PPG-Sport: A Dataset for Reliable Heart Rate Monitoring from Wrist PPG Under Dynamic Sports Conditions

CHANGSHUO HU, Singapore Management University, Singapore
HUNG MANH PHAM, Singapore Management University, Singapore
YIMING ZHANG, Singapore Management University, Singapore
GUANRU YAN, Singapore Management University, Singapore
XIAO MA, Singapore Management University, Singapore
YUEZHONG WU, Fuzhou University, China
THIVYA KANDAPPU, Singapore Management University, Singapore
ARCHAN MISRA, Singapore Management University, Singapore
DONG MA*, University of Cambridge, United Kingdom

Photoplethysmography (PPG) has become a cornerstone of physiological sensing in wearable devices, enabling non-invasive monitoring of heart rate and related biomarkers. However, its reliability deteriorates sharply under dynamic, high-intensity, or non-periodic motions such as those in sports, where existing datasets fail to capture realistic wrist dynamics. To address this gap, we introduce PPG-Sport, the first large-scale dataset designed for heart rate monitoring from wrist-worn PPG under real sports conditions. The PPG-Sport dataset includes synchronized PPG, inertial measurement unit (IMU), and electrocardiography (ECG) recordings from both wrists of 30 participants across six representative activities: stationary, walking, running, badminton, table tennis, and basketball, amounting to 48 hours of multimodal data. PPG-Sport uniquely captures three critical properties absent in prior datasets: (1) non-periodicity, reflecting irregular and broadband motion patterns; (2) high intensity, with frequent, large-magnitude accelerations; and (3) bilateral asymmetry, caused by distinct functional roles of the dominant and non-dominant hands during sports. We further establish a deep-learning-based benchmark that combines both temporal and spectral representations of PPG and IMU signals to evaluate heart rate estimation performance. Experimental results show that models trained only on conventional periodic activities fail drastically in sports scenarios. Although incorporating sports data mitigates the degradation, significant errors remain. These findings highlight PPG-Sport as an essential and challenging benchmark for developing motion-robust physiological sensing algorithms during sports activities. Dataset and benchmark code are available at: https://github.com/LaserHu/PPG-Sport.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

*Corresponding author

Authors' Contact Information: Changshuo Hu, Singapore Management University, Singapore, Singapore, cs.hu.2023@phdcs.smu.edu.sg; Hung Manh Pham, Singapore Management University, Singapore, Singapore, hm.pham.2023@phdcs.smu.edu.sg; Yiming Zhang, Singapore Management University, Singapore, Singapore, yimingzhang@smu.edu.sg; Guanru Yan, Singapore Management University, Singapore, Singapore, guanru.yan.2023@mitb.smu.edu.sg; Xiao Ma, Singapore Management University, Singapore, Singapore, xiaoma.2022@phdcs.smu.edu.sg; Yuezhong Wu, Fuzhou University, Fuzhou, China, yuezhong.wu@fzu.edu.cn; Thivya Kandappu, Singapore Management University, Singapore, Singapore, thivyak@smu.edu.sg; Archan Misra, Singapore Management University, Singapore, Singapore, archanm@smu.edu.sg; Dong Ma (corresponding author), University of Cambridge, Cambridge, United Kingdom, dm878@cam.ac.uk.

ACM 2474-9567/2026/6-ART43
https://doi.org/10.1145/3810216

ACM Reference Format:
Changshuo Hu, Hung Manh Pham, Yiming Zhang, Guanru Yan, Xiao Ma, Yuezhong Wu, Thivya Kandappu, Archan Misra, and Dong Ma. 2026. PPG-Sport: A Dataset for Reliable Heart Rate Monitoring from Wrist PPG Under Dynamic SportsConditions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 2, Article 43 (June 2026), 28 pages. https://doi.org/10. 1145/3810216

1 INTRODUCTION

Photoplethysmography (PPG) has become one of the most fundamental sensing modalities for non-invasive physiological monitoring, owing to its convenience and low cost. It enables the estimation of key vital signs, including heart rate (HR) [46], respiration rate [18], and blood oxygen saturation [15]. Consequently, PPG has been widely adopted in commercial wearable devices, such as smartwatches (e.g., Apple Watch [1]), fitness trackers (e.g., Garmin series [8]), and even wireless earbuds (e.g., Powerbeats Pro 2 [10]). PPG has demonstrated remarkable performance in HR estimation under low-intensity daily scenarios, such as resting, walking, or light jogging, where stable skin contact and limited motion allow reliable signal acquisition. However, it continues to face challenges under intense or dynamic conditions [2]. Major manufacturers, including Apple and Huawei, officially acknowledge that their PPG sensors perform optimally during rhythmic exercises but may yield unreliable readings in irregular or high-impact activities like tennis or basketball [5, 6]. Consistent with these reports, for instance, a Reddit user documented a HR discrepancy of up to 60 beats per minute between a Garmin smartwatch and a chest strap during a squash game [3]. Despite these challenges, in sports scenarios, accurate HR monitoring with PPG is highly valuable for users, as it provides critical insights into training load, recovery status, and potential health risks [24, 40].

The poor PPG sensing performance during sports stems primarily from the distinct motion dynamics that differ fundamentally from the typical routine exercises like walking or running. Specifically, during walking or running, wrist movements are *rhythmic* and *repetitive*, following a well-defined motion cycle that produces consistent kinematic patterns and a narrowband motion energy spectrum. In contrast, sports such as badminton, table tennis, and basketball involve *diverse*, *non-repetitive* actions with rapid swings, abrupt direction changes, and alternating hand coordination, breaking this regularity. These complex motions generate trajectories with high variability in both amplitude and duration, leading to irregular temporal structures and broadband frequency components. Consequently, the corresponding PPG signals exhibit dispersed and asymmetric spectral characteristics, in stark contrast to the clean, periodic waveforms observed in walking or running. To quantitatively characterize this notion of motion non-periodicity, we later introduce the *Harmonic Spectral Concentration Ratio* (HSCR), which measures how strongly the IMU spectral energy concentrates around the dominant frequency and its harmonics (see Section 4.2).

These distinct motion patterns pose fundamental challenges to existing PPG-based heart rate estimation algorithms. Most existing methods, whether handcrafted or data-driven, are designed and trained under low-intensity, cyclic conditions. Therefore, they implicitly rely on the signal's frequency-domain regularity [39, 63]. Their core principle involves identifying and removing motion-induced frequency components that overlap or interfere with the cardiac rhythm. While effective for rhythmic activities like walking or running, this strategy fails when motion becomes irregular and broadband as in sports. In such scenarios, the interference cannot be mitigated through simple spectral filtering, since the relationship between motion and PPG distortion is highly nonlinear and time-varying. Consequently, more advanced algorithms and models are required to robustly handle complex motion artifacts and achieve accurate heart rate estimation.

Achieving this goal requires a PPG dataset that comprehensively captures diverse motion patterns inherent in sports activities. However, existing public PPG datasets are inadequate for this purpose. As summarized in Table 1, clinical datasets, such as MIMIC-III [45] and PulseDB [59], were collected in controlled hospital environments using fingertip pulse oximeters, producing high-fidelity signals with minimal motion artifacts, rendering them unsuitable for studying motion-related distortions. Laboratory datasets, including WESAD [56] and TROIKA [63], shift data collection to the wrist and introduce structured activities such as walking or running. Nevertheless, these repetitive motion patterns fail to represent the irregular and high-intensity dynamics of sports. Daily activity datasets, such as PPG-DaLiA [52] and WildPPG [42], further extend to naturalistic routines like working, eating, driving, or commuting. However, these activities involve only light to mild wrist movements, far from the rapid and multidirectional motion found in sports. Consequently, no existing dataset captures the *intensive*, *dynamic*, and *irregular* motion diversity necessary to evaluate and improve PPG reliability in genuine sports scenarios. Moreover, the lack of standardized benchmarks and publicly available code further hinders fair comparison and reproducible evaluation of heart rate estimation algorithms under such complex motion conditions.

To bridge this gap, we present the PPG-Sport dataset, a first collection specifically designed for wrist-based heart rate sensing in sports contexts. The data set comprises synchronized PPG, inertial (IMU), and ECG signals from 30 participants, with ECG providing the heart rate ground truth. Each participant wore two Empatica Embrace Plus watches on both wrists and performed six representative activities: badminton, table tennis, basketball, walking, running, and stationary, which together capture a broad spectrum of motion conditions. The three sports were selected to represent highly dynamic and non-repetitive upper-limb movements, where the arms exhibit large motion amplitudes, rapid swings, and diverse action patterns involving sudden changes in direction and intensity. Walking and running sessions were included to ensure continuity and comparability with prior studies focused on periodic locomotion, enabling fair cross-activity evaluation under identical sensor setups. Finally, the stationary condition serves as a clean baseline, establishing an upper bound of PPG sensing accuracy in the absence of motion artifacts. Along with the dataset, we provide a unified benchmark framework with released code to facilitate reproducible evaluation and comparison of heart rate estimation methods across different motion regimes (Link to the code repository will be added upon acceptance).

The PPG-Sport dataset captures three key properties that distinguish it from previous collections: (1) **Non-periodicity**: Sports activities generate dispersed and irregular spectral distributions rather than narrowband rhythmic patterns, reflecting the complex and highly variable limb trajectories observed in real sport movements; (2) **High intensity**: In sports, wrist undergoes frequent, large-magnitude accelerations with rapid transitions in both direction and amplitude, introducing substantial motion artifacts that challenge the stability of optical sensing and expose algorithms to a broader range of signal degradations; and (3) **Bilateral asymmetry**: The distinct functional roles of the dominant and non-dominant hands in sports (e.g., striking vs. balancing in table tennis and dribbling versus guarding in basketball) produce markedly different motion and distortion patterns on each wrist. This provides valuable insights for developing robust algorithms that generalize across wearing sides.

In summary, this work makes three key contributions:

- We introduce PPG-Sport, the first-of-its-kind large-scale dataset for wrist-based heart rate sensing in real-world sports scenarios, featuring synchronized PPG, IMU, and ECG recordings from both wrists across 30 participants and six diverse activities. This dataset is tailored for studying PPG sensing in non-periodic and high-intensity sports. The dataset is available at https://github.com/LaserHu/PPG-Sport.
- We develop and release a multimodal benchmark framework with open-source code that jointly leverages temporal and spectral features from PPG and IMU signals. This framework establishes a reproducible baseline for heart rate estimation on PPG-Sport, enabling fair comparison across methods under complex motion conditions.
- Through extensive experiments, we demonstrate that models trained solely on conventional periodic activities fail to generalize to highly dynamic sports. Even incorporating the sports data, performance degradation persists, underscoring the need for more advanced, motion-resilient HR estimation algorithms.

2 RELATED WORKS

2.1 PPG Applications and Algorithms

Photoplethysmography (PPG) is one of the most widely adopted sensing modalities in wearable devices due to its compact form factor, low power consumption, and rich hemodynamic information. A single PPG signal encodes multiple aspects of cardiovascular and autonomic function, supporting a broad spectrum of downstream tasks, including heart rate (HR) and heart rate variability (HRV) estimation [46], blood oxygen saturation (SpO_2) [15][16], respiratory rate (RR) detection [18], blood pressure (BP) estimation [25][62], vascular stiffness analysis [27], arrhythmia detection [48], as well as stress, emotion, and sleep stage recognition [16]. Across these applications, the shared objective is to learn a stable mapping from PPG waveforms to target physiological indicators that generalizes across users, devices, and real-world conditions [16].

Early PPG studies relied on handcrafted time-, morphology-, and frequency-domain features [16], combined with statistical or classical machine-learning models such as linear regression, SVMs, and random forests [11][25][48]. While interpretable and computationally efficient, these approaches struggled to capture complex temporal dynamics. Recent advances have shifted PPG analysis toward end-to-end deep learning and multimodal fusion [47], where CNNs, RNNs, and Transformer-based models directly learn latent representations from raw signals. The integration of complementary modalities such as IMUs and ECG improves robustness under motion-intensive scenarios [52], while attention mechanisms, self-supervised pretraining [22], and domain adaptation [13] enhance cross-subject and cross-device generalization. Combined with lightweight architectures and deployment-aware optimization [62], deep multimodal pipelines have consistently outperformed traditional methods in both robustness and transferability across daily-life and exercise benchmarks [52][13].

2.2 Reliable PPG Monitoring in Real-world Scenarios

The optical nature of PPG makes it inherently sensitive to external perturbations and individual physiological variability, leading to degraded reliability in real-world settings. Key factors include variations in contact pressure that alter optical coupling and vascular perfusion [54][55][30]; motion artifacts (MA) that induce amplitude instability, baseline drift, and waveform distortion [47][54]; sensor placement and wearing configuration (e.g., site, wrist dominance, strap tightness, curvature), which modulate both optical path length and motion coupling [16][54]; and inter-individual and environmental variability such as skin properties, perfusion level, temperature, and sweat [54]. These factors collectively distort waveform morphology and reduce signal-to-noise ratio (SNR), thereby constraining the accuracy and generalizability of downstream physiological estimation models.

Among them, MA are the dominant source of performance degradation during daily activities and sports. MAs arise from skin–sensor displacement, soft-tissue deformation, wrist rotation, and abrupt acceleration or impact, with spectral components (0.1–10 Hz) that substantially overlap the cardiac band (0.5–4 Hz), rendering conventional filtering ineffective. While periodic activities (e.g., running or cycling) generate quasi-harmonic artifacts that are partially predictable, non-periodic, high-intensity, and multidirectional motions (e.g., ball sports or resistance training) introduce irregular distortions that challenge both traditional signal processing and modern deep models [54][47]. To mitigate these effects, prior work has explored algorithmic and hardware co-design strategies, including accelerometer-assisted adaptive filtering [52], ICA-based separation, signal decomposition and reconstruction methods such as EEMD, SSA [63], and joint sparse spectral modeling [58], as well as deep multimodal PPG–IMU frameworks [52][12] and artifact-aware or self-supervised learning objectives [22]. On the hardware side, multi-channel or multi-wavelength sensors [37], adjustable contact-pressure mechanisms [30], and flexible multi-point sensor arrays [17] have been proposed to improve mechanical stability and motion robustness. Despite these advances, most evaluations remain focused on periodic and structured activities [52][63], leaving non-periodic, high-intensity, and multidirectional motions underexplored [54].

2.3 PPG Datasets

Table 1 summarizes representative publicly available PPG datasets in terms of release year, sensing modalities, participant scale, activity types, motion non-periodicity and intensity, and sensor placement. These datasets have been widely used to benchmark PPG algorithms under different motion conditions. For clarity, we categorize

Table 1. Summary of representative PPG datasets. Non-Periodicity specifically refers to motion-induced non-periodicity, categorized as periodic motion (×), non-periodic motion (✓), or not applicable (NA) for datasets without explicit movement. Intensity reflects the level of hand movement and impact intensity associated with the activities, and Placements indicate sensor wearing locations.

Dataset	Year	Modalities / Sensors	#Part.	Activity Types	Non-Periodicity	Intensity	Placements
Clinical Datasets							
BIDMC [50]	2017	PPG, ECG	32	Resting	NA	Low	Finger
PPG-BP [38]	2018	PPG, ECG	219	Resting	NA	Low	Finger
MIMIC-III [45]	2022	PPG, ECG, ABP, RESP	5596	Resting	NA	Low	Finger
VitalDB [36]	2022	PPG, ECG	5866	Resting	NA	Low	Finger
Aurora [43]	2022	PPG, ECG, IMU, Tonometry	1125	Resting	NA	Low	Non-dominant Wrist
PulseDB [59]	2023	PPG, ECG, ABP	5381	Resting	NA	Low	Finger
Laboratory Datasets							
SDB [29]	2014	PPG, ECG, EEG	146	Sleep	NA	Low	Finger
TROIKA [63]	2014	PPG, ECG, IMU	12	Walking, Running	×	Medium	Single-Wrist
NIBP [26]	2017	PPG, ECG, FSR, PCG	32	Running	×	Medium	Finger
BAMI [20]	2018	PPG, ECG, IMU	24	Resting, Walking, Running	×	Medium	Single-Wrist
WESAD [56]	2018	PPG, ECG, IMU, EDA, EMG	15	Amusement, Stress, Meditation, Recovery	NA	Low	Non-dominant Wrist
MESA [61]	2018	PPG, ECG, EEG	2056	Sleeping	NA	Low	Finger
Graphene-HGCPT [34]	2022	PPG, Bio-Z	7	Sedentary Activities	NA	Low	Finger
PTT-PPG [41]	2022	PPG, ECG, IMU, TEMP	22	Resting, Walking, Running	×	Medium	Fingers
EarSet [44]	2023	PPG, IMU	30	Walking, Running	×	Medium	Ear
Lokomat-PPG [21]	2024	PPG, IMU	46	Robot-Assisted Gait Rehab	✓	Medium	Non-dominant Wrist
ECSMP [28]	2024	PPG, ECG, IMU, EEG, TEMP	89	Resting, Cognitive Assessments	✓	Low	Non-dominant Wrist, Ear
WF-PPG [30]	2025	PPG, ECG	27	Sedentary Activities	NA	Low	Dominant Wrist, Finger
UTSA-PPG [60]	2025	PPG, ECG, IMU	12	Sitting, Sleeping, Walking	×	Low	Non-dominant Wrist, Finger
Daily Activity Datasets							
PPG-DaLiA [52]	2019	PPG, ECG, IMU	15	Resting, Walking, Stairs, Driving, Cycling, Soccer	✓	Medium	Non-dominant Wrist
WildPPG [42]	2024	PPG, ECG, IMU, TEMP	16	Walking, Hiking, Stair Climbing, Eating, Resting	✓	Medium	Dominant Wrist, Ankle, Sternum, Head
PPG-Sport (Ours)	**2025**	**PPG, ECG, IMU**	**30**	**Resting, Walking, Running, Badminton, Table Tennis, Basketball**	✓	**High**	**Dual Wrist**

existing datasets into three groups according to their recording environments and motion characteristics: (1) *clinical datasets*, (2) *laboratory datasets*, and (3) *daily activity datasets*, forming a continuum from stationary, highly controlled settings to increasingly naturalistic scenarios.

Clinical datasets. Early datasets such as MIMIC-III [45], VitalDB [36], and PulseDB [59] were collected in hospitals or intensive care units, typically using fingertip pulse oximeters that provide high-fidelity PPG signals with minimal motion interference. These datasets are large in scale and include synchronized physiological signals such as PPG, ECG, arterial blood pressure (ABP), and respiration (RESP), and have served as the foundation for studies on cuffless blood pressure estimation and waveform transformation. However, their resting or low-motion recording conditions and medical-grade form factors make them unsuitable for studying motion artifacts or wearable sensing in everyday environments.

Laboratory datasets. With the adoption of wrist-worn devices, laboratory datasets such as WESAD [56], TROIKA [63], and PTT-PPG [41] extended clinical settings by introducing structured activities like walking or running under controlled indoor conditions. These datasets often combine PPG with accelerometer and ECG signals, enabling evaluation under mild motion artifacts. Nevertheless, the activities are typically short-term and highly periodic, emphasizing reproducible motion patterns rather than natural variability, and mainly address the transition from static to mildly dynamic scenarios.

Daily activity datasets. More recent datasets, including PPG-DaLiA [52] and WildPPG [42], move toward free-living data collection and capture longer-term daily routines such as sitting, eating, working, or cycling. While these datasets improve ecological validity, the recorded motions remain predominantly low-to-moderate in intensity and largely regular in rhythm, falling short of representing the irregular, asymmetric, and high-impact movements common in vigorous sports or non-repetitive upper-limb activities.

Limitations of existing datasets. Overall, existing PPG datasets exhibit limited motion diversity. Clinical datasets focus on resting fingertip measurements, while laboratory and daily activity datasets shift to wrist-worn sensing but still concentrate on periodic or low-intensity movements. Moreover, most datasets record from a single wrist, neglecting bilateral motion asymmetry that naturally arises in sports and daily tasks. In contrast, the proposed *PPG-Sport* dataset systematically captures bilateral wrist PPG together with synchronized ECG and IMU signals across diverse sports, including badminton, basketball, and table tennis. By covering complex, non-periodic, and high-intensity motions in realistic settings, *PPG-Sport* provides a unique benchmark for developing and evaluating motion-robust physiological sensing algorithms in sports scenarios.

3 PPG-SPORT DATASET

As mentioned earlier, our primary goal is to construct a dataset that supports the development of algorithms capable of reliable heart rate estimation under realistic sports conditions. These scenarios are dominated by non-periodic and high-intensity body movements that introduce irregular baseline shifts, signal dropouts, and severe motion artifacts—conditions under which conventional models often fail. To emphasize this challenge, we focus primarily on three complex sports (badminton, table tennis, and basketball) that feature rapid, asymmetric, and unstructured wrist motion. In addition, we include two periodic reference activities, walking and running, as well as a stationary condition for fair comparison and cross-dataset compatibility. Together, these six activities cover a broad spectrum of motion intensity, rhythm, and upper-limb involvement.

3.1 Activity selection

3.1.1 Badminton, Table Tennis, and Basketball. These three sports were chosen to represent non-periodic and high-intensity activities that introduce complex and irregular wrist motion. They align with our design objectives for two main reasons. First, these sports involve rapid, asymmetric, and unstructured upper-limb movements such as swinging, striking, dribbling, or throwing, which generate substantial and unpredictable motion artifacts in wrist-worn signals. Second, they typically cause large and rapid heart-rate fluctuations [23], making HR monitoring particularly valuable for evaluating exercise intensity and physiological response during dynamic sports. Below, we briefly describe the motion features of each sport.

Badminton. Badminton features frequent, large-amplitude swings of the dominant arm, including forehand and backhand drives, drop shots, smashes, clears, and deceptive flicks. These strokes vary greatly in speed and direction, producing sharp, non-repetitive wrist accelerations. The non-dominant arm mainly aids balance and body rotation, thereby generating moderate but continuous movement. Players also move rapidly across the court, combining jumping, lunging, and short-distance sprints. This combination of arm-swing and whole-body motion layers multiple sources of movement, resulting in highly dynamic and irregular signal patterns [57].

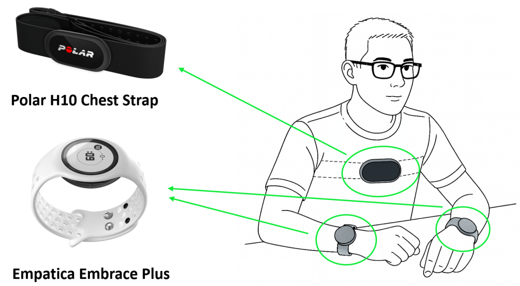

Fig. 1. Hardware setup for the PPG-Sport dataset collection. Each participant wore two Empatica Embrace Plus devices on both wrists and a Polar H10 chest strap around the lower chest.

Table Tennis. Compared with badminton, table tennis is performed within a smaller area but at a much faster tempo. Players execute short and powerful strokes such as topspin drives, blocks, flicks, and chops, often alternating between forehand and backhand within fractions of a second. The quick switching of stroke type and direction, together with continuous micro-adjustments of posture and balance, yields frequent bursts of high-frequency wrist motion superimposed on subtle lateral body movement. Overall, table-tennis actions exhibit high temporal variability but limited spatial displacement, forming a distinct non-periodic motion pattern compared to other sports [19].

Basketball. Basketball differs markedly from racket sports. Both hands are actively involved, alternating between dribbling, passing, and shooting; however, the ratio of dominant to non-dominant hand use varies across players. Notably, a large portion of play time also involves off-ball motion, including running, defending, and rebounding. These alternating phases of fine hand control and full-body acceleration occur over a large court area and feature frequent rhythm changes, from slow positional play to rapid transitions or sudden direction shifts [35].

3.1.2 Walking and Running. Walking and running are incorporated as periodic reference activities, providing a bridge between the resting state and high-intensity sports. While similar data exist in prior PPG datasets [42, 52], their experiment setups, sensors, and participant populations differ considerably from ours. Including these activities ensures compatibility for cross-dataset evaluation and allows a fair comparison of periodic versus non-periodic motion effects on both PPG morphology and HR estimation performance under consistent conditions.

3.1.3 Stationary. Although stationary poses do not involve physical movement, they are included as a necessary reference condition. They represent the upper bound of PPG signal quality achievable under our experimental setup and hardware configuration. The recordings collected during stationary sessions serve as a baseline for quantifying the intrinsic performance limit of the device, independent of motion interference.

3.2 Hardware Setup

To collect synchronized multimodal physiological and motion data, we employed a sensing platform composed of two wrist-worn smartwatches and one chest-worn reference strap. Each participant wore a pair of off-the-shelf Empatica EmbracePlus [4], one on each wrist, to simultaneously record PPG and IMU signals.

As illustrated in Figure 1, both smartwatches were positioned on the outer side of the wrists, approximately one finger's width above the wrist bone, consistent with typical consumer usage. Participants adjusted the strap tightness to ensure stable skin contact while maintaining comfort, providing reliable optical coupling for the PPG sensors. The IMU embedded in the smartwatch follows a dominant-handed coordinate system, where the x-axis points toward the fingers, the y-axis points outward from the palm, and the z-axis points upward, perpendicular to the back of the hand. In addition, a Polar H10 chest strap [9] was used to collect ECG signals, serving as the ground truth for heart-rate estimation. The Polar H10 was worn horizontally around the lower chest, just below the pectoral muscles, following the manufacturer's recommended placement to maintain firm electrode contact and minimize motion-induced noise.

The Empatica EmbracePlus watch recorded PPG and triaxial IMU signals at 64 Hz, whereas the Polar H10 chest strap recorded ECG at 130 Hz. Data were captured using dedicated iOS applications: the Empatica Care Lab App for the smartwatches and ECGLogger for the chest strap. After each session, recorded files were retrieved from the Empatica cloud server and from the ECGLogger app on the iPhone for offline processing. Although the devices were not hardware-synchronized, each recording contained precise timestamps generated by its companion app. These timestamps enabled accurate temporal alignment across modalities during subsequent data processing.

3.3 Experimental Protocol and Environment

The data collection was conducted in realistic sports environments to capture physiological and motion signals under conditions close to everyday physical exercise. Rather than a fully controlled laboratory study, we intentionally allowed participants to perform each activity in their natural manner and at self-selected intensity levels. As illustrated in Figure 2, the six activities were recorded in two on-campus locations covering both outdoor and indoor environments. The first three activities, stationary, walking, and running, were performed on the outdoor athletics track, while the remaining three, badminton, table tennis, and basketball, were conducted in a multi-purpose indoor sports hall equipped with standard badminton courts, table tennis tables, and a half basketball court.

For the stationary activity, participants sat quietly on a bench beside the track with both wrists relaxed and placed naturally in a comfortable position. During walking and running, participants moved around the track at their preferred pace and stride length, maintaining a steady rhythm. For the non-periodic sports, a skilled partner assisted each participant to ensure realistic gameplay and consistent motion engagement. In badminton and table tennis, over 80% of the session time was spent rallying, with short breaks for picking up balls. In basketball, participants spent the majority of their time dribbling and shooting, interleaved with brief off-ball movements such as defense or positioning. All participants used their own athletic apparel and shoes, but were provided with standardized rackets and basketballs for consistency. Before each session, participants were equipped with all sensors as described in Section 3.2. Each participant completed all six activities in sequence, with 8 minutes per activity and 2-minute rest intervals between sessions to allow sufficient recovery. The principal investigator (PI) annotated the start and end times of each activity in real time, enabling accurate segmentation of activity-specific data based on timestamps during post-processing. Participants could request longer breaks when needed.

3.4 Ground Truth Reference

In PPG-Sport, we use chest-worn ECG signals recorded by the Polar H10 as the reference for heart rate estimation. The Polar H10 has been widely adopted in prior wearable sensing and physiological monitoring studies [14, 30, 31] as a practical reference device for heart rate measurement, particularly in mobile and exercise settings. For activities such as stationary sitting, walking, and running, which involve limited upper-body motion, the reliability of chest ECG has been well established in prior work and was also consistently observed in our recordings. However, during non-periodic sports such as badminton, table tennis, and basketball, large and rapid upper-limb movements can induce secondary chest wall motion through torso rotation and arm–shoulder coupling. These movements

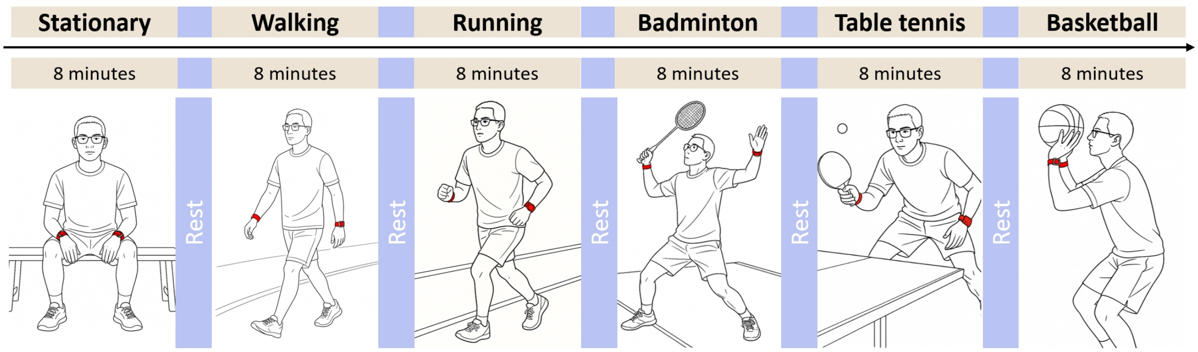

Fig. 2. Overview of the data collection protocol illustrating the six recorded activities and their durations.

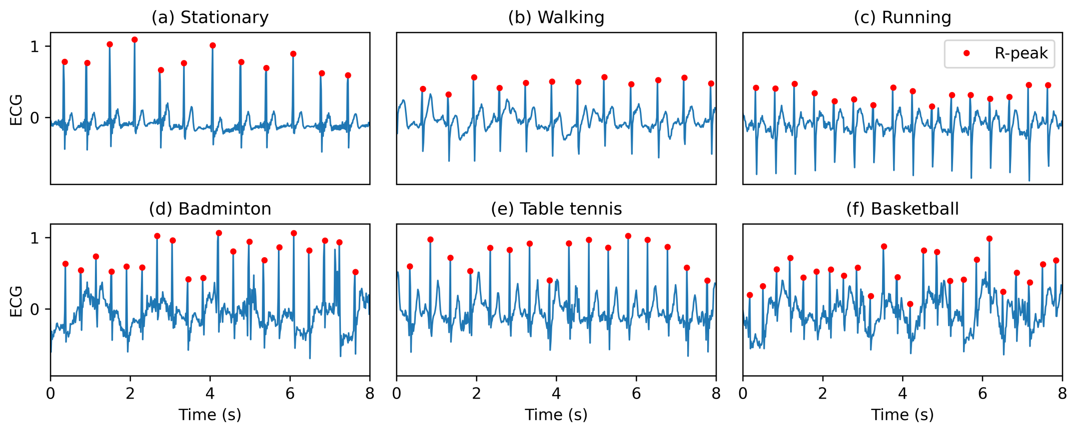

Fig. 3. Chest ECG ground-truth signals from Polar H10 and detected heartbeat R-peaks across six activities.

may intermittently affect the electrode–skin contact of the chest strap, leading to visible waveform deformation or baseline fluctuation in the ECG signal.

Importantly, our use of ECG as ground truth focuses on deriving heart rate via heartbeat R-peak timing rather than on precise ECG waveform morphology. While upper-body motion can distort the ECG waveform shape, the temporal structure of R-peaks remains largely preserved in our data. Figure 3 presents ECG signals from a representative participant across all six activities, together with the detected heartbeat R-peaks. From Figure 3, we observe that stationary ECG signals exhibit near-ideal morphology, while walking and running introduce only minor variations. For the three non-periodic sports, waveform deformation becomes more pronounced due to intensive arm and torso motion; nevertheless, the R-peaks remain clearly identifiable, enabling reliable heart rate derivation even under vigorous activity. Therefore, despite potential waveform distortion during high-impact sports, chest ECG recorded by a strap-based device remains a reasonable and sufficiently robust reference for heart rate estimation in realistic sports scenarios. We acknowledge that ECG is not entirely immune to motion-induced artifacts, but it provides the most reliable reference under these conditions.

3.5 Participants

Thirty healthy volunteers (16 males and 14 females) participated in the data collection. Participants' ages ranged from 21 to 34 years (27 ± 3.2 years, mean ± SD), with a mean height of 172 ± 6.8 cm and a mean weight of 65.1 ± 13.3 kg, respectively. All participants reported no known cardiovascular or musculoskeletal disorders and were capable of safely performing moderate- to high-intensity exercise.

To characterize inter-participant diversity, we collected each participant's self-reported experience level for the three non-periodic sports—badminton, table tennis, and basketball. The experience was rated on a three-point scale: novice (little or no prior experience), intermediate (occasional recreational play), or experienced (regular practice or competition). Across all three sports, most participants reported intermediate familiarity (approximately 50%), with the remaining participants split between novice (20–30%) and experienced (20–30%) levels. Reporting these experience levels helps contextualize variations in motion smoothness and exertion intensity that may influence both IMU and PPG signal quality, and also enables future subgroup analyses based on skill level.

All participants were recruited from the university community, including both undergraduate and graduate students. They wore their own comfortable athletic clothing and shoes during the experiment and could rest or withdraw at any time if discomfort occurred. Before participation, each individual was informed about the study procedure and provided written consent. The experimental protocol was reviewed and approved by the Institutional Review Board (IRB) of the authors' institution, ensuring compliance with ethical research standards. Detailed demographic and self-reported experience information for all participants is provided in the dataset's accompanying file `participants.csv`.

3.6 Data Processing and Organization

After the data collection, all raw signals were downloaded and processed offline following a unified pipeline, including (1) data parsing and extraction, (2) timestamp-based alignment and trimming, (3) signal resampling, and (4) standardized organization into the final repository structure.

3.6.1 Data Extraction and Preprocessing. The wrist-worn recordings were retrieved from the Empatica cloud server, where each session was automatically stored as an `.avro` file containing timestamps, PPG samples, and IMU channels. We parsed these Avro files to extract relevant fields and saved them as structured `.csv` files while preserving the original timestamps for subsequent alignment. The ECG recordings from the Polar H10 were directly exported from its companion iOS application in CSV format.

As described in Section 3.3, the start time of each activity was recorded during the experiment. Using these timestamps, we precisely extracted the 8-minute segments corresponding to each activity from the continuous recordings. This trimming procedure ensured temporal alignment across modalities (PPG, IMU, and ECG) within the same duration while retaining the original sampling resolution and timing accuracy of each device.

3.6.2 Signal Resampling. To facilitate synchronized analysis, we resampled the ECG data from the Polar H10 (originally at 130 Hz) to match the 64 Hz sampling rate of the Empatica EmbracePlus devices. The resampling was performed based on the recorded timestamps using linear interpolation to achieve uniform temporal spacing after alignment. Each activity segment was then saved as a NumPy file containing exactly *64 (Hz)* × *60 (seconds)* × *8 (minutes)* = *30,720 samples* per signal channel. For reproducibility, both the processed `.npy` files and the original `.csv` files (parsed directly from Avro data) are included in the released dataset.

3.6.3 Repository Structure and File Naming. All processed data are organized hierarchically by participant and activity. Each participant has a folder named with a subject ID (*S01*, *S02*, ..., *S30*), and within each folder, data are grouped by activity. Every activity folder contains synchronized recordings from both wrists (dominant and non-dominant) along with a single ECG reference. The file naming pattern is summarized in Table 2. The dataset and accompanying preprocessing scripts are publicly available at our anonymous repository[1] for reproducibility and benchmarking purposes.

As most participants were right-handed, the dominant wrist is defined as the right wrist by default, while for left-handed participants, recordings from the left wrist are consistently designated as the dominant wrist. To maintain a unified IMU coordinate convention across participants, the x- and y-axes of the IMU signals from left-handed participants are inverted, while the z-axis is kept unchanged.

For example, S10N_badminton_ppg.npy represents the non-dominant wrist PPG recording of participant 10 during the badminton session, while S20_running_ecg.npy corresponds to the chest-strap ECG signal of participant 20 during running. This structure allows each file to be uniquely identified by subject, wrist, activity, and signal modality, ensuring convenient access for model training or analysis.

Table 2. File organization and naming convention.

Modality	**File Name Pattern**
Non-dominant wrist PPG	SXXN_<activity>_ppg.npy
Non-dominant wrist IMU	SXXN_<activity>_imu.npy
Dominant wrist PPG	SXXD_<activity>_ppg.npy
Dominant wrist IMU	SXXD_<activity>_imu.npy
Chest ECG (reference)	SXX_<activity>_ecg.npy

3.6.4 Dataset Statistics and Supplementary Files. The final dataset contains synchronized multimodal recordings from 30 participants performing six activities of 8 minutes each. In total, this corresponds to *30 (participants)* × *6 (activities)* × *8 (minutes)* = *1,440 minutes (or 24 hours)* of ECG reference data, and twice that amount for each wrist modality (PPG and IMU), yielding approximately *48 hours* of recordings per signal type from both the dominant and non-dominant wrists.

Participant demographic attributes are provided in a separate file, participants.csv, which lists each participant's gender, age, height, weight, and familiarity levels with badminton, table tennis, and basketball.

4 PRELIMINARY VISUALIZATION AND DATA CHARACTERISTICS

4.1 Signal Overview

To provide an intuitive sense of how different activities manifest in physiological and motion signals, we visualize the chest ECG reference, dominant wrist PPG, and IMU recordings in the time domain. Figure 4 presents 8-second segments of synchronized signals, where the IMU magnitude is computed as the vector norm of the three acceleration axes. As shown in the figure, the three modalities exhibit distinct patterns across activity types. In (a) Stationary, both PPG and ECG show highly regular waveforms, with the PPG closely aligned with the ECG-derived ground truth, indicating clear cardiac coupling under minimal motion. As the activity transitions to Walking and Running, rhythmic oscillations emerge in both PPG and IMU signals with increasing amplitude, consistent with regular arm swings. Interestingly, although Running produces larger acceleration magnitudes, the PPG waveform remains relatively stable because the arm moves primarily in a plane perpendicular to the optical sensor's contact surface, reducing direct mechanical deformation at the skin interface. For the three vigorous and non-repetitive sports, Badminton, Table Tennis, and Basketball, the IMU signals show sharp, irregular bursts, while the PPG waveforms become heavily distorted and highly irregular. These observations illustrate the wide

[1]https://github.com/LaserHu/PPG-Sport

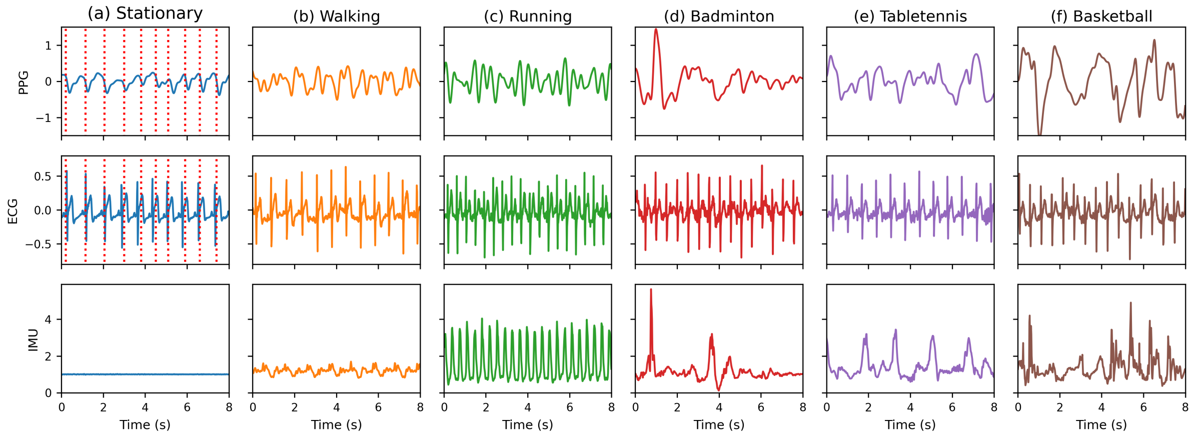

Fig. 4. Time-domain visualization of dominant wrist PPG, ECG and IMU signals across six activities.

range of motion intensity and signal variability captured in PPG-Sport, which forms the basis for the following analyses on motion periodicity, intensity, and symmetry.

4.2 Motion Periodicity

Figure 5 illustrates the frequency-domain representations of PPG, ECG, and IMU signals across six representative activities, computed over 120-second windows to capture long-term temporal dynamics. For *Stationary*, the IMU spectrum is nearly flat, while both PPG and ECG exhibit a distinct heartbeat peak (highlighted in red), showing strong spectral alignment between the two physiological signals. In *Walking* and *Running*, the IMU energy is sharply concentrated around a single dominant frequency and its harmonics, reflecting highly periodic arm swings. After conceptually removing these motion-related components from the PPG spectrum, the remaining prominent peak still aligns with the ECG-derived heart rate, suggesting that cardiac frequency remains separable under regular, rhythmic locomotion. In contrast, for *Badminton*, *Table Tennis*, and *Basketball*, the IMU spectra become widely dispersed, implying rapid transitions between gestures with varying swing rates and impact intensities. Correspondingly, the PPG spectra are dominated by broadband motion-induced energy that fully masks the cardiac peak seen in the ECG, highlighting the substantial challenge of extracting heart rate under vigorous, non-periodic activities.

To quantitatively describe the degree of motion periodicity, we define the *Harmonic Spectral Concentration Ratio* (HSCR), which evaluates how strongly the IMU spectral energy concentrates around the dominant frequency and its harmonics within the 0–5 Hz range:

$$\mathrm{HSCR} = \frac{E(f_0 \pm \Delta f) + E(2f_0 \pm \Delta f) + E(3f_0 \pm \Delta f)}{E_{\text{total, 0–5 Hz}}}, \tag{1}$$

where f_0 denotes the dominant frequency, Δf is a small bandwidth (e.g., 0.1 Hz), and $E(\cdot)$ represents the spectral energy integrated over the specified range. The denominator $E_{\text{total, 0–5 Hz}}$ is the total spectral energy within 0–5 Hz. *A higher HSCR indicates strong spectral concentration and thus more periodic motion, whereas a lower HSCR corresponds to dispersed spectral energy and irregular, non-repetitive motion patterns.* The HSCR values in Figure 5 clearly differentiate rhythmic activities such as walking and running from the irregular, dynamic sports.

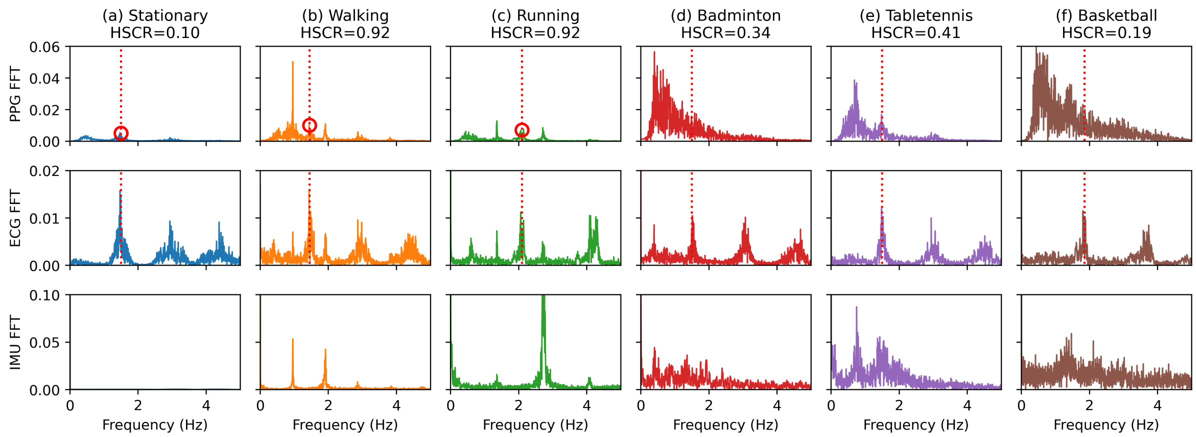

Fig. 5. Frequency-domain visualization of dominant wrist PPG, ECG and IMU signals across six activities.

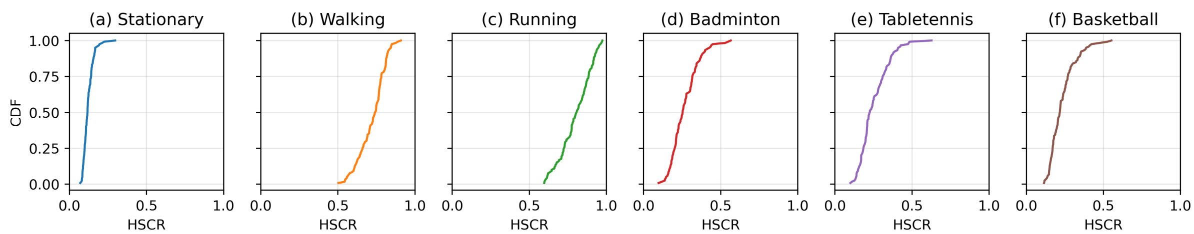

Fig. 6. Cumulative distribution functions of HSCR across six activities and the overall dataset.

Figure 6 further illustrates the cumulative distribution functions (CDFs) of HSCR across the six activities as well as the overall dataset, providing a distributional view of motion periodicity under different activity patterns.

4.3 Motion Intensity

To better illustrate the increased motion intensity in our dataset, we compare PPG-Sport with PPG-DaLiA [51], a widely used daily-activity dataset that includes a broad range of everyday scenarios. Figure 7 presents representative dominant wrist IMU signals from six activities in PPG-DaLiA (a–f) and PPG-Sport (g–l).

For the shared activities, *Stationary* and *Walking*, both datasets exhibit similar temporal patterns, confirming comparable signal quality and sensor configurations. Comparing *Stairs* from PPG-DaLiA with *Running* from PPG-Sport reveals that while both are periodic motions, the amplitude of acceleration in running is considerably larger, reflecting greater hand movement and impact intensity.

In contrast, PPG-DaLiA's three non-periodic activities, *Driving*, *Cycling*, and *Soccer*, show only occasional motion bursts separated by long stationary intervals. These signals have relatively low amplitude, indicating mild arm movements. Moreover, the main body motion in cycling or soccer originates from the lower limbs rather than the wrist, resulting in weaker IMU responses.

In comparison, the three sports in PPG-Sport, *Badminton*, *Table Tennis*, and *Basketball*, involve frequent and vigorous hand movements with large amplitude and short inter-motion intervals. The corresponding IMU signals exhibit dense, high-energy fluctuations that continuously vary in both magnitude and timing. Such intense and

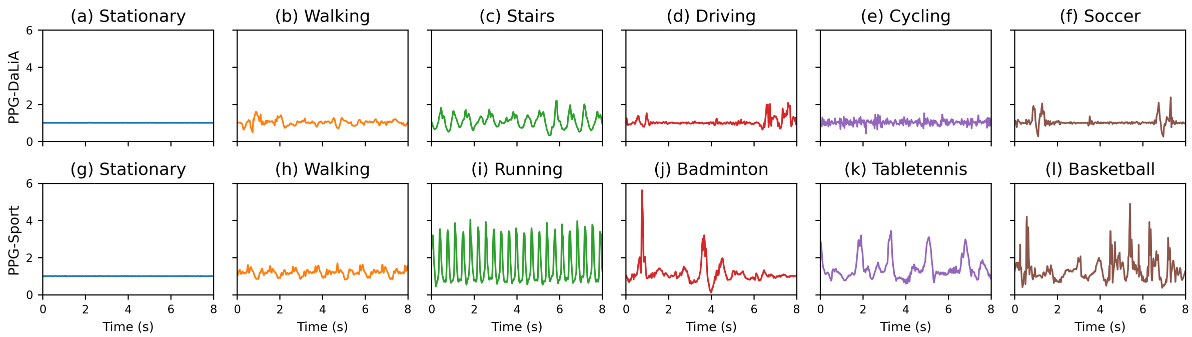

Fig. 7. Comparison of dominant wrist IMU time-domain signals between PPG-DaLiA (a–f) and PPG-Sport (g–l).

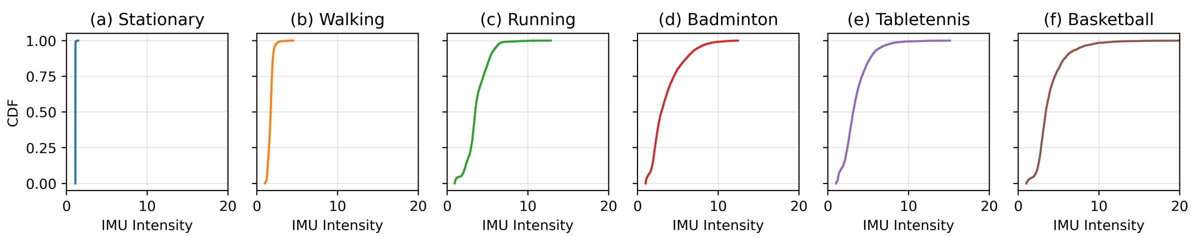

Fig. 8. Cumulative distribution functions of motion intensity across six activities and the overall dataset.

irregular wrist dynamics introduce strong motion artifacts into the PPG signal, posing significant challenges for robust heart rate estimation during high-impact sports.

To further quantify these differences in motion intensity, Figure 8 visualizes the distribution of window-level wrist motion intensity across activities in PPG-Sport. Here, motion intensity is defined as the mean-squared acceleration magnitude within each window, expressed in units of g^2. The cumulative distribution functions clearly show that running exhibits modestly higher acceleration intensity than daily activities, while racket sports and basketball demonstrate even higher and more variable intensity levels due to frequent and irregular hand movements.

4.4 Motion Symmetry

Another distinguishing feature of PPG-Sport is its *bilateral wrist recording*, which enables the analysis of motion symmetry across activities. Unlike previous datasets that typically include signals from only one wrist, PPG-Sport simultaneously captures PPG and IMU data from both sides. This design allows a more comprehensive view of whole-body dynamics and helps examine how similar or different the motion patterns of the two hands are under different conditions.

As shown in Figure 9, we present representative dominant and non-dominant PPG (upper part) and IMU (lower part) signals across six activities. We note that motion symmetry is inherently a relational property between the two wrists and is therefore more appropriately illustrated using paired bilateral signals rather than summarized by a single scalar distribution. For periodic activities such as *Walking* and *Running*, the signals exhibit not only clear rhythmicity but also high bilateral similarity. Although the two hands move alternately rather than synchronously, they follow the same temporal rhythm, resulting in overall waveform patterns that

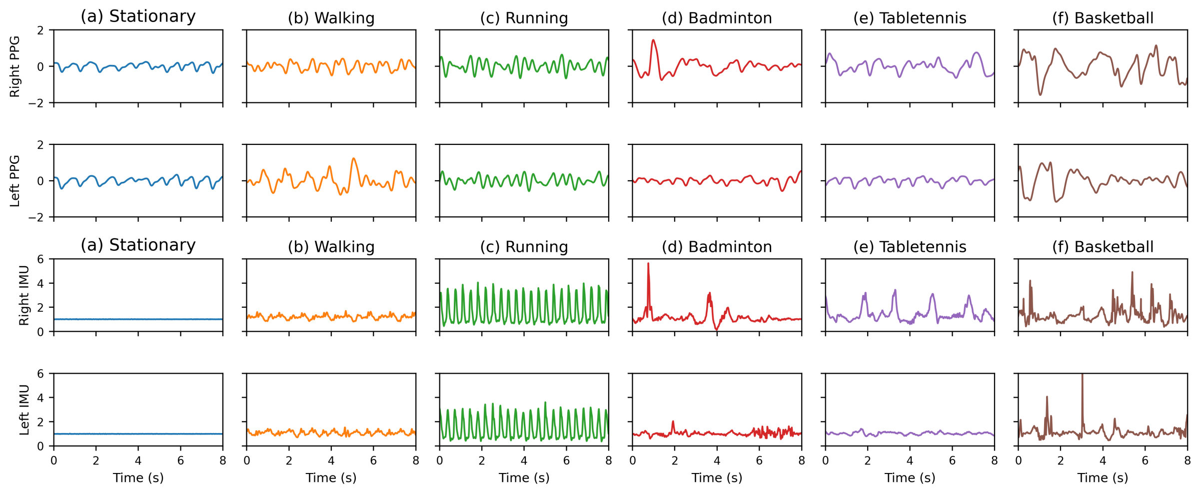

Fig. 9. Representative dominant and non-dominant wrist signals across six activities, where the upper panels show PPG and the lower panels show IMU recordings.

closely resemble each other. In contrast, sports such as *Badminton*, *Table Tennis*, and *Basketball* reveal strong asymmetry between the two sides: the dominant hand exhibits higher acceleration amplitudes and denser motion bursts, introducing stronger motion artifacts into the PPG signal, while the non-dominant hand moves with lower intensity but follows distinct coordination patterns. Although these auxiliary motions are less vigorous, they still alter the PPG waveform through subtle shifts in wrist posture and muscle tension.

These observations indicate that single-wrist data may not adequately represent the overall motion characteristics during asymmetric activities like sports. The signal features and motion-induced distortions can differ substantially between the two wrists, meaning that models trained on one hand may not generalize well to the other. The bilateral design of PPG-Sport effectively addresses this limitation by providing synchronized dominant and non-dominant wrist recordings for comprehensive motion analysis.

4.5 Heart Rate Distribution

Heart rate varies substantially across activities and intensity levels. A dataset with a limited or highly skewed HR distribution can bias model learning, as the network may only experience a narrow physiological range during training. Consequently, models trained on such data often fail to generalize to unseen conditions, for example, underestimating HR during vigorous exercise, since high-intensity samples are underrepresented or absent. Ensuring sufficient coverage across the full HR spectrum is thus crucial for training robust and generalizable HR estimation models.

To examine the HR coverage of our dataset, we first segmented the ECG reference into 8-second windows and then applied a standard peak-detection algorithm to compute the heart rate for each window. We then plotted the cumulative distribution functions (CDFs) of HR values across all six activities and the overall dataset, as shown in Figure 10. The distributions exhibit clear separation between activities: stationary and walking sessions concentrate below 100 bpm, while running and high-intensity sports (e.g., badminton, table tennis, basketball) span a much broader range up to 180–200 bpm. This balanced and continuous coverage across low, moderate, and high HR regions indicates that our dataset captures diverse physiological states, providing a solid foundation for developing models that generalize well across varying physical intensities.

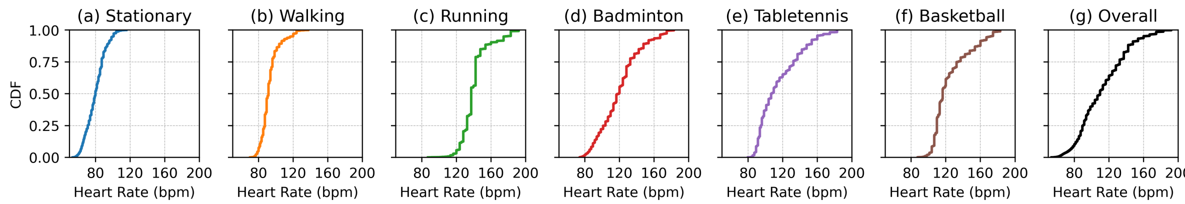

Fig. 10. Cumulative distribution functions of ground-truth heart rate (HR) across six activities and the overall dataset

5 BENCHMARK TASKS AND BASELINE MODELS

5.1 Benchmark Task Definition

We define a single benchmark task in PPG-Sport: motion-robust heart rate estimation from wrist-worn PPG and IMU signals collected across diverse sports activities. The ground-truth heart rate is derived from the chest-worn ECG using peak detection. To further examine the consistency and symmetry of motion coupling between the two wrists, we train two independent models, one using the non-dominant wrist signals and the other using the dominant wrist signals.

5.2 Data Preparation and Segmentation

All PPG and IMU signals are segmented into overlapping windows of 8 seconds with a 2-second hop size, following common practices in prior heart rate estimation studies [39, 63]. This configuration provides sufficient time to capture multiple cardiac cycles while ensuring adequate data volume for model training. The PPG signals collected from the wrist-worn devices are already DC-coupled, and thus no additional filtering is applied. For the IMU data, we retain the raw three-axis IMU signals without gravity removal to preserve the complete motion dynamics associated with each activity. To maintain consistency between dominant and non-dominant hand data, we identify participants who are left-handed and swap their left/right wrist data. Correspondingly, the x- and y-axes of the IMU signals are inverted to align motion directions across participants, while the z-axis is kept unchanged.

We adopt a subject-wise random split, where data from 20, 4, and 6 participants are used for training, validation, and testing, respectively. The split is generated using a fixed random seed for reproducibility. Compared with leave-one-subject-out (LOSO) evaluation, this setting significantly reduces evaluation time while still effectively assessing model generalization to unseen users. The ground-truth heart rate for each window is computed from the ECG reference using a simple peak-detection algorithm, averaging the detected inter-beat intervals within the same 8-second window.

5.3 Baseline Model

The overall architecture of our baseline model is shown in Figure 11. The network adopts a multimodal end-to-end design that jointly leverages wrist PPG and IMU signals for heart rate estimation. It consists of four components: (1) parallel PPG and IMU encoders, (2) an *Artifact Cancellation Unit* for motion suppression, (3) a *Fusion and Attention* module for multimodal interaction, and (4) an *Auxiliary Spectral Head* that regularizes frequency-domain consistency during training.

Feature Encoding. Two parallel encoders process the PPG and IMU modalities, respectively. The *PPGEncoder* employs a series of depthwise-separable temporal convolutional blocks with exponentially increasing dilations (e.g., 1, 2, 4, 8, 16), a standard design in temporal convolutional networks for capturing multi-scale temporal dependencies while keeping the model compact. The initial convolution layer uses a relatively large kernel (k=41,

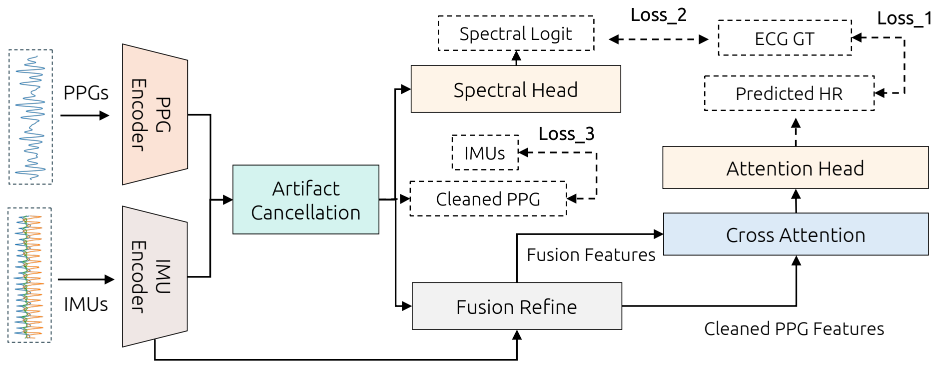

Fig. 11. Overall architecture of the baseline model for heart rate estimation from single wrist PPG and IMU signals.

padding=20) to cover approximately 0.6 s of signal at 64 Hz, roughly corresponding to one cardiac cycle under normal heart rates. This design allows the encoder to perceive complete pulse morphologies and short-term waveform variations from the very first layer. In contrast, the *ACCEncoder* adopts a similar structure but with a smaller kernel (k=5, padding=2) to emphasize rapid, non-periodic motion fluctuations captured by the tri-axial accelerometer. Such asymmetric kernel choices effectively balance the physiological and motion timescales, enabling the model to extract complementary features from the two modalities.

Artifact Cancellation Unit. The Artifact Cancellation Unit (ACU) jointly processes the encoded PPG and IMU features to perform adaptive artifact suppression through two complementary mechanisms. First, it employs a *Cross-FiLM* [49] modulation, where the IMU features are transformed into channel-wise scaling and shifting coefficients that dynamically modulate the PPG feature maps; this design enables efficient conditional modulation with low computational overhead, making it well-suited for a compact baseline model compared to more complex attention-based fusion strategies. Second, the module predicts a residual noise component directly from the IMU stream using a small convolutional subnetwork with dilated kernels to capture both short- and mid-range temporal dependencies. The predicted artifact is then subtracted from the raw PPG waveform to achieve direct denoising in the time domain. To prevent over-suppression or distortion of genuine pulse components, a learnable *gating branch* produces an element-wise weight map that adaptively balances the contributions of the modulated PPG and the residual-subtracted signal. Finally, a lightweight temporal projection layer aggregates the cleaned representation back into a one-channel time-domain signal.

Fusion and Attention. After motion artifact suppression, the cleaned PPG signal is fused with IMU features to form a joint representation that captures both physiological and kinematic cues. The cleaned PPG signal is first projected into the IMU latent space via a 1×1 convolution and concatenated with IMU features along the channel dimension. The fused representation is then refined by two depthwise-separable temporal convolutional blocks with short dilation factors (1 and 2) to model local temporal correlations. To enhance inter-modal interaction and temporal reasoning, we introduce a *LiteCrossAttnBlock*, a streamlined multi-head attention layer that uses PPG features as the query and the fused multimodal representation as key–value pairs. This design allows the model to focus on motion-affected segments while preserving intervals dominated by clean physiological rhythms. Finally, an *attention pooling head* aggregates the temporal features into a single heart rate prediction per window. A learnable query assigns adaptive attention weights across time steps, emphasizing pulse-dominant regions and down-weighting motion-contaminated segments, followed by a small MLP to regress the mean heart rate.

Auxiliary Spectral Head and Losses. While the primary regression head predicts mean heart rate in the time domain, we additionally introduce an *Auxiliary Spectral Head* to provide frequency-domain supervision. The

Table 3. Architecture summary of the baseline model (input window: 4×512 at 64 Hz; total parameters: 972,452).

Module	Configuration (main layers)	Output
PPGEncoder	Stem Conv1D: $1 \to 96$, $k = 41$; 5× DepthwiseSeparableTCN ($k = 9$, dilations $\{1, 2, 4, 8, 16\}$);	
	1×1 projection: $96 \to 192$	192×512
ACCEncoder	Stem Conv1D (grouped): $3 \to 96$, $k = 5$, groups= 3; 4× DepthwiseSeparableTCN ($k = 9$, dilations $\{1, 2, 4, 8\}$);	
	1×1 projection: $96 \to 192$	192×512
ACU	Cross-FiLM from IMU: $192 \to 128 \to 384$ (to produce γ, β);	
	artifact predictor: Conv1D $192 \to 128$ ($k = 3$) $\to$ Conv1D $128 \to 128$ ($k = 3$, dilation= 2) $\to$ Conv1D $128 \to 1$;	
	gating: 1×1 Conv on concat(192 + 192)$\to 64 \to 1$; time projection: 1×1 Conv $192 \to 1$	1×512
FusionRefine	clean projection: 1×1 Conv $1 \to 192$; concat with IMU feat (192 + 192) then 1×1 reduce $384 \to 192$;	
	2× DepthwiseSeparableTCN ($k = 9$, dilations $\{1, 2\}$)	192×512
LiteCrossAttnBlock	MultiheadAttention ($d = 192$, heads= 4), PPG as query and fused features as key/value;	
	FFN: Linear $192 \to 768 \to 192$ with GELU; residual + LayerNorm after attention and FFN	192×512
AttnPoolHead	learnable query attention pooling over time + MLP $192 \to 96 \to 1$	HR: scalar
SpectralHead	STFT ($n\_fft = 256$, hop=32), mean magnitude spectrum, interpolated to 30–210 bpm grid (181 bins)	181-d logits

spectral head applies a short-time Fourier transform (STFT) to the cleaned PPG waveform output by the ACU and computes its mean magnitude spectrum, which is interpolated onto a predefined 30–210 bpm grid aligned with the physiological heart-rate range. During training, the network predicts an unnormalized spectral distribution and is supervised by a soft Gaussian target centered at the ground-truth heart rate using a KL-divergence loss. This auxiliary constraint encourages spectral energy to concentrate around the correct frequency and mitigates common half-rate or double-rate errors. In addition, we introduce a lightweight *decoupling regularizer* that minimizes residual correlation between the cleaned PPG and the raw accelerometer magnitude, encouraging the model to separate motion-related artifacts from physiological content rather than re-encoding motion cues. The final training objective combines three terms: a time-domain regression loss ($\mathcal{L}$L1), a spectral alignment loss ($\mathcal{L}$spec), and a decoupling loss ($\mathcal{L}_{\text{dec}}$). The loss weights are empirically selected to balance the auxiliary objectives and ensure stable convergence:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{L1}} + 0.2\,\mathcal{L}_{\text{spec}} + 0.1\,\mathcal{L}_{\text{dec}}.$$

Table 3 summarizes the detailed architectural configuration of the baseline model, including the depth of each encoder, kernel, and dilation settings, attention design, and overall model scale. These design choices are informed by prior work [13, 32, 33] in PPG-based sensing and motion artifact mitigation, where multi-scale temporal modeling and conditional feature modulation are effective and robust in practice.

6 RESULTS AND ANALYSIS

We adopt a subject-wise 5-fold cross-validation protocol to ensure robust and unbiased performance evaluation across participants. Specifically, the 30 subjects are randomly partitioned into five mutually exclusive folds, each containing six subjects. For each fold, models are trained on data from four folds (24 subjects) and evaluated on the remaining fold (6 subjects), ensuring that training and testing subjects are strictly disjoint.

Within each training split, a small subset of subjects is further held out for validation to monitor convergence and select the best-performing checkpoint. All models are trained for 30 epochs with a batch size of 64 and an initial learning rate of 3×10^{-4} using the AdamW optimizer with a cosine learning-rate schedule. Mixed-precision training is enabled to accelerate convergence.

Performance is evaluated using the mean absolute error (MAE) and Standard Deviation (SD), measured in beats per minute (bpm), between the predicted heart rate and the ECG-derived ground-truth heart rate. For each fold, we select the model checkpoint with the lowest validation MAE and evaluate it on the held-out test subjects.

Table 4. Heart rate estimation performance (MAE ± SD in bpm) across six activities under different training data coverage and hand placement settings, averaged over 5-fold subject-wise cross-validation. (a)–(c) compare periodic-only versus all-activity training on the dominant and non-dominant hand;

Stationary	Walking	Running	Badminton	Table Tennis	Basketball	Overall MAE
(a) Periodic-Only Training, Dominant Hand						
1.12 ± 1.57	2.59 ± 1.91	3.18 ± 2.45	22.76 ± 9.32	16.54 ± 7.79	16.86 ± 9.71	**10.51**
(b) Periodic-Only Training, Non-dominant Hand						
1.09 ± 2.29	2.53 ± 2.31	3.75 ± 4.35	8.86 ± 8.59	9.16 ± 5.98	12.70 ± 12.55	**6.35**
(c) All-Activity Training, Dominant Hand						
1.43 ± 1.77	2.86 ± 2.54	4.28 ± 3.54	10.42 ± 9.99	9.09 ± 14.90	13.25 ± 7.42	**6.89**

All reported results in this paper are obtained by averaging performance across the five folds, providing a more stable estimate of model generalization across unseen subjects.

6.1 Performance on Periodic-Only Training

As we mentioned in Section 1, models trained on only periodic and low-intensity activities often struggle to handle complex sports scenarios involving rapid and irregular motions. To verify this limitation, we conducted an experiment using only periodic activities for training and evaluated the model across all six activities. Specifically, we trained and validated our baseline model described in Section 5.3 using only three periodic or light activities, stationary, walking, and running. The testing stage, however, covered all six activities, including three additional high-intensity sports: badminton, table tennis, and basketball. This configuration simulates a realistic scenario where the model (trained with existing public PPG dataset without involving our collected sports activities) encounters unseen, non-periodic motions at deployment.

Table 4 (a) reports the results for the dominant hand, which primarily performs the sporting actions. When trained only on periodic activities and evaluated across all six activities, the model performs well on the seen activities, achieving MAE values generally below 3 bpm. In contrast, performance degrades sharply once the model encounters the three unseen sports, where errors increase by an order of magnitude. Figure 12 illustrates a representative example from one subject's dominant hand, showing both the predicted and ground-truth heart-rate traces. While the prediction closely follows the ground truth during walking and running, it exhibits large fluctuations during badminton, table tennis, and basketball. This stark contrast indicates that models trained solely on periodic, low-intensity motions suffer severe degradation when exposed to unseen, irregular activities. Importantly, this degradation arises not only because these activities are unseen during training, but also because their motion patterns differ fundamentally from those in the training set, being more non-periodic and impact-heavy. These findings highlight the necessity of including such challenging motion patterns in training datasets to enable robust heart-rate estimation in realistic sports scenarios.

6.2 Performance on Non-dominant Hand

In non-periodic sports, the non-dominant hand typically plays a supporting role, contributing to balance and assisting force generation rather than executing the primary actions. As shown in Figure 9 previously, the non-dominant hand generally exhibits lower motion intensity than the dominant, suggesting that motion artifacts might also be less pronounced. This raises an intuitive question: could heart rate monitoring during sports be improved simply by wearing the device on the non-dominant hand?

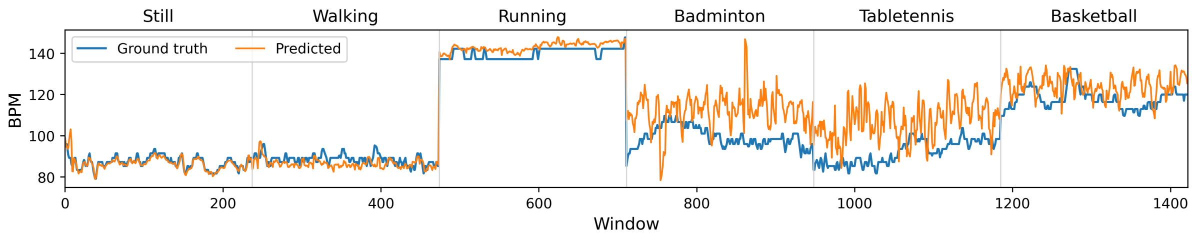

Fig. 12. Comparison between predicted and ground-truth heart rate across six activities for a representative subject (dominant hand).

To examine this hypothesis, we followed the same experimental setup as in Section 6.1, training the model using only periodic activities (stationary, walking, and running) and evaluating it across all six activities with data from the non-dominant hand. The results are summarized in Table 4(b). On average, the non-dominant-hand model achieves an overall MAE of 6.35 bpm, which is only marginally better than the corresponding dominant-hand result under the same training setting (10.51 bpm in Table 4(a)). This indicates that while reduced motion intensity offers some benefit, estimation errors still increase sharply once activities become non-periodic. This is primarily because irregular and multidimensional wrist movements—such as rotation, flexion, and changes in contact pressure—alter optical coupling and local blood perfusion at the wrist, introducing complex and time-varying artifacts that differ fundamentally from those observed in periodic motions. As a result, despite its lower motion intensity, the non-dominant hand does not eliminate the core challenges of wearable PPG sensing during sports.

From a practical perspective, wearing the device on the non-dominant hand can serve as a situational mitigation strategy, consistent with the recommendations of many commercial smartwatches for achieving more stable physiological monitoring. However, many sports-oriented applications [7] explicitly require placement on the dominant wrist to capture performance-related motion patterns, such as racket swings or striking techniques. Consequently, robustness under intense and non-periodic motion remains unavoidable in real-world sports scenarios. In this context, our results suggest that incorporating realistic sports data during training addresses the underlying motion regimes more directly, whereas changing the wearing side alone provides only limited and activity-dependent benefits.

6.3 Performance on Extended Activities Training

Although PPG-Sport introduces three additional sports beyond conventional walking and running, an immediate question arises: **is simply adding sports data sufficient to achieve robustness in complex motion scenarios?** To examine this, we retrained the baseline model using all six activities for training, validation, and testing. Table 4 (c) summarizes the MAE across six test subjects and six activity types. Compared to the previous setting (Section 6.1), the inclusion of non-periodic sports data substantially improves performance on those activities, especially for badminton and table tennis, whose average MAE decreases by roughly half. This demonstrates that incorporating diverse motion patterns indeed helps the model generalize better to dynamic and irregular movements. However, this improvement comes with a trade-off. The MAE for the previously well-performing periodic activities (stationary, walking, and running) slightly increases. One possible reason is that the expanded training data introduces larger variability in motion patterns, forcing the model to balance between fundamentally different motion domains. Another contributing factor could be that the optimization process now emphasizes robustness to motion artifacts rather than precision under stable conditions.

These results reveal an important insight: adding data alone is not enough. Although the extended training set improves robustness, the model still fails to achieve the ideal accuracy level observed in Section 6.1 (below 3 bpm

Table 5. Heart rate estimation performance (MAE ± SD in bpm) across six activities under cross-hand generalization.

Stationary	Walking	Running	Badminton	Table Tennis	Basketball	Overall MAE
(a) Cross-Hand Evaluation (Dominant → Non-dominant)						
1.37 ± 1.81	5.95 ± 3.96	4.60 ± 4.75	14.72 ± 17.39	14.26 ± 16.75	23.88 ± 21.48	**10.80**
(b) Cross-Hand Evaluation (Non-dominant → Dominant)						
2.94 ± 2.58	4.90 ± 4.58	6.27 ± 5.02	29.04 ± 23.29	31.40 ± 14.82	22.72 ± 23.37	**16.21**

for periodic motions), and the performance for sports activities is still lower than that for walking and running. This suggests that the key challenge lies not merely in data coverage but in model design. Future approaches should incorporate domain-aware architectures and tailored loss functions to explicitly model how motion artifacts distort PPG signals, enabling the network to truly disentangle physiological rhythms from complex movement interference.

6.4 Performance on Cross-Hand Evaluation

During sports activities, the two hands often play very different roles. The right hand (dominant hand) typically performs the main actions, such as swinging or dribbling, while the left hand mainly assists with balance and coordination. This asymmetry leads to distinct motion patterns between the two hands. In contrast, activities like walking or running feature rhythmic, alternating arm movements that are inherently symmetric. We therefore hypothesize that models trained on one hand may not generalize well to the other, especially for non-periodic sports. To verify this, we followed the same training protocol as in Section 6.3 and trained two separate models using all six activities—one with right-hand data and the other with left-hand data. Each model was then evaluated on the opposite hand's data after reversing the IMU axes for alignment.

Table 5 summarizes the results for (a) training on the right hand and testing on the left, and (b) training on the left hand and testing on the right. The results show that for periodic activities such as walking and running, performance remains relatively stable, suggesting that their motion and physiological patterns are symmetric across hands. However, for the three sports, performance drops drastically, indicating that a model trained on one hand cannot directly transfer to the other. Interestingly, the model trained on the right hand generalizes slightly better to the left than vice versa. This may be because the right-hand data contain a wider range of motion intensities—including moderate movements, whereas the left hand predominantly exhibits auxiliary and low-intensity motions, providing fewer transferable patterns. These findings highlight the pronounced asymmetry between hands in complex sports and suggest that future algorithm designs should explicitly account for this factor. Possible directions include collecting dual-hand data for training, introducing hand-aware representations, or leveraging data augmentation strategies to improve cross-hand generalization, since users may wear devices on either wrist in real-world use.

6.5 Performance on Cross-Dataset Evaluation

To examine the distributional differences between our sports-oriented and previous daily-life PPG datasets, we conduct a series of cross-dataset evaluations using PPG-Sport and PPG-DaLiA. Our goal is to use a unified heart rate estimation model as a probing tool to quantify the domain gap between the two datasets. Specifically, models are trained on one dataset and directly evaluated on the other without any fine-tuning.

To ensure a fair comparison, we align the experimental settings across datasets. Since PPG-DaLiA was collected exclusively from the non-dominant wrist, we use only non-dominant wrist data from PPG-Sport. We also match the total training size by randomly selecting data from 15 participants in PPG-Sport, consistent with the number

Table 6. Cross-dataset heart rate estimation performance (MAE ± SD in bpm) using models trained on periodic activities only.

(a) PPG-Sport training → PPG–DaLiA testing						
Stationary	**Walking**	**Stepping**	**Driving**	**Cycling**	**Soccer**	**Overall MAE**
4.71 ± 2.52	6.35 ± 4.60	6.82 ± 5.22	11.94 ± 13.87	13.18 ± 17.73	21.67 ± 18.56	**10.78**
(b) PPG–DaLiA training → PPG-Sport testing						
Stationary	**Walking**	**Running**	**Badminton**	**Basketball**	**Table Tennis**	**Overall MAE**
5.22 ± 4.66	7.11 ± 6.57	8.73 ± 8.68	26.13 ± 25.09	27.63 ± 19.71	24.26 ± 21.58	**16.51**

of subjects in PPG-DaLiA. In addition, the IMU signals in PPG-DaLiA are upsampled to 64 Hz to match the sampling rate of PPG-Sport, while all other preprocessing steps and model configurations remain identical.

Both datasets contain a mixture of periodic and non-periodic activities, although the specific activity types differ. The periodic activities in PPG-Sport include stationary, walking, and running, while those in PPG-DaLiA include stationary, walking, and stepping. For non-periodic activities, PPG-Sport focuses on high-intensity sports such as badminton, table tennis, and basketball, whereas PPG-DaLiA includes comparatively milder activities such as driving, cycling, and soccer.

6.5.1 Periodic-only Training. We first consider a restrictive training setting in which models are trained using only periodic activities from one dataset and then evaluated on all activities from the other dataset. This setting reflects a common practice where models are trained primarily on easily collectable, rhythmical daily-life activities and deployed more broadly. As shown in Table 6, when trained on periodic activities only, models exhibit relatively stable cross-dataset performance on periodic test activities, despite differences in data collection protocols and environments between the two datasets. This suggests that periodic motion patterns provide a reasonably transferable training signal for heart rate estimation, even across datasets.

When evaluated on non-periodic activities, performance degrades noticeably. For PPG-DaLiA, the degradation is moderate for driving and cycling, which involve relatively low-intensity and smoother motion patterns. These activities partially overlap with the motion characteristics observed in running, allowing models trained on periodic data to generalize to a limited extent. However, performance on soccer deteriorates substantially, indicating that increased motion irregularity and intensity significantly challenge models trained only on periodic activities. A more pronounced performance drop is observed when evaluating on PPG-Sport. Models trained solely on periodic activities from PPG-DaLiA perform poorly across all three sports activities in PPG-Sport. Unlike daily-life activities, these sports involve high-impact, non-periodic, and upper-limb-dominated motion patterns that are entirely absent from the periodic training data. As a result, the learned representations fail to generalize to these motion regimes, leading to consistently high estimation errors.

Taken together, these observations indicate that while training on periodic activities is sufficient for handling other periodic motions and can marginally generalize to low-intensity non-periodic activities, it is fundamentally inadequate for high-intensity, non-periodic sports scenarios. This cross-dataset behavior mirrors the trends observed within PPG-Sport itself and underscores the importance of incorporating realistic, high-impact non-periodic activities when developing and benchmarking motion-robust heart rate estimation models.

6.5.2 Extended Activities Training. While the previous experiment focuses on training with periodic activities only to examine cross-dataset generalization under limited motion diversity, we next consider a more permissive training setting that includes all available activities. Specifically, we train models using all six activities from one dataset and evaluate them on all activities from the other dataset under the same experimental setup as before.

Table 7. Cross-dataset heart rate estimation performance (MAE, bpm) using models trained on all six activities.

(a) PPG-Sport training → PPG–DaLiA testing						
Stationary	**Walking**	**Stepping**	**Driving**	**Cycling**	**Soccer**	**Overall**
4.89	6.51	6.72	10.26	9.07	13.13	**8.43**
(b) PPG–DaLiA training → PPG-Sport testing						
Stationary	**Walking**	**Running**	**Badminton**	**Table Tennis**	**Basketball**	**Overall**
5.74	7.92	9.12	21.76	23.47	22.14	**15.03**

As shown in Table 7, when trained on all activities from PPG-Sport, the model achieves improved performance on PPG-DaLiA compared to the periodic-only training setting. Notably, this improvement is observed not only for periodic activities but also for non-periodic activities that are not explicitly present in PPG-Sport. This can be attributed to the relatively low intensity of non-periodic activities in PPG-DaLiA, whose motion patterns may overlap with lower-intensity segments embedded within high-impact sports, such as brief pauses between strokes, ball retrieval, or off-ball movements during basketball play.

In contrast, training on all activities from PPG-DaLiA yields only marginal improvement over the periodic-only setting when evaluated on PPG-Sport, particularly for high-intensity sports. This suggests that the non-periodic activities in PPG-DaLiA, despite increasing activity diversity, do not adequately capture the motion characteristics and artifact patterns induced by high-impact, upper-limb-dominated sports. As a result, the learned representations fail to generalize to the more challenging motion regimes present in PPG-Sport.

Overall, these results further reinforce the unique value of PPG-Sport. The dataset captures high-intensity, non-periodic motion patterns that are largely absent from existing datasets, and models trained on PPG-Sport exhibit better downward compatibility to milder non-periodic activities in other datasets. In contrast, simply adding low-intensity non-periodic activities during training does not sufficiently bridge the domain gap to high-impact sports, highlighting the necessity of including realistic sports motions when developing and benchmarking motion-robust heart rate estimation models.

7 DISCUSSION

7.1 Participant Diversity

All participants involved in our dataset collection were healthy young adults aged between 21 and 34. This recruitment was intentionally restricted to minimize potential safety risks during sports activities and to ensure stable and reliable data collection. However, it also limits the representativeness of the dataset for other populations, including older adults, children, and individuals with cardiovascular conditions. Extending the dataset to cover a broader age range remains an important direction for future work, as it would better capture inter-individual variability in physiological responses and movement behaviors. In addition, only two participants in the current dataset were left-handed. Although their IMU signals were mirrored to align with the dominant and non-dominant hand categories, the potential influence of handedness on motion dynamics and PPG distortion is not fully understood and warrants further investigation.

7.2 Sports Experience Level

Our dataset includes self-reported sports experience levels (novice, intermediate, and experienced), which may influence motion intensity and, consequently, motion artifacts in wrist-based PPG signals. In principle, more experienced participants may produce faster and more forceful movements, leading to stronger IMU signals and increased motion artifacts. However, in our evaluation, we did not observe a clear or consistent difference in heart rate MAE across different experience levels.

This can be explained by two factors. First, we observed substantial variability in motion intensity and acceleration patterns even among participants with similar experience levels, making it difficult to attribute signal quality differences to experience alone. Second, the selected sports activities already involve large and rapid arm movements, which introduce substantial motion artifacts regardless of experience level. As a result, even novice participants may generate sufficient motion-induced noise, while additional experience does not necessarily lead to systematically worse signal quality. Finally, our analysis is based on a relatively simple baseline model. More advanced algorithms may better leverage subtle differences across experience levels, potentially improving performance for novice participants.

7.3 Activity Granularity

Unlike walking or running, which involve relatively uniform and continuous motion patterns, the three sports in PPG, Sport—badminton, table tennis, and basketball, naturally contain heterogeneous sub-actions such as picking up the ball, defending, or brief idle pauses. In the current version of the dataset, these sub-actions are not explicitly segmented and are treated as part of continuous gameplay. In practice, more than 80% of the recorded time corresponds to primary sports activities characterized by sustained and repetitive arm movements, which dominate the motion patterns of interest in this study. Based on our inspection of the collected data, brief sub-actions typically involve lower-intensity but still active movements, rather than introducing qualitatively different motion dynamics or signal artifacts. These actions therefore remain representative of realistic sports conditions and do not substantially alter the overall motion characteristics captured in the dataset.

We emphasize that PPG-Sport is not intended for fine-grained human activity recognition or action segmentation, but rather for studying robust heart rate estimation under realistic and unconstrained sports scenarios. For applications that require more detailed temporal structure, future extensions of the dataset could incorporate finer-grained annotations, for example by segmenting activity phases using IMU dynamics, video-assisted labeling, or self-supervised clustering of motion embeddings to identify distinct sub-actions within each sport.

7.4 Dataset Scale and Generalization

Although PPG-Sport provides synchronized recordings from both wrists totaling approximately 48 hours of data, we acknowledge that larger-scale collections are generally beneficial for training highly parameterized deep learning models. However, the primary goal of PPG-Sport is not to maximize data volume, but to characterize and benchmark the impact of complex, sports-induced motion on wrist PPG signals, an aspect that is largely underrepresented in existing datasets. Accordingly, the dataset emphasizes motion diversity and realism rather than sheer participant count.

In this context, the current scale is sufficient to support controlled evaluation and comparison of motion-robust modeling strategies under realistic sports conditions. Expanding the dataset with more participants would further increase inter-subject variability and improve model robustness, while complementary approaches such as motion-aware data augmentation (e.g., motion synthesis or cross-modal perturbation) could help mitigate data scale limitations. In addition, self-supervised pretraining on large-scale PPG datasets (e.g., foundation models [53]) offers a practical pathway to leverage PPG-Sport for deep model training despite its moderate size. Future extensions may also include a broader range of sports to enhance generalization to unseen activities and motion patterns.

7.5 Beyond Heart Rate Estimation

In this work, we focused primarily on heart rate estimation. However, the PPG-Sport dataset can be extended to support broader physiological monitoring once additional labels are incorporated. Beyond heart rate, heart rate variability reflects autonomic balance and recovery, while respiration rate relates to endurance and ventilatory efficiency. Other parameters, such as oxygen saturation and vascular compliance, though harder to validate during motion, are critical for assessing oxygen delivery and vascular response in high-intensity training. Enriching the dataset with such annotations would enable a more comprehensive understanding of cardiorespiratory and hemodynamic dynamics in sports, supporting applications in load management, fatigue detection, and safety monitoring. Furthermore, rather than targeting a single parameter, a promising direction is to reconstruct the clean PPG waveform itself from motion-corrupted signals. This waveform-level restoration would allow a unified model to recover physiologically meaningful signals, from which diverse metrics could be derived without training separate models for each task.

8 CONCLUSION

This paper presented PPG-Sport, the first dataset dedicated to heart-rate monitoring from wrist PPG under dynamic sports conditions. By capturing synchronized PPG, IMU, and ECG signals across six representative activities, PPG-Sport fills the critical gap between existing low-motion datasets and real-world, high-intensity scenarios. Through comprehensive analyses, we demonstrated how sports activities introduce non-periodic, high-intensity, and asymmetric motion artifacts that significantly degrade PPG reliability. Our benchmark results further reveal that current models trained on conventional datasets fail to generalize to these complex conditions, and while incorporating sports data improves robustness, substantial errors remain. These findings underscore the pressing need for new algorithmic designs that can disentangle physiological rhythms from motion interference. We believe PPG-Sport will serve as a valuable foundation for advancing motion-robust sensing and stimulating future research on reliable physiological monitoring in sports and beyond.

Acknowledgments

This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2 grant (Fund Code: T2EP20124-0046).

References

[1] Online. Apple Watch Series 11 - Apple (SG). https://www.apple.com/sg/apple-watch-series-11/.
[2] Online. Complaint about smartwatch. https://www.reddit.com/r/Coros/comments/1i578if/hf_measuring_way_off_pace_3/.
[3] Online. Complaint about smartwatch. https://www.reddit.com/r/Garmin/comments/15zcqhe/accurate_heart_rate_for_racquet_sports/.
[4] Online. EmbracePlus | The world's most advanced smartwatch for continuous health monitoring. https://www.empatica.com/en-int/embraceplus/.
[5] Online. Factors that affect heart rate monitoring. https://consumer.huawei.com/en/support/content/en-us00737221/.
[6] Online. Get the most accurate measurements using your Apple Watch. https://support.apple.com/en-sg/105002.
[7] Online. GoodShot - Badminton tracker on the App Store. https://apps.apple.com/us/app/goodshot-badminton-tracker/id1622183934.
[8] Online. Instinct Crossover AMOLED - Tactical Edition | Outdoor Recreation | Garmin Singapore. https://www.garmin.com.sg/products/outdoor/instinct-crossover-amoled-tactical/.
[9] Online. Polar H10 | Polar Singapore. https://www.polar.com/sg-en/sensors/h10-heart-rate-sensor?srsltid=AfmBOopY03E-BrBmXC6e4VXINJEXqw3RGB_vzFbQOli5ab6FEAKpynnR.
[10] Online. Powerbeats Pro 2 - Wireless Sports and Running Earbuds - Beats. https://www.beatsbydre.com/sg/earbuds/powerbeats-pro-2.
[11] John Allen. 2007. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement* 28, 3 (2007), R1.
[12] Erick Javier Argüello-Prada and Javier Ferney Castillo García. 2024. Machine learning applied to reference signal-less detection of motion artifacts in photoplethysmographic signals: a review. *Sensors* 24, 22 (2024), 7193.

[13] Luca Benfenati, Sofia Belloni, Alessio Burrello, Panagiotis Kasnesis, Xiaying Wang, Luca Benini, Massimo Poncino, Enrico Macii, and Daniele Jahier Pagliari. 2025. EnhancePPG: Improving PPG-based Heart Rate Estimation with Self-Supervision and Augmentation. In *2025 IEEE 7th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 1–5.

[14] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. heart: Motion-resilient heart rate monitoring with in-ear microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209.

[15] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. 2018. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics* 4, 4 (2018), 195.

[16] Peter H Charlton, Panicos A Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, and Jordi Alastruey. 2022. Wearable photoplethysmography for cardiovascular monitoring. *Proc. IEEE* 110, 3 (2022), 355–381.

[17] Shuwen Chen, Jiaming Qi, Shicheng Fan, Zheng Qiao, Joo Chuan Yeo, and Chwee Teck Lim. 2021. Flexible wearable sensors for cardiovascular health monitoring. *Advanced Healthcare Materials* 10, 17 (2021), 2100116.

[18] Wee Jian Chin, Ban-Hoe Kwan, Wei Yin Lim, Yee Kai Tee, Shalini Darmaraju, Haipeng Liu, and Choon-Hian Goh. 2024. A novel respiratory rate estimation algorithm from photoplethysmogram using deep learning model. *Diagnostics* 14, 3 (2024), 284.

[19] Che-Yu Chou, Zheng-Hao Chen, Yung-Hoh Sheu, Hung-Hsuan Chen, Min-Te Sun, and Sheng K Wu. 2025. ttSwing: a Dataset for table tennis Swing and Racket Kinematics analysis. *Scientific Data* 12, 1 (2025), 339.

[20] Heewon Chung, Hoon Ko, Hooseok Lee, and Jinseok Lee. 2020. Deep learning for heart rate estimation from reflectance photoplethysmography with acceleration power spectrum and acceleration intensity. *Ieee Access* 8 (2020), 63390–63402.

[21] Simone Costantini, Emilia Biffi, Fabio A Storm, Giuseppe Andreoni, and Anna M Bianchi. 2024. Automatic Detection of Artifacts in Photoplethysmography Signals Through Convolutional Neural Networks During Robot-Assisted Gait Rehabilitation. In *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*. IEEE, 91–96.

[22] Cheng Ding and Chenwei Wu. 2024. Self-Supervised Learning for Biomedical Signal Processing: A Systematic Review on ECG and PPG Signals. *medRxiv* (2024), 2024–09.

[23] D Docherty. 1982. A comparison of heart rate responses in racquet games. *British Journal of Sports Medicine* 16, 2 (1982), 96–100.

[24] Jin-Guo Dong. 2016. The role of heart rate variability in sports physiology. *Experimental and therapeutic medicine* 11, 5 (2016), 1531–1536.

[25] Chadi El-Hajj and Panayiotis A Kyriacou. 2020. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomedical Signal Processing and Control* 58 (2020), 101870.

[26] Amirhossein Esmaili, Mohammad Kachuee, and Mahdi Shabany. 2017. Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time. *IEEE Transactions on Instrumentation and Measurement* 66, 12 (2017), 3299–3308.

[27] Daniel Franklin, Andreas Tzavelis, Jong Yoon Lee, Ha Uk Chung, Jacob Trueb, Hany Arafa, Sung Soo Kwak, Ivy Huang, Yiming Liu, Megh Rathod, et al. 2023. Synchronized wearables for the detection of haemodynamic states via electrocardiography and multispectral photoplethysmography. *Nature biomedical engineering* 7, 10 (2023), 1229–1241.

[28] Zhilin Gao, Xingran Cui, Wang Wan, Wenming Zheng, and Zhongze Gu. 2021. ECSMP: A dataset on emotion, cognition, sleep, and multi-model physiological signals. *Data in Brief* 39 (2021), 107660.

[29] Ainara Garde, Parastoo Dehkordi, Walter Karlen, David Wensley, J Mark Ansermino, and Guy A Dumont. 2014. Development of a screening tool for sleep disordered breathing in children using the phone Oximeter™. *PloS one* 9, 11 (2014), e112959.

[30] Matthew Yiwen Ho, Hung Manh Pham, Aaqib Saeed, and Dong Ma. 2025. WF-PPG: A wrist-finger dual-channel dataset for studying the impact of contact pressure on PPG morphology. *Scientific Data* 12, 1 (2025), 200.

[31] Changshuo Hu, Hung Manh Pham, and Dong Ma. 2025. Morphology-Aware HRV Estimation from Wrist PPG in Sedentary Scenarios. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 745–750.

[32] Panagiotis Kasnesis, Lazaros Toumanidis, Alessio Burrello, Christos Chatzigeorgiou, and Charalampos Z Patrikakis. 2023. Feature-level cross-attentional ppg and motion signal fusion for heart rate estimation. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 1731–1736.

[33] Christodoulos Kechris, Jonathan Dan, Jose Miranda, and David Atienza. 2024. Kid-ppg: Knowledge informed deep learning for extracting heart rate from a smartwatch. *IEEE Transactions on Biomedical Engineering* (2024).

[34] Dmitry Kireev, Kaan Sel, Bassem Ibrahim, Neelotpala Kumar, Ali Akbari, Roozbeh Jafari, and Deji Akinwande. 2022. Continuous cuffless monitoring of arterial blood pressure via graphene bioimpedance tattoos. *Nature nanotechnology* 17, 8 (2022), 864–870.

[35] Takeshi Koyama, Akira Rikukawa, Yasuharu Nagano, Shogo Sasaki, Hiroshi Ichikawa, and Norikazu Hirose. 2022. Acceleration profile of high-intensity movements in basketball games. *The Journal of Strength & Conditioning Research* 36, 6 (2022), 1715–1719.

[36] Hyung-Chul Lee, Yoonsang Park, Soo Bin Yoon, Seong Mi Yang, Dongnyeok Park, and Chul-Woo Jung. 2022. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data* 9, 1 (2022), 279.

[37] Jongshill Lee, Minseong Kim, Hoon-Ki Park, and In Young Kim. 2020. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. *Sensors* 20, 5 (2020), 1493.

[38] Yongbo Liang, Zhencheng Chen, Guiyong Liu, and Mohamed Elgendi. 2018. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Scientific data* 5, 1 (2018), 1–7.

[39] Akash Kumar Maity, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2022. PPGMotion: Model-based detection of motion artifacts in photoplethysmography signals. *Biomedical Signal Processing and Control* 75 (2022), 103632.
[40] Bojan Makivić, Marina Nikić Djordjević, and Monte S Willis. 2013. Heart Rate Variability (HRV) as a tool for diagnostic and monitoring performance in sport and physical activities. *Journal of Exercise Physiology Online* 16, 3 (2013).
[41] Philip Mehrgardt, Matloob Khushi, Simon Poon, and Anusha Withana. 2022. Pulse Transit Time PPG Dataset.
[42] Manuel Meier, Berken Utku Demirel, and Christian Holz. 2024. WildPPG: A Real-World PPG Dataset of Long Continuous Recordings. *Advances in Neural Information Processing Systems* 37 (2024), 2246–2266.
[43] Rebecca Mieloszyk, Hope Twede, Jonathan Lester, Jeremiah Wander, Sumit Basu, Gabe Cohn, Greg Smith, Dan Morris, Sidhant Gupta, Desney Tan, et al. 2022. A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting. *IEEE journal of biomedical and health informatics* 26, 7 (2022), 2864–2875.
[44] Alessandro Montanari, Andrea Ferlini, Ananta Narayanan Balaji, Cecilia Mascolo, and Fahim Kawsar. 2023. Earset: A multi-modal dataset for studying the impact of head and facial movements on in-ear ppg signals. *Scientific data* 10, 1 (2023), 850.
[45] Benjamin Moody, George Moody, Mauricio Villarroel, Gari D. Clifford, and Ikaro Silva. 2020. MIMIC-III Waveform Database Matched Subset (version 1.0). https://doi.org/10.13026/c2294b. doi:10.13026/c2294b
[46] Aravind Natarajan, Alexandros Pantelopoulos, Hulya Emir-Farinas, and Pradeep Natarajan. 2020. Heart rate variability with photoplethysmography in 8 million individuals: a cross-sectional study. *The Lancet Digital Health* 2, 12 (2020), e650–e657.
[47] Junyung Park, Hyeon Seok Seok, Sang-Su Kim, and Hangsik Shin. 2022. Photoplethysmogram analysis and applications: an integrative review. *Frontiers in physiology* 12 (2022), 808451.
[48] Tania Pereira, Nate Tran, Kais Gadhoumi, Michele M Pelter, Duc H Do, Randall J Lee, Rene Colorado, Karl Meisel, and Xiao Hu. 2020. Photoplethysmography based atrial fibrillation detection: a review. *NPJ digital medicine* 3, 1 (2020), 3.
[49] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
[50] Marco AF Pimentel, Alistair EW Johnson, Peter H Charlton, Drew Birrenkott, Peter J Watkinson, Lionel Tarassenko, and David A Clifton. 2016. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering* 64, 8 (2016), 1914–1923.
[51] Attila Reiss, Ina Indlekofer, and Philip Schmidt. 2019. PPG-DaLiA. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C53890.
[52] Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. 2019. Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors* 19, 14 (2019), 3079.
[53] Mithun Saha, Maxwell A Xu, Wanting Mao, Sameer Neupane, James M Rehg, and Santosh Kumar. 2025. Pulse-ppg: An open-source field-trained ppg foundation model for wearable applications across lab and field settings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–35.
[54] Francesco Scardulla, Gloria Cosoli, Susanna Spinsante, Angelica Poli, Grazia Iadarola, Riccardo Pernice, Alessandro Busacca, Salvatore Pasta, Lorenzo Scalise, and Leonardo D'Acquisto. 2023. Photoplethysmograhic sensors, potential and limitations: Is it time for regulation? A comprehensive review. *Measurement* 218 (2023), 113150.
[55] Francesco Scardulla, Leonardo D'Acquisto, Raffaele Colombarini, Sijung Hu, Salvatore Pasta, and Diego Bellavia. 2020. A study on the effect of contact pressure during physical activity on photoplethysmographic heart rate measurements. *Sensors* 20, 18 (2020), 5052.
[56] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction*. 400–408.
[57] Daniel Yong Wen Tan, Huong Yong Ting, and Simon Boung Yew Lau. 2016. A review on badminton motion analysis. In *2016 International Conference on Robotics, Automation and Sciences (ICORAS)*. IEEE, 1–4.
[58] Min Wang, Zhe Li, Qirui Zhang, and Guoxing Wang. 2019. Removal of motion artifacts in photoplethysmograph sensors during intensive exercise for accurate heart rate calculation based on frequency estimation and notch filtering. *Sensors* 19, 15 (2019), 3312.
[59] Weinan Wang, Pedram Mohseni, Kevin L Kilgore, and Laleh Najafizadeh. 2023. PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health* 4 (2023), 1090854.
[60] Jingye Xu, Yuntong Zhang, Wei Wang, Mimi Xie, and Dakai Zhu. 2025. A Comprehensive PPG-based Dataset for HR/HRV Studies. *arXiv preprint arXiv:2505.18165* (2025).
[61] Guo-Qiang Zhang, Licong Cui, Remo Mueller, Shiqiang Tao, Matthew Kim, Michael Rueschman, Sara Mariani, Daniel Mobley, and Susan Redline. 2018. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association* 25, 10 (2018), 1351–1358.
[62] Yiming Zhang, Congcong Zhou, Xianglin Ren, Qing Wang, Hongwei Wang, Ting Xiang, Shirong Qiu, Yuan-Ting Zhang, and Xuesong Ye. 2025. Personalized Continuous Blood Pressure Tracking through Single channel PPG in Wearable Scenarios. *IEEE Journal of Biomedical and Health Informatics* (2025).

[63] Zhilin Zhang, Zhouyue Pi, and Benyuan Liu. 2014. TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on biomedical engineering* 62, 2 (2014), 522–531.