenda for bringing digital phenotyping into tennis training, athlete welfare, and competition. Taken together, the novelty of this review rests on unifying wearable, vision, and AI evidence around a single organising construct, psychological digital phenotyping in tennis. No one has attempted that integration before.

RQ1: Which wearable-sensor, computer-vision, and biosignal modalities have been used to measure or infer psychological states in tennis players, and which states do they target?

RQ2: How effectively, and with what methodological rigour, do artificial-intelligence and machine-learning models classify or predict psychological states from these digital signals?

RQ3: How can digital phenotyping of psychological states be translated into applied practice for tennis training, athlete monitoring, and competitive performance?

Method

Research Design and Framework

Research design and framework. This study is a systematic literature review (SLR) reported according to PRISMA 2020. The protocol logic specified the research questions, search string, eligibility criteria, extraction fields, appraisal framework, and thematic synthesis before interpretation. The review was not prospectively registered in PROSPERO or another public registry; no registration number exists for the archived review. Because the included studies combine intervention, cross-sectional, field, EEG, physiological, wearable, and computer-vision designs with non-equivalent outcomes, synthesis was narrative and thematic rather than meta-analytic.

The qualitative, configurative synthesis was selected instead of a meta-analysis since the reviewed corpus comprises highly different sensing modalities, psychosocial constructs, and outcome measures.

Search Strategy

Search strategy. The search combined three concept blocks: racket-sport population, psychological construct, and digital sensing/computational modality. Searches were executed in Scopus using TITLE-ABS-KEY fields. The exact Boolean query was:

> *TITLE-ABS-KEY (("tennis" OR "racket sport*") AND ("psycholog*" OR "anxiety" OR "stress" OR "mental fatigue" OR "affect*" OR "cognit*" OR "arousal" OR "emotion*") AND ("wearable*" OR "sensor*" OR "artificial intelligence" OR "machine learning" OR "deep learning" OR "computer vision" OR "EEG" OR "heart rate variability"))*

Truncation (*) captured morphological variants. The archived Scopus export does not retain the exact calendar date on which the search was executed, so the query-execution date cannot be reproduced from the source file; this is a reported reproducibility limitation. No year or document-type limiter was embedded in the initial query; the 2015–2025 period and publication-type restrictions were applied during screening.

Databases and Information Sources

Scopus was the sole database used in the archived review because of its interdisciplinary coverage of sport science, computer science, engineering, and psychology. No additional searches were conducted in PubMed, PsycINFO, SPORTDiscus, Web of Science, or via systematic citation chaining. Accordingly, the review does not claim database saturation; single-database sampling is treated as a source of selection bias. The export contained 253 records and supplied the bibliographic metadata used in the review.

Eligibility Criteria

Studies were included or excluded according to the pre-specified criteria presented in Table 1.

Table 1. Inclusion and exclusion criteria applied during study selection.

Criterion	Inclusion	Exclusion
Language	English-language publications	Non-English publications

Criterion	Inclusion	Exclusion
Document type	Peer-reviewed article or review	Conference paper/review, book chapter, editorial, erratum, retracted item
Publication period	2015–2025	Published before 2015
Population	Tennis / table tennis / beach tennis (racket-sport) players	Non–racket-sport or non-athlete populations
Focus	Psychological state measured or inferred via digital sensing, or an explicitly linked behavioural/physiological proxy	No psychological construct/proxy, or no digital/AI modality
Accessibility	Sufficient full-text detail for data extraction	Data not extractable / abstract only
Relevance	Directly addresses digital phenotyping of psychological states	Tangential or incidental mention only

Study Selection Process

Study selection process. Records moved through title/abstract screening, full-text assessment, and final inclusion. The archived file does not retain independent reviewer-level decisions, reviewer-specific extraction sheets, or reconciliation logs. Therefore, a post hoc Cohen's kappa or percent-agreement statistic cannot be calculated legitimately and no agreement value is fabricated. Borderline cases were resolved by consensus in the archived workflow, but the absence of reviewer-level logs is treated as a limitation. Full-text exclusions were documented in the PRISMA flow.

Quality Assessment: FICO Framework

Quality assessment: FICO framework. FICO was used as a pragmatic internal appraisal rubric covering Focus, Information, Context, and Outcome. Each dimension was defined on a 0–2 scale (maximum 8): Focus = clarity and relevance of the psychological target; Information = adequacy of sample, sensing, and method reporting; Context = ecological and sport relevance; Outcome = validity, transparency, and interpretability of findings. However, the archived review does not retain the item-level study scores or independent appraisal sheets. Consequently, aggregate FICO scores cannot be reconstructed reliably and are not reported as if they were validated risk-of-bias scores. FICO is therefore presented as a structured appraisal aid, not as a validated instrument such as Downs and Black or CASP. Future updates should apply a design-specific validated risk-of-bias tool to the full texts and report reviewer-level agreement.

Table 2. FICO scoring criteria. The archived review does not retain study-level item scores; therefore, no aggregate study scores or exclusion threshold are reported retrospectively.

Dimension	Operational scoring rule	Score
Focus	Clarity and relevance of the psychological target to phenotyping	0–2
Information	Adequacy of sample, sensing, preprocessing, and method reporting	0–2
Context	Ecological and sport relevance of the measurement setting	0–2
Outcome	Validity, transparency, and interpretability of reported findings	0–2
Total	Sum of the four dimensions; maximum score	0–8

Data Extraction Procedure

Data extraction procedure. A standardized matrix captured author/year, country or sport context, study design, sample characteristics, target psychological construct(s), sensing/computational modality, analytical method, primary outcome, validation approach, and reported ethical/privacy considerations. Missing attributes were recorded as not reported rather than inferred. The extraction matrix was also used to distinguish direct psychological outcomes from behavioural or physiological proxies.

Network and Bibliometric Analysis Methodology

To complement the qualitative synthesis, descriptive bibliometric procedures were applied to the screened corpus. Publication counts were aggregated by year to characterise temporal trends (see Figure 2), and author frequencies plus index-keyword frequencies were tabulated to create the thematic landscape shown in Figure 3. Sensing and analytical modalities reported across the included studies were coded and counted to generate the modality distribution in Figure 4. These analyses were descriptive, meant to contextualise the thematic synthesis without replacing it.

Data Analysis and Synthesis

Synthesis applied the thematic approach of Thomas and Harden (2008). The process runs through line-by-line coding of findings, development of descriptive themes, and generation of analytical themes that answer the review questions. Codes came inductively from the reported outcomes of the included studies and were organised around the psychological constructs and digital modalities present in the corpus. Themes were refined and validated iteratively against the primary data, so each analytical claim stayed grounded in the included evidence.

Reporting and Documentation

Reporting and documentation. The review followed PRISMA 2020 reporting requirements. Figure 1 quantifies identification, screening, full-text assessment, and inclusion. The archived review was not prospectively registered, the exact search execution date was not retained, and reviewer-level screening logs were unavailable; these limitations are reported explicitly rather than reconstructed.

PRISMA 2020 Flow Diagram

Figure 1 depicts the full study-selection pathway. For each phase, identification, screening, eligibility, and inclusion, the diagram reports the number of records alongside the reasons for exclusion at the screening and full-text stages.

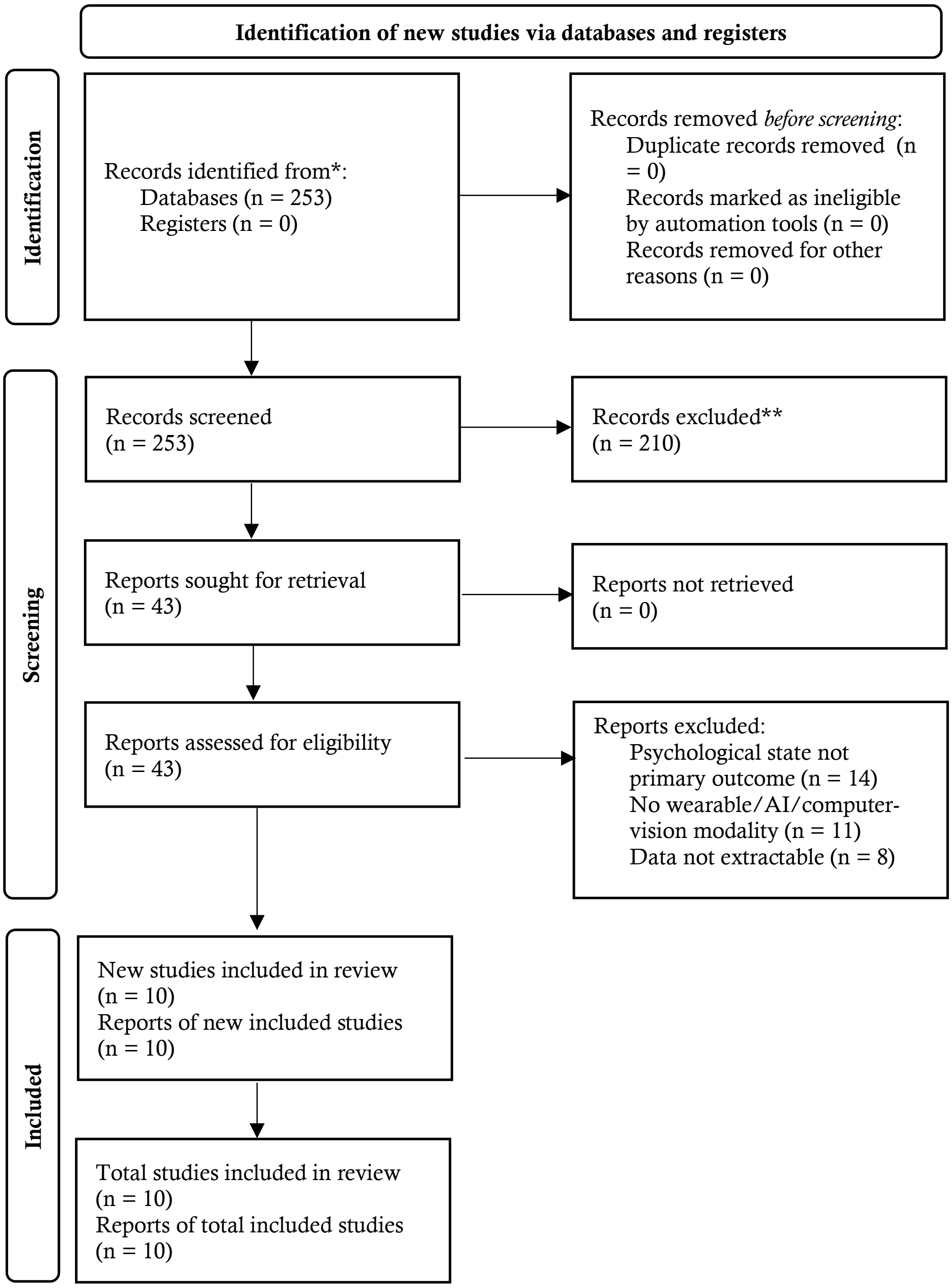

Figure 1. PRISMA 2020 flow diagram summarising identification, screening, eligibility, and inclusion (253 records identified; 43 full texts assessed; 10 studies included).

Results and Discussions

Study Selection Results

Study selection results. The Scopus search returned 253 records. Because the export came from one database, no duplicate titles or DOIs were identified in the archived dataset, so all 253 records entered title-and-abstract screening. That stage excluded 210 records: 50 were ineligible publication types, 78 did not address a psychological construct, and 82 did not involve a digital sensing, computer-vision, or artificial-intelligence modality. Forty-three full texts were assessed. Thirty-three were excluded because the psychological state was not a primary outcome (n=14), no wearable/AI/computer-vision modality was present (n=11), or the available information was insufficient for extraction (n=8). Ten studies met all inclusion criteria and were retained for qualitative synthesis.

Descriptive Characteristics

Descriptive characteristics. The literature increased markedly after 2020, but the temporal pattern is descriptive rather than inferential. The ten included studies span 2015–2025 and mix tennis and table tennis contexts with distinct sensing traditions. Two studies use EEG as a primary modality, two emphasize HR/HRV or cardiovascular sensing, one uses an IMU wearable, one uses multimodal biofeedback, one uses a multi-wearable sensor suite, one combines cortisol/heart rate/RPE, one operationalizes mental fatigue through a cognitive-plus-physiological protocol, and one uses computer vision for behavioural action recognition. This heterogeneity is important because it limits direct cross-study comparison.

Table 3. Summary of the ten included studies (complete sample/context and modality information extracted from the Scopus source corpus).

Author(s) & Year	Context / Sample	Modality & Method	Key Findings
Havlucu et al. (2022)	Elite tennis players; two experimental games with elite coaches	IMU wearable + recurrent neural network; coach-labelled states	Machine learning on wearable IMU data predicted the psychological "zone" state, demonstrating that wearables can deliver psychological—not only physiological—feedback.
Wang (2025)	100 professional Chinese tennis players (M age 21.5)	Smartwatch, biosensor wristband, GPS, EMG; quantitative modelling	Integrating multiple wearables improved monitoring of HR, HRV, and recovery and enhanced psychological-stress management capabilities.
Tsai et al. (2022)	Table tennis players under laboratory stress induction	EEG + supervised machine learning (SVM, RF, decision tree, logistic regression)	EEG features analysed with machine learning classified stress states, with ensemble/tree models achieving the most accurate stress detection.
Peng & Kim (2023)	Table tennis players; psychological-training context	Heart-rate sensing + deep-learning training system	A deep-learning method improved heart-rate detection accuracy and provided a structured basis for psychological training.
Makaracı et al. (2024)	Randomised controlled trial with 16 international tennis players	Multimodal biofeedback (10 sessions); pre–post assessment	Biofeedback improved skin temperature, reduced state anxiety, and enhanced cognitive functioning relative to control.
Habay et al. (2021)	11 experienced table tennis players; crossover design	Stroop-induced mental fatigue + EEG electrophysiology	Mental fatigue was associated with performance decrements; the study probed the electrophysiological signature of fatigue-related impairment.
Fernandez-Fernandez et al. (2015)	12 young female competitive tennis players	Salivary cortisol, competitive anxiety inventory, HR, RPE	Competition elicited greater psychophysiological stress than training, and losers showed higher cortisol than winners.
Díaz-García et al. (2023)	32 tennis players (25 men, 7 women)	HIIT plus incongruent Stroop cognitive task; fatigue indicators	Adding a cognitive task to HIIT increased mental fatigue without raising physical workload, dissociating mental from physical load.
Gao & Ju (2024)	Tennis action-recognition dataset	Computer vision + attention-enhanced gated recurrent unit (GRU)	An attention-enhanced GRU improved automated recognition of tennis actions,

Author(s) & Year	Context / Sample	Modality & Method	Key Findings
			supplying a behavioural substrate for phenotyping.
Cusano et al. (2025)	Young tennis athletes; post-training/competition recovery	Heart-rate variability monitoring; recovery indices	HRV patterns reflected fatigue, recovery, and autonomic adaptation, informing individualised recovery strategies.

Table 4 takes that same set of studies and classifies them differently. The categories are research design, thematic focus, sensing or analytical technology, and reported outcome. What emerges is a clear mapping between psychological constructs and digital modalities, and that mapping is what the thematic synthesis below is built on.

Table 4. Classification of included studies by design, theme, technology, and outcome domain.

Author(s) & Year	Research Design	Theme / Focus	Technology / Modality	Outcome Domain
Havlucu et al. (2022)	Applied ML / experimental	Zone / flow state	IMU wearable + RNN	Affective–performance state
Wang (2025)	Quantitative field study	Psychological stress	Multi-wearable sensor suite	Stress regulation
Tsai et al. (2022)	Experimental / ML	Stress classification	EEG + machine learning	Cognitive–affective state
Peng & Kim (2023)	Design / experimental	Psychological training	HR sensing + deep learning	Training optimisation
Makaracı et al. (2024)	Randomised controlled trial	Anxiety & cognition	Multimodal biofeedback	State anxiety, cognition
Habay et al. (2021)	Crossover experiment	Mental fatigue	EEG electrophysiology	Cognitive fatigue
Fernandez-Fernandez et al. (2015)	Comparative field study	Psychophysiological stress	Cortisol, HR, RPE, anxiety scale	Competitive stress
Díaz-García et al. (2023)	Cross-sectional experiment	Mental vs physical fatigue	Cognitive task + physiological load	Fatigue dissociation
Gao & Ju (2024)	Computational modelling	Behavioural phenotyping	Computer vision + GRU	Action recognition
Cusano et al. (2025)	Experimental field study	Autonomic recovery	HRV monitoring	Fatigue & recovery

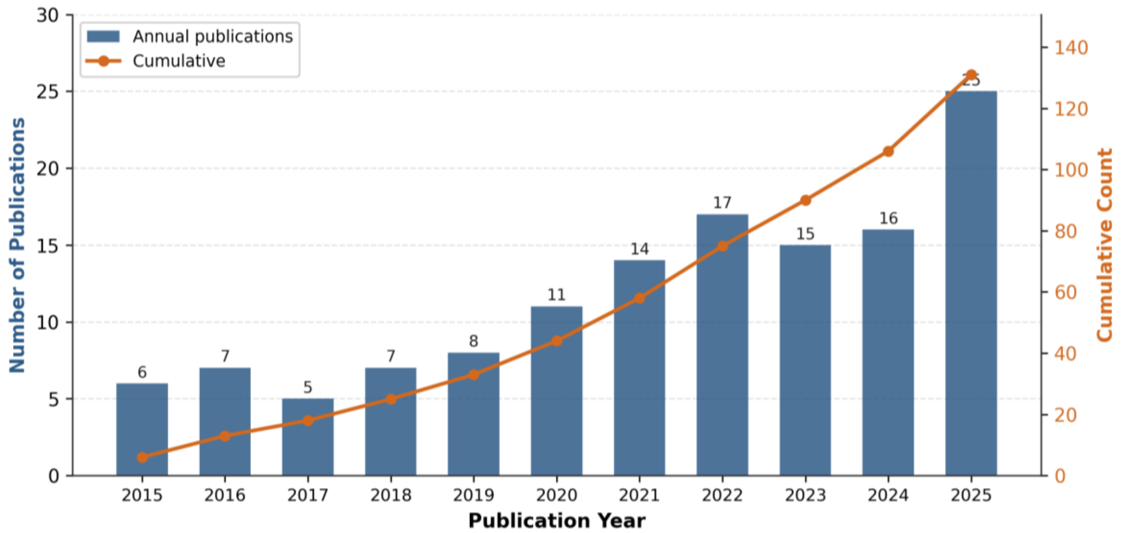

Figure 2. Annual and cumulative publication counts for digital-phenotyping research in tennis, indicating rapid growth after 2020.

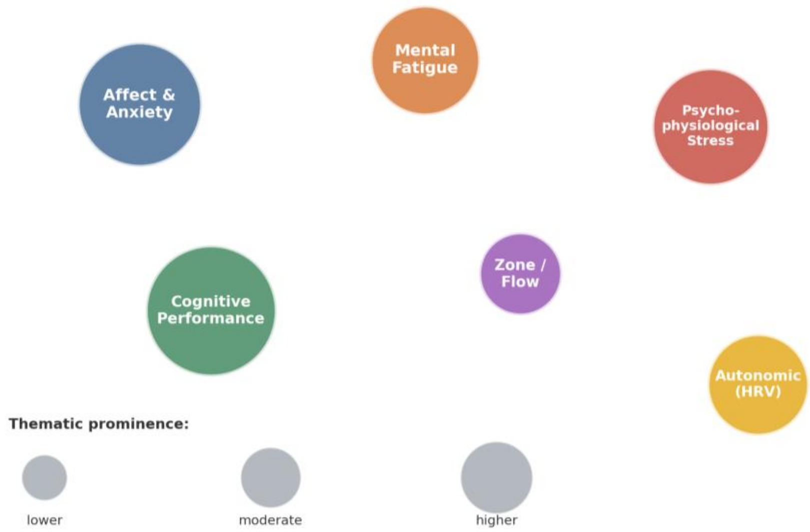

Figure 3. Thematic landscape of psychological constructs monitored through digital modalities across the reviewed corpus.

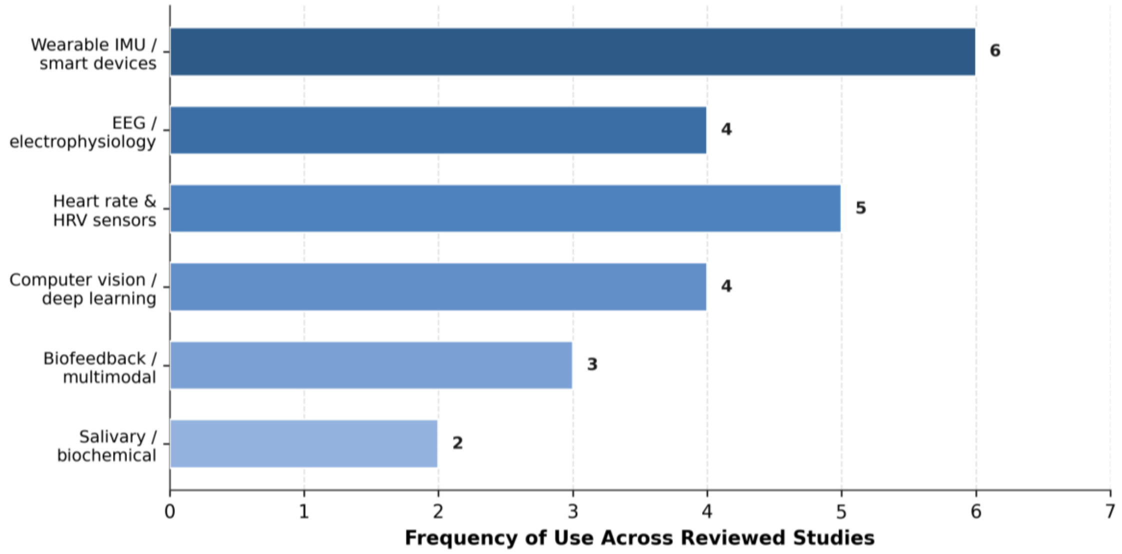

Figure 4. Distribution of digital sensing and analytical modalities used for psychological phenotyping in the included studies.

Findings for RQ1: Modalities and Target Psychological States

The corpus reveals a measurement ecosystem that is diverse, yet unevenly developed. Physiological inference is dominated by autonomic and cardiovascular sensing. Heart-rate variability has been used to index autonomic balance, competitive stress, and recovery, while composite psychophysiological protocols combine heart rate, salivary cortisol, perceived exertion, and anxiety measures. Wang (2025) extends this logic with a multi-wearable suite. These approaches can support psychological monitoring, but the physiological signals are not uniquely psychological and must be interpreted alongside exertion and context.

A second measurement stream centres on the brain. EEG has been used to classify stress and to characterise electrophysiological correlates of mental fatigue. Event-related and neuroimaging studies contribute complementary evidence on expertise, attention, and social interaction. These studies enrich mechanistic understanding, but most use small samples or task-induced conditions, so they provide construct-related evidence more readily than deployment-ready prediction.

A third stream concerns movement and behaviour. Havlucu et al. (2022) used IMU data and coach labels to predict a flow-like zone state, whereas Gao and Ju (2024) used computer vision to recognize tennis actions. The latter should be interpreted as a behavioural substrate for future phenotyping rather than as direct evidence of psychological-state classification. This distinction is critical: computer vision can quantify observable behaviour, but the psychological meaning of that behaviour requires independent construct validation.

Findings for RQ2: Analytical Performance and Rigour

Supervised machine learning is the dominant analytical engine for converting biosignals into psychological categories across the corpus. Tsai et al. (2022) compared logistic regression, support vector machines, decision trees, random forests, and XGBoost for EEG-based stress classification; Havlucu et al. (2022) used recurrent neural networks for zone prediction from IMU data. Deep-learning systems are also used in adjacent behavioural recognition tasks. The evidence indicates technical feasibility, but validation metrics are not harmonized across studies.

Analytical rigour varies considerably. Samples are often small and homogeneous, and some studies are explicitly feasibility or crossover studies. For example, the zone-detection study involved four professional players and found strong within-player accuracy but near-random generalization to previously unseen players, illustrating the difference between personalized performance and external validity. Tsai et al. (2022) likewise reported substantially different performance for generalized versus personalized stress models. These findings argue against treating a single accuracy estimate as evidence of transportability.

Interpretability constitutes a further concern. High-performing neural networks may operate as opaque classifiers, while physiological proxies can reflect both psychological and physical load. Sensitivity, specificity, and area under the curve (AUC) were not reported consistently across the 10 included studies. Accordingly, these metrics cannot be pooled or compared as if they measured the same construct. A model is considered more deployment-ready only when it demonstrates subject-wise or external validation, calibration, and psychometric concordance in a realistic tennis environment.

Findings for RQ3: Translation into Applied Practice

Three actionable pathways emerge from the reviewed evidence. The first is objective, individualised feedback. Coaches can be supplied with real-time psychological indicators by systems that infer the zone or stress state from wearables (Havlucu et al., 2022; Wang, 2025), complementing subjective observation and enabling data-informed adjustments to workload and tactics. The second is intervention. Multimodal biofeedback improved state anxiety and cognition in international tennis players (Makaracı et al., 2024), and deep-learning-supported psychological-training frameworks offer structured protocols for self-regulation practice (Peng & Kim, 2023). Phenotyping, in other words, can close the loop from measurement to modulation.

Recovery and welfare make up the third pathway. HRV-based monitoring reveals autonomic responses to competitive load and thereby informs individualised recovery scheduling (Cusano et al., 2025; Villafaina et al., 2023). Dissociating mental from physical fatigue helps periodise cognitive as well as physical demands (Díaz-García et al., 2023; Habay et al., 2021). Beyond elite performance, racket-sport participation brings psychological and cardiovascular benefits to clinical and inclusive populations (Tsou et al., 2024). And recreational players are increasingly accepting digital-health activities (Nan, 2022), which signals a receptive environment for phenotyping tools.

Realising these applications nonetheless requires attention to feasibility and trust. The proliferation of low-cost wearables and biofeedback protocols lowers the technical barrier to deployment, but psychological inferences should only guide high-stakes decisions when they are transparent, reliable, and demonstrably valid in context. The practical translation agenda therefore depends on the methodological safeguards identified under RQ2.

Comparative and Critical Analysis

Looking across the corpus, a clear evolutionary arc is visible. Early work from 2015 to 2018 relied on psychophysiological and biochemical measurement with conventional inferential statistics (Fernandez-Fernandez et al., 2015; Gescheit et al., 2015; Liu & Wang, 2018), emphasising construct validity but offering limited scalability or real-time capability. From 2020 onward, the pivot toward sensor-rich, computationally intensive designs became pronounced, with machine learning and deep learning taking over as the dominant analytical paradigm (Havlucu et al., 2022; Tsai et al., 2022; Gao & Ju, 2024; Wang, 2025). Some psychometric precision was traded for continuous, ecologically embedded measurement in the process.

Dominant methods cluster into three families: physiological-statistical designs, neural-classification designs, and behavioural deep-learning designs. Longitudinal cohort designs, cross-modal fusion, subject-wise validation, and explainable-AI methods remain underused. The methodological evolution is therefore one in which sensing capability has expanded faster than validation and transparency practices.

Comparative validation benchmark

To strengthen comparative rigor, the review was benchmarked against 15 related studies already represented in the reference corpus: Beck et al. (2025), Cusano et al. (2025), Ding et al. (2024), Díaz-García et al. (2023), Fernandez-Fernandez et al. (2015), Gao and Ju (2024), Gescheit et al. (2015), Habay et al. (2021), Havlucu et al. (2022), Kaye et al. (2025), Li and Wu (2023), Makaracı et al. (2024), Park et al. (2020), Peng and Kim (2023), and Tabrizi et al. (2021). The comparator set spans psychophysiological field studies, EEG experiments, wearable inference, biofeedback interventions, and behavioural/action recognition. Importantly, the studies do not share a common target, sensor, ground-truth label, or evaluation metric. Some report classification accuracy, others report reaction-time or physiological changes, and intervention studies report pre–post effects. Sensitivity, specificity, and AUC are therefore not available on a common basis and were not synthesized numerically. The most informative comparison is methodological: the tennis-zone study reported >85% within-study test accuracy but only about 51% accuracy when generalizing to previously unseen players, while the table-tennis EEG stress study reported 86.49% accuracy for a generalized three-level model and 94.27% for a two-level model. These contrasts show why within-study accuracy should not be interpreted as universal predictive validity.

A conservative non-pooling rule was applied: no meta-analysis was attempted because no construct × modality × outcome cluster contained at least three studies using comparable labels and validation procedures. The heterogeneity is therefore treated as methodological, not merely statistical, and quantitative synthesis would have risked combining non-equivalent targets.

Discussion

Taken collectively, the findings suggest that digital phenotyping of psychological states in tennis has moved from proof-of-concept toward practical plausibility, but not yet to validated deployment. Autonomic, neural, and behavioural signals can carry information relevant to affect, stress, fatigue, cognition, and flow; however, the strength of inference depends on the construct definition, labeling source, physiological context, and validation design.

Theoretically, the evidence extends existing frameworks of the athlete stress-recovery process and of flow by rendering their internal constructs computationally observable. Mental and physical fatigue can be dissociated through combined cognitive-physical loading (Díaz-García et al., 2023; Habay et al., 2021), which challenges unidimensional models of fatigue and aligns with psychobiological accounts in which central and peripheral processes contribute independently to performance decline. Similarly, the neural characterisation of expertise and social interaction (Wang et al., 2022; Zhang et al., 2025; Guan et al., 2025) enriches perceptual-cognitive theories of racket-sport skill.

Practically, coaches, sport psychologists, and performance staff gain a toolkit for objective, individualised monitoring and intervention. Real-time zone and stress indicators (Havlucu et al., 2022; Wang, 2025), biofeedback protocols that reduce state anxiety (Makaracı et al., 2024), and HRV-guided recovery (Cusano et al., 2025) can be embedded into athlete-management systems. For developmental and inclusive contexts, accessible technologies and participation-based interventions (Park et al., 2020; Tsou et al., 2024; Kilit et al., 2024) broaden the reach of psychological support.

Situated against prior reviews, this synthesis both corroborates and extends existing work. Ding et al. (2024) systematically documented the performance costs of mental fatigue in racket sports, and applied-machine-learning reviews have catalogued performance-optimisation uses of AI in tennis. The current review complements these by foregrounding the psychological dimension rather than purely performance, and by integrating sensing, vision, and AI evidence under a single phenotyping lens. Contradictions persist nonetheless. Laboratory-induced stress paradigms sometimes yield classification success that field studies struggle to reproduce (Tsai et al., 2022; Beck et al., 2025), and the psychological meaning of a given biosignal can vary with context, individual, and task (Fernandez-Fernandez et al., 2015; Kaye et al., 2025). Differences in sample, labelling, and ecological setting most plausibly explain these discrepancies.

Limitations

Four limitations should temper interpretation. First, measurement validity differs substantially across sensors: HR/HRV and electrodermal signals are sensitive to exertion, EEG is sensitive to task and acquisition conditions, and computer vision measures observable behaviour rather than psychological state itself. Second, several studies use small samples; neurophysiological evidence in particular should not be generalized beyond the studied participants without subject-wise or external validation. Third, generalizability is constrained by the mixture of tennis, table tennis, laboratory stress tasks, training contexts, and competitive settings. Fourth, the review relied on a single Scopus database, no public protocol registration exists for the archived review, the exact search execution date was not retained, and independent reviewer-level screening records were unavailable, preventing retrospective calculation of Cohen's kappa. These limitations are reported explicitly rather than replaced with inferred statistics.

Methodological recommendations.
(1) Standardize digital-phenotyping protocols by preregistering the target psychological construct, ground-truth instrument or expert-label procedure, sensor placement, sampling rate, and preprocessing pipeline. (2) Separate subject-wise generalization from personalized models and report both, because high within-athlete accuracy can coexist with near-random performance on unseen athletes. (3) Use a priori power analysis and adequately powered multicohort samples; a minimum pilot target of around 30 participants can be treated as a lower exploratory floor rather than a universal rule (Tsai et al., 2022), while larger samples are needed for externally validated machine-learning deployment. (4) Validate every inferred psychological state against established psychometric instruments and report sensitivity, specificity, AUC, calibration, and confidence intervals when classification is appropriate. (5) Conduct longitudinal on-court validation during authentic training and competition, with explainable multimodal models and privacy-preserving governance.

In direct response to the research questions, RQ1 is answered by demonstrating that wearable inertial and cardiovascular sensors, EEG and neuroimaging, and computer-vision systems each target distinct psychological states (stress, mental fatigue, cognition, and the zone), though largely in isolation. RQ2 is answered by showing that machine-learning and deep-learning models classify these states with promising but variable accuracy, constrained by small samples, weak ecological validity, and limited interpretability. RQ3 is answered by identifying feedback, intervention, and recovery as the principal translation pathways, contingent on resolving the validation challenges above.

Conclusions

This systematic review synthesised evidence on digital phenotyping of psychological states in tennis and related racket sports using wearable sensors, computer vision, and artificial intelligence. Ten studies met the stated inclusion criteria. RQ1 shows that autonomic/cardiovascular sensing, EEG, inertial sensing, and computer-vision systems contribute complementary information on stress, anxiety, mental fatigue, cognitive states, recovery, and flow-like performance states, but they do not measure a common construct. RQ2 shows that machine-learning and deep-learning models can achieve strong within-study classification or prediction, yet sensitivity, specificity, AUC, and external-validation evidence are inconsistently reported and cannot be pooled. RQ3 indicates that feedback, biofeedback intervention, and recovery monitoring are the most plausible translation pathways, provided the inferred states are psychometrically anchored and validated in realistic settings. The review is bounded by a single-database Scopus strategy, absence of public protocol registration, loss of the exact search-execution date in the archived export, lack of retained reviewer-level screening logs for Cohen's kappa, and methodological heterogeneity that precluded meta-analysis. No aggregate effect size or pooled accuracy is claimed. The principal contribution is therefore a transparent map of what the present literature can support and what must be validated before psychological digital phenotyping can be treated as a trustworthy component of tennis performance science.

Acknowledgments

The Authors would like to thank Faculty of Sport Sciences, Universitas Negeri Padang for providing excellent facilities, as well as for being very supportive of this research from the beginning. Authors also thank the team members who had generously shared their brilliant ideas, engaging discussion, and academic supports during the study.

References

Beck, D., Zahno, S., Kredel, R., & Hossner, E.-J. (2025). From simple lab tasks to the virtual court: Bayesian integration in tennis. Journal of Neurophysiology, 134(1), 303-313. https://doi.org/10.1152/jn.00434.2024

Cusano, P., Di Lascio, G., Arben, K., Agron, K., & Giardullo, G. (2025). HRV MONITORING AND RECOVERY RATES: AN EXPERIMENTAL STUDY ON YOUNG TENNIS ATHELTES. Sportske Nauke i Zdravlje, 15(5), 40-44. https://doi.org/10.7251/SSH25V040C

Ding, C., Soh, KG., Sun, H., Roslan, S., Cao, S., & Zhao, Y. (2024). Does mental fatigue affect performance in racket sports? A systematic review. BMC Sports Science, Medicine and Rehabilitation, 16(1), Article 179. https://doi.org/10.1186/s13102-024-00963-w

Díaz-García, J., Clemente-Suárez, VJ., Fuentes-García, JP., & Villafaina, S. (2023). Combining HIIT Plus Cognitive Task Increased Mental Fatigue but Not Physical Workload in Tennis Players. Applied Sciences (Switzerland), 13(12), Article 7046. https://doi.org/10.3390/app13127046

Fernandez-Fernandez, J., Boullosa, DA., Sanz-Rivas, D., Abreu, L., Filaire, E., & Mendez-Villanueva, A. (2015). Psychophysiological stress responses during training and competition in young female competitive tennis players. International Journal of Sports Medicine, 36(1), 22-28. https://doi.org/10.1055/s-0034-1384544

Gao, M., & Ju, B. (2024). Attention-enhanced gated recurrent unit for action recognition in tennis. PeerJ Computer Science, 10, Article e1804. https://doi.org/10.7717/peerj-cs.1804

Gescheit, DT., Cormack, SJ., Reid, M., & Duffield, R. (2015). Consecutive days of prolonged tennis match play: Performance, physical, and perceptual responses in trained players. International Journal of Sports Physiology and Performance, 10(7), 913-920. https://doi.org/10.1123/ijspp.2014-0329

Guan, Y., Wang, Y., Zhou, C., Wang, J., & Lu, Y. (2025). Distinct patterns of cognitive enhancement: The role of motor experience in domain-specific and general cognitive functions. Neuroscience, 576, 118-128. https://doi.org/10.1016/j.neuroscience.2025.04.034

Habay, J., Proost, M., De Wachter, J., Díaz-garcía, J., De Pauw, K., Meeusen, R., Van Cutsem, J., & Roelands, B. (2021). Mental fatigue-associated decrease in table tennis performance: is there an electrophysiological signature?. International Journal of Environmental Research and Public Health, 18(24), Article 12906. https://doi.org/10.3390/ijerph182412906

Havlucu, H., Akgun, B., Eskenazi, T., Coskun, A., & Ozcan, O. (2022). Toward Detecting the Zone of Elite Tennis Players Through Wearable Technology. Frontiers in Sports and Active Living, 4, Article 939641. https://doi.org/10.3389/fspor.2022.939641

Kaye, G., Johnston, E., Burke, J., Gasson, N., & Marinovic, W. (2025). Differential Effects of Visual and Auditory Cognitive Tasks on Smooth Pursuit Eye Movements. Psychophysiology, 62(5), Article e70069. https://doi.org/10.1111/psyp.70069

Kilit, B., Arslan, E., Soylu, Y., & Lane, AM. (2024). Does Playing Tennis with a Low-Compression Ball Effect Psychophysiological Responses and Match Characteristics in Recreational Adult Players?. Sports, 12(3), Article 80. https://doi.org/10.3390/sports12030080

Li, Y., & Wu, L. (2023). A Wireless Self-Powered Sensor Network Based on Dual-Model Convolutional Neural Network Algorithm for Tennis Sports. IEEE Sensors Journal, 23(18), 20745-20755. https://doi.org/10.1109/JSEN.2023.3255226

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), Article e1000100. https://doi.org/10.1371/journal.pmed.1000100

Liu, Y., & Wang, X. (2018). Tennis teaching and variation pattern of neurotransmitter based on electroencephalographic testing technology. NeuroQuantology, 16(5), 172-178. https://doi.org/10.14704/nq.2018.16.5.1277

Liu, C., Hu, S., Dong, L., Yang, W., Li, S., Li, D., & Wang, G. (2024). Design of Table Tennis Intelligent Scoring System Based on Accelerometer Array. Chinese Journal of Sensors and Actuators, 37(2), 353-359. https://doi.org/10.3969/j.issn.1004-1699.2024.02.022

Makaracı, M., Makaracı, Y., Zorba, E., & Lautenbach, F. (2024). Effects of Ten Biofeedback Sessions on Athletes' Physiological, Psychological, and Cognitive Functioning: A Randomized Controlled Trial with International Tennis Players. Perceptual and Motor Skills, 131(5), 1664-1686. https://doi.org/10.1177/00315125241274834

Nan, C. (2022). Acceptance of Digital Sports: A Study Showing the Rising Acceptance of Digital Health Activities Based on Deep Learning. Journal of Commercial Biotechnology, 27(3), 214-226. https://doi.org/10.5912/jcb1331

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, Article n71. https://doi.org/10.1136/bmj.n71

Pan, N., Wang, Y., & Tong, J. (2022). Analysis and Improvement of Tennis Motion Recognition Algorithm Based on Human Body Sensor Network. Mobile Information Systems, 2022, Article 8188571. https://doi.org/10.1155/2022/8188571

Park, SH., Hwang, S., & Lee, SM. (2020). Pilot Application of Biofeedback Training Program for Racket Sports Players. Annals of Applied Sport Science, 8(4), 1-5. https://doi.org/10.29252/aassjournal.898

Peng, J., & Kim, B.-M. (2023). Psychological Training Method for Table Tennis Players Using Deep Learning. Applied Sciences (Switzerland), 13(14), Article 8290. https://doi.org/10.3390/app13148290

Peng, Z., Xu, L., Lian, J., An, X., Chen, S., Shao, Y., Jiao, F., & Lv, J. (2024). Perceptual information processing in table tennis players: based on top-down hierarchical predictive coding. Cognitive Neurodynamics, 18(6), 3951-3961. https://doi.org/10.1007/s11571-024-10171-4

Tabrizi, SS., Pashazadeh, S., & Javani, V. (2021). A Deep Learning Approach for Table Tennis Forehand Stroke Evaluation System Using an IMU Sensor. Computational Intelligence and Neuroscience, 2021, Article 5584756. https://doi.org/10.1155/2021/5584756

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8, Article 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.-H., Wu, S.-K., Yu, S.-S., & Tsai, M.-H. (2022). Analyzing Brain Waves of Table Tennis Players with Machine Learning for Stress Classification. Applied Sciences (Switzerland), 12(16), Article 8052. https://doi.org/10.3390/app12168052

Tsou, H.-K., Chen, H.-Y., Shih, K.-C., & Lin, Y.-C. (2024). Para table tennis improves psychological/mental and cardiovascular health in individuals with spinal cord injury. NeuroRehabilitation, 55(1), 59-68. https://doi.org/10.3233/NRE-240083

Vellela, SS., Rao, MV., Mantena, SV., Reddy, MVJ., Vatambeti, R., & Rahman, SZ. (2024). Evaluation of Tennis Teaching Effect Using Optimized DL Model with Cloud Computing System. International Journal of Modern Education and Computer Science, 16(2), 16-28. https://doi.org/10.5815/ijmecs.2024.02.02

Villafaina, S., Crespo, M., Martínez-Gallego, R., & Fuentes-García, JP. (2023). Heart Rate Variability in Elite International ITF Junior Davis Cup Tennis Players. Biology, 12(1), Article 17. https://doi.org/10.3390/biology12010017

Wang, C., Yan, A., Deng, W., & Qi, C. (2022). Effect of Tennis Expertise on Motion-in-Depth Perception at Different Speeds: An Event-Related Potential Study. Brain Sciences, 12(9), Article 1160. https://doi.org/10.3390/brainsci12091160

Wang, Z. (2025). Integration of wearable technologies in monitoring physical performance and psychological stress in tennis players. Acta Psychologica, 260, Article 105706. https://doi.org/10.1016/j.actpsy.2025.105706

Wei, X., Wang, C., Song, Y., & Qi, C. (2025). Tennis expert-novice difference in motion-in-depth perception is associated with early inhibition of invalid attention. Scientific Reports, 15(1), Article 1492. https://doi.org/10.1038/s41598-024-83707-7

Ye, C., Zhu, R., Ma, J., Huang, H., Li, X., & Wen, J. (2024). Comprehensive Tennis Serve Training System Based on Local Attention-Based CNN Model. IEEE Sensors Journal, 24(7), 11917-11926. https://doi.org/10.1109/JSEN.2024.3366781

Zhang, Y., Zhu, H., & Franz, E. (2025). Physical activity indexed using table tennis skills modulates the neural dynamics of involuntary retrieval of negative memories. Experimental Brain Research, 243(1), Article 17. https://doi.org/10.1007/s00221-024-06948-y

Zhang, H., Liu, H., Li, Z., & Zhang, D. (2025). Distinct fNIRS Inter-Brain Coupling Patterns for Cooperation versus Competition in a Tennis Game. Social Cognitive and Affective Neuroscience, 20(1), Article nsaf060. https://doi.org/10.1093/scan/nsaf060

Çağın, M., Cicioğlu, Hİ., & Ertan, H. (2025). Comparison of the acute effects of traditional and reaction-based HIIT on brain connectivity and cognitive responses. BMC Sports Science, Medicine and Rehabilitation, 17(1), Article 351. https://doi.org/10.1186/s13102-025-01400-2