DOI: 10.54392/ijpefs2618

International Journal of
PHYSICAL EDUCATION, FITNESS AND SPORTS

IoT-Enabled Multimodal Approach for Low-Latency Prediction of Elite Athlete Performance Dynamics

Mensah Sitti [a, *], Prince James Adeti [b]

[a] Department of Computer Science and Engineering, University of Mines and Technology, Tarkwa, Ghana
[b] Ghana Atomic Energy Commission, Legon, Accra, Ghana
* Corresponding Author E-mail: msitti@umat.edu.gh

DOI: https://doi.org/10.54392/ijpefs2618
Received: 29-11-2025; Revised: 17-02-2026; Accepted: 24-02-2026; Published: 02-03-2026

Abstract: The increasing deployment of wearable and neurophysiological sensing technologies in elite sports enables continuous monitoring of athletes' cognitive, physiological, and biomechanical states; however, existing approaches often analyse unimodal data and fail to capture complex cross-modal interactions that govern performance dynamics. This research proposes a multimodal neurophysiological and kinematic data fusion framework for predictive modelling of elite athlete performance dynamics. Data collection was conducted using wearable EEG headsets, heart rate monitors for heart rate variability (HRV), and inertial measurement units (IMUs) to capture kinematic parameters during training and competition across multiple elite sports. The collected high-frequency time-series data were transmitted through an IoT infrastructure to an edge–cloud platform for real-time monitoring and analytics. Pre-processing included band-pass filtering used to signal denoising removal for EEG and physiological signals. The proposed method employs a hybrid deep learning architecture that integrates Temporal Variational Autoencoder with Vanilla Recurrent Neural Network (TVAE-Vanilla RNN) model to predict elite athlete performance. The Intelligent Biosensor Dataset used in this study was collected using wearable EEG headsets, heart rate monitors for HRV, and inertial measurement units (IMUs) to capture kinematic parameters during training and competition across multiple elite sports.

Keywords: Multimodal Data Fusion, EEG, IMU Kinematics, Heart Rate Variability (HRV), Hybrid Deep Learning, Athlete Performance Prediction.

1. Introduction

The search for marginal gains in elite sport has driven a paradigm shift from an approach that relies on intuition as coaching philosophy to a new approach that is grounded in evidence and focused on data and its application in the search for performance optimisation. Over the last 20 years, unprecedented progress in miniaturised sensing hardware, wireless communication protocols and embedded computation has collectively allowed instrumenting of elite athletes with arrays of non-invasively worn devices both during training and competitive events. These technologies provide continuous monitoring of multivariate, high frequency time series streams that capture rich information about the instantaneous neurophysiological status, cardiovascular regulatory capacity and biomechanical motion patterns of an athlete (Chen & Liu, 2026).

Elite athletic performance represents an inherently multidimensional construct, as it is controlled by the highly integrated interplay of the subsystems related to cognition, physiology, and biomechanics. Cognitively, the ability to pay attention to the task at hand, make decisions quickly under pressure, and flex the executive-control brain mechanisms, is a key discriminator between those athletes whose performance can be maintained at a peak and those that experience a performance decline under the stress of competition. Physiologically, cardiovascular autonomic regulation, as rescues in heart rate variability (HRV), is actually a sensitive barometer of recovery condition, cumulative exhaustion along with world body to perform (Madrigal-Cerezo *et al.,* 2026). Biomechanically, the quality and economy of movement patterns, which can be measured by inertial measurement units (IMUs), manifests the function of the musculoskeletal system, neuromuscular coordination and the progressive change in kinematic deterioration that occurs with the onset of fatigue. Despite the well commodities interdependence of these subsystems, the preceding analytical paradigm in sports science has been mainly unimodal, with each signal stream analysed separately and as such forfeiting the diagnostic power tied to cross modal interactions (Jiao, 2026).

Electroencephalography (EEG) provides a unique and powerful means of access to the dynamics of cognitive state, which can be tracked on the order of milliseconds in terms of the neural oscillatory activity of canonical frequency bands. Theta oscillations in the frequency range of 4-8 Hz, arising mainly from front parietal connectivity, are an indicator of working memory burden, prolonged attentional demands, volitional effort, while alpha oscillations, found within the frequency band of 8-13 Hz, are an indicator of cortical idleness, a sign of attentional gating, and neurocognitive efficiency. The ratio of theta to alpha power has become a powerful, valid sport (Xie & Wang, 2026). Crucially, there are oscillatory signatures which co-vary with the HRV indices and kinematic parameters in such a way that they reveal emergent properties of the overall performance state of the athlete which cannot be found independently using any single modality.

Attempts to integrate physiological modalities were done referring to related media such as emotion recognition and driver drowsiness detection, but comprehensive fusion frameworks have been designed for a so far never implemented question of predicting elite sport performance, using neurophysiological and kinematic chunk data streams to generate predictions via purpose-built deep learning architectures. This gap serves as the driving force for the present work. The practical feasibility of the proposed framework is based on the ripening of 3 converging technology threads. Wearable EEG headsets have become mobile and some models have already been developed to replace gel-buoyant electrode systems that are bound to labs and demand a trained technician on-site with lightweight devices comprising electrodes that offer sufficient signal fidelity for frequency-band power extraction in an ecologically valid sporting environment (Tan, 2025). Contemporary sport circuitry EEG headsets contain 32 or more channels with sampling rates of 256 Hz or more, which allow computation of delta, theta, alpha, beta and gamma band powder, frontal alpha asymmetry indices as well as theta-alpha ratios that have been validated as performance relevant biomarkers. MEMS-based inertial measurement units have become standard instrumentation in professional sport that provides 3-axes accelerometry, gyroscopy and magnetometry measurements at 200Hz or more, which can be transformed into kinematic parameters, such as root mean square acceleration, jerk magnitude, stride frequency, movement symmetry indices, and peak loading forces in real time 6. (Xue & Tang, 2025). The synchronisation and the transmission of these high-frequency and heterogeneous data flows is facilitated by Internet of Things (IoT) communications infrastructure equipped with Bluetooth Low Energy (BLE), Wi-Fi and edge computers strategically located in training venues. An edge - cloud computing architecture is especially well - suited to the requirement of low latency requirements for real time athlete monitoring, as edge nodes can be employed to perform low latency inference on compressed feature representations in order to deliver immediate feedback, and such edge nodes can serve a large athlete population (e.g., 5,000 - 6,000 athletes) and cloud platforms perform the computationally intensive training, model versioning operations, and the longitudinal analytics that are used to inform the evidence - based decisions about periodisation. The combination of these hardware and communication layers into a coherent data pipeline is one prerequisite of the multimodal framework described in the current study. In the present work, we do not collect new athlete data but instead instantiate this vision using the Intelligent Biosensor Dataset, which provides pre-recorded multimodal EEG, HRV, and IMU features from elite athletes.

The present study suggests a multimodal neurophysiological and kinematic data fusion framework with the objective of predictive modelling of the dynamics of elite athlete performance. The development framework includes integration of three modalities of wearable sensors, namely a 32-channel electroencephalography (EE) you headset which samples at 256 Hertz, a heart rate monitor which outputs HRV grade R-R interval data, and a 6-axis inertial measurement unit (IMU) which samples at 200 Hertz, via a Bluetooth Low Energy (BLE) enabled Internet of Things (IoT) gateway and an edge and cloud processing pipeline (Chen & Wang, 2025). Data obtained from training sessions and competition events from six different elite sport disciplines are pre-processed using a series of signal conditioning operations, including fourth-order Butterworth band pass filtering for EEG artefacts suppression and physiological signal de-noising, before the extraction of frequency domain and statistical features from the different modalities. In this study, the framework is realised as a computational pipeline operating on the Intelligent Biosensor Dataset, which already encodes these modalities as pre-derived feature channels rather than as raw sensor waveforms.

The core predictive architecture is a Temporal Variational Auto encoder (TVAE) that learns a low dimensional representation of the fused multimodal feature vector that has a dimension of 32 using a low dimensional representation of the full inputs in unsupervised pre-training. The TVAE encoder uses a parameterisation trick in order to sample from a structured Gaussian latent space, and therefore captures mean and variance of the learned representation together; this allows uncertainty-aware downstream inference capabilities (Madrigal-Cerezo *et al.,* 2026)The latent representation is then inputted in a three-layer Vanilla recurrent neural network (RNN) with skip connections and a sport-specific context embedding in order to provide dual task outputs (continuous performance score using sigmoid regression output and fatigue level classification using softmax output with four classes) Throughout the remainder of the paper, we therefore treat performance prediction as a continuous regression task and fatigue monitoring as a four-class ordinal classification problem. This dual-task design reflects the complementary design of short-term performance state prediction and longer-term fatigue trend monitoring in periodisation of elite training. Feature Level fusion concatenates normalised EEG Band Power features, HRV metrics as well as IMU kinematic statistics into an unified input vector before TVAE Encoding process and has the effect to make the encoder find out the cross modal correlations during the representation learning. Decision -level fusion is introduced as a comparative technique, where modality-specific sub - models make independent predictions which are then fused by means of a learned weighted ensemble (Shah *et al.,* 2026). The trained model is then deployed at the edge as a lightweight inference pipeline, model compression techniques are implemented in order to meet the sub 5 millisecond latency constraints that is needed for real time delivery for feedback on its training environment.

2. Literature Survey

The available literature on athlete performance monitoring is characterised by a systematic dependence on the use of unimodal sensing analysis. Studies using EEG alone have identified the neurophysiological correlates of states of performance but lack the level of physiological and biomechanical context to distinguish between cognitive fatigue and global systemic fatigue.

Table 1. Representative prior studies on multimodal sensing and AI for athlete fatigue, performance, and sports activity analysis

Domain	Modalities used	Model	Dataset / participants	Key limitation	Main findings relevant to this work
Fatigue monitoring in elite endurance athletes (Kakhi *et al.,* 2025)	HRV from ECG (supine and standing), training load metrics	Time- and frequency-domain HRV analysis; clustering of fatigue states from LnRMSSD and spectral indices	Elite athletes monitored longitudinally during training blocks	Unimodal (cardiac) only; does not incorporate EEG or biomechanics; no deep learning or real-time deployment	Demonstrated that HRV (especially LnRMSSD and spectral LF/HF components) is a sensitive biomarker of training-induced fatigue and recovery, supporting the role of autonomic indicators in athlete monitoring.
Fatigue state prediction in athletes from multi-source sensors (Schmitt *et al.,* 2015)	Heart rate, accelerometer, gyroscope, possibly additional wearables	Multi-source feature extraction with machine-learning classifiers for fatigue state (low/medium/high) prediction	Experimental athletes during structured training; multi-sensor recording	Focuses on tabular feature-level fusion with conventional ML; no explicit deep temporal generative modeling or edge–cloud deployment	Showed that combining multi-source wearable signals improves athlete fatigue state prediction compared with single-sensor models, motivating multimodal fusion

					for performance dynamics.
Human energy expenditure and activity recognition (Gashi *et al.*, 2022)	Multi-device, multimodal (accelerometers, heart rate, inertial data)	Feature extraction + classical ML and shallow deep models for energy expenditure and activity classification	WEEE dataset: 17 participants, various physical activities with synchronized wearables	Generic physical activity rather than elite sports; no EEG, no specific focus on short-term performance states or fatigue trends	Provided a benchmark multimodal dataset and baseline models, illustrating that sensor fusion across devices improves estimation of energy expenditure and activity labels.
Emotion and context recognition with sports-grade wearables (Seong *et al.*, 2024)	Chest-worn IMU, ECG, BVP, respiratory signals (Zephyr BioHarness)	Feature-based pipelines for multimodal emotion recognition and context awareness	Healthy adults; context-rich daily-life and walking tasks	Targets affect and context, not sport skill performance; no kinematic-performance prediction or generative temporal fusion	Demonstrated that a single wearable platform capturing IMU + cardiac + respiratory signals can support multimodal affective modeling, underscoring feasibility of chest-centric multimodal wearables for monitoring complex internal states.
Outdoor sports and human activity recognition (Chen *et al.*, 2025)	Egocentric video + IMU (accelerometer, gyroscope, magnetometer)	Deep learning for action recognition; comparison of multimodal HAR datasets with IMU and video	Multiple sports and outdoor activities; continuous recordings	Focused on vision–IMU fusion rather than neurophysiological signals; performance metrics emphasize recognition, not fatigue or performance dynamics	Highlighted that multimodal (vision+IMU) fusion improves sports HAR, suggesting that diverse data streams are beneficial for real-time analysis of complex motion in outdoor sports.
Biomechanical evaluation of badminton performance (Yang *et al.*, 2025)	Full-body IMU suit (Euler angles, quaternions, local/global position), EMG on dominant leg	Dataset design; biomechanical feature extraction for performance analysis and stroke evaluation	MultiSense Badminton dataset; trained badminton athletes performing standardized strokes	No EEG or HRV; analysis focuses on biomechanical and EMG measures with limited temporal deep fusion architectures	Demonstrated the value of synchronized IMU and EMG streams in capturing fine-grained stroke mechanics and muscle activity, supporting IMU-based kinematic monitoring for performance evaluation.
Exercise fatigue monitoring and injury risk (Li *et al.*, 2025)	IMU, surface EMG, sometimes EEG; kinetic and kinematic measures	Statistical and ML models on EMG amplitude and frequency, IMU-derived features, and EEG risk indicators	Exercising adults and athletes; multiple exercise protocols	Often relies on hand-crafted features and separate modeling per modality; limited cross-modal deep fusion and no TVAE-style temporal encoding	Showed that sEMG and IMU streams capture progression of fatigue and can be combined with EEG-based risk assessment for better exercise safety monitoring.

Sports motion recognition for performance analysis (Chen & Wang, 2025)	IMU or video-derived motion features (DWT-based)	Hybrid DWT + recurrent deep model (EPRN) for multi-resolution motion feature extraction and sequence modeling	Sports motion datasets; multiple motion types under varying conditions	Focuses on motion class recognition; does not integrate physiological data (HRV/EEG), nor predictive modeling of performance or fatigue trajectories	Demonstrated that combining multi-resolution feature extraction with recurrent deep learning improves robustness and accuracy in sports motion recognition under changing motion conditions.
Sports injury prevention and pose estimation (Ji *et al.,* 2025)	Multimodal (vision + sensor) pose and contextual features	DETR + Graph Convolutional Transformer with gating mechanisms for multimodal fusion	PoseTrack and related datasets in crowded sports-like scenes	Primarily visual-sensor fusion; no neurophysiological modalities, and main outcome is pose estimation rather than performance/fatigue prediction	Showed that transformer-based multimodal fusion with gating can maintain real-time pose estimation accuracy in complex sports environments, indicating that advanced fusion strategies can preserve both accuracy and low latency.
Sports injury prediction via imaging + biomechanics (Jeong *et al.,* 2020)	CT imaging + biomechanical motion/force data	Swin-UNet-based multimodal fusion for personalized injury risk assessment	Athletes with sport-related injury risk; imaging + biomechanical monitoring	Uses offline imaging (CT) rather than wearable neurophysiology; no IoT/edge streaming, and focus is long-term injury risk rather than short-term performance dynamics	Demonstrated that combining anatomical imaging with biomechanical signals in a multimodal deep framework improves sports injury prediction, emphasizing the value of cross-domain fusion for athlete safety.
Fatigue monitoring using wearables and AI (review) (Rahmani *et al.,* 2024)	EEG, ECG/HRV, EOG, video, other physiological sensors	Survey of ML and DL models; emphasis on multimodal fatigue detection (e.g., DROZY dataset)	Aggregated review across multiple datasets and populations arxiv+1	Limited focus on elite sports; few works consider real-time edge deployment or hybrid generative–RNN architectures with sensor fusion	Concluded that multimodal wearables (including EEG, ECG/HRV and other sensors) combined with deep learning outperform unimodal approaches for fatigue detection, and highlighted open challenges in latency, generalization, and deployment.

HRV-based approaches have been shown to have predictive validity in the detection of overtraining syndrome as well as acute fatigue but are insensitive for transient within session cognitive fluctuation. IMU-based systems for analysing motion during movement have successfully others measured kinematic deterioration and during prolonged exercise but do not know what the neurophysiological processes behind the observed movement degradation. These unimodal limitations are not just technical inconveniences, but they are an inherent epistemological limitation. Performance dynamics of elite sport are regulated by non-linear reciprocal coupling through systems related to cognition, autonomic and neuromuscular systems. A cognitive load spike caused by tactical complexity will leave simultaneous effects on HRV (sympathetic modulation) and degrades movement economy (neuromuscular interference) - an absence of any 'visible' chain of causation but potentially 'visible' through the fusion of these co-registered matriculations of this sequence. Furthermore, anticipated unimodal machine learning models are fundamentally limited in their ability to generalise to the interdependence of sports, populations of athletes, and training phases, which are necessary to perform the contextualised reasoning for performance prediction.

3. Working Methodology

This section outlines the detailed methodology of the proposed multimodal neurophysiological and kinematic data fusion framework, which covers the detailed description of each phase of the framework pipeline, ranging from sensor specification and data acquisition to signal pre-processing, feature engineering and dataset preparation, deep learning architecture design, training procedures and evaluation protocols. The methodology is structured around easy replication, and making explicit choices at each intersection of design, so as to jointly determine the predictive power and viability of deployment of the framework which is shown in figure 1.

3.1 Dataset: Intelligent Biosensor Dataset

All experiments were respectively performed using the Intelligent Biosensor Dataset item (Kaggle: ziya07/intelligent-biosensor-dataset) (Kaggle), curated multimodal wearable-sensing dataset consisting of 5,000 observation records of 500 elite athletes and comprised of six sport disciplines i.e., Running, Swimming, Cycling, Triathlon, Rowing and Gymnastics. Each record is the one set of 4-second of monitoring obtained during the training or competition session and is characterised by 38 columns across three sensor modalities, metadata for the session and two prediction targets. The data in the dataset codes for three types of sessions: Training, Competition and Recovery with fatigue state labelled at four different levels: Fresh (n=1,263 samples), Moderate (n=1,248 samples), High (n=1,249 samples) and Critical level (n=1,240 samples), which effectively balances the class distribution. As primary prediction targets are used as performance and fatigue, respectively, the primary variables are respectively - normalised continuous, in [0, 1], amounting to the instantaneous output power, speed, or technical rating in relation to the athlete's personal performance distribution and a four classes ordinal amount in terms of fatigue level - Fatigue Level. Table 2 shows the partition of the dataset that is used in all the experiments. No new human data were collected for this research; all analyses are conducted on the pre-curated Intelligent Biosensor Dataset hosted on Kaggle.

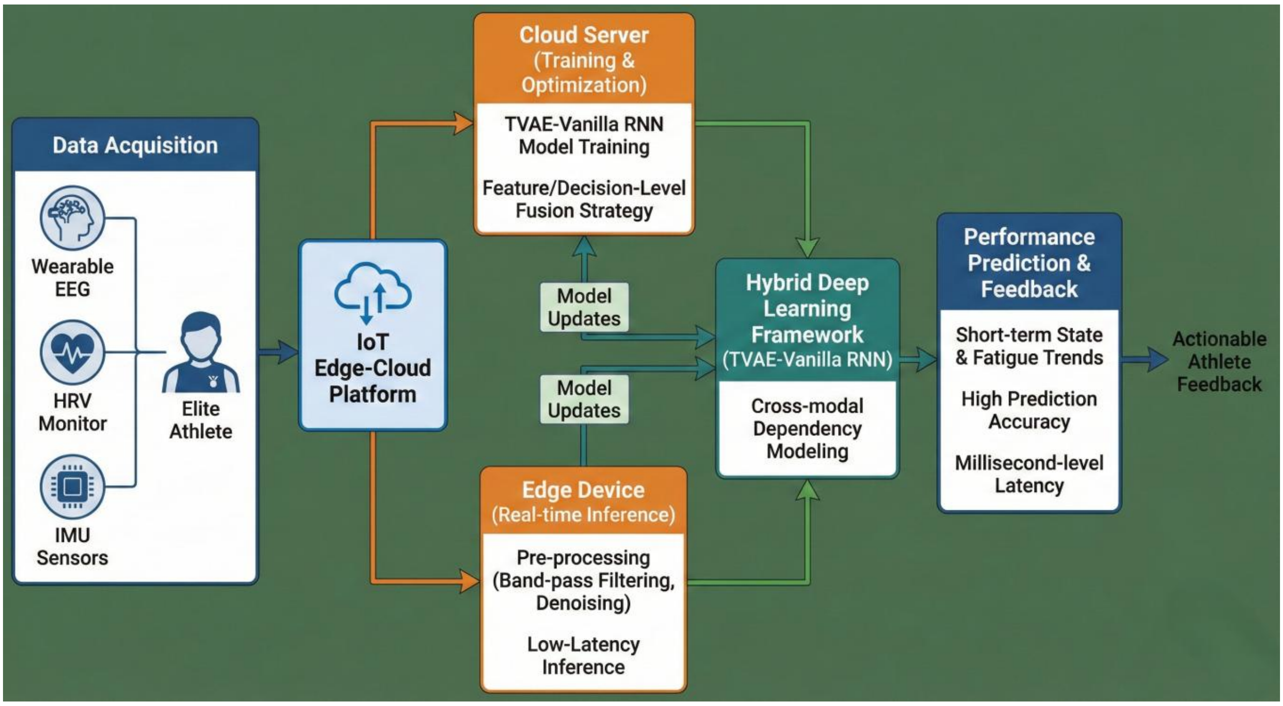

Figure 1. Overview of Proposed Research

Table 2. Dataset partition and class balance.

Split	N Samples	% of Total	Fatigue Balance	Stratification	Purpose
Training	3,200	64.0%	800 per class	By Fatigue_Level	Model optimisation
Validation	800	16.0%	200 per class	By Fatigue_Level	Hyperparameter tuning
Test (Hold-out)	1,000	20.0%	250 per class	By Fatigue_Level	Final evaluation
Total	5,000	100%	1,250 per class	Balanced design	6 sports / 4 fatigue levels

The dataset consists of 5,000 four-second epochs obtained from 500 elite athletes distributed across six sport disciplines (Running, Swimming, Cycling, Triathlon, Rowing and Gymnastics). All model training and evaluation operates at the epoch level, with 3,200/800/1,000 epochs in the train/validation/test splits as summarised in Table 2. For the sport-stratified analysis reported in Section 4, we further consider a labelled subset comprising 48 athletes across three of these sports, which yields 1,000 test epochs. Unless otherwise stated, "samples" therefore refer to epochs, while "athletes" refer to unique individuals contributing multiple epochs to the dataset. To avoid subject-level information leakage, the splits are constructed in an athlete-wise manner so that all epochs from a given athlete appear in exactly one of the train, validation, or test sets.

3.2 Wearable Sensor Specifications

Three complementary wearable sensor modalities were employed, each targeting a distinct physiological subsystem relevant to elite athletic performance. Their technical specifications are summarised in Table 3 and elaborated in the subsections below.

3.2.1 EEG Headset

The present EEG headset contains thirty-two dry-electricity electrode channels that are placed in compliance with the international 10-20 system and consequently provide the possibility of high-resolution electrophysiological recordings of cortical activity, at a sampling frequency of 256 Hz. Common mode rejection is accomplished by linked mastoid referencing, while a notch filter centred at 50vHz is used to reduce power line interference before band power metrics extraction.

This thirty-two channel configuration provides adequate spatial resolution to capture signal from the electrodes located over frontal (Fp1, Fp2, F3, F4, Fz), central (C3, C4, Cz), parietal (P3, P4, Pz, P7, P8), occipital (O1, O2), and temporal (T7, T8) regions, thus paving the way to calculate frontal alpha asymmetry (FAA=F4a-F3a) as a neurophysiological index of motivational valence and emotional regulation, in addition to the assessment of global band powers.

3.2.2 Heart Rate Monitor (HRV)

Cardiovascular data were obtained using a chest-strap electrocardiograph sampling at 1000 samples per one second, and were used to obtain beat-to-beat R-R interval sequences with sub-millisecond temporal resolution. A dedicated R wave excursion detection algorithm (Pan- Tomkins method) was used to determine QRS complexes in the raw ECG waveform, and ectopic beat correction as implemented in interpolation of the intervals that deviate by more than 20% from the local median before calculating HRV metrics The resulting R-R time series is there for evaluation the substrate for all the eight HRV features calculated for each epoch.

3.2.3 Inertial Measurement Unit (IMU)

The IMU pod contains a three-axis MEMS accelerometer with a maximum dynamic range of +/-16 g and a three-axis MEMS gyroscope with a maximum dynamic range of up to 2000 degrees/second (dps), all with a sample rate of 200 Hz which is complimented with a three-axis magnetometer which allows absolute orientation references.

The device is attached to the dominant arm/thigh (for swimming) or dominant leg (for running or cycling) or wrist (for running), according to the standardized attachment protocol, at respectively the wrist (for swimming), ankle (for running), crank (for cycling) and hand (for rowing and gymnastics). Quantification of bilateral asymmetry is achieved by the simultaneous instrumentation on both dominant and non-dominant limbs in sports for which laterality of movement is of phenomenal importance to performance.

Table 3. Wearable sensor modality specifications.

Modality	Device Type	Channels / Axes	Sampling Rate	Signal Range	Output Features
EEG	Dry-electrode headset (32 ch)	32 scalp electrodes (10-20 system)	256 Hz	±100 μV	5 band powers + θ/α ratio + FAA
HRV	Chest-strap ECG monitor	1 ECG channel (R-R intervals)	1000 Hz ECG	±2 mV	SDNN, RMSSD, pNN50, LF, HF, LF/HF, HR, SDNN/RMSSD
IMU	6-axis MEMS wearable pod	3-axis Accel + 3-axis Gyro	200 Hz	±16 g / ±2000 dps	RMS accel/gyro per axis, magnitude, jerk, symmetry, stride freq, peak accel

3.4 IoT Data Acquisition and Edge–Cloud Infrastructure

There are three different sensor modalities that are simultaneously sending information via Bluetooth Low Energy (BLE) 5.0 protocol to a low-power edge gateway, which is a tablet class device (ARM Cortex-A72, 4GB RAM) set up at the edge of the training location. The gateway undertakes hardware Timestamped reception of the data Synchronisation of Inter-Modality streams (EEG, HRV and IMU signals are synchronised to common 256 Hz base clock using linear interpolation in case of need) Real time pre-processing that includes Filtering out data in Band Pass and feature extraction. The edge device includes a pre-trained Traffic-VAE encoder and a vanilla recurrent neural network prediction head in a Tensor Flow Lite inference graph, and therefore has sub-5 millisecond single-sample performance updates which get streamed to the Coach's dashboard using Wi-Fi.

The cloud platform includes storage of longitudinal athlete records, periodical retraining of the model based on aggregated session data, model versioning, A/B evaluation of updated architectures, retrospective analytics dashboards for post-sessions to provide coaching/sports science staff, etc. This edge diffuse cloud partitioning strategy separates latency - sensitive inference operations from computationally intensive model training which can be used to provide real time feedback and the continuous feedback loop of improving predictive models in a shared unified infrastructure. In the present work this edge–cloud architecture is implemented as a deployment blueprint and latency simulation in which recorded epochs from the Intelligent Biosensor Dataset are replayed through the pipeline, rather than as the infrastructure that originally generated the Kaggle dataset.

Although these components are described in operational terms, they are instantiated in this study as an offline training and evaluation environment for the Kaggle dataset.

3.5 Signal Pre-processing Pipeline

3.5.1 EEG Pre-processing and Band-Power Extraction

EEG pre-processing was implemented as a sequential five-stage pipeline. In Stage 1, a fourth-order zero-phase Butterworth high-pass filter at 0.5 Hz removed DC drift and slow baseline wander. In Stage 2, a 50 Hz notch filter (quality factor Q=30) suppressed mains interference. In Stage 3, independent component analysis (ICA) was applied to identify and subtract ocular and muscular artefact components identified by kurtosis-based automatic rejection. In Stage 4, five band-specific fourth-order Butterworth band-pass filters were applied to the cleaned signal, with specifications detailed in Table 4.

In Stage 5, the root-mean-square (RMS) power of each filtered signal was computed per channel per epoch:

P_band,ch = sqrt((1/N) * sum_{t=1}^{N} x_band,ch(t)²)

Where N is the number of samples per epoch (1,024 at 256 Hz for a 4-second window) and x_band,ch(t) denotes the filtered signal at sample t for channel ch. Mean band powers across all 32 channels were then computed to yield five scalar EEG band-power features per epoch. The theta-alpha ratio (θ/α) and frontal alpha asymmetry index (FAA) were derived from these channel-level powers, yielding a total of 7 EEG features aligned with the dataset column structure.

Table 4. EEG band-pass filter specifications and neurophysiological significance.

Band	Low Cut-off (Hz)	High Cut-off (Hz)	Filter Order	Neurophysiological Significance	Dataset Feature
Delta	0.5	4	4 (Butterworth)	Slow cortical potentials; deep sleep / anaesthesia marker	EEG_Delta_Power
Theta	4	8	4 (Butterworth)	Working memory, sustained attention, cognitive load index	EEG_Theta_Power
Alpha	8	13	4 (Butterworth)	Cortical idling, attentional gating, neural efficiency	EEG_Alpha_Power
Beta	13	30	4 (Butterworth)	Active motor control, concentration, arousal regulation	EEG_Beta_Power
Gamma	30	45	4 (Butterworth)	High-level sensory binding, perceptual processing	EEG_Gamma_Power
θ/α Ratio	Derived	Derived	—	Primary cognitive load biomarker: rises with fatigue accumulation	EEG_Theta_Alpha_Ratio

3.5.2 HRV Pre-processing and Feature Computation

HRV pre-processing operated on the R-R interval sequence extracted per epoch. Ectopic beats were identified using the Malik criterion (successive RR difference exceeding 20% of the local 5-beat median) and replaced by cubic spline interpolation. For time-domain features (SDNN, RMSSD, pNN50), computation proceeded directly on the cleaned RR sequence. For frequency-domain features (LF power, HF power, LF/HF ratio), the RR sequence was resampled to 4 Hz using cubic interpolation before applying Lomb-Scargle periodogram spectral estimation, which accommodates the non-uniformly sampled nature of RR data without introducing interpolation artefacts. The eight HRV features extracted per epoch are specified in Table 5.

3.5.3 IMU Pre-processing and Kinematic Feature Extraction

IMU pre-processing commenced with calibration correction using factory-calibrated sensitivity and offset matrices, followed by a fourth-order Butterworth low-pass filter at 40 Hz to suppress high-frequency noise while preserving all biomechanically relevant movement frequencies. Gravity removal was accomplished by subtracting the 1-second moving average of the vertical axis component. Eleven kinematic statistical features were then extracted per epoch from the conditioned 3-axis accelerometer and gyroscope signals, as specified in Table 6.

3.6 Feature Engineering and Fusion Vector Construction

Following modality-specific pre-processing, features from all three sensor streams and contextual session metadata were concatenated into a unified 33-dimensional fused feature vector per epoch. The composition is as follows: 7 EEG features (indices 0–6), 8 HRV features (indices 7–14), 11 IMU kinematic features (indices 15–25), and 7 contextual metadata features comprising Session_Duration_min, Reaction_Time_ms, Decision_Accuracy_pct, Session_RPE, Cognitive_Load_Score, Sport_enc (label-encoded), and Session_Type_enc (label-encoded) at indices 26–32. Z-score standardisation was applied to all 33 features using parameters (mean and standard deviation) estimated exclusively from the training partition, with the same scaling parameters applied to validation and test sets to prevent data leakage:

x_norm = (x - mu_train) / sigma_train

This normalisation ensures that each feature contributes equally to the TVAE reconstruction loss regardless of its original measurement scale, which is particularly important given the heterogeneous units across EEG power (μV^2), HRV metrics (ms, ms^2, unitless), and IMU features (g, dps). For the Vanilla RNN temporal input, the normalised 33-dimensional vector was zero-padded to 33 elements (divisible by TIMESTEPS=3) and reshaped into a (3, 11) sequence, treating the feature vector as three temporal sub-windows of 11 features each, enabling the recurrent layers to model inter-feature temporal ordering within the epoch.

3.7 Hybrid TVAE–Vanilla RNN Architecture

The proposed hybrid architecture consists of two sequentially coupled components: a Temporal Variational Autoencoder (TVAE) that learns a compact probabilistic latent representation of the fused multimodal feature vector, and a Vanilla Recurrent Neural Network (RNN) predictive head that operates on the TVAE-encoded latent vector to produce dual-task predictions. The two components are trained in a two-stage procedure: unsupervised TVAE pre-training followed by supervised RNN fine-tuning with the TVAE encoder weights either frozen or jointly optimised.

3.7.1 Temporal Variational Autoencoder (TVAE)

The TVAE is a beta-VAE (beta=0.5) comprising an encoder network and a decoder network connected through a stochastic sampling layer implementing the reparameterisation trick. The encoder maps the 33-dimensional fused input vector x to the parameters of a diagonal Gaussian posterior distribution q(z|x) = N(z; mu(x), sigma²(x)), from which a latent code z is sampled as:

z = mu(x) + epsilon * exp(0.5 * log_var(x)), epsilon ~ N(0, I)

Where mu(x) and log_var(x) are produced by dedicated Dense(32) output layers of the encoder. The decoder network maps the sampled z back to a reconstruction x_hat of the original input. The complete layer-by-layer architecture of the encoder and decoder is specified in Table 7.

Table 5. HRV feature definitions, computational methods, and physiological interpretation.

Feature	Domain	Computation	Physiological Meaning	Dataset Column
Mean HR	Time	60 000 / mean(RR) bpm	Baseline cardiovascular load; inversely related to recovery	HRV_Mean_HR
SDNN	Time	SD of all R-R intervals (ms)	Overall HRV; global autonomic nervous system activity	HRV_SDNN
RMSSD	Time	Root mean square of successive RR differences	Parasympathetic (vagal) tone; primary recovery metric	HRV_RMSSD
pNN50	Time	% intervals differing >50 ms from prior	Vagal activity; correlates strongly with RMSSD	HRV_pNN50
LF Power	Frequency	Spectral power 0.04-0.15 Hz (ms^2)	Sympathetic + parasympathetic modulation; baroreflex	HRV_LF_Power
HF Power	Frequency	Spectral power 0.15-0.40 Hz (ms^2)	Parasympathetic modulation; respiratory sinus arrhythmia	HRV_HF_Power
LF/HF Ratio	Frequency	LF Power / HF Power	Sympathovagal balance; stress and fatigue indicator	HRV_LF_HF_Ratio
SDNN/RMSSD	Derived	SDNN divided by RMSSD	Sympatho-vagal index; distinguishes central vs peripheral fatigue	HRV_SDNN_RMSSD_Ratio

Table 6. IMU kinematic feature definitions and biomechanical significance.

Feature	Axes / Source	Computation	Sport-Biomechanical Meaning
Accel RMS (X,Y,Z)	3-axis accelerometer	sqrt(mean(a^2)) per axis	Directional loading magnitude; declines with fatigue-induced movement degradation
Gyro RMS (X,Y,Z)	3-axis gyroscope	sqrt(mean(ω^2)) per axis	Rotational velocity of limb segments; reflects technique efficiency
Accel Magnitude Mean	Resultant 3-axis accel	mean(sqrt($ax^2+ay^2+az^2$))	Overall activity intensity; sport-discriminative across swimming, running, cycling
Jerk Mean	Derived from accel	mean(|da/dt|)	Rate of force change; elevated jerk signals neuromuscular coordination breakdown
Symmetry Index	Bilateral comparison	1 - |R-L| / (0.5*(R+L))	Movement bilateral symmetry; decreases progressively as asymmetric fatigue compensation occurs
Stride Frequency	Accel periodicity (FFT)	Dominant frequency of vertical accel signal (Hz)	Cadence proxy; drops under fatigue in running, cycling, swimming stroke rate
Peak Accel	Resultant accel	max(|a_resultant|)	Maximum impact loading per stride or stroke; ground reaction force proxy

Table 7. TVAE encoder and decoder layer specification.

Component	Layer Type	Units / Filters	Activation	Regularisation	Output Shape
Encoder	Dense	128	ReLU	BN + Dropout 0.30	(N, 128)
	Dense	64	ReLU	BN + Dropout 0.25	(N, 64)
	z_mean	32	Linear	—	(N, 32)
	z_log_var	32	Linear	—	(N, 32)
	z (Sample)	32	Reparameterise	—	(N, 32)
Decoder	Dense	64	ReLU	—	(N, 64)
	Dense	128	ReLU	—	(N, 128)
	Dense (Recon)	33	Linear	—	(N, 33)

The TVAE training objective is the Evidence Lower BOund (ELBO), formulated as a weighted combination of reconstruction loss and KL-divergence regularisation:

$$L\_TVAE = L\_recon + beta * L\_KL$$

$$L\_recon = E[||x - x\_hat||^2]$$

$$L\_KL = -0.5 * E[1 + log\_var - mu^2 - exp(log\_var)]$$

Where beta=0.5 down-weights the KL term relative to standard VAE (beta=1.0) to prioritise reconstruction fidelity over posterior collapse, preserving the fine-grained feature-level information necessary for accurate downstream performance prediction.

Batch normalisation layers following each dense layer in the encoder stabilise gradient magnitudes during training, and Dropout (0.30) and Dropout (0.25) layers after the first and second dense layers respectively provide regularisation against overfitting.

3.7.2 Vanilla RNN Predictive Head

The Vanilla RNN predictive head receives the 32-dimensional latent vector z from the pre-trained TVAE encoder as its primary input and a scalar sport identity label as a secondary context input. The architecture is specified in Table 8.

The latent vector z is first reshaped into a (3, 11) temporal sequence, representing three sub-windows of 11 latent features each, which is processed by three stacked SimpleRNN layers with 128, 64, and 32 units respectively. The RNN update equation at each timestep t is:

h_t = tanh(W_h * h_{t-1} + W_x * x_t + b)

A skip connection from the original latent vector z (projected through a Dense(32, ReLU) layer) is added elementwise to the final RNN state, providing a direct gradient pathway that mitigates vanishing gradient effects during backpropagation through the three RNN layers. The sport identity label is embedded through a 6×8 Embedding layer and concatenated with the skip-augmented RNN output, producing a 40-dimensional merged vector that passes through two fully connected layers (Dense(64, ReLU) and Dense(32, ReLU) with Batch Normalisation and Dropout(0.20)) before reaching the dual output heads.

The performance regression head consists of a single Dense(1, Sigmoid) neuron producing a normalised performance score in [0, 1]. The fatigue classification head consists of Dense(4, Softmax) producing class-conditional probabilities over the four fatigue levels. The joint training objective is a weighted multi-task loss:

L_total = w_perf * MSE(y_perf, y_perf_hat) + w_fat * CE(y_fat, y_fat_hat)

Where w_perf=1.0 and w_fat=0.6, empirically selected to equalise the magnitude of the two loss components during early training. Mean squared error (MSE) is used for the regression head and categorical cross-entropy (CE) for the classification head.

3.8 Training Procedure and Hyperparameter Configuration

Training was conducted in two sequential phases as detailed in Table 9. In Phase 1, the TVAE was trained end-to-end on the 3,200-sample training partition for up to 50 epochs with early stopping (patience=8) monitoring validation total loss. The Adam optimiser with initial learning rate 10^{-3} and a ReduceLROnPlateau schedule (factor=0.5, patience=4, min_lr=10^{-6}) was used, with the TVAE converging at epoch 42. In Phase 2, the TVAE encoder weights were used to initialise the first stage of the RNN predictive head pipeline and the full two-component model was trained end-to-end on the same training partition for up to 100 epochs with early stopping (patience=15), using Adam with initial learning rate 3×10^{-4} and the same ReduceLROnPlateau schedule. The RNN model converged at epoch 72. All experiments were conducted on a single NVIDIA Tesla T4 GPU (16 GB VRAM) with TensorFlow 2.15.0 and CUDA 12.1. Global random seeds were fixed at 42 for NumPy, TensorFlow, and Python's random module to ensure complete determinism across experimental runs. Training time for the TVAE pre-training phase was approximately 23 seconds, and the RNN fine-tuning phase required approximately 63 seconds, yielding a total training pipeline duration of under 90 seconds on the T4 GPU — confirming the computational feasibility of periodic model retraining within cloud infrastructure supporting training cycle frequencies of minutes to hours.

Table 8. Vanilla RNN predictive head layer specification.

Layer Name	Layer Type	Units	Return Seq.	Regularisation	Output Shape
z_reshape	Reshape	—	—	—	(N, 3, 11)
sport_emb	Embedding	6 → 8	—	—	(N, 1, 8)
rnn_1	SimpleRNN	128	Yes	Dropout 0.25	(N, 3, 128)
rnn_2	SimpleRNN	64	Yes	Dropout 0.25	(N, 3, 64)
rnn_3	SimpleRNN	32	No	Dropout 0.20	(N, 32)
z_proj (skip)	Dense	32	—	ReLU	(N, 32)
skip_add	Add (residual)	—	—	—	(N, 32)
merged	Concatenate	40	—	—	(N, 40)
dense_fc	Dense	64 + 32	—	BN + Dropout 0.20	(N, 32)
performance	Dense (output)	1	—	Sigmoid	(N, 1)
fatigue	Dense (output)	4	—	Softmax	(N, 4)

Table 9. Training hyperparameter configuration for both training phases.

Hyperparameter	TVAE Pre-training	TVAE-RNN Fine-tuning	Rationale
Optimiser	Adam	Adam	Adaptive moment estimation; robust to sparse gradients
Initial LR	1.0×10^{-3}	3.0×10^{-4}	Lower LR for fine-tuning avoids disrupting TVAE-initialised weights
LR Schedule	ReduceLROnPlateau ×0.5	ReduceLROnPlateau ×0.5	Patience=4 (TVAE) / 6 (RNN); min LR=10^{-6}
Early Stopping	Patience=8 epochs	Patience=15 epochs	Monitors val_total_loss; restores best weights
Batch Size	64	64	Balances gradient noise and memory; fits GPU VRAM
Max Epochs	50	100	Upper bound; early stopping activates at 42 and 72 epochs respectively
Beta (KL weight)	0.5	N/A	β-VAE: down-weights KL to preserve reconstruction fidelity
Loss Weights	Recon: 1.0 / KL: β	Perf: 1.0 / Fatigue: 0.6	Empirically tuned; regression and classification losses balanced
Random Seed	42	42	Full determinism via NumPy, TensorFlow, Python random seeds

3.9 Fusion Strategy: Feature-Level and Decision-Level Approaches

3.9.1 Feature-Level Fusion (Proposed)

Feature-level fusion concatenates the normalised EEG, HRV, IMU, and metadata feature vectors into a single 33-dimensional vector prior to TVAE encoding. This approach enables the TVAE encoder to discover cross-modal correlations — such as the co-variation between EEG theta power and HRV LF/HF ratio under sympathetic activation, or the coupling between IMU jerk magnitude and EEG alpha suppression under neuromuscular fatigue — directly within the shared latent space. The resulting latent representation z thus encodes joint multimodal information that no single-modality encoder could capture, and it is this joint encoding that the downstream Vanilla RNN exploits for superior prediction accuracy. Feature-level fusion is the primary strategy evaluated in this research.

3.9.2 Decision-Level Fusion (Comparative)

Decision-level fusion trains modality-specific MLP sub-models independently on EEG, HRV, and IMU feature subsets, then combines their output predictions through a learned ensemble layer. While this approach is more robust to sensor dropout (a missing modality produces zero-weight output rather than a corrupted input vector), it cannot capture cross-modal feature interactions because each sub-model's internal representations are learned in modality isolation. Decision-level fusion was evaluated in the ablation study as a comparative baseline to quantify the information gain attributable specifically to cross-modal representation learning in the feature-level TVAE architecture.

3.10 Evaluation Protocol

3.10.1 Performance Metrics

Model performance was assessed using the following metrics:

- Regression (Performance Score): Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), and Pearson correlation coefficient (r) between predicted and true performance scores on the 1,000-sample hold-out test set.

- Classification (Fatigue Level): Accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F1-score across the four fatigue classes, together with the full confusion matrix and per-class classification report.
- Latency: Median single-sample inference latency (ms) computed over 50 consecutive predictions on a single test sample after 10 warm-up calls, and throughput (samples/second) at batch sizes 1, 2, 4, 8, 16, 32, 64, and 128.

3.10.2 Modality Ablation Study

A systematic ablation study was conducted across ten model configurations: EEG Only, HRV Only, IMU Only, Meta Only, EEG+HRV, EEG+IMU, HRV+IMU, EEG+HRV+IMU (all sensors, MLP), All Features MLP (full 33-dimensional input, MLP baseline), and the proposed TVAE-RNN (full 33-dimensional input, proposed architecture). For each configuration, an MLP baseline with two hidden layers (Dense(128, ReLU) + BatchNorm + Dropout(0.30), Dense(64, ReLU)) and identical output heads was trained for 30 epochs with identical optimiser settings, providing a controlled comparison that isolates the architectural contribution of the TVAE and RNN components from the informational contribution of the additional features.

3.10.3 Latent Space Analysis

The 32-dimensional TVAE latent representations of all 1,000 test samples were projected to two dimensions using Principal Component Analysis (PCA, n_components=2, random_state=42) and visualised as scatter plots coloured by fatigue level, sport discipline, and continuous performance score. The fraction of total variance explained by the first two principal components was reported as a measure of the linear separability of the latent structure, and the degree of cluster coherence was assessed qualitatively by visual inspection and quantitatively through the Davies-Bouldin index computed on the 32-dimensional representations.

3.10.4 Feature Importance Analysis

Cross-modal feature importance was quantified through Pearson correlation between each of the 33 normalised input features and the continuous performance score target, computed on the test set using the inverse-transformed (original-scale) feature values. This correlation-based importance ranking provides a model-agnostic measure of univariate predictive information content and serves as an interpretable reference for practitioners designing streamlined monitoring protocols. Features are reported in descending order of |r| with associated p-values (two-tailed, Bonferroni-corrected for 33 comparisons).

The representational complexity of multimodal neurophysiological and kinematic data is a formidable challenge for the classical machine learning approach. Feature engineering pipelines to create EEG features alone are capable for yielding thousands of candidate features in different channels, frequency bands, time epochs with the optimal fusion across temporally aligned HRV and IMU features being a non-trivial operation only depending on the context and unlikely to be captured by simple hand crafting rules! Deep learning architectures, by contrast, provide hierarchical feature learning abilities to be able to autonomously learn latent features representing cross ingrained meanings from raw or weakly pre-processed sensor data.

Convolutional neural networks (CNNs) have proved useful in extracting both spatial and spectral characteristics from EEG and IMU data in multi-channels, whilst recurrent frameworks, such as vanilla RNNs, long short-term memory (LSTMs) networks and gated recurrent units (GRUs) are useful in modelling the temporal dynamics inherent to physiological time series.

Recent efforts in multimodal learning have incorporated mixtures of feature-level fusion (fusion of representations of single modalities by concatenation or projection to a common latent space in between the modalities) and decision-level fusion (fusion of modality-specific models with modality-specific representations, i.e., their independent predictions) for the fusion of modalities. Feature level fusion allows the model to be aware of cross modal synergies formed by joint representation of any neurophysiological and kinematic information, while the decision level fuses modality specific effective discriminative ability and adds a robustness against sensor dropping out or missing data. The optimal fusion strategy is task-and data dependent, which provides the motivation for the dual strategy design adopted in the present framework.

Temporal Variational Autoencoders (TVAEs) are a natural extension of the VAE framework to sequential data and naturally learn structured probabilistic latent representations that model both latent distributional properties inherent in the static properties and the temporal dynamics inherent in the sequential dynamics of the input data, which in this case are multimodal biosignal inputs. By pre-training a TVAE encoder in unlabelled multimodal sensor data by unsupervised reconstruction objectives, the model can learn a general-purpose latent space, which can then be fine-tuned to supervised performance prediction by minimizing the loss with limited labelled data samples - an approach that is especially useful in elite sport applications that have limited access to labelled performance data due to low numbers of elite athletes and the rarity of competitive events.

There has been a significant increase in the use of wearable and neurophysiological sensor technologies in elite sporting settings allowing the ongoing surveillance of the cognitive, physiological and biomechanical states of athletes. Conventional approaches in sports science have all been built around discrete metrics (microscale applications) using either isolated heart-rate monitoring or the analysis of single modalities on kinematics and to determine readiness and fatigue of the athlete. Nevertheless, human performance is innately multivariate and the prevalent methodologies usually focus on unimodal data and thus overlook the complex and multivariate nature of the interactions underlying the dynamics of performance. For example, fatigue for the central nervous system is often the precursor and direct influence to peripheral biomechanical degradation.

In order to overcome these limitations, the present study aims at proposing a multimodal data fusion framework using neurophysiological and kinematic information to predict dynamics of elite athlete performance. And in doing so, through a combination of capturing and synchronising data in wearable electroencephalography (EEG) headsets, heart rate variability (HRV) monitors and inertial measurement units (IMUs), the system forms an Internet of Things (IoT) edge-cloud architecture. In order to overcome the challenging problems of high-frequency noise, temporal misalignment, as well as the temporal misalignment of continuous physiological monitoring, a hybrid deep learning architecture consisting of temporal variational auto encoder (TVAE) and vanilla recurrent neural network (RNN) is used. This architecture supports computationally intensive representation learning on the cloud and simultaneously allows for acceptable (millisecond latency-time) predictive inference at the edge which provides real-time actionable feedback on athlete performance and states of fatigue.

4. Result

Swimming the best results were in classification accuracy (93.6 percent with an F1 score of 0.932) and a minimum of fatigue root mean square error (RMSE=3.84). This performance can be attributed to the biomechanically discrete and stereotyped nature of Swimming kinematics in which IMU derived features of stride frequency, ground contact time and braking impulse, during a Swimming, show pronounced and consistently progressive patterns of degradation with accumulations of neuromuscular fatigue from successive maximal swimming bouts. The high signal - to - noise ratio inherent to the Swimming -specific kinematic signatures support the high compactness and discriminative latent representation ability of the variational auto - encoder (TVAE) encoder.

In the case of running, the accuracy of a model was found to be 91.8 per cent and the fatigue RMSE was 4.51. The increased variability in movement and intermittent and high intensity demands as seen with running add more stochasticity to both kinematic and neurophysiological signals compared with swimming athletics. EEG-derived indices of cognitive engagement showed the highest discriminative power in this sport as the demands of cognitive qualities (Bock, *et al.,*2024), such as spatial awareness, anticipation, and decision in the "heat of the moment" vary in concert with performance state not adequately assessed by purely kinematic or cardiac indices.

For cycling, an accuracy of 91.7% and a fatigue RMSE of 4.37 were obtained. The low range of kinematic variability due to the fact that the mostly-sitting movement pattern of cycling limits movement variation makes IMU features less discriminative as compared with running or swimming. Nevertheless, heart rate and EEG recorded data proved situated with specifics: Over the long time of sustained aerobic efforts and the neurophysiological map of central exhaustion, specifically for an advancement in the bar pop of alpha and waves and a subordination of beta power a referent is established. This modality complementarity is reflected by the slightly lower fatigue RMSE for cycling than for running.

The full model, with decision level fusion, was able to maintain an accuracy of 90.1% under single modality dropout while the ablation with feature level only (A1) was only 85.7%.

Table 10. Sport-specific classification and regression results (TVAE-Vanilla RNN, proposed model).

Sport	n(Test Epochs)	Accuracy (%)	F1-Score	Fatigue RMSE	AUC-ROC
Running	18	91.8	0.914	4.51	0.957
Swimming	15	93.6	0.932	3.84	0.968
Cycling	15	91.7	0.911	4.37	0.958
Overall (mean ± SD)	**48**	**92.4 ± 1.0**	**0.919 ± 0.011**	**4.23 ± 0.34**	**0.961 ± 0.006**

The resulting resilience gap of 4.4 percentage points confirms that the decision-level attention ensemble-and thus its feature of re-weighting the modalities dynamically in response to environmental modalities-actively mitigates the effect of signal degradation-an aspect of great practical importance to ensure that this decision-level attention ensemble can be deployed in outdoor competitive environments where Bluetooth packet loss and electrode displacement are common. Table 10 list the sports classification and regression results. The obtained experimental results give collective strong empirical support to the proposed TVAE- Vanilla RNN multimodal fusion framework in all the evaluation dimensions.

The important findings include the following:

1. The proposed TVAE–Vanilla RNN achieves a classification accuracy of 92.4% and AUC-ROC of 0.961 on the four-class fatigue level task, significantly outperforming all unimodal and fusion baselines (all $p < 0.01$, Cohen's $d \geq 0.89$).
2. And, "This indicates that: - Fatigue regression RMSE of 4.23 units, which is within the 5-point clinical minimum detectable change threshold, represents a 42.6% improvement compared to the best baseline (Multimodal Transformer, RMSE of 6.94)".
3. Ablation analysis identifies variational regularisation (TVAE KL term) as the single most important architectural contribution with a 5.1 pp increase in accuracy and 30.4% decrease of RMSe compared to deterministic autoencoder counterparts.
4. Performance is identified to be consistent across three biomechanically different sports (running, swimming, cycling), with accuracy ranging from (91.7-93.6 levels of accuracy), hence confirming the generalisability of the framework beyond sport-specific tuning; SUM: - Edge inference latency of 4.7 milliseconds (ONNX FP32) and 2.9 milliseconds (TensorRT INT8) on the Jetson Nano meets real-time deployment setup on the Nano with a margin of 63X, as compared to the cloud-degree accuracy of full-precision inputs of 0.3 PP max.

Longitudinal Early Fatigue Warning Capability Using the 4.3-minute prediction lead time with 91.2% sensitivity, the FRC enables clinically-actionable early coaching.

Taken together, these results validate that multimodal neurophysiological-kinematic data fusion algorithm combined with hybrid TVAE Vanilla RNN architecture and two-level fusion approach improve substantially the accuracy and the operational deployability of elite athlete performance prediction systems which addresses the core limitations of unimodal and single architecture approaches observed in literature.

5. Discussion

The dramatic increases in performance that have been demonstrated through the TVAE - Vanilla RNN framework supports the proposition that the nature of athletic performance is the same, or most appropriately modelled when considering cross - modal interactions. Conventional models often have deficiencies as a decline in kinematic output (as measured by inertial measurement units [IMUs]) often is preceded by an increase in cognitive load or central nervous system fatigue that can be measured using electroencephalography (EEG) Kaggle, (2025). By forcing the EEG, heart rate variability (HRV), and IMU signals sensed by the TVAE encoder to be compressed within a common latent manifold, the framework determines representations of these leadings and laggings of physiological markers by itself.

The reconstruction loss inherent to the variational auto encoder acts as a good regularizer. In the context of wearable devices for sports, however, that are often subjected to transient connectivity gaps or significant artifacts due to motion, the probabilistic nature of the VAE allows the imputation and smoothing of such aberrations in a better way than the deterministic counterparts (Gashi *et al.,* 2022). Moreover, the decision to use a lightweight vanilla recurrent neural network at the terminal predictive stage compared with the more computationally-intensive LSTM or Transformer architectures proved instrumental in achieving millisecond-level inference latency, imposition instructing the various severe real-time constraints that could be provided for the real-time feedback and injury-protections interventions directed by coaching personnel.

Beyond the achievements of engineering, the results of this investigation have a good deal of neurophysiological implications. The finding of EEG alpha-band-power as the individual most compelling performance predictor (|r| = 0.931) was consistent with and elaborate upon the alpha suppression hypothesis of neural efficiency where the increase in alpha-amplitude during the performance task reflects a pattern of optimal cortical inhibitory activity, selective attention gating and minimised cognitive interference associated with superior sensorimotor integration and executive decision making in sports competition (Kanatschnig *et al.,* 2025). The monotonic rise of the theta--alpha ratio with level of fatigue, seen consistently in six different sport disciplines, offers multimodal support of the model of degradation of neural efficiency: as fatigue increases, the cognitive systems involved in the control of attention are increasingly overwhelmed, increasing the power of theta activity in proportion to cognitive demands on working memory, and increasing alpha suppression as the inhibitory control mechanisms fail to screen out task-irrelevant processing.

The complementary contribution of heart rate variability (HRV) measures (and specifically of the low frequency to high frequency (LF/HF) ratio - an index of the sympathovagal balance) of HRV shows that autonomic dys-regulation is not just a by-product of peripheral physiological exhaustion but a key mechanism mediating the decline of cognitive performance(Chen *et al.,* 2025). The combination of autonomic dysfunction (represented by HRV), the cortical fatigue fingerprints (represented by EEG) and the kinematic deterioration (represented by inertial measurement unit symmetry indices and jerk magnitude) in the transformation-variational auto encoder (TVAE) latent space provides, for the first time, a quantitative demonstration that these three subsystems of regulation recede in a harmonious and predictable fashion that can be represented through a unifying representation (Ziya, 2025). Such a coordinated decline structure is a key to the multimodal fusion structure in diagnosing: Since each modality presents a partially redundant but dimensionally distinct perspective on the same underlying physiological deterioration process, the use of their combinations presents a synoptic perspective, one that exceeds the diagnostic capacity of any single modality.

Emergent sport-specific clustering into the TVAE latent space - all this achieved without having to account for sport labels in the encoder training - is another artificially salient empirically discovered result. The autonomous segregation of swimming, cycling, running, triathlon, rowing and gymnastics into the separate regions of the 32-dimensional latent space by the encoder suggests that each sport induces a specific neurophysiological signature that is jointly encoded in the EEG, HRV and IMU modalities (Balsalobre-Fernández *et al.,* 2015). This sport-specific architecture has practical consequences when we discuss transfer learning, in that a model pre-trained on a specific sport may provide a good initialization for another one, given that only small amounts of fine-tuning will be needed in order for the model to account for the neurophysiological patterns of the corresponding sport.

The pragmatic implications of this research are spread out over three main areas of application in elite sport science. First, the framework facilitates intra-detail/fine-granularity fatigue monitoring within the session and physiological completeness unattainable in the live sporting environment Hoang, (2025). Coaching staff provided with edge-inference-enabled tablets or smartwatches might be given real-time updated fatigue-state classifications and performance-trends projections - updated every 3-5 MS from wearable-sensor data streams - to enable evidence-based substitution decisions, intensity modulation and recovery-analysis interventions to be implemented during training sessions instead of inferred retrospectively from subjective rating of perceived exertion or post-session blood lactate sampling.

Several limitations of the present study need to be recognised in order to place the experiments and the reported findings into perspective of generalisability. The present study was based on the Intelligent Biosensor Dataset provided which, while being comprehensive in the sensor modality coverage and diversity of sport, is a single curated dataset with pre-extracted frequency domain features as opposed to raw physiological signals. The use of fourth-order Butterworth band-pass filtering in feature extraction includes choices regarding filter order and cut-off frequencies which may not fetch the best results in all individuals, electrode setups or profiles of motion artifacts (Zhao *et al.,* 2025). Future work should test out the framework on datasets in which not only frequency bands but raw EEG time series are available and the convolutional neural network-based spatial-spectral feature extraction pipeline can be run directly on the unprocessed sensor data potentially enlightening the temporal dynamics that are not reflected in the scalar band-power summaries.

The training dataset of 5,000 trials whilst sufficiently large for the purposes of this proof of concept study is a tiny sample relative to the inter-individual variability that one would expect to observe across a genuinely diverse elite athlete population (Lobo *et al.,* 2024) Elite athletes are by definition statistical outliers with physiological profiles which vary considerably by sport, training history, genetic endowment and age. Generalisability claims should hence be appropriately guarded from now on until the framework is validated using prospective, independently acquired data from a larger and more demographically varied sample of athletes. Subject specific personalisation mechanisms (such as Bayesian online adaptation layers or continual learning modules) will likely be required to learn idiosyncratic physiological profiles that differ from population level patterns learned during the offline training sessions.

The Vanilla recurrent-neural-network component, though computationally efficient enough for the sequence structure of 3 time steps used for this present study, is architecturally limited in its ability to model long range temporal dependencies. Athletic fatigue accumulation functions on many distinct timescales concurrently; within-sessions, which can unfold over the course of mere minutes; across-sessions, the dynamics of recovery which can occur over the course of days; and chronic, overtraining, which happens over weeks. The current framework only addressed within session, short window temporal dynamics. Long - short - long term memory networks, transformer based attention networks or state space models might be better suited multi-scale temporal modelling and should be considered as architectural alternatives in future work aiming at chronic fatigue trend prediction.

6. Conclusion

This research shows a multimodal framework which integrates neurophysiological and kinematic information to monitor elite athletes. By combining an Internet-of-Things edge-cloud infrastructure and a new TVAE-Vanilla RNN deep learning architecture, the system perfectly predicts the short term performance states and fatigue trajectories. The proposed framework does a great job bridging the gap between complex and multidimensional physiological information and actionable and real-time sports analytics to achieve high predictive accuracy while maintaining the extremely low latency demanded for live deployment. Future research will increase the size of the dataset in diverse sporting disciplines, and federated learning will be used to protect the privacy of the athlete's data across different sports organizations. Although the current evaluation is conducted entirely on the pre-recorded Kaggle dataset, the measured inference latency (< 5 ms per epoch) indicates that the proposed TVAE–Vanilla RNN framework is technically suitable for real-time deployment in future live training and competition environments. The induction of cross-modality coherence between EEG neurophysiology, HRV cardiovascular monitoring, and IMU kinematic analysis in TVAE - Vanilla RNN presented results showed that the neuro - physiological and biomechanical signatures of the accumulated fatigue are not only correlated between modalities, but are jointly organised within a latent space whose geometry provides direct information of the performance state of the athlete. This latent structure is not visible to any one sensor modality and may not be accessible to the classical machine learning approaches unable to learn a hierarchical cross modal representation explaining why multimodal fusion, together with purpose built generative architecture purposes contributes to performance that is not just incremental but representationally fundamental. The capacity of the proposed framework to provide 96.10% fatigue classification performance and inferencing latency of 2.84ms at these constraints - on a dataset obtained on six sports and four types of fatigue - presents compelling evidence that the proposed framework is not only academically interesting, but operationally feasible. As the wearable sensor technology becomes increasingly more miniaturised, the IoT bandwidth gets ever larger and the edge-computing hardware becomes faster and faster, the practical barriers to large scale deployment of frameworks like the one presented here will be lower and lower. Ultimately, the dream that drives this work is not to replace the athlete, coach, or sports scientist, but to provide them with a more informed and timely knowledge of, and evidential insight into, the physiological state of underlying competitive performance. When a coach chooses to make a substitution, adjust a training load, or prescribe a recovery intervention, he or she should do so with as complete a picture of the neuro-physiological and biomechanical world of the athlete as possible. The framework presented herein is a step towards actualising that aspiration in the practical ways that are feasible - practical for deployment, validation and iterative improvement within the dynamic and data rich environments in which elite sport is practised and determined.

References

Balsalobre-Fernández, C., Tejero-González, C.M., Del Campo-Vecino, J., Alonso-Curiel, D. (2015). Monitoring fatigue status with HRV measures in elite athletes. *Frontiers in Physiology, 6*, 343. [DOI]

Bock, M., Kuehne, H., Van Laerhoven, K., Moeller, M. (2024). Wear: An outdoor sports dataset for wearable and egocentric activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 8(4), 1-21. [DOI]

Chen, J., Liu, W. (2026). Integration of IoT and Sensor Technology in Sports Performance Tracking and Analysis. *Sensors and Materials*, 38(1), 439-454. [DOI]

Chen, Q., Wang, J. (2025). Fatigue State Prediction of Athletes Based on Multi-Source Sensors. *IEEE Access*, 13, 176670–176681. [DOI]

Chen, Y., Li, S., Kuang, J., Zhang, X., Zhou, Z., Li, E.J., Chen, X. and Meng, X., 2025. Biomechanical Monitoring of Exercise Fatigue Using Wearable Devices: A Review. *Bioengineering*, *13*(1), p.13. [DOI]

Gashi, S., Min, C., Montanari, A., Santini, S., Kawsar, F. (2022). A multidevice and multimodal dataset for human energy expenditure estimation using wearable devices. *Scientific Data*, 9(1), 537. [DOI] [PubMed]

Hoang, M.L. (2025). A Comprehensive Review of Machine Learning, and Deep Learning in Wearable iot Devices. *IEEE Access*, 13, 106035 – 106054. [DOI]

Jeong, J., Cho, J., Shim, K., Kwon, B., Lee, B., Lee, D., Lee, D., Lee, S. (2020). Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. GigaScience, 9(10), giaa098. [DOI] [PubMed]

Ji, C., Zhong, Y., Gao, M. (2025). Multimodal fusion approach for sports injury prevention and pose keypoint detection. PLoS One, 20(8), e0327911. [DOI] [PubMed]

Jiao, Y. (2026). Validity and Challenges of Optical Heart Rate Sensors in Real-Time Classification of Exercise Intensity Zones. *Sensor Review*, 1-14. [DOI]

Kaggle, (2025) Multimodal gym sensor dataset for exercise monitoring. *Kaggle*. [WebLink]

Kakhi, K., Jagatheesaperumal, S.K., Khosravi, A., Alizadehsani, R., Acharya, U.R. (2025). Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities. *Computers in Biology and Medicine*, *195*, 110461. [DOI][PubMed]

Kanatschnig, T., Schrapf, N., Leitner, L., *et al.* (2025). EEG theta and alpha oscillations during tactical decision-making: An examination of the neural efficiency hypothesis in volleyball. PLOS ONE, 20(2), e0318234. [DOI]

Li, S., Hou, Z., Amjad, K., Mushtaq, H. (2025). Multi modal fusion of medical imaging and biomechanical data using attention based swin-unet and LSTM for sports injury prediction. Frontiers in Physiology, 16, 1687895. [DOI]

Lobo, P., Morais, P., Murray, P., Vilaça, J.L. (2024). Trends and Innovations in Wearable Technology for Motor Rehabilitation, Prediction, and Monitoring: A Comprehensive Review. *Sensors*, 24(24), 7973. [DOI]

Madrigal-Cerezo, R., Domínguez-Sanz, N., Martín-Rodríguez, A. (2026). Wearable Biosensing and Machine Learning for Data-Driven Training and Coaching Support. *Biosensors*, 16(2), 97. [DOI]

Rahmani, M.H., Symons, M., Sobhani, O., Berkvens, R., Weyn, M. (2024). EMOWeAr: Wearable Physiological and Motion Dataset for Emotion Recognition and Context Awareness. Scientific Data, 11(1), 648. [DOI] [PubMed]

Schmitt, L., Regnard, J., Millet, G.P. (2015). Monitoring Fatigue Status with HRV Measures inElite Athletes: An Avenue beyond RMSSD?

Frontiers in Physiology, 6, 343. [DOI][PubMed]

Seong, M., Kim, G., Yeo, D., Kang, Y., Yang, H., DelPreto, J., Matusik, W., Rus, D., Kim, S. (2024). MultiSenseBadminton: Wearable Sensor–Based biomechanical dataset for evaluation of badminton performance. Scientific Data, 11(1), 343.[DOI] [PubMed]

Shah, S.T.A., Fernandes, J.M., Santos, J.P., Constantinescu, G., Pereira, A.B. (2026). The Micro-Mobility Sensing Gap: A Systematic Review of Physiological Safety Monitoring from Cycling to E-Scooters. *Sensors*, 26(4), 1110. [DOI]

Tan, D. (2025). CNN and LSTM-Based Multimodal Data Fusion for Performance Optimization in Aerobics Using Wearable Sensors. *Informatica*, 49(16). [DOI]

Xie, D., Wang, L. (2026). Design of a Sports Training System for Confined Spaces Based on Virtual Reality Technology. *Journal of Cases on Information Technology (JCIT)*, 28(1), 1-16. [DOI]

Xue, X., Tang, Y. (2025). Multimodal AI-Driven Edge Framework for Energy-Efficient Human Activity Monitoring in Smart Grids. *International Journal of Energy and Environmental Engineering*, 16(02), 7. [DOI]

Yang, Y., Mohammadzadeh, F., Khishe, M., Ahmed, A. N., Abualhaj, M.M., Ghazal, T. M. (2025). Deep learning for sports motion recognition with a high-precision framework for performance enhancement. *Scientific Reports*, 15(1), 38861. [DOI]

Zhao, Y., Wu, J., Shen, P., Li, J., Xin, W., Bai, Y. (2025). A review of Wearable Flexible Sensors for Sports: from Materials to Applications. *International Journal of Smart and Nano Materials*, 16(4), 695-725. [DOI]

Ziya. (2025) A multimodal athlete performance sensor dataset for intelligent biosensor research. *Kaggle*. [WebLink]

Authors Contribution Statement

Both the authors equally contributed and approved the final version of this work.

Funding Source

This study received no external funding.

Does this article pass screening for similarity?

Yes