AthleTwin: Towards AI-Driven Digital Twins For Personalized Sports Injury Prevention And Rehabilitation

[1]Harleen Kaur, [2]Ashish Kumar
[1,2]GreyB Analytics Private Limited, Mohali, Punjab, India

Abstract—Sports injuries impose substantial physiological, psychological, and economic burdens on athletes worldwide. Although artificial intelligence (AI) and machine learning (ML) have shown increasing potential for predicting injury risk, most existing approaches remain retrospective, population-based, and reliant on a single data modality, limiting their ability to capture the complex, individualized, and dynamic nature of sports injuries. Digital twin (DT) technology, which creates a continuously evolving virtual representation of a physical entity using real-time sensor data, offers a promising solution to address these limitations. This study presents AthleTwin, the first comprehensive conceptual framework for AI-driven digital twins specifically developed for sports injury prevention and rehabilitation. AthleTwin comprises five key players: (i) multi-modal data acquisition from wearable IoT sensors, computer vision, and medical imaging; (ii) personalized biomechanical and physiological modeling; (iii) AI-based injury risk prediction using hybrid physics-informed neural networks; (iv) an explainable AI (XAI) layer to support clinicians and coaches in decision-making; and (v) a closed-loop adaptive intervention engine for real-time load management and rehabilitation optimization. Further, it introduces a taxonomy of digital twin maturity levels in sports medicine, examine the major technical challenges spanning data, modelling, AI, and deployment, and outline a structured five-priority research agenda for future work. Thus, this work establishes a conceptual foundation and practical roadmap for the development of personalized, proactive, and clinically grounded AI systems for athlete health management.

Index Terms—*Digital Twins; Artificial Intelligence (AI); Sports Injury Prevention; Wearable Sensors; Physics-Informed Neural Networks; Explainable AI; Personalized Medicine; Rehabilitation; IoT; Biomechanics*

I. INTRODUCTION

Sports injuries constitute one of the most prevalent and consequential challenges in elite and recreational athletics. Epidemiological data consistently report incidence rates of 2-19 injuries per 1,000 athlete-exposure hours across contact and endurance sports, with anterior cruciate ligament (ACL) tears, hamstring strains, stress fractures, and rotator cuff pathologies imposing the largest burden of time loss and long-term morbidity [1,2]. The global economic cost of sports injury encompassing medical treatment, rehabilitation, lost productivity, and insurance expenditure has been estimated at hundreds of billions of dollars annually [3].

Over the past decade, artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for sports injury risk prediction, with ensemble tree-based models (XGBoost, Random Forest) and deep learning (DL) architectures (Convolutional Neural Networks (CNN), long-Short-Term Memory (LSTM), Transformer) demonstrating AUC scores of 0.80-0.92 across multiple prospective studies [4,5]. Despite this progress, three fundamental limitations restrain current AI approaches from achieving their full preventive potential. First, population-averaged models- most published models learn from aggregated cohort data, producing injury risk estimates that reflect population-level statistical patterns but cannot capture the highly individual biomechanical and physiological signatures that determine a specific athlete's risk at a specific moment. Second, temporal discontinuity- current AI systems are typically applied at discrete assessment time points (weekly wellness checks, pre-season screenings) rather than continuously across the full training and competition cycle, missing dynamic intra-session risk fluctuations. Third, one-way prediction- existing systems predict injury risk but do not close the feedback loop to adaptively modify training programs or rehabilitation protocols in response to evolving risk profiles. Digital twin (DT) technology originating in aerospace and manufacturing engineering and increasingly applied in healthcare [7] addresses all three limitations simultaneously. A digital twin is a continuously updated virtual replica of a physical entity that synchronizes in real time with sensor data from its physical counterpart, enabling simulation, prediction, and optimization without disrupting the physical system. Applied to athletes, a DT creates a personalized, dynamic computational model of an individual athlete's biomechanical, physiological, and psychological state thus, enabling individual-specific injury risk assessment, prospective what-if simulation of training load scenarios, and closed-loop adaptive intervention.

Contributions of Paper: This paper makes four contributions: -

1) A novel AthleTwin framework is proposed which is to the best of our knowledge- the first comprehensive layered conceptual framework for AI-driven digital twins in sports injury prevention and rehabilitation;
2) A five-level maturity taxonomy for athletic DT systems is introduced;
3) Open challenges across data, modeling, AI, and deployment dimensions are systematically analyzed; and
4) A structured five-priority research agenda is proposed to advance the field toward clinical deployment.

This work is a conceptual framework developed through a comprehensive synthesis of the existing literature and state-of-the-art (SOTA) technologies in the field, serving as a foundation for future implementation and experimental validation.

II. BACKGROUND

A. Digital Twins: Concepts and Evolution

The digital twin concept was s by Grieves [6] in the context of product lifecycle management and subsequently extended to cyber-physical systems, smart manufacturing, and precision medicine. A canonical DT architecture comprises three components: the physical entity (the real-world athlete and their environment); the virtual entity (the computational model); and the data connection layer (the bidirectional real-time data flows linking physical and virtual components). The defining property that distinguishes a digital twin from a static simulation model is continuous synchronization. The virtual model updates in real time as new sensor data arrives from the physical entity, and can feed control actions back to the physical system through actuators or human decision-support interfaces.

In healthcare, DT applications have expanded rapidly, encompassing cardiac digital twins for cardiovascular medicine [8] and surgical planning twins for orthopedic procedures [9]. Diniz et al. [7] provided the first systematic review of digital twin systems for musculoskeletal applications, documenting emerging implementations in gait analysis, implant design, and ACL reconstruction planning thus establishing the technical feasibility of DT approaches for the biomechanical domain directly relevant to sports injury.

B. AI in Sports Injury: Current State and Gaps

Current AI applications in sports injury prediction span supervised ML for training-load-based risk modeling [4], DL for medical image analysis [10], computer-vision-based estimation for biomechanical screening [11], and multimodal fusion frameworks integrating sensor, imaging, and clinical data [12]. Wearable IoT sensors including GPS trackers, inertial measurement units (IMUs), and heart rate variability (HRV) monitors provide the continuous data streams that underpin real-time monitoring systems [13,14].

Despite these advances, no existing system integrates the full pipeline from continuous multi-modal sensing through personalized physiological and biomechanical simulation to closed-loop adaptive intervention- the capability that a purpose-built DT architecture would enable. The convergence of DT technology with sports AI is therefore a natural and timely research priority, first articulated by Nobari [15] and receiving growing attention in the sports science literature [16].

III. THE ATHLETWIN FRAMWEWORK

AthleTwin is proposed which is a five-layer architecture for AI-driven digital twins in sports injury prevention and rehabilitation (see Fig. 1 and Table I). Each layer is described below.

A. Layer 1: Multi-Modal Data Acquisition

The data acquisition layer continuously ingests heterogeneous sensor streams from the physical athlete. External workload data streams originate from GPS/GNSS trackers (total distance, high-speed running, sprint metrics, accelerations/decelerations) and optical tracking systems. Internal load and recovery data streams arise from wearable IMUs (joint kinematics, impact transients), photoplethysmography-based HRV monitors (autonomic recovery status), surface EMG patches (muscle activation fatigue indicators), and sleep trackers (duration, efficiency, REM proportion). Imaging data taken from MRI, ultrasound, or markerless motion capture are integrated episodically during scheduled assessments. Athlete-reported outcome measures (wellness scores, pain ratings, readiness perception) complete the input data landscape. Thus, a standardized athlete health data ontology analogous to Fast Healthcare Interoperability Resources (FHIR) in clinical medicine is essential for harmonized multi-source data ingestion.

B. Layer 2: Personalized Biomechanical and Physiological Modeling

The modeling layer constructs and maintains a personalized computational model of the athlete's musculoskeletal system, continuously calibrated against incoming sensor data. Musculoskeletal (MSK) models implemented using established multibody dynamics frameworks (OpenSim, AnyBody) and calibrated to individual anthropometrics, muscle-tendon unit properties, and joint geometry from imaging thus enable estimation of internal tissue loads (muscle forces, ligament strains, bone stress distributions) from the external kinematic and kinetic measurements available from wearables. Physics-informed neural networks (PINNs) [17] which embed physical laws (equations of motion, constitutive material models) as regularization constraints within neural network training provide a computationally efficient mechanism for personalizing biomechanical model parameters from sparse individual data while preserving physical plausibility. Physiological compartment models of fatigue, recovery, and training adaptation (e.g., Banister impulse-response models, Margaria's phosphocreatine model) are incorporated to simulate internal physiological state trajectories.

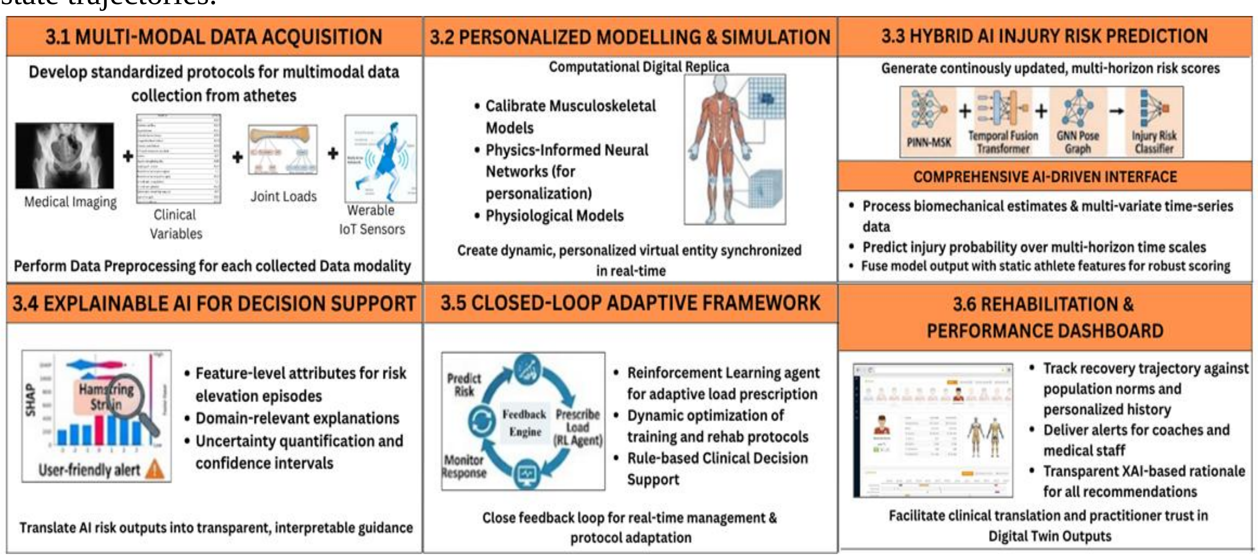

Fig. 1. AtheTwin: AI-Driven Digital Twin Framework

TABLE I. ATHLETWIN PROPOSED FRAMEWORK

Layer / Level	Name	Sensing	Modeling	AI Capability	Feedback
L1	Monitoring	Single wearable	-	Descriptive stats	-
L2	Predictive	Data from multiple athletes	Statistical	ML risk scores	Passive Alerts
L3	Simulant	Multi-Modal IoT	MSK[a] Model	DL TFT[c]	Load Advisories
L4	Adaptive	IoT + Imaging	PINN[b]-MSK	Hybrid + XAI	RL[d] Prescription
L5	Autonomous	Seamless fusion	Full physiome	Foundation Model	Closed-Loop

[a.] MSK: Musculoskeletal, [b.] PINN: Physics-Informed Neural Network, [c.] TFT: Temporal Fusion Transformer, [d.] RL: Reinforcement Learning

C. Layer 3: AI-Driven Injury Risk Prediction

This layer integrates outputs from the modeling layer with raw sensor features using a hybrid AI architecture. A temporal fusion transformer (TFT) [18] processes multi-variate time series of biomechanical estimates, learning attention-weighted risk factors over variable time horizons. The TFT output is combined with static athlete features (age, injury history, anthropometrics) through a gradient boosting classifier that provides calibrated injury probability estimates and confidence intervals. Graph neural networks (GNNs) encode body-pose graphs derived from markerless motion capture for real-time biomechanical risk scoring during training sessions. The combined output is a continuously updated, multi-horizon (24-hour, 7-day, 28-day) injury risk probability surface, decomposed by anatomical region and injury mechanism.

D. Layer 4: Explainable AI Interface

The explainability layer translates AI-generated risk estimates into actionable, interpretable outputs for coaches, physiotherapists, and athletes. SHAP TreeExplainer provides feature-level attribution for risk elevation episodes, expressed in domain-relevant language (e.g., ACWR exceeding 1.5 for three consecutive days contributes 23% to hamstring injury risk). Attention heatmaps from the TFT visualize which historical time windows are most influential for current risk elevation. Counterfactual explanations which answer ‘what training load reduction would bring this athlete's risk below the alert threshold?’- provide directly actionable prescriptions. An uncertainty quantification module reports prediction confidence intervals, distinguishing well-characterized risk profiles from uncertain out-of-distribution cases requiring enhanced clinical oversight.

E. Layer 5: Closed-Loop Adaptive Intervention Engine

The intervention engine closes the feedback loop from risk prediction to load management and rehabilitation protocol adaptation. A rule-based clinical decision support layer translates XAI-generated risk alerts into standardized traffic-light notifications (green/amber/red) delivered to coaching and medical staff dashboards with specific recommended actions. A reinforcement learning (RL) agent [19] trained via simulation on the digital twin's virtual athlete model learns optimal training load prescription policies that jointly minimize injury risk and maximize training adaptation, operating within boundaries established by human clinicians. During rehabilitation, the DT tracks recovery trajectory against population-normative and individual-historical benchmarks, triggering protocol progression or regression recommendations with transparent XAI-based rationale.

IV. OPEN CHALLENGES

A. Data Challenges

Realizing the AthleTwin vision requires unprecedented data richness, continuity, and integration. The heterogeneity of wearable sensor technologies with differing sampling rates, coordinate systems, and error characteristics across manufacturers demands robust sensor fusion pipelines and standardized data schemas. Continuous multi-modal data collection generates terabyte-scale longitudinal datasets per athlete, raising challenges of efficient storage, compression, and retrieval. Missing data from sensor drop-out, non-compliance, and device failure must be handled gracefully, with imputation strategies that preserve temporal structure. Critically, the data governance challenges of athlete health data including DPDP compliance, competitive confidentiality, and informed consent for secondary data use in DT model training require dedicated legal and ethical frameworks [20].

B. Modeling and Computational Challenges

Personalized MSK model calibration from sparse wearable data (rather than full laboratory motion capture) remains an unsolved problem. Individual-specific model parameters (muscle-tendon unit properties, joint contact geometry, cartilage material characteristics) are not directly measurable in field settings and must be inferred through computational identification procedures that are computationally expensive and prone to identifiability problems. PINNs offer a promising mechanism for embedding biomechanical constraints into data-driven models [17], but their application to complex, high-degree-of-freedom MSK systems during dynamic athletic movements requires further methodological development. Model uncertainty arising from both parameter uncertainty and model structural error must be propagated through to downstream injury risk estimates with appropriate confidence quantification.

C. AI and XAI Challenges

The hybrid physics-AI architecture of AthleTwin introduces novel training and inference challenges. Training PINNs and TFTs jointly on heterogeneous data from multi-modal sensors requires careful loss function design and gradient balancing. The non-stationary nature of athlete physiological state (as fitness, fatigue, and injury history evolve across a season) challenges the assumption of distributional stationarity underpinning most ML models. Continuous online learning- updating DT models as new athlete-specific data accumulates, is required but must guard against catastrophic forgetting of population-level priors. XAI explanations must be generated in milliseconds for real-time dashboard delivery, necessitating efficient approximation of exact SHAP computation.

D. Clinical Translation and Trust

Even technically excellent DT systems will fail to achieve clinical impact if practitioners do not trust, understand, and act on their outputs. The automation bias risk where practitioners over-rely on AI recommendations without critical evaluation, must be counterbalanced by XAI designs that promote active engagement rather than passive acceptance. Alert fatigue from high false-positive rates can lead practitioners to disengage from DT system outputs. Regulatory pathways for AI-based clinical decision support systems add compliance requirements that must be incorporated from the design stage [20].

V. RESEARCH AGENDA

Based on the framework and challenge analysis, we identify five high-priority research directions for the community as given in Table II.

TABLE II. ATHLETWIN RESEARCH AGENDA

Priority	Key Research Questions	Methods
Standarized Atlete Health Data Collection	How do we harmonize heterogeneous wearable and imaging data across devices and institutions?	FHIR extension; DPDP compliance
PINN-MSK Model Personalization	How can biomechanical model parameters be identified from wearable data only in field conditions?	Bayesian PINN; Sequential data assimilation
Hybrid Physics-AI Inference	How do we train and deploy computationally efficient hybrid models for real-time DT inference?	Edge inference deployment
RL-Based Closed-Loop Intervention	How do we safely learn optimal load management policies from DT simulation without harming athletes?	Clinical trial validation
Prospective Clinical Validation	Do AthleTwin-guided interventions reduce injury rates in randomized controlled trials?	Multi-Site RCT

VI. POSITIONING RELATIVE TO EXISTING WORK

The AthleTwin framework is explicitly positioned to complement, not replace existing AI-based sports injury prediction approaches. Population-level ML models (XGBoost, RF, LSTM) [4,5] provide the statistical priors and feature importance insights that inform AthleTwin's inference engine. Medical imaging AI systems [10] supply structural tissue health information that calibrates MSK model geometry and material parameters. Multimodal fusion frameworks [12] provide architectural precedents for integrating heterogeneous data streams. IoT wearable monitoring platforms [13,14] supply the continuous sensor infrastructure on which AthleTwin's data acquisition layer depends.

What AthleTwin uniquely contributes is the integration of these components within a unified, personalized, bidirectional, and continuously updating digital twin architecture thereby enabling individual-level, multi-horizon, causally grounded, and intervention-capable injury risk management that no existing system provides. The closest prior work- the FC Barcelona digital twin initiative for player injury prevention [16] and the TCS Future Athlete Project creating digital heart twins [21] demonstrates real-world institutional interest in the DT paradigm but lacks the full five-layer AI integration proposed here. The systematic review by Diniz et al. [7] confirms that existing MSK digital twin systems remain primarily at Levels L1–L2 of our proposed maturity taxonomy, validating the need for the AthleTwin framework to guide advancement toward L3–L5.

VII. CONCLUSION

AthleTwin is presented which is a comprehensive conceptual framework for AI-driven digital twins in sports injury prevention and rehabilitation. By integrating multi-modal IoT sensing, personalized physics-informed biomechanical modeling, hybrid deep learning inference, SHAP-based explainability, and closed-loop adaptive intervention within a unified five-layer architecture, AthleTwin addresses the fundamental limitations of current population-averaged, temporally discrete, prediction-only AI approaches. The accompanying five-level maturity taxonomy provides a practical scaffold for evaluating and planning DT system development across the full spectrum from basic monitoring to autonomous adaptive management. The five-priority research agenda identifies the concrete technical, methodological, and clinical translation challenges that the research community must address to realize the AthleTwin vision. As wearable sensing, edge computing, physics-informed machine learning, and safe reinforcement learning continue to mature, AI-driven digital twins represent the most promising pathway towards truly personalized, proactive, and clinically effective sports injury prevention- a future in which every athlete benefits from a continuously watching, always-learning, individually tailored computational guardian of their musculoskeletal health.

Future work will involve the implementation and experimental validation of the proposed AthleTwin framework using real-world multimodal athlete data to evaluate its effectiveness for personalized injury prediction, prevention, and rehabilitation.

ACKNOWLEDGEMENT

The authors are thankful to GreyB Analytics Private Limited, India for funding this project.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] G. V. Gurau, G. Gurau, D. C. Voinescu, L. Anghel, G. Onose, D. A. Iordan, C. Munteanu, I. Onu, and C. L. Musat, "Epidemiology of injuries in men's professional and amateur football (Part I)," *J. Clin. Med.*, vol. 12, no. 17, Art. no. 5569, 2023, doi: 10.3390/jcm12175569.

[2] D. Hernandez, S. J. Dixon, N. F. Collier, and D. J. Saxby, "Epidemiology of sports-related injuries and associated risk factors in adolescent athletes: An injury surveillance," *Int. J. Environ. Res. Public Health*, vol. 18, no. 9, Art. no. 4857, 2021.

[3] *Global Sports Medicine Market Report*. Grand View Research, 2023. [Online]. Available: https://www.grandviewresearch.com/Filters?search=sports+medicine&search_submit=

[4] H. Van Eetvelde, L. D. Mendonça, C. Ley, R. Seil, and T. Tischer, "Machine learning methods in sport injury prediction and prevention: A systematic review," *J. Exp. Orthop.*, vol. 8, no. 1, Art. no. 27, 2021, doi: 10.1186/s40634-021-00346-x.

[5] A. Majumdar, R. Bakirov, D. Hodges, S. Scott, and T. Rees, "Machine learning for understanding and predicting injuries in football," *Sports Med. Open*, vol. 8, no. 1, Art. no. 73, 2022, doi: 10.1186/s40798-022-00465-4.

[6] M. Grieves, *Digital Twin: Manufacturing Excellence Through Virtual Factory Replication*, White Paper, 2015.

[7] P. Diniz, B. Grimm, F. Garcia, J. Fayad, C. Ley, C. Mouton, J. F. Oeding, M. T. Hirschmann, K. Samuelsson, and R. Seil, "Digital twin systems for musculoskeletal applications: A current concepts review," *Knee Surg. Sports Traumatol. Arthrosc.*, vol. 33, no. 5, pp. 1892-1910, 2025, doi: 10.1002/ksa.12627.

[8] J. Corral-Acero, F. Margara, M. Marciniak, C. Rodero, F. Loncaric, Y. Feng, A. Gilbert *et al.*, "The 'Digital Twin' to enable the vision of precision cardiology," *Eur. Heart J.*, vol. 41, no. 48, pp. 4556-4564, 2020, doi: 10.1093/eurheartj/ehaa159.

[9] F. C. Oettl, K. Samuelsson, and J. A. Pruneski, "The use of artificial intelligence and digital twins: Revolutionizing orthopedic surgery from joint arthroplasty to improving evaluation and treatment options," in *Handbook of Tissue Reconstruction and Regeneration*. Singapore: Springer Nature, 2026, pp. 1-18, doi: 10.1007/978-981-96-7448-0_40-2.

[10] M. Yin, "AI-driven medical image analysis for sports injury diagnosis and prevention," *Sci. Rep.*, vol. 15, no. 1, Art. no. 41484, 2025, doi: 10.1038/s41598-025-20580-y.

[11] M. Souaifi, W. Dhahbi, N. Jebabli, H. İ. Ceylan, M. Boujabli, R. I. Muntean, and I. Dergaa, "Artificial intelligence in sports biomechanics: A scoping review on wearable technology,

motion analysis, and injury prevention," *Bioengineering*, vol. 12, no. 8, Art. no. 887, 2025, doi: 10.3390/bioengineering12080887.

[12] S. Li, Z. Hou, K. Amjad, and H. Mushtaq, "Multi modal fusion of medical imaging and biomechanical data using attention based Swin-UNet and LSTM for sports injury prediction," *Front. Physiol.*, vol. 16, Art. no. 1687895, 2025, doi: 10.3389/fphys.2025.1687895.

[13] D. R. Seshadri, R. T. Li, J. E. Voos, J. R. Rowbottom, C. M. Alfes, C. A. Zorman, and C. K. Drummond, "Wearable sensors for monitoring the physiological and biochemical profile of the athlete," *NPJ Digit. Med.*, vol. 2, no. 1, Art. no. 72, 2019, doi: 10.1038/s41746-019-0150-9.

[14] A. Rebelo, D. V. Martinho, J. Valente-dos-Santos, M. J. Coelho-e-Silva, and D. S. Teixeira, "From data to action: A scoping review of wearable technologies and biomechanical assessments informing injury prevention strategies in sport," *BMC Sports Sci. Med. Rehabil.*, vol. 15, no. 1, Art. no. 169, 2023, doi: 10.1186/s13102-023-00783-4.

[15] H. Nobari, "Takes two to tango: Digital twins and AI revolutionize sports science and medicine," *Acta Kinesiologica*, vol. 18, no. 2, pp. 56-57, 2024.

[16] Barça Innovation Hub, "From the laboratory to the pitch: Barça test digital twins to anticipate injuries," FC Barcelona, 2025. [Online]. Available: https://barcainnovationhub.fcbarcelona.com/blog/from-the-laboratory-to-the-pitch-barca-test-digital-twins-to-anticipate-injuries/

[17] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems," *J. Comput. Phys.*, vol. 378, pp. 686-707, 2019, doi: 10.1016/j.jcp.2018.10.045.

[18] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748-1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[20] A. Čartolovni, A. Tomičić, and E. Lazić Mosler, "Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review," *Int. J. Med. Inform.*, vol. 161, Art. no. 104738, 2022, doi: 10.1016/j.ijmedinf.2022.104738.

[21] Tata Consultancy Services (TCS), "The Future Athlete Project." [Online]. Available: https://www.tcs.com/who-we-are/sports-sponsorships/digital-twins-transforming-potential-into-performance