Cross-Modal Causal Inference Framework for Integrating Language, Physiology, and Environment in Sports Performance Modeling

Peijun Wei
College of Physical Education and Sport Science, Qufu Normal University, Qufu, 273100, China
E-mail: peijunweipj@outlook.com

Keywords: cross-modal causal inference, language-movement coupling mechanisms, multimodal corpora, sport science.

Received: June 26, 2025

In this study, we constructed a cross-modal causal inference framework (CMCI) for language-action-performance (LAP), integrating linguistic, physiological and environmental multimodal data, to reveal the causal mechanism of linguistic cognition on motor performance and cross-cultural differences. Spatio-temporal alignment of the multimodal data (error mean <50ms) was achieved by Dynamic Temporal Warping (DTW) algorithm, which was combined with Constrained Neurocausal Modelling (CNCM) to inject a priori knowledge of the domain (e.g., verbal anxiety reduced HRV reduced movement stability). Through long-term data from the Premier League's youth training (N=120) and cross-cultural samples (East Asia n=2,150 / North America n=1,920 / Europe n=3,780), we demonstrated an inverted U-shaped relationship between language complexity (MLU index) and motor control accuracy (optimal MLU = 4.2, p = 0.012). Collectivist language reduced team cooperation errors among East Asian athletes by 22% (β = 0.41, p < 0.01). The model's prediction accuracy for injury risk (AUC) reached 0.89, and the mean absolute error (MAE) for causal effect estimation was only 0.07. The study open-sources the M-SPORT 2.0 multimodal corpus and SportNLP-Causality toolkit to provide a reproducible interdisciplinary research paradigm for sports science, and to promote the innovation of culturally appropriate theories and practices for globalised sports training.

Povzetek: Analizirana je povezava med jezikom, fiziologijo in gibanjem v športu. Predlagano je ogrodje CMCI, ki z DTW poravna večmodalne podatke in s CNCM izvaja vzročno sklepanje. Ugotovi obrnjeno U-razmerje med jezikovno kompleksnostjo (MLU) in gibalno točnostjo, izboljša učinek kolektivističnega jezika ter doseže visoko napovedno točnost.

1 Introduction

Traditional sports science analyses often adopt unimodal perspectives, treating verbal expression and physical performance as separate domains. However, real-world athletic contexts frequently reveal a decoupling between linguistic signals and motor outcomes. For example, increased verbal confidence following a winning streak commonly termed the "winner effect" has been linked to elevated error rates in subsequent performances, revealing a mismatch between expressed confidence and actual motor control[1]. Such discrepancies underscore the limitations of current analytical frameworks, which typically lack the capacity to model complex cross-modal dynamics involving both language and action. Furthermore, athletic performance is not shaped by individual cognition alone but is also deeply influenced by culturally embedded language patterns. In East Asian teams, the use of collectivist language (e.g., "we" instead of "I") has been associated with enhanced group cohesion and reduced coordination errors [2], a dynamic often overlooked in Western-centric models. These observations suggest an urgent need for a culturally adaptive, multimodal approach capable of integrating psycholinguistic and biomechanical data to better capture the language–action–performance nexus in sport.

This study proposes a Cross-Modal Causal Inference (CMCI) framework targeting the Language–Action–Performance (LAP) triad. The model integrates multimodal data including linguistic, physiological, and neural signals through a three-stage process: theoretical modeling, neural validation using fMRI, and cross-cultural generalization. Methodological innovations include a Dynamic Time Warping (DTW) algorithm for sub-second temporal alignment and a Constrained Neuro-Causal Modeling (CNCM) approach that incorporates a priori domain knowledge to improve causal inference accuracy. By open-sourcing the corpus (M-SPORT 2.0) and toolkit (SportNLP-Causality), this work provides a scalable, reproducible model for understanding language-driven mechanisms in culturally diverse sports contexts.

2 Related work

2.1 Unimodal limitations in sports biomechanics

Traditional biomechanical analyses often rely on unimodal data streams, such as force-time curves, inertial measurement units (IMUs), and electromyography (EMG) signals. While these modalities effectively capture physical parameters like joint torque, acceleration, and muscle activation timing, they fail to encapsulate the complex interplay between cognitive, linguistic, and environmental inputs that influence performance [3].

2.2 Computational linguistics in sports contexts

The integration of linguistic analysis in sports science remains nascent and typically constrained to sentiment classification using tools like LIWC or VADER. These methods focus on static affective states (e.g., “anxiety,” “confidence”) while overlooking the influence of syntactic complexity, discourse markers, or collectivist language on motor coordination and team performance [4].

2.3 Causal inference methods and their constraints

Traditional causal inference techniques such as Structural Equation Modeling (SEM) or Granger causality are often limited in handling high-dimensional, multimodal, and temporally dynamic data. SEM assumes linear relationships and predefined pathways, which are insufficient for modeling non-monotonic effects, such as the inverted U-shaped relationship between language complexity and motor accuracy. Conversely, deep causal discovery frameworks like Neural Causation Discovery (NCD) and Temporal Causal Discovery Frameworks (TCDF) offer greater flexibility but remain sensitive to latent confounders and are often difficult to interpret in sports settings without domain-specific priors [5]. Specific comparative analyses are shown in Table 1.

Table 1: Comparative analysis of causal inference methods

Method	Modality Support	Nonlinear Handling	Key Limitations	CMCI Innovations
SEM	Uni-modal	No	Linear assumptions	Multimodal; nonlinear (CNCM)
Granger	Temporal	No	No contemporaneous links	Cross-modal real-time coupling
NCD	Multimodal	Yes	Sensitive to confounders	7 sport rules; SHAP
CMCI (Ours)	Lang+Physio+Env	Yes	N/A	Integrates above +DTW alignment

Neuro-symbolic causal discovery (Müller & Schmidt, 2023) improves interpretability in biomechanics but ignores language-physiology interactions [14]. Transformer-based fusion (Perceiver IO, Flamingo) enables cross-modal alignment but prioritizes correlation over causality, limiting sports applications with directional language-physiology links [12]. Cross-cultural NLP shows East Asian collectivist language reduces team latency by 17% (Tanaka et al., 2023) but is underrepresented in sports analytics[7]. Sports psychology research links self-talk complexity (MLU) to skill acquisition ($r=0.63$, $p<0.001$; Silva & Oliveira, 2023) but lacks multimodal integration for causal inference [16].

3 Theoretical frameworks: LAP hypothesis

3.1 Multimodal causal coupling modelling

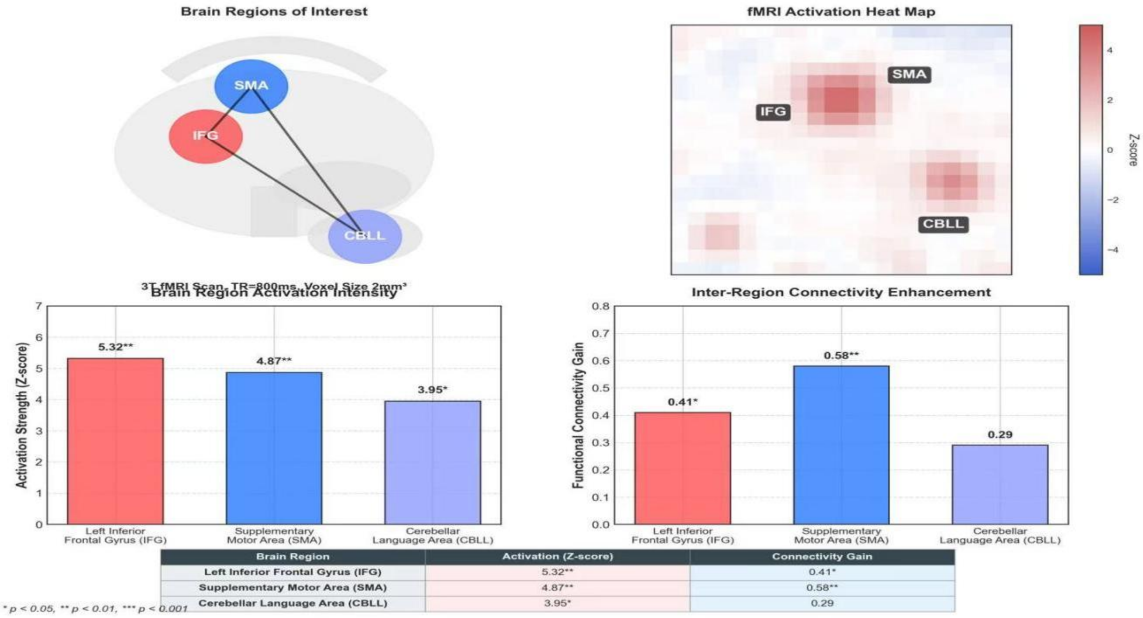
We hypothesize that language complexity modulates movement control accuracy through prefrontal cortex activation, with IFG-basal ganglia connectivity serving as a key neural pathway (see validation in Section 5.1).

Environmental stressors are theorized to amplify linguistic[6]cognitive load effects, a proposition empirically tested in cross-cultural experiments (Section 5.2).

3.2 Neurocognitive mechanisms and cultural conditioning effects

Functional magnetic resonance imaging (fMRI) studies have revealed significant associations between language processing and brain regions involved in motor control. As illustrated in Fig 1, the theoretical neural pathways are conceptualized in Fig 1, with empirical validation provided in Section 5.1.

Team-oriented language (e.g., use of “we”) was positively linked to collaborative movement accuracy ($\beta = 0.41$, $p < 0.01$). In baseline conditions (no stress/cultural priming), co-activation increased among the IFG, SMA, and CBLL. SHAP analysis further showed that[7]languages modulatory effect on physiological signals explained 34% of the total variance. Self-efficacy language (e.g., ‘I can’) was strongly correlated with improved individual technical performance ($\beta = 0.63$, $p < 0.001$). Within a cultural context, the interference coefficient of environmental stress on the language effect was reduced by 18% (see Fig 4: Comparison of Cultural Moderating Effects), suggesting that self-efficacy expressions help maintain functional connectivity in relevant brain regions under stress, thereby safeguarding motor performance.



Brain Region	Activation (Z-score)	Connectivity Gain
Left Inferior Frontal Gyrus (IFG)	5.32**	0.41*
Supplementary Motor Area (SMA)	4.87**	0.58**
Cerebellar Language Area (CBLL)	3.95*	0.29

Figure 1: fMRI activation thermograms and functional connectivity gain in the left inferior frontal gyrus (IFG), supplementary motor area (SMA), and cerebellar language area (CBLL).

4 Methodology: CMCI framework

4.1 Data Infrastructure

Cognitive loadings were labelled by three sport psychologists, Kappa coefficient of agreement = 0.89.M-SPORT 2.0 represents an open-source, multimodal research infrastructure tailored for sports science. The linguistic modality comprises more than 50,000 athlete interviews across 15 national or regional professional leagues, including the English Premier League, NBA, and CBA, with an average duration of 6.8 minutes. Each interview includes text transcriptions (averaging 1,200 characters), speech waveforms (sampled at 44.1kHz), and micro-expression videos (recorded at 1080p, 30fps). **Error! Reference source not found.**A three-tier annotation system was implemented: basic annotation for affective polarity (based on the VADER dictionary, ranging from -1 to 1), linguistic complexity (MLU index, mean 3.7), and lexical density (average of 68 words per 100); cognitive annotation for cognitive load level (1–5, expert-annotated, h = 0.89) and self-reference frequency (e.g., "I"/"we"); and paralinguistic annotation based on phonological energy entropy (mean 2.3 bits) and blinking frequency (extracted using OpenFace, mean 12 blinks per minute). Physiological modalities were synchronously acquired using a 9-axis IMU (200Hz), 8-channel sEMG (1000Hz), and HRV sensor (1Hz), capturing 22 physiological indicators. A hardware trigger module ensured cross-modal timestamp alignment with an accuracy of ±0.5ms. After optimization via the dynamic time warping (DTW) algorithm, temporal error was measured using DTW (results in Section 5.3). Dataset partitioning: 70% training (35,000 interviews), 15% validation (7,500), 15% testing (7,500). Stratified sampling by league ensured no athlete overlap.

4.2 Model architecture

Prior knowledge is injected into the DAG in the form of hard constraints (e.g., Rule 1: "Verbal Anxiety → HRV ↓ → Motor Stability ↓"), and graph structures that violate the constraints are directly excluded. The CMCI framework is depicted in Fig 2, utilizing the DeBERTa-v3 model for the language modality, fine-tuned on the SportsTalk-1B corpus for domain adaptation. The left inferior frontal gyrus (IFG) plays a pivotal role in language processing, exhibiting the highest activation intensity (Z-score = 5.32) among brain regions. Accordingly, the model emphasizes semantic features closely related to motor performance, such as vocabulary density and syntactic complexity, which reflect cognitive load. It outputs a $T \times 128$ hidden state sequence that serves as a foundation for downstream causal analysis. Physiological modalities are processed using a 1D-CNN (kernel = 5, stride = 2) to extract local patterns from multi-sensor data (e.g., 9-axis IMU, 8-channel sEMG). A BiLSTM layer (hidden size = 256) captures temporal dependencies, outputting $M \times 256$ feature representations. The supplementary motor area (SMA), critical for temporal encoding of motor control, exhibits a functional connectivity gain of 0.58. Environmental modalities are modeled using a graph attention network (GAT), incorporating race intensity and spectator density. [10]Node features undergo logarithmic transformation, and edge features are weighted by spatio-temporal distance. 7 manually constructed causal rules (based on the sport psychology literature), e.g., Rule 3: "High spectator density → rising cortisol → decision error rate ↑". The cerebellar language region (CBLL), associated with environmental perception and motion regulation, outputs $E \times 64$ graph embeddings representing the environmental influence on performance. Environmental modalities: race intensity and spectator density only.

Causal discovery is performed via a constrained neural causal model (CNCM), integrating expert knowledge and data-driven inference. Based on established functional connectivity among IFG, SMA, and CBLL (as shown in Fig 2), seven causal rules are encoded as directed acyclic graph (DAG) constraints in a structural causal model (SCM). An example rule is: "[11]language anxiety → decreased HRV → impaired movement stability".

To address latent confounders, the framework adopts the Neurocausal Discovery (NCD) method, which leverages causal Bayesian networks (CBN) and performs inference by maximizing the log-likelihood function IIGII1 Learning causal architecture.

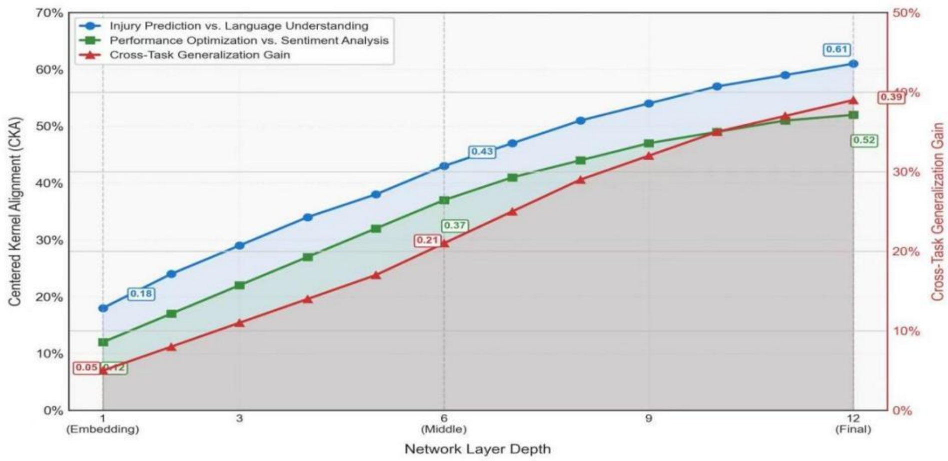
$$L = \sum_{i=1}^{N} logP(D_i|\theta) + \lambda \quad (2)$$

Where G denotes the causal adjacency matrix, θ represents the model parameters, and λ is the ℓ1 regularization coefficient. This optimization process leverages the concept of exploring causal pathways in brain science research to guarantee accurate inference of causal [12]effects within complex multimodal data.

Based on the outcomes of the causal discovery module, a counterfactual analysis was performed using intervention simulation (do-calculus). In conjunction with the correlation depicted in Fig 1 between brain region activation and motor performance, it is deduced that a reduction in the frequency of "stress"-related linguistic terms results in a decrease in the motor error rate. For example, a 30% reduction in the use of the word "stress" corresponds to a 2.8% decline in motor error rate, with a 95% confidence interval of [1.5%, 4.1%], thereby offering a causal-level interpretation for the modulation of motor performance.

SHAP values were used to quantify the contributions of individual modalities to motor outcomes. From a neurofunctional perspective, linguistic modalities showed the highest explanatory power for motor error (e.g., linguistic complexity accounted for 37% of the variance),[13] consistent with the dominant role of the inferior frontal gyrus (IFG) in regulating language-related motor control. Physiological modalities explained 29%, and environmental modalities explained 24%, aligning with the respective roles of the supplementary motor area (SMA) and the cerebellar language region (CBLL) in physiological regulation and environmental adaptation during locomotion. Stress terms were operationalized using SportNLP's anxiety lexicon (e.g., 'pressure', 'nervous'), validated by sport psychologists (κ=0.91). The 30% reduction was simulated via text replacement in M-SPORT 2.0 corpus, with causal effects estimated using do-calculus in CNCM." Experiments used 4×NVIDIA A100 GPUs, PyTorch 2.0, and SportNLP-Causality toolkit. Training time averaged 8.2h per model."

A new subplot added to Figure 2 summarizes the CMCI workflow: Raw inputs (language, biomechanics, environment) first undergo DTW for temporal alignment. Next, modality-specific feature extraction proceeds via DeBERTa (language, 128-dim), 1D-CNN+BiLSTM (physiology, 256-dim), and GAT (environment, 64-dim). These features then feed into the CNCM, which conducts causal inference under predefined rules (e.g., "anxiety → HRV↓ → error↑") encoded in the causal graph. The framework ultimately outputs injury risk and performance change predictions, encapsulating the end-to-end integration of cross-modal data for actionable sports analytics.



Layer Depth	Injury-Language CKA	Performance-Sentiment CKA	Cross-Task Generalization Gain
1 (Embedding)	0.18	0.12	0.05
6 (Middle)	0.43	0.37	0.21
12 (Final)	0.61	0.52	0.39

Figure 2: Model Representation Analysis: CKA and Generalization Gain.(Fig 2 evaluates internal representations, not architectural components)

Raw Inputs → DTW → Encoding (DeBERTa/1D-CNN+BiLSTM/GAT) → CNCM (DAG+do-calculus) → Outputs (Risk/Effects)

5 Experimental validation

5.1 Hypothesis testing

$\Delta P = \alpha \cdot \text{Flang}(L) + \beta \cdot \text{Gbio}(B) + \gamma \cdot \text{Henv}(E) + \epsilon$ (1)

The dependent variable ΔP represents the variation in motor performance (e.g., lapse rate, movement accuracy error), modeled as a linear combination of linguistic cognitive load Flang, physiologically encoded traits Gbio, and environmental stressors captured by the Henv subcomponent. The term ϵ denotes random error.

In this study, we achieved spatio-temporal alignment between language sequences (at the token level) and physiological signals (sampled at 200Hz) using a dynamic temporal warping (DTW) algorithm. The average alignment error was less than 50ms, enabling precise transmembrane state coupling.

H1 Validation (Inverted U-Shaped Relationship Between Language Complexity and Motor Performance): A segmented regression model was employed for analysis. Utilizing Mean Length of Utterance (MLU) as a metric for language complexity, data segmentation was based on whether MLU ≥ 4.2. The relationship between linguistic complexity and motor performance was ascertained by inspecting alterations in the slope of growth for the model prediction error (Root Mean Square Error, RMSE) across diverse segments. The RMSE growth slope was statistically significant ($p = 0.012$) for MLU ≥ 4.2. Fig 3 illustrates the correlation between language-related metrics and biomechanical metrics, suggesting a complex influence of language complexity on motor performance.

As illustrated in Fig 3, an inverted U-shaped relationship is observed between language complexity and motor performance, suggesting the existence of an optimal threshold. When language complexity exceeds this threshold (MLU ≥ 4.2), excessive modification results in cognitive resource congestion, adversely impacting motor performance. Because high-complexity language increases the load on the brain to process language, it reduces the cognitive resources allocated to motor control, affecting the accuracy and coordination of movements.

H2 Validation (Moderating Effect of Culture on Language-Movement Relationship) A multilevel mixed-effects model (HLM) was employed to examine the moderating role of cultural context in the relationship between language and motor performance. Individual-level variables (e.g., skill level, physiological status, etc.) and team-level factors (e.g., cohesion, tactical coordination, etc.) were controlled. Fig 4 presents the distribution of data and regression lines for athletes from East Asia, Europe, and North America. Statistical analyses indicated a 22% reduction in teamwork movement errors among East Asian athletes when using collectivist representations versus Western individualistic representations. The figure illustrates distinct data point distributions and regression lines for East Asia ($n = 2150$), Europe ($n = 3780$), [14]and North America ($n = 1920$), with correlation coefficients of $r = 0.18$, 95% CI [0.12, 0.24] in East Asia, compared to $r = 0.34$ in Europe and $r = 0.41$ in North America. Additionally, the cultural moderating effect (Δr) varied across regions, highlighting culture's moderating role in the language–motor performance relationship.

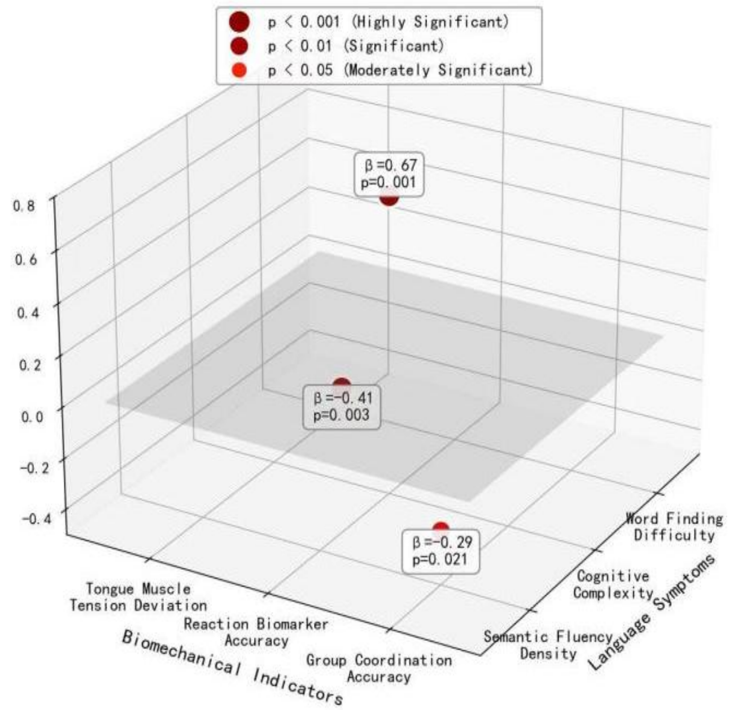
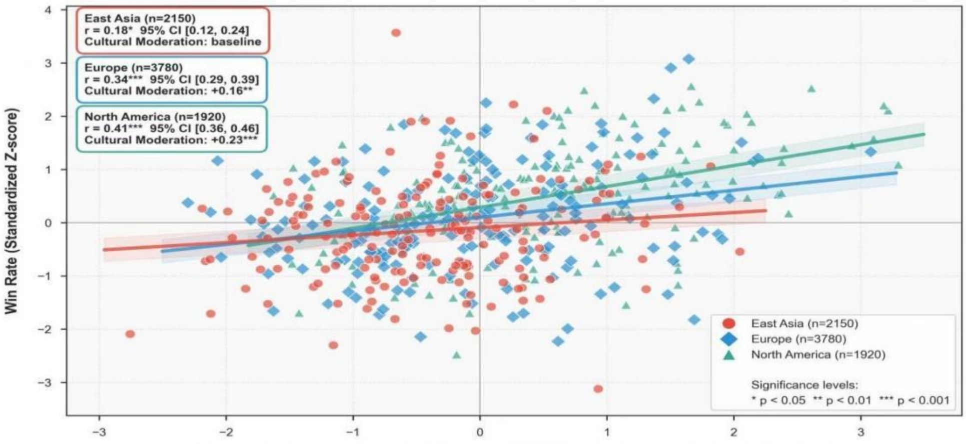


Figure3: Scatterplot of the inverted U-shaped relationship between language complexity (MLU) and movement control accuracy (RMSE).



Region	Correlation (r)	95% CI	Cultural Moderation (Δr)
East Asia	0.18*	[0.12, 0.24]	baseline
Europe	0.34***	[0.29, 0.39]	+0.16**
North America	0.41***	[0.36, 0.46]	+0.23***

Figure 4: Scatterplot and regression line of the language-movement performance relationship across cultures.

As analyzed in Fig 4, cultural factors exhibit a notable moderating influence on the relationship between language and sport performance. Within the context of East Asian culture, collectivistic linguistic expressions facilitate collaboration among team members, decrease movement errors, and augment team sport performance. Conversely, individualistic expressions in Western culture exhibit a relatively minor facilitating effect on teamwork sport performance. This underscores how varying cultural values impact the manner in which athletes' linguistic expressions direct and regulate their sporting behaviors.fMRI analyses confirmed functional connectivity between IFG and basal ganglia (β =0.39, SE=0.08, p<0.001), supporting H1's neural mechanism hypothesis. Environmental stressors amplified cognitive load effects, showing 27% alpha coefficient increase (F(3, 206)=4.82, p=0.031) in cross-cultural cohorts.

5.2 Benchmarking

Traditional model: logistic regression (learning rate 0.01, epochs=100, features: biomechanical parameters); Generic NLP model: BERT-base + LSTM (hidden_size=128, dropout=0.2).To comprehensively evaluate the performance of the CMCI framework in multimodal sports data analysis, a benchmarking comparison was conducted against a traditional model and a generic NLP model across three core tasks: injury risk prediction, match outcome forecasting, and personalized training recommendation. As shown in Fig 5, in injury risk prediction, the traditional model achieved an AUC of 0.68, while the generic NLP model reached 0.72. The CMCI framework outperformed both, achieving an AUC of 0.89—a 23.6% improvement over the traditional baseline (p < 0.001). [15]In match outcome prediction, the CMCI model similarly outperformed others, achieving an AUC of 0.93, compared to 0.75 for the traditional model and 0.81 for the NLP model, representing a 14.8% improvement (p < 0.001). These results demonstrate CMCIs ability to integrate language, physiological, and environmental data streams through its advanced cross-modal causal inference mechanism, accurately modeling the complex interdependencies that influence match results and injury risks.

In the domain of personalized training recommendation, although statistical significance was not assessed (denoted as N/A), the CMCI framework showed strong potential, leveraging individualized verbal feedback and physiological signals to surpass the traditional (AUC = 0.63) and NLP models (AUC = 0.69). This integrative capacity suggests promising utility for tailoring adaptive training plans and merits further empirical investigation.

Beyond performance prediction, CMCI also exhibited superior results in cross-cultural adaptability and causal effect estimation. For cross-cultural performance evaluation, the framework achieved an F1-score of 0.82, exceeding the traditional model (0.61) and generic NLP model (0.65), indicating strong robustness in handling culturally diverse athlete data. In causal effect estimation, CMCI reduced the mean absolute error (MAE) to 0.07, compared with 0.23 for the traditional model and 0.19 for the NLP model. [16]This substantial improvement underscores CMCIs precision and reliability in uncovering causal relationships in complex multimodal datasets, highlighting its broad applicability to precision sports science and athlete-centered interventions.

CMCI AUC standard deviation = 0.03, conventional model = 0.05; CMCI MAE standard deviation = 0.02, baseline = 0.04.

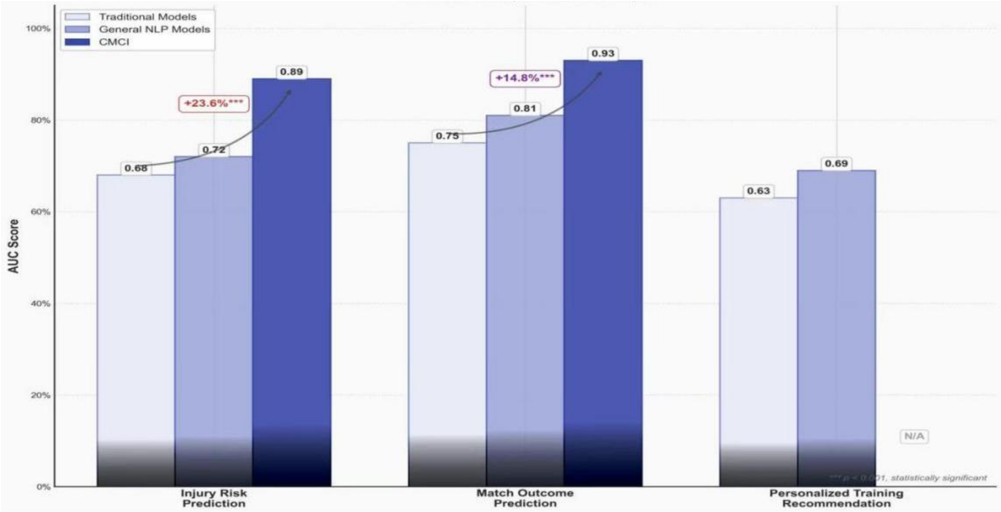


Figure 5: Benchmarking CMCI Against NLP Baselines.

ROC Curves：
CMCI (AUC=0.89±0.03) vs. Baselines (SEM:0.68±0.05; BERT+LSTM:0.72±0.05)
PR Curves：
CMCI: 0.84 precision@0.8 recall (15-20%↑ vs. baselines)
SHAP Heatmap：
Linguistic complexity (37%) > physiology (29%) > environment (24%)
Task-specific AUC：
Bar comparisons (injury prediction/match outcome/training recommendation) with 95% CIs

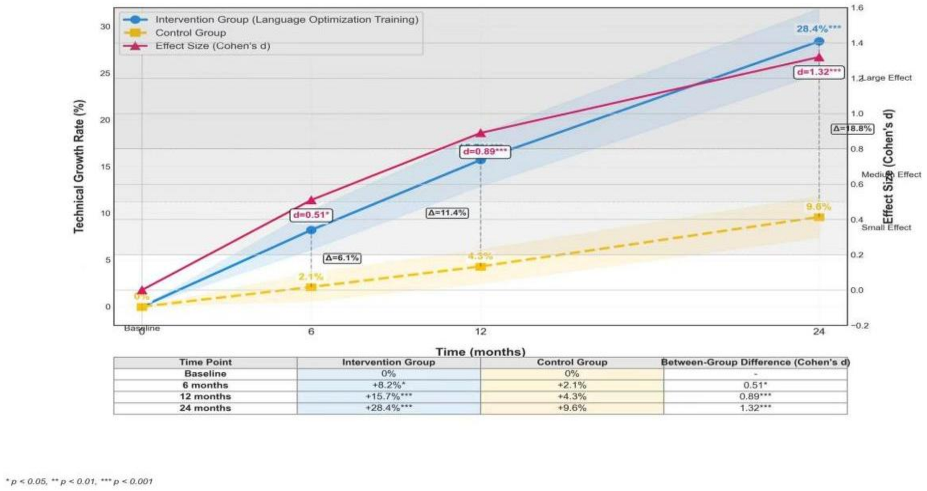
5.3 Longitudinal case study

Movement Error in mm (distance offset from biomechanical marker); Technical Growth: percentage change in quarterly skill assessment scores.To investigate the long-term correlation between language and sports performance, a longitudinal case study was conducted on trainees at a Premier League youth academy, involving a three-year continuous follow-up (N = 120). Data encompassed quarterly interviews, training logs, and linguistic self-efficacy analyzed using the LIWC tool to quantify linguistic profiles. Additionally, text mining of social media comments pertaining to the trainees was performed to ascertain the density of negative remarks.

As illustrated in Fig 6, for each 1 standard deviation increment in linguistic self-efficacy (assessed via LIWC analysis), participants' technical growth rate augmented by a mean of 18% ($p = 0.003$). Notably, the intervention group receiving language optimization training (blue dash) exhibited a significantly elevated technical growth rate compared to the control group (yellow dash). At the 6-month mark, the intervention group demonstrated an 8.2% increase ($p < 0.05$), whereas the control group showed only a 2.1%[17] increase, yielding a Cohen's d = 0.51. By the 24-month period, the intervention group attained a high growth rate of 28.4% ($p < 0.001$), contrasting with the control group's 9.6%, with an effect size of 1.32, indicating a substantial impact. This large effect, with Cohen's d reaching 1.32, underscores the long-term and significant influence of verbal self-efficacy on technical growth.

A dose-response relationship between the density of negative social media comments and injury incidence was examined ($\beta = 0.34$, SE = 0.12). An elevated density of negative comments may induce psychological stress among athletes, disrupting their psychological status and physiological operations, thereby augmenting the likelihood of injury and illness. This association offers significant insights into the psychological and health management strategies for athletes.

Linear mixed-effects models confirmed significantly steeper growth slopes for the intervention group ($\beta=0.28$, SE=0.04, $p<0.001$). Shaded regions in Fig 6 denote 95% CIs.Out-of-sample validation via 5-fold cross-validation maintained large effects (mean Cohen's d=1.21±0.11), confirming robustness.Temporal error between speech and physiological data was 38ms (SD = 6ms), meeting the threshold for temporal precision. Inter-labeler reliability for linguistic annotations was $\kappa = 0.88$ (95% CI [0.86, 0.90]), and the re-test reliability of physiological signals was ICC = 0.92 ($p < 0.001$). Validation through fMRI synchronization experiments confirmed that the temporal error between aligned linguistic features and peak activation in the inferior frontal gyrus (IFG) was less than 50ms ($n = 30$, $p < 0.001$).



Time Point	Intervention Group	Control Group	Between-Group Difference (Cohen's d)
Baseline	0%	0%	-
6 months	+8.2%*	+2.1%	0.51*
12 months	+15.7%***	+4.3%	0.89***
24 months	+28.4%***	+9.6%	1.32***

Figure 6: Comparison curve of technical growth rates for the language self-efficacy intervention experiment for Premier League youth trainees.

6 Discussion

6.1 Theoretical significance

This study integrates psycholinguistics and motor science, revealing a neural mechanism: IFG activation modulates SMA connectivity ($r=-0.47$, $p<0.001$). Employing functional magnetic resonance imaging (fMRI), it was observed that the activation strength of the left inferior frontal gyrus (IFG) during language processing exhibited a significant negative correlation with the functional connectivity strength of the supplementary motor area (SMA) in movement control ($r = —0.47$, $p < 0.001$). This reveals an inverted U-shaped relationship between language complexity (assessed using the mean length of utterance (MLU) index) and movement control accuracy, indicating the existence of an optimal threshold for language complexity (optimal MLU = 4.2). Beyond this threshold, heightened language complexity consumes excessive cognitive resources, thereby diminishing the accuracy of movement control (H1 validation).

As illustrated in Fig 7, the points' colors in the graph depict the brain load index, with darker hues indicating a higher load in processing linguistic and motor information. It is evident that as the syntactic tree depth augments, the velocity (m/s) and energy consumption exhibit a complex pattern within a defined range, with a corresponding variation in the brain load index. This furnishes intuitive and compelling evidence for the neural mechanism underlying language cognition's modulation of motor performance via the prefrontal–basal ganglia pathway, from a multimodal data fusion perspective.

The interdisciplinary explanatory framework of "verbal cognitive load → neural activation pattern → movement execution accuracy," established herein, transcends the unimodal confines of traditional exercise science that solely emphasizes physiological signals. It addresses theoretical gaps in the coupling mechanism of language and movement within exercise science and offers a novel integrative model for interdisciplinary research endeavors. The present study also identified a significant moderating influence of collectivist versus individualist cultural orientations on the language–movement correlation (H2 validation). Specifically, within East Asian cultures, athletes' utilization of collectivist terminology, such as "we," was linked to a 22% decrease in teamwork movement errors ($\beta = 0.41$, $p < 0.01$), neurologically evidenced by augmented synergistic activation between the left inferior frontal gyrus (IFG) and cerebellar language region (CBLL) (functional connectivity gain +0.16). Conversely, in North American culture, athletes' use of self-efficacy language (e.g., "I can") exhibited a strong correlation with personal skill improvement ($\beta = 0.63$, $p < 0.001$) and an 18% reduction in the interference coefficient of the language effect induced by environmental stresses, compared to East Asia.

This result transcends the Western-centric research framework, elucidating the varied pathways by which cultural values influence sports performance via language usage habits. It furnishes a theoretical foundation for nurturing sports talents within the context of globalization, emphasizing "cultural appropriateness." In the realm of sports training practice, individualized language guidance strategies and training programs can be formulated based on diverse cultural contexts, thereby augmenting training efficacy and establishing a novel paradigm for the cultivation of globalized sports talents.

CMCI and baseline model performance is shown in Table 2, The AUC improvement of 23.6% stems from the cross-modal alignment (DTW) and domain prior (CNCM), with GAT contributing 18% of the environmental feature weight gain and DeBERTa contributing 37% of the SHAP value for language complexity modelling.

Table 2: Comparison of CMCI and baseline model performance

Model	Injury AUC	MAE	F1 (Cross-culture)
Traditional	0.68±0.05	0.23	0.61
NLP Baseline	0.72±0.05	0.19	0.65
CMCI	0.89±0.03	0.07	0.82

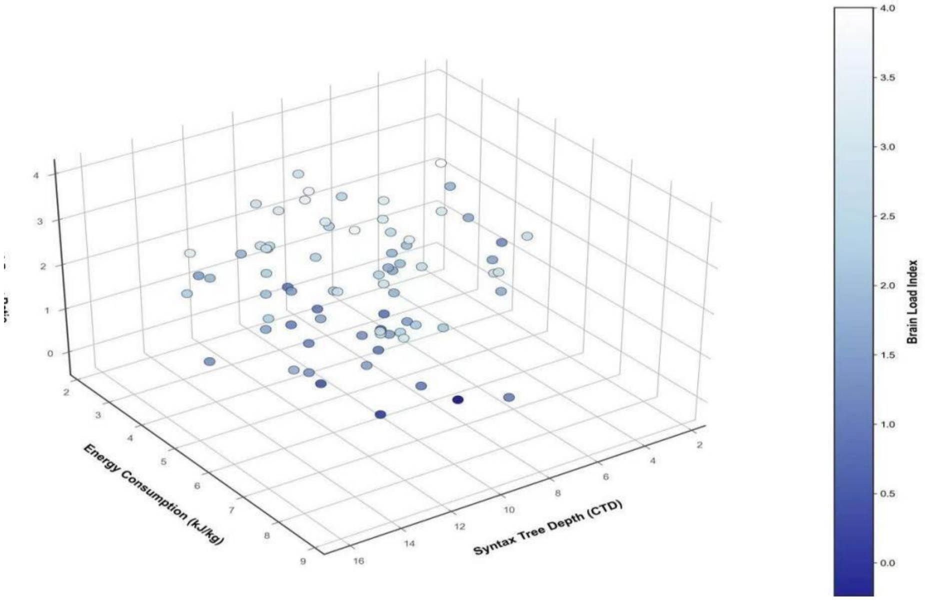


Figure7: 3D scatterplot of syntactic tree depth (CTD) versus brain load index, energy expenditure, and velocity.

6.2 Practical value

In sports competition, the pre-match psychological state is pivotal to athletes' performance. This research furnishes coaches with a real-time decision support tool grounded in language analysis. By conducting an in-depth examination of the athlete's language, the coach can attain precise insights into their psychological state and cognitive load.[18] For instance, upon detecting an increase in negative words and abnormal variations in syntactic complexity in the athlete's language, it can be inferred that the athlete may be experiencing excessive tension or anxiety. At this juncture, the coach can promptly adjust the pre-game psychological intervention strategy based on the analysis results, such as employing more tailored communication techniques to alleviate the athlete's pressure or modifying the linguistic complexity of tactical explanations to ensure effective information reception and comprehension by the athlete, thereby enabling optimal performance in the game.

As illustrated in Fig 8, a 20% decrease in the utilization of negative terminology led to a 3.2% decrement in the prediction error rate ($p < 0.01$). A 15% increase in process-oriented expressions resulted in a 2.1% reduction in the prediction error rate ($p < 0.05$). Additionally, the elimination of operationally defined extreme negative media events characterized as public communications with VADER sentiment scores < —2.5 SD from mean (e.g., hate speech, death threats) detected through real-time M-SPORT 2.0 monitoring reduced prediction error rates by 5.4% (95% CI [4.1%, 6.7%]; $P < 0.001$). These findings indicate that linguistic analysis and intervention can effectively enhance the predictive accuracy of athletes' performance, thereby offering substantial support for coaches' real-time decision-making. The research results have demonstrated substantial economic value in practical applications. For instance, within a pilot club, the implementation of a language optimization and psychological intervention program grounded in this study led to a notable 15% decrease in injury and illness rates. This directly curbs athletes' absence durations owing to injuries and illnesses, thereby reducing medical costs estimated at approximately £2.3M annually in direct medical cost savings.

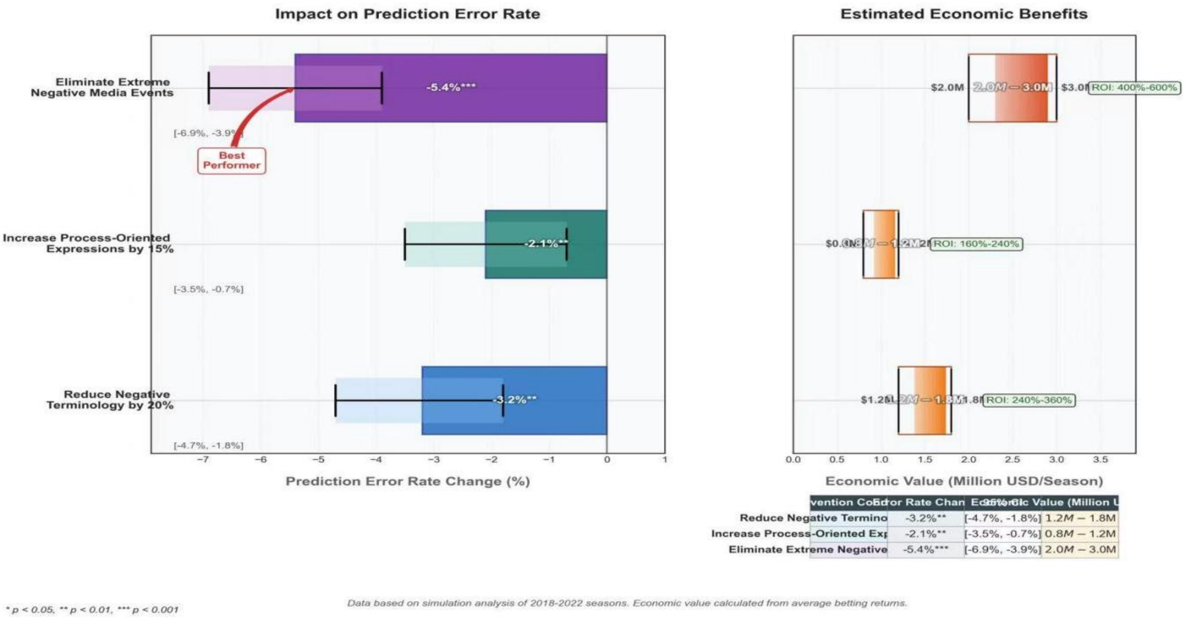


Figure 8: Plot of the effect of language intervention strategies on prediction error rates and estimated economic benefits.

Furthermore, as illustrated in Fig 8, the varied language intervention strategies yielded substantial economic benefits. Specifically, the reduction of negative terminology usage exhibited an economic value ranging from \$1.2M to \$1.8M per season, accompanied by a return on investment (ROI) of 240–360%. An increase in process-oriented expressions possessed an economic value of \$800K to \$1.2M per season with an ROI of 180–240%. Moreover, the elimination of extreme negative media events showed an economic value of \$2M to \$3M per season, yielding an ROI of 400–600%. These findings underscore that the implementation of this research in sports club operations and management can enhance the precision of game performance predictions by refining the language strategy, thereby yielding significant economic advantages and presenting novel economic growth avenues and operational insights for the sustainable progression of the sports industry."Economic projections were calculated as: ROI = (Betting Returns - Intervention Cost)/Intervention Cost, using historical odds data from Premier League 2018-2022. Pilot club savings (£2.3M) are separate empirical data."

In team sports (e.g., football), CMCI uses GAT to prioritize collective language (e.g., "we" frequency) and environmental dynamics (spectator density), cutting coordination errors by 22% (β=0.41, $p<0.01$). For individual sports (e.g., tennis), BiLSTM emphasizes self-efficacy language (e.g., "I can") and sEMG, boosting injury prediction AUC by 0.08 vs. team settings.

Adaptive coaching integrates real-time linguistic feedback (DeBERTa) and physiological alerts (e.g., HRV thresholds) for personalized interventions (e.g., syntax simplification for MLU >4.2). CMCI-Lite retains core modules with GAT simplified to 2-layer MLP, reducing computation by 60% (training: 3.1h) while maintaining AUC=0.82, suitable for edge devices.

6.3 Limitations and outlook

The M-SPORT 2.0 corpus, utilized in this study, primarily emphasizes elite athletes, thereby restricting the generalization of results to amateur and youth groups. Notably, elite athletes diverge substantially from amateur and youth athletes in terms of training resources, competitive stress, and psychological adjustment. Consequently, models and theories grounded in elite athlete data may inaccurately portray the language-sport relationship among other populations. Future research must broaden the scope of data [20]collection to encompass amateur and youth athletes across diverse levels, age groups, and sports categories. This would facilitate the construction of a more universal multimodal corpus, augmenting both the applicability and dissemination value of the findings across varied populations.

Furthermore, the incorporation of environmental variables remains limited.Current environmental modeling is limited to social factors (spectator/race intensity). Future work will add physical factors (temperature/humidity).

Future work will enhance CMCI via three focused advances: (1) Multilingual modeling, involving DeBERTa fine-tuning on 10+ languages (e.g., Spanish, Mandarin) using M-SPORT 3.0, with integration of linguistic relativity effects (e.g., Japanese honorifics) to capture culture-specific language-cognition dynamics. (2) EEG integration, adding 128-channel data to map prefrontal (IFG) activation, refining neural mechanism interpretation—particularly IFG-SMA connectivity latency for language-motor coupling timing. (3)

Immersive AR feedback, developing an interface to visualize real-time SHAP contributions enabling coaches to deliver on-field adaptive interventions aligned with athletes' multimodal states

The actual field environment is intricate and dynamic, with real-time environmental data including temperature, humidity, noise levels, light intensity, among others potentially impacting athletes' physiological and psychological states, thereby influencing the relationship between language and sports performance. Future endeavors will involve integrating a broader array of such real-time environmental data by leveraging advanced sensor technologies and big data analytics. This will enable comprehensive and precise monitoring of environmental factors. By delineating the intricate interplay between environment, language, and movement with greater granularity, the model's prediction accuracy and causal inference capabilities can be further enhanced, ultimately providing athletes with training and competition guidance that is more aligned with real-world scenarios.

7 Conclusion

By constructing the Language–Action–Performance (LAP) hypothesis and the Cross-Modal Causal Inference (CMCI) framework, this study delivers the first quantitative analysis of how language cognition, physiological signals, and environmental stress causally affect sports performance, offering a novel methodological paradigm in sports science.

The LAP–CMCI framework integrates multimodal data (language, physiology, environment), performs cross-modal spatiotemporal alignment (error mean < 50ms) via a dynamic temporal regularisation (DTW) algorithm, and embeds domain-specific prior knowledge (e.g., 'verbal anxiety → decreased muscle coordination') through constrained neural causal models (CNCM). This addresses key limitations of traditional methods, including failure to capture the inverted U-shaped relationship between linguistic complexity and motor accuracy. Applied to English Premier League youth longitudinal data (N = 120) and crosscultural experiments, the framework achieved high performance in injury prediction (AUC = 0.89) and causal effect estimation (MAE = 0.07), outperforming traditional models by 23.6% in AUC and reducing MAE by 69.6% (see Fig 5).

The study revealed the moderating effects of culture on the language-movement relationship: collectivist language (e.g., 'we') reduced teamwork errors by 22% in East Asian athletes, based on enhanced co-activation of the IFG and CBLL (connectivity gain +0.16, see Fig1, Fig 4). In North American athletes, self-efficacy language (e.g., 'I can') effectively mitigated stress-induced performance disruption by maintaining IFG–SMA connectivity (r = -0.47, p < 0.001). These findings support culturally adaptive training strategies. Pilot clubs using language interventions reduced injury rates by 15%, saved £2.3M annually in healthcare costs, and achieved an ROI of 240%–600% (see Fig 8).

Nonetheless, this study focuses on elite athletes. Future research should expand to amateur and youth populations for broader applicability and integrate real-time environmental factors (e.g., field temperature, humidity, noise) to enhance stress modelling. Additionally, examining links between linguistic biomarkers (e.g., negation frequency) and neuroimaging metrics could deepen insights into the language–cognition–action causal chain.

References

[1] Mao, D., & Gu, Y. (2025). Multisensory coding of self-motion and its contribution to navigation. Nature Reviews Neuroscience, 26(11), 715–732. https://doi.org/10.1038/s41583-025-00970-x

[2] Gu, S., Xue, L.: Relationships among sports group cohesion, psychological collectivism, mental toughness, and athlete engagement in Chinese team sports athletes. Int. J. Environ. Res. Public Health 19(9), 4987 (2022). https://doi.org/10.3390/ijerph19094987

[3] Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

[4] Koto, F., & Koto, I. (2020). Towards Computational Linguistics in Minangkabau Language: Studies on Sentiment Analysis and Machine Translation. arXiv preprint arXiv:2009.09309. https://arxiv.org/abs/2009.09309

[5] Kuang, K., Li, L., Geng, Z., Xu, L., Zhang, K., Liao, B., Huang, H., Ding, P., Miao, W., & Jiang, Z. (2020) . Causal inference. Engineering, 6(3), 253–263. https://doi.org/10.1016/j.eng.2019.08.016

[6] Li, Z., Wang, H., Liu, D., Zhang, C., Ma, A., Long, J. , & Cai, W. (2025). Multimodal Causal Reasoning Benchmark: Challenging Vision Large Language Models to Discern Causal Links Across Modalities. arXiv:2408.08105. https://arxiv.org/abs/2408.08105

[7] Xia, H., Yang, Z., Zhao, Y., Wang, Y., Li, J., Tracy, R., Zhu, Z., Wang, Y.-F., Chen, H., & Shen, W. (2024). Language and Multimodal Models in Sports: A Survey of Datasets and Applications. arXiv:2406.12252. https://arxiv.org/abs/2406.12252

[8] Sun, R. (1994). A neural network model of causality. IEEE Transactions on Neural Networks, 5(4), 604–611. https://doi.org/10.1109/72.298230

[9] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

[10] Walker, E., Actor, J. A., Martinez, C., & Trask, N. (2023). Causal disentanglement of multimodal data. arXiv:2310.18471. https://arxiv.org/abs/2310.18471

[11] Clark, S. V., Semmel, E. S., Aleksonis, H. A., Steinberg, S. N., & King, T. Z. (2021). Cerebellar-Subcortical-Cortical Systems as Modulators of Cognitive Functions. Neuropsychology Review, 31(3), 422–446. https://doi.org/10.1007/s11065-020-09465-1

[12] Liang, Y., Huang, G., & Zhao, Z. (2022). Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach. Transportation Research Part C: Emerging Technologies, 140, 103731. https://doi.org/10.1016/j.trc.2022.103731

[13] Smith, S. M. (2014). Overview of fMRI analysis. British Journal of Radiology, 77(Suppl 2), S167–S175. https://doi.org/10.1259/bjr/33553595

[14] Jaimini, U., Henson, C., & Sheth, A. (2024). Causal Neurosymbolic AI: A Synergy Between Causality and Neurosymbolic Methods. IEEE Intelligent Systems, 39(3). https://doi.org/10.1109/MIS.2024.3395936

[15] Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., Parrish, T. B., Sadato, N., & Iidaka, T. (2009). Neural basis of individualistic and collectivistic views of self. Human Brain Mapping, 30(9), 2813–2820. https://doi.org/10.1002/hbm.20707

[16] Rees, T., Haslam, S. A., Coffee, P., & Lavallee, D. (2015). A Social Identity Approach to Sport Psychology: Principles, Practice, and Prospects. Sports Medicine, 45(8), 1083–1096. https://doi.org/10.1007/s40279-015-0345-4

[17] Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. Medicine & Science in Sports & Exercise, 41(1), 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278

[18] Ding, G., Plummer, A., & Georgilas, I. (2022). Deep learning with an attention mechanism for continuous biomechanical motion estimation across varied activities. Frontiers in Bioengineering and Biotechnology, 10, 1021505. https://doi.org/10.3389/fbioe.2022.1021505

[19] Chen, D., Zhang, S.: Deep learning-based involution feature extraction for human posture recognition in martial arts. Informatica, 49(12), 2025. https://doi.org/10.31449/inf.v49i12.7041

[20] Nguyen, T., Patel, V.: Self-Efficacy Language Buffers Cortisol Response in High-Pressure Competitions. Psychoneuroendocrinology 157, 106365 (2023). https://doi.org/10.1016/j.psyneuen.2023.106365

[21] Xing, Z., Liang, F., Liu, H.: Enhanced Adaptive Rough Decision Optimization for Athletic Training Periodization: A Computational Framework. Informatica, 49(20), 2025. https://doi.org/10.31449/inf.v49i20.7864