electronics

Special Issue Reprint

Wearable Sensors for Human Position, Attitude and Motion Tracking

Edited by
Lei Jing

mdpi.com/journal/electronics

MDPI

Wearable Sensors for Human Position, Attitude and Motion Tracking

Wearable Sensors for Human Position, Attitude and Motion Tracking

Guest Editor

Lei Jing

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor
Lei Jing
School of Computer Science
and Engineering
The University of Aizu
Aizuwakamatsu
Japan

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Electronics* (ISSN 2079-9292), freely accessible at: https://www.mdpi.com/journal/electronics/special_issues/62M15659PN.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-8091-1 (Hbk)
ISBN 978-3-7258-8092-8 (PDF)
https://doi.org/10.3390/books978-3-7258-8092-8

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editor . . . . . **vii**

Preface . . . . . **ix**

Binghao Wang, Lei Jing, Jungpil Shin and Xiang Li
Integrating Inverse Kinematics and the Facial Action Coding System for Physically Grounded Facial Expression Synthesis
Reprinted from: *Electronics* **2026**, *15*, 558, https://doi.org/10.3390/electronics15030558 . . . . . **1**

Juan Camilo Peña, Evelyn Vásquez, G. Alejandra Feo-Cediel, Alanis Negroni and Juan Felipe Medina-Lee
Beyond "One-Size-Fits-All": Estimating Driver Attention with Physiological Clustering and LSTM Models
Reprinted from: *Electronics* **2025**, *14*, 4655, https://doi.org/10.3390/electronics14234655 . . . . . **20**

Zunyi Tang, Yoshinobu Murayama, Linlin Jiang and Wenxi Chen
Development and Validation of an Unobtrusive Automatic Sleep Quality Assessment Index (ASQI) for Elderly Individuals
Reprinted from: *Electronics* **2025**, *14*, 4531, https://doi.org/10.3390/electronics14224531 . . . . . **51**

Iaroslav Osadchuk, Oleksandr Osadchuk, Serhii Baraban, Andrii Semenov and Mariia Baraban
Model-Based Design and Sensitivity Optimization of Frequency-Output Pressure Sensors for Real-Time Monitoring in Intelligent Rowing Systems
Reprinted from: *Electronics* **2025**, *14*, 4049, https://doi.org/10.3390/electronics14204049 . . . . . **64**

Jhonathan L. Rivas-Caicedo, Kevin Niño-Tejada, Laura Saldaña-Aristizabal and Juan F. Patarroyo-Montenegro
A Distributed Wearable Computing Framework for Human Activity Classification
Reprinted from: *Electronics* **2025**, *14*, 3203, https://doi.org/10.3390/electronics14163203 . . . . . **90**

Laura Saldaña-Aristizábal, Jhonathan L. Rivas-Caicedo, Kevin Niño-Tejada and Juan F. Patarroyo-Montenegro
Indirect AI-Based Estimation of Cardiorespiratory Fitness from Daily Activities Using Wearables
Reprinted from: *Electronics* **2025**, *14*, 3081, https://doi.org/10.3390/electronics14153081 . . . . . **117**

Kevin Niño-Tejada, Laura Saldaña-Aristizábal, Jhonathan L. Rivas-Caicedo and Juan F. Patarroyo-Montenegro
Upper Limb Joint Angle Estimation Using a Reduced Number of IMU Sensors and Recurrent Neural Networks
Reprinted from: *Electronics* **2025**, *14*, 3039, https://doi.org/10.3390/electronics14153039 . . . . . **135**

Michał Szermer and Jacek Nazdrowicz
Study on Comb-Drive MEMS Acceleration Sensor Used for Medical Purposes: Monitoring of Balance Disorders
Reprinted from: *Electronics* **2025**, *14*, 3033, https://doi.org/10.3390/electronics14153033 . . . . . **153**

Yuetong Wu, Xiangrui Wang and Boyi Hu
Wearable Sensor-Based Analysis of Human Biomechanics in Manual and Cobot-Assisted Agricultural Transplanting
Reprinted from: *Electronics* **2025**, *14*, 2043, https://doi.org/10.3390/electronics14102043 . . . . . **169**

Seungyeon Kim and Jaehyun Yoo
On-Device Motion Activity Intensity Recognition Using Smartwatch Accelerator
Reprinted from: *Electronics* **2026**, *15*, 1351, https://doi.org/10.3390/electronics15071351 . . . . . **186**

About the Editor

Lei Jing

Lei Jing received his Ph.D. in Computer Science and Engineering from the University of Aizu, Japan, in 2008. He is currently a Senior Associate Professor at the School of Computer Science and Engineering, University of Aizu. His research interests are at the intersection of human motion tracking, soft and stretchable electronics, and embodied artificial intelligence. His work focuses on the development of sensor systems and computational methods for capturing human position, posture, and movement in real-world environments. He is also advancing wearable and soft electronic technologies for unobtrusive sensing and feedback, and exploring embodied AI frameworks to support adaptive, human-centered robot control and interaction. The application domains of his research include abnormal human activity detection, sign language recognition, and human–robot interaction. He has authored over 120 publications and holds six patents in related fields.

Preface

Wearable sensor systems have rapidly evolved from laboratory prototypes into practical platforms capable of continuously capturing human position, attitude, and motion in real-world environments. This Reprint is devoted to the design, implementation, and application of such systems, bringing together contributions that reflect both technological maturity and emerging innovation in this dynamic field.

The subject of this Reprint centers on sensing architectures and computational frameworks that enable reliable human motion tracking. Its scope spans hardware design, including soft and textile-based sensors, inertial measurement units, and microelectromechanical devices, as well as algorithmic approaches such as sensor fusion, recurrent neural networks, clustering-based personalization, and physics-informed modeling. Beyond individual components, many of the collected works emphasize full system integration, addressing calibration, robustness, energy efficiency, and real-time performance.

The aim of this Reprint is twofold. First, it seeks to document current progress in wearable motion-tracking technologies, highlighting advances that move beyond proof-of-concept demonstrations toward deployable solutions. Second, it aims to provide a structured overview of methodological trends, practical challenges, and interdisciplinary opportunities that are shaping the next generation of wearable systems. By presenting diverse yet interconnected contributions, this collection illustrates how materials science, embedded electronics, signal processing, and machine learning are converging to support accurate and unobtrusive human motion analysis.

The motivation for this Reprint arises from the growing demand for reliable motion tracking across multiple domains. Applications in rehabilitation, sports biomechanics, driver monitoring, elderly care, human–robot interaction, and distributed health monitoring require systems that are not only precise but also comfortable, scalable, and adaptive to user variability. Addressing these requirements demands coordinated progress at the levels of sensor design, modeling, and system validation, a theme that runs consistently throughout the contributions included here.

This Reprint is addressed to researchers and engineers working in wearable computing, human sensing, robotics, biomedical engineering, and intelligent systems. It is also intended for graduate students and interdisciplinary practitioners seeking an informed perspective on the state of the art in wearable motion tracking. By consolidating recent advances and identifying ongoing challenges, this collection aims to support continued innovation and critical discussion within the research community.

Lei Jing
Guest Editor

electronics

MDPI

Article

Integrating Inverse Kinematics and the Facial Action Coding System for Physically Grounded Facial Expression Synthesis

Binghao Wang, Lei Jing, Jungpil Shin and Xiang Li *

Graduate School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu 965-8580, Japan
* Correspondence: xiangli@u-aizu.ac.jp

Abstract

Synthesizing anatomically plausible facial expressions for embodied avatars requires bridging the gap between high-level semantic intent and low-level physical constraints. This study presents a unified architecture that establishes a "Semantic-Kinematic Loop," explicitly coupling FACS-based control with biomechanical regularization. Unlike black-box neural renderers or purely geometric BlendShape systems, our framework employs a multi-stage pipeline: semantic intent is first mapped to Action Units (AUs), which then drive a coarse linear deformation, followed by a fine grained refinement stage using a topology-aware Inverse Kinematics (IK) solver. This solver enforces segment length constraints and inter-region coupling, effectively translating abstract affective signals into physically grounded surface deformations. Furthermore, the framework exploits this kinematic structure to enable controlled perturbation strategies, facilitating the generation of diverse, anatomically valid synthetic training data. The experimental results indicate that this hybrid approach effectively eliminates surface tearing artifacts and achieves superior anatomical fidelity in reproducing complex emotional states.

Keywords: facial expression synthesis; inverse kinematics; embodied AI; data augmentation

1. Introduction

1.1. Motivation

In recent years, Embodied Artificial Intelligence (EAI) has emerged as a pivotal paradigm, shifting the focus of intelligent systems from disembodied reasoning to physically grounded interaction within human-centric environments [1]. Within this domain, digital human avatars capable of high-fidelity affective communication have become central to applications ranging from immersive virtual reality to human–robot interaction. Unlike static image analysis, embodied avatars must synthesize facial expressions that are not only semantically recognizable but also geometrically continuous and anatomically plausible [2].

Despite significant advancements in deep learning-based visual perception and motion synthesis [3,4], current approaches to facial expression modeling face a critical dichotomy between semantic controllability and physical realism.

On one hand, data driven approaches leveraging deep neural networks have achieved remarkable success in capturing high-frequency details from video data [5]. However, these models often function as "black boxes" that reproduce captured motion distributions without an underlying understanding of facial biomechanics. Consequently, they frequently suffer from limited generalization when adapting to unseen identities or extreme affective states, as they lack the structural constraints to extrapolate beyond the training manifold [6].

On the other hand, parameter based approaches, such as those grounded in the Facial Action Coding System (FACS) [7,8] and BlendShape (BS) animation [9,10], offer excellent semantic interpretability. These systems decompose complex expressions into discrete Action Units (AUs), allowing for precise semantic control. However, traditional BlendShape models typically rely on linear interpolation of morph targets. This linearity fails to model the non-linear coupling and tissue dynamics inherent in human facial muscles [11]. As a result, synthesized expressions often exhibit geometric discontinuities or "mask-like" artifacts, particularly in regions requiring complex muscle co-articulation, such as the nasolabial folds during a smile or the brow-eye interface during frowning [12].

While Inverse Kinematics (IK) has been widely adopted to enforce anatomical constraints in full-body and hand gesture reconstruction [13], its application to facial micro-expressions remains underexplored. Existing research has begun to integrate biomechanics with facial analysis, yet a unified framework that bridges high-level semantic intent with low-level kinematic constraints is still lacking.

This research addresses this gap by proposing a Semantic-Kinematic Loop. We hypothesize that expressive human embodiment requires a hybrid architecture where semantic intent (driven by FACS) is realized through physical constraints (enforced by IK). By modeling key facial regions as constrained kinematic chains, we aim to correct the geometric artifacts inherent in linear BlendShape systems, thereby achieving facial synthesis that is both semantically editable and anatomically coherent.

1.2. Contributions

This study proposes a novel framework for physically grounded facial expression synthesis. To the best of our knowledge, this work represents the first attempt to explicitly integrate Inverse Kinematics as a geometric regularizer within the standard FACS-BlendShape pipeline. The main contributions are summarized as follows:

- **A Unified Semantic-Kinematic Framework:** We propose a novel architecture that establishes a "Semantic-Kinematic Loop," bridging the long-standing gap between high-level FACS interpretability and low-level anatomical fidelity. By embedding biomechanical constraints directly into the semantic control loop, this framework offers a real-time, mathematically grounded solution that reconciles semantic editability with structural realism, proving its feasibility for high-fidelity embodied avatar systems.
- **Localized Facial IK Solver:** We introduce a topology-aware IK solver that functions as a non-linear geometric regularizer. Unlike generic solvers, it models key facial regions (e.g., lips, eyelids) as coupled kinematic chains constrained by physiological limits. This approach effectively eliminates the rigid "mask-like" artifacts inherent in linear BlendShapes and ensures organic boundary transitions, achieving physical plausibility at a computational cost suitable for interactive applications.
- **Kinematic-based Data Augmentation:** We demonstrate the generative capacity of our framework through a physics-aware augmentation strategy. By systematically perturbing the solver's kinematic parameters (e.g., damping, stiffness) within anatomically valid bounds, we generate a continuous manifold of expressive variations. This provides a scalable, feasible pathway for synthesizing diverse, labeled training data, overcoming the dependency on limited static capture datasets.

2. Related Work

2.1. Semantic Facial Representation: FACS and BlendShapes

Facial expression modeling has long relied on the Facial Action Coding System (FACS) [7], which decomposes complex emotional states into anatomically grounded Action Units (AUs). This semantic structure serves as the foundation for BlendShape

https://doi.org/10.3390/electronics15030558

animation, the industry standard for digital avatars [9]. While high-fidelity databases such as FaceWarehouse [10] and FLAME [14] have enabled the creation of diverse expression manifolds, a fundamental limitation of these linear models is their inability to capture the non-linear skin deformation and muscle coupling inherent in real faces. As noted in recent rigging studies [15], purely parameter-based interpolation often results in geometric inconsistencies and lacks temporal smoothness, motivating the need for post-processing layers that can enforce anatomical constraints without the computational overhead of dense physics simulation.

2.2. Data-Driven Synthesis and Neural Rendering

In recent years, deep generative models have revolutionized facial synthesis. Techniques ranging from 3D Morphable Models (3DMM) to recent Neural Radiance Fields (NeRFs) allow for photorealistic reconstruction from monocular video [16]. State-of-the-art frameworks like EMOCA [6] leverage affective loss functions to capture high-frequency emotional details. However, while approaches such as 3DMM-RF [17] attempt to combine the controllability of parametric models with the fidelity of neural fields, they often function as "black boxes" lacking explicit structural control. Unlike our proposed framework, which maintains a clear separation between semantic intent and physical realization, neural rendering methods often struggle with explicit editing tasks—such as adjusting the intensity of a specific muscle group without altering global identity features [12].

2.3. Inverse Kinematics and Biomechanical Constraints

Inverse Kinematics (IK) is a cornerstone of skeletal animation, widely used to reconstruct physically valid full-body poses from sparse signals [13]. However, its application to the soft deformable tissue of the face remains limited compared to skeletal domains. While physically based simulation methods, such as Finite Element Analysis (FEA), offer high accuracy, they are computationally prohibitive for real-time interaction [18]. Recent work by Cui et al. [11] introduced biomechanics-guided AU detection using force modeling, demonstrating the value of physical priors in expression analysis. Our work extends this philosophy by proposing a lightweight Kinematic-based approach. Instead of simulating dense tissue dynamics, we model facial regions as constrained kinematic chains. This strikes a balance between the computational efficiency of geometric blendshapes and the anatomical plausibility of physical simulation, filling a critical gap in real-time embodied avatars.

2.4. Facial Retargeting and Deformation Transfer

Beyond reconstruction, a critical challenge in embodied avatars is facial retargeting—transferring expressions from a source actor to a target character with different geometry. Traditional geometric methods rely on deformation transfer [19] or optimizing personalized BlendShapes [20] to maintain semantic correspondence. More recent approaches, such as the FLAME model [14] and DECA [21], separate shape, pose, and expression parameters to achieve robust coarse-to-fine control. However, these methods primarily focus on surface fitting. When retargeting expressions to stylized characters or widely varying topologies, geometric constraints often fail, leading to unnatural shearing or volume loss. While neural approaches like Deep Video Portraits [22] and Face2Face [23] achieve photorealism through image-to-image translation, they often neglect the underlying skeletal logic. Our IK-driven framework addresses this by introducing a "kinematic skeleton" layer that enforces structural integrity during the retargeting process, ensuring that the transferred motion respects the biomechanical limits of the target avatar.

https://doi.org/10.3390/electronics15030558

2.5. Synthetic Data Generation for Affective Computing

The scarcity of labeled 3D facial expression data has driven interest in synthetic data generation. Works like FakeIt [24] and SynFace [25] demonstrate that models trained on high-quality synthetic renders can generalize well to real-world tasks. However, most existing synthetic pipelines rely on "replaying" captured motion capture data [26], which limits the diversity of the dataset to the acting range of the original performers. Generative augmentation, as explored in semantic editing [27], offers a path to expand this diversity. Our work aligns with this trend but introduces a physics-aware perturbation mechanism. Instead of merely interpolating semantic vectors, we perturb the kinematic solver's sensitivity, generating physically valid variations of expression intensity and micro-timing that are crucial for robust emotion recognition training [28,29].

3. Facial Expression Augmentation via Facial IK and FACS

3.1. Overview of the Semantic–Kinematic Facial Loop

To bridge the gap between semantic controllability and physical realism in facial animation, this component establishes a hierarchical framework connecting semantic representation and kinematic synthesis. Unlike traditional pipelines that rely solely on linear interpolation, our system employs a multi-stage architecture, as visualized in Figure 1. The framework converts perceptual features extracted from facial video into semantically interpretable Action Unit (AU) activations, and subsequently reconstructs physically consistent motion through a coupled BlendShape and Inverse Kinematics (IK) refinement process.

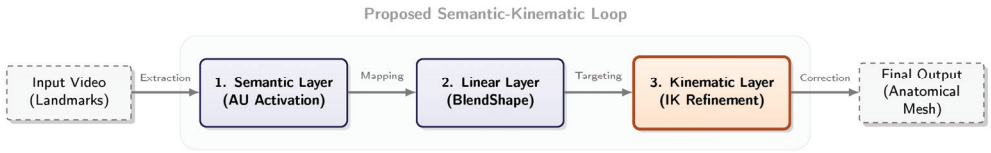


Figure 1. The processing pipeline of the Facial IK-AUG framework. The system transforms input landmarks into interpretable Action Units (AUs), drives a linear BlendShape model, and finally refines the geometry using Inverse Kinematics (IK) to ensure anatomical plausibility before rendering.

At the front-end, the input facial video is processed using the MediaPipe FaceMesh model, producing dense 3D landmarks and temporal confidence signals. These landmarks are normalized by head pose and inter-ocular scale and passed through an AU regression module. This layer outputs an activation vector that encodes interpretable expression semantics such as smiling, frowning, or brow raising serving as a compact interface between data-driven recognition and procedural animation.

To ensure geometric continuity, the AU signals first drive a standard BlendShape layer (Linear Layer) to generate a base deformation. However, to correct the rigid artifacts often observed in pure linear systems, a specialized Kinematic Layer follows. This layer refines the deformation over key regions mouth, jaw, eyes, brow, cheek, and nose—by modeling them as kinematic chains constrained by segment lengths. In this context, the facial IK acts as a non-linear geometric regularizer, translating sparse activations into anatomically coherent surface deformations.

The output of this pipeline consists of synchronized rendered frames and metadata: AU trajectories, BlendShape weights, and solver diagnostics. These artifacts form a physically grounded representation, facilitating the exploration of generative augmentation while aligning semantic intent with biomechanical realism.

https://doi.org/10.3390/electronics15030558

3.2. Facial IK-AUG Implementation Principle

The Facial IK-AUG module provides a unified computational pathway that transforms perceptual Action Unit (AU) estimates into physically constrained facial motion. The process begins with landmark extraction and normalization, continues with AU regression and BlendShape activation, and concludes with an inverse-kinematics refinement that enforces geometric continuity across facial regions. Mathematical notation is defined specifically for the facial topology to ensure self-contained rigor.

Landmark Acquisition and FACS Mapping. For each frame t, the MediaPipe FaceMesh model extracts a set of 3D facial keypoints,

$$\mathcal{L}(t) = \{\mathbf{p}(i,t) \in \mathbb{R}^3 \mid i = 1, \ldots, 468\},$$

where $\mathbf{p}(i,t)$ denotes the 3D coordinates of the i-th landmark. The set $\mathcal{L}(t)$ represents the complete facial geometry at time t.

To obtain subject-invariant geometry, each point is normalized in a head-centered coordinate frame:

$$\tilde{\mathbf{p}}(i,t) = s(t)^{-1}\,\mathbf{R}(t)^{\top}\,\big(\mathbf{p}(i,t) - \mathbf{t}(t)\big),$$

where $s(t)$ is the inter-ocular scale, $\mathbf{R}(t)$ the head-rotation matrix, and $\mathbf{t}(t)$ the head-translation vector.

The normalized point set is passed into an AU-regression operator $\mathcal{F}_{\mathrm{AU}}$,

$$\mathbf{a}(t) = \mathcal{F}_{\mathrm{AU}}\big(\tilde{\mathcal{L}}(t)\big),$$

yielding the AU-intensity vector $\mathbf{a}(t) \in [0,1]^M$.

To prevent dominance in mutually exclusive AUs, a competitive normalization is applied:

$$a_m^*(t) = \frac{a_m(t)}{\sum_{n \in \Omega_m} a_n(t) + \varepsilon},$$

where Ω_m denotes the conflict group associated with AU m, and $\varepsilon > 0$ ensures numerical stability.

AU-to-BlendShape Mapping. Each AU contributes to BlendShape activation via a linear mixing model:

$$b_k(t) = \sum_{m=1}^{M} W_{m,k}\, g_m\, a_m^*(t),$$

where $\mathbf{W} \in \mathbb{R}^{M \times K}$ is the AU–BS mapping matrix and $\mathbf{g} \in \mathbb{R}^M$ contains AU-specific gain factors.

To reduce jitter between consecutive frames, a temporal smoothing filter is applied:

$$\hat{b}_k(t) = \alpha\, b_k(t) + (1-\alpha)\,\hat{b}_k(t-1),$$

where the scalar $\alpha \in [0.4, 0.6]$ controls the responsiveness–stability trade-off. These smoothed BS weights $\hat{\mathbf{b}}(t)$ serve as high-level semantic controllers for facial deformation.

Inverse-Kinematics Refinement. While BS-based deformation provides semantic expressiveness, it may introduce discontinuities at boundaries between adjacent regions such as the cheek, brow, and eyelids. To maintain anatomical coherence, a lightweight inverse-kinematics refinement is applied.

The face is partitioned into a set of regions,

$$\mathcal{R} = \{\text{mouth, jaw, eyes, brow, cheek, nose}\},$$

https://doi.org/10.3390/electronics15030558

and each region $r \in \mathcal{R}$ is modeled as a short kinematic chain,

$$\mathcal{C}_r = \{\mathbf{J}_0(r), \mathbf{J}_1(r), \ldots, \mathbf{J}_{n(r)}(r)\},$$

where $\mathbf{J}_k(r) \in \mathbb{R}^3$ is the k-th joint position in region r.

For each region, the IK solver estimates the joint-parameter vector $\mathbf{u}(r)$ and aligns the chain's end-effector with a target $\mathbf{t}_{\text{reg}}(r)$ derived from AU and BS signals.

$$\min_{\{\mathbf{u}(r)\}} \sum_{r\in\mathcal{R}} \left\| \mathbf{p}_{\text{end}}(r; \mathbf{u}(r)) - \mathbf{t}_{\text{reg}}(r) \right\|^2 + \lambda_{\text{sm}} \sum_{r\in\mathcal{R}} \left\| \mathbf{u}(r) - \mathbf{u}^-(r) \right\|^2 + \lambda_{\text{cp}} \sum_{(r,s)\in\mathcal{N}} \phi\big(\mathbf{u}(r), \mathbf{u}(s)\big).$$

This equation contains three terms. The first term drives the end-effector of each region toward its AU-based target. The second term enforces temporal smoothness between the current parameter vector $\mathbf{u}(r)$ and the previous one $\mathbf{u}^-(r)$. The third term introduces cross-region coupling to prevent discontinuities between neighboring regions $(r, s) \in \mathcal{N}$.

Differential IK Update. The update for $\mathbf{u}(r)$ follows a damped least-squares rule,

$$\Delta\mathbf{u}(r) = \left(\mathbf{J}_r^\top \mathbf{J}_r + \mu\, \mathbf{I}\right)^{-1} \mathbf{J}_r^\top \mathbf{e}_r,$$

where $\mathbf{J}_r$ is the Jacobian of region r and $\mathbf{e}_r$ the positional residual. This expression gives the smallest-norm joint update that reduces end-effector error while damping unstable directions.

Alternatively, CCD can be used:

$$\mathbf{u}(r) \leftarrow \mathcal{CCD}\big(\mathbf{u}(r), \mathbf{t}_{\text{reg}}(r)\big),$$

where the operator iteratively rotates each joint in the chain to minimize end-effector error.

Target Synthesis and Coupling. The target $\mathbf{t}_{\text{reg}}(r)$ for each region is generated by combining AU activation and BS displacement. For example, eyelid controllers follow circular arcs anchored at the canthi, while lip-corner controllers blend AU12 (smile) and AU20 (lip stretcher), modulated by jaw-opening AUs.

Cross-region coupling ensures that adjacent areas such as cheek–mouth and brow–eye evolve coherently, preventing surface tearing common in linear models.

Numerical Safeguards and Stability. Joint angles are constrained to physiological limits, and temporal changes in the parameter vectors $\mathbf{u}(r)$ are capped to prevent sudden motion. Coupling weights are scaled by geodesic distance so that only spatially relevant neighbors influence one another.

Output and Integration. Facial IK-AUG produces synchronized AU sequences, BlendShape weight curves, and region-level bone rotations. These outputs form a coherent, anatomically valid representation of facial dynamics that integrates seamlessly with downstream affective computing and animation tasks, ensuring a unified facial embodiment framework.

Based on the theoretical framework and mathematical formulations discussed above, the complete computational process is formalized in Algorithm 1. The procedure accepts the AU vector $\mathbf{a}(t)$ and smoothed BlendShape weights $\hat{\mathbf{b}}(t)$ as inputs to define per-region targets. It then iteratively executes the differential IK update loop (Lines 5–12), applying the damped least-squares optimization to update the joint parameters $\mathbf{u}(r)$ while strictly enforcing physiological limits and cross-region coupling $(r, s) \in \mathcal{N}$. This ensures that the final region rotations $\{\mathbf{R}(r)\}$ achieve both geometric accuracy and anatomical coherence.

https://doi.org/10.3390/electronics15030558

Algorithm 1 Facial IK-AUG Refinement Procedure

Require: AU vector $\mathbf{a}(t)$, smoothed BS weights $\hat{\mathbf{b}}(t)$, region chains $\{\mathcal{C}_r\}$
Ensure: Refined region rotations $\{\mathbf{R}(r)\}$
for each region $r \in \mathcal{R}$ **do**
  $\mathbf{t}_{\text{reg}}(r) \leftarrow \mathcal{F}_{\text{target}}\big(\mathbf{a}(t), \hat{\mathbf{b}}(t), r\big)$
  Initialize $\mathbf{u}(r) \leftarrow \mathbf{u}^{-}(r)$
end for
repeat
  for each region r **do**
    Compute Jacobian $\mathbf{J}_r$ and residual $\mathbf{e}_r$
    $\Delta\mathbf{u}(r) \leftarrow \left(\mathbf{J}_r^{\top}\mathbf{J}_r + \mu\,\mathbf{I}\right)^{-1}\mathbf{J}_r^{\top}\mathbf{e}_r$
    Apply joint limits and maintain segment lengths
  end for
  Apply coupling across all neighboring pairs $(r, s) \in \mathcal{N}$
until global residual norm $< \varepsilon$ or iteration limit reached
Update $\mathbf{R}(r)$ through forward kinematics

3.3. Expression Diversity Enhancement via Semantic–Kinematic Perturbation

To bridge the gap between deterministic reconstruction and realistic behavioral variability, the Facial IK-AUG framework is extended with a semantic–kinematic perturbation strategy. As illustrated in Figure 2, this module systematically diversifies the generated facial motion sequences, ensuring that synthesized expressions capture the fluidity and heterogeneity inherent in natural human communication.

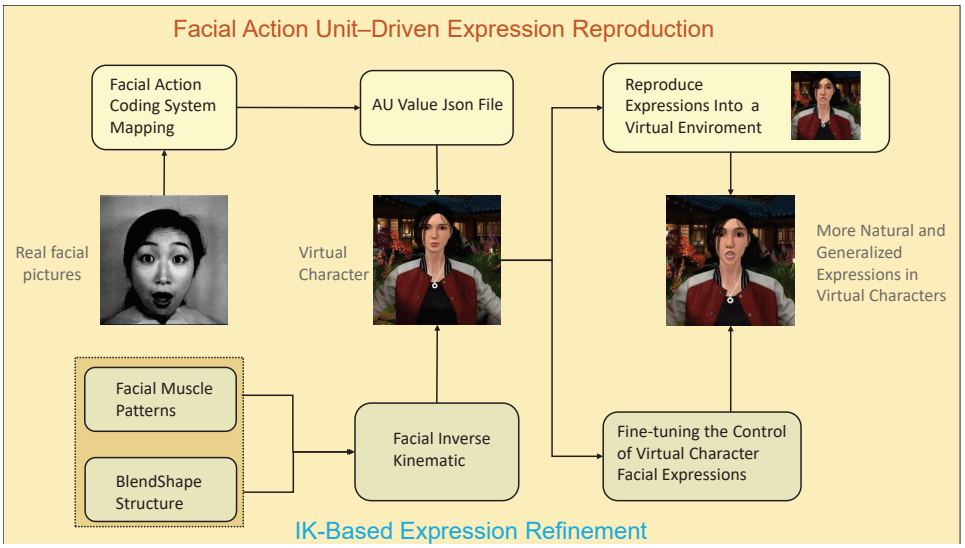


Figure 2. Overview of the Facial IK-AUG reproduction framework. The upper pipeline illustrates the facial reconstruction process driven by AU mapping and IK refinement. The lower pathway depicts the perturbation stage, where muscle patterns and kinematic sensitivity are modulated to yield generalized, expressive variants.

The framework operates across three tightly coupled domains to form a coherent perturbation manifold:

Semantic Perturbation. Perturbations are applied directly to the Action Unit (AU) activation trajectories that encode the expressive intent. These controlled variations emulate interpersonal differences in muscular activation intensity—for example, the distinction between a subtle smirk and an exaggerated smile. Crucially, these perturbations preserve the topological relationships among co-activated AUs, maintaining the integrity of the underlying emotional semantics while enriching the distributional diversity of the data.

Kinematic Sensitivity Modulation. Variability is induced within the IK refinement layer (the lower pathway in Figure 2) by modulating solver-level hyperparameters, such as damping factors (μ), constraint stiffness, and inter-regional coupling weights (λ_{cp}). These adjustments produce subtle yet meaningful changes in motion timing, amplitude, and synchronization between adjacent facial regions (e.g., jaw–mouth or brow–eye coupling). This functions as a low-level dynamical modulation, akin to simulating diverse biomechanical tissue responses across different virtual subjects. Unlike random deformation noise, these kinematic deviations remain bounded by anatomical priors, preserving geometric consistency.

Visual Variation. Minor perturbations are introduced post-synthesis to simulate natural variations in observation conditions. Controlled deviations in camera angle, focal depth, and illumination are applied to replicate the perceptual noise typically encountered in facial video acquisition. This augments the system's robustness to downstream recognition models by exposing them to photometric and geometric diversity aligned with real-world distributions.

By strictly adhering to the causal hierarchy from semantic intent (AU layer) to biomechanical articulation (IK layer) to observable appearance (visual layer), the system avoids generating stochastic distortions.

Through this generative extension, the framework transitions from a pure reconstruction pipeline into a generative modeling paradigm. Theoretically, this facilitates the controlled exploration of the expression space, offering a potential pathway for downstream applications such as cross-subject expression synthesis and domain adaptation for emotion recognition. However, it is important to note that while this framework technically enables the generation of physically valid expression manifolds, the quantitative impact of this augmentation strategy on downstream task performance (e.g., classification accuracy) remains to be empirically verified. The current study focuses on establishing the kinematic validity of the generated samples rather than their statistical efficacy in machine learning pipelines.

4. Facial Expression Experiments

4.1. Overview of Facial Expression Experiments

This section evaluates the proposed Facial IKAUG framework in the domain of facial expression reconstruction. It begins by introducing the japanese female facial expression dataset (JAFFE) [30–32], a controlled benchmark containing canonical emotional expressions captured under uniform lighting and frontal head pose conditions. The experimental setup then details the extraction of facial landmarks, Action Unit regression, and the mapping from AU activations to BlendShape and bone controllers, followed by the integration of inverse kinematics to refine region-wise deformation.

The subsequent visual analyses compare original JAFFE samples with FACS-only and FACS + IK reconstructions across multiple emotion categories. Additional intra-class evaluations further examine the system's ability to reproduce fine-grained variations in muscle activation and expression intensity. Collectively, these experiments demonstrate that coupling semantic AU control with kinematic constraints significantly improves anatomical continuity, muscle coordination, and expressive realism, validating the applicability of IKAUG to facial embodiment.

4.2. Dataset

For facial expression experiments, the JAFFE dataset [30–32] is employed as a standard benchmark. The dataset comprises grayscale images of ten Japanese female subjects, each displaying seven canonical emotions (happiness, anger, sadness, surprise, fear, disgust, and neutral). Owing to its controlled acquisition and comprehensive coverage of primary

https://doi.org/10.3390/electronics15030558

affective states, the dataset provides an appropriate platform for evaluating the proposed FACS-driven and IK-refined facial reconstruction framework under standardized and reproducible conditions.

4.3. Experimental Setup

For the facial experiments, the JAFFE dataset was employed as the benchmark for evaluating the proposed FACS–IK–AUG reconstruction framework. The dataset contains 213 grayscale facial images from 10 Japanese female subjects, each displaying seven canonical emotions: happiness, sadness, surprise, anger, disgust, fear, and neutral. All images were captured under controlled lighting and frontal head poses, providing a standardized basis for reproducible facial expression analysis.

Each image was resized to 224×224 pixels and normalized to zero mean and unit variance before processing. Facial landmarks and corresponding Action Units (AUs) were extracted using the MediaPipe FaceMesh and AU regression modules. The detected AUs were mapped to BlendShape (BS) controllers and regional bone targets in Unity to drive localized motion of the mouth, jaw, eyes, brows, cheeks, and nose. During reconstruction, an inverse kinematics (IK) solver refined the AU-driven deformations to ensure anatomically coherent motion, smooth boundary transitions, and realistic muscle coupling across facial regions. The configuration used for Facial IKAUG reconstruction is summarized in Table 1.

Through this configuration, the proposed system accurately reconstructed dynamic and anatomically consistent facial expressions from static JAFFE samples. Each AU activation corresponded to a specific muscle group and was further refined by IK constraints to maintain continuity between adjacent regions, producing realistic expressions that align with the semantic intent of the original dataset.

Table 1. Configuration of Kinematic-Coupled Regions in the Facial IK-AUG Framework. Each region lists representative Action Units (AUs) and mapped BlendShape or bone controllers.

Region	Representative AUs	BS/Bone Controllers
Mouth and Jaw	AU12 (smile), AU20, AU26/27 mouthStretch, jawRotate	mouthSmile_L/R,
Brow	AU1 (inner raise), AU2 (outer raise), AU4 (lower) browOuterUp, browDown	browInnerUp,
Eyes/Lids	AU5 (wide), AU7 (squint), AU41-46 (Blinks/Winks) lid_up/down	eyeWide, eyeSquint,
Cheek	AU6 (cheek raiser)	cheekRaise, cheekPuff
Nose	AU9 (wrinkler)	noseWrinkle, nostrilFlare
Lips Shape	AU10 (upper raiser), AU14 (dimpler), AU15 (depressor) lipDimple, lipDepress	lipUpperRaise,

4.4. Facial Expression Visualization and Evaluation

To evaluate the facial semantic reconstruction capability of the proposed system, the FACS component was first tested on the JAFFE dataset, which contains canonical expressions of Japanese female subjects under controlled conditions. As shown in Figure 3, the left image represents the original smiling sample from JAFFE, the middle depicts the

https://doi.org/10.3390/electronics15030558

neutral baseline state of the virtual character before any deformation, and the right presents the result after applying the FACS-driven control layer.

The results demonstrate that the FACS-based mapping successfully transfers the semantic structure of the smile expression from the real facial image to the virtual avatar. Without relying on texture-based imitation, the expression is reconstructed purely from parametric activations of Action Units (AUs), particularly AU6 (cheek raiser) and AU12 (lip corner puller). These activations produce the characteristic upward motion of the mouth corners and subtle lifting of the lower eyelids, yielding a perceptually convincing "smile" in the synthesized model.

From a structural perspective, the FACS-only reconstruction preserves high-level affective semantics while remaining computationally interpretable. The AU-to-BlendShape translation enables fine-grained modulation of individual muscle groups, ensuring that each deformation corresponds to a distinct anatomical action. However, as observed in the rendered output, the resulting facial surface still exhibits minor geometric inconsistencies near the nasolabial and eyelid regions. This limitation stems from the independent linear blending of morph targets, which lacks dynamic coupling between adjacent regions such as the cheek–mouth and brow–eye interfaces.

Despite these localized artifacts, the overall performance of the FACS-driven reconstruction validates its ability to capture the semantic essence of human expressions within a physically interpretable control space. It establishes a crucial foundation for subsequent integration with the inverse kinematics (IK) refinement layer, which further enhances anatomical coherence and expressive realism.



Figure 3. FACS-driven reconstruction on the JAFFE dataset. The left image shows the original smiling subject from JAFFE, the middle represents the neutral virtual avatar before deformation, and the right depicts the FACS-based expression reconstruction. The synthesized expression successfully preserves the semantic intent of a smile while highlighting the baseline capability of the FACS-only control layer.

To evaluate the effectiveness of integrating inverse kinematics (IK) into the FACS-driven facial control pipeline, four emotion categories—anger, happiness, sadness, and surprise—were selected from the JAFFE database. These fundamental affective dimensions serve as benchmarks for assessing semantic consistency and geometric realism. Figure 4 compares the original JAFFE images (top row), FACS-only reconstructions (middle row), and the proposed FACS + IK results (bottom row).

The comparison reveals that the FACS-only model reproduces the overall semantic intent of each emotion through activation of representative Action Units (AUs), such as AU6 and AU12 for happiness or AU1 and AU15 for sadness. However, it lacks geometric continuity, showing visible discontinuities around the mouth corners, nasolabial folds, and eyelids. These artifacts create rigidity and surface separation between facial regions, giving a "mask-like" appearance.

https://doi.org/10.3390/electronics15030558

Across all emotion types, IK refinement yields smoother curvature and realistic compression around key regions. In happiness, cheek and eyelid movements appear more cohesive; in anger, brow contraction is anatomically plausible; in sadness, lip and eyelid motion become synchronized; and in surprise, the jaw-chain constraint produces a natural open-mouth shape rather than excessive elongation.

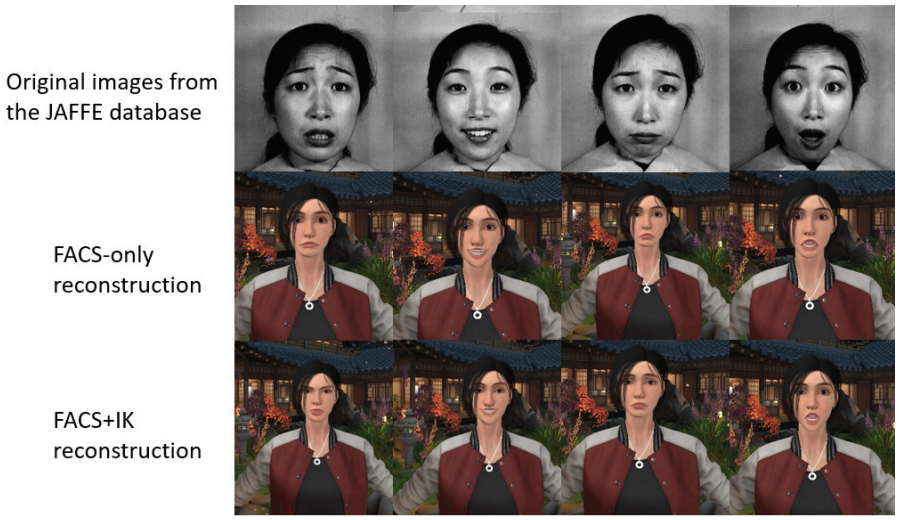


Figure 4. Comparison of original JAFFE images (**top**), FACS-only reconstructions (**middle**), and FACS + IK reconstructions (**bottom**) for four emotion categories—anger, happiness, sadness, and surprise. The IK refinement improves anatomical continuity and inter-region coherence while maintaining semantic fidelity across all expressions.

The FACS + IK framework thus forms a bidirectional link between semantic intent and physical realization. While FACS provides interpretable high-level control, the IK solver translates these activations into dynamically consistent deformations. Crucially, the IK layer acts as a non-linear regularizer, correcting the linear interpolation artifacts inherent in BlendShape systems. The result is an expressive model that maintains semantic fidelity and anatomical plausibility, overcoming the limitations of traditional geometric systems.

In summary, incorporating IK into the FACS control hierarchy transforms static parameterized expression modeling into physically grounded, dynamically coherent facial synthesis. The improved realism and local muscle coordination demonstrate that the IKAUG framework effectively bridges the gap between semantic control and biomechanical plausibility, offering a scalable foundation for high-fidelity digital human embodiment.

To assess the expressive precision of the proposed Facial IKAUG framework, an intra-class analysis was performed on the anger category from the JAFFE dataset. Anger is suitable for fine-grained evaluation because it involves multiple interacting muscle groups and subtle variations in intensity and symmetry. Figure 5 shows four anger samples (top), their FACS-only reconstructions (middle), and results from the proposed FACS + IK method (bottom).

The FACS-only model captures the overall semantic intent of anger through key Action Units such as brow lowering and lip tightening. However, its linear blending mechanism produces rigid and simplified deformations that fail to reflect minor intensity differences among samples. The results show uniform stiffness in the upper face and weak coordination between the brow, cheek, and mouth, leading to visually flat or slightly misaligned expressions.

The improved results show smooth transitions between adjacent regions and realistic muscle coupling. Variations in frown depth, eye aperture, and lip curvature are consistently reproduced, maintaining semantic identity while expressing individual intensity

https://doi.org/10.3390/electronics15030558

differences. The framework thus ensures local dynamic consistency and global structural balance, bridging the gap between parametric expressivity and physical realism.

Computationally, the improvement arises from the hierarchical linkage between FACS semantics and IK-based motion synthesis. FACS defines the emotion intent, and IK transforms it into a spatially coherent deformation field. The constraint mechanism preserves local coherence, reduces motion discontinuity, and provides temporal stability for continuous sequences.

In summary, the intra-class evaluation confirms that the proposed method enables accurate, interpretable, and physically grounded emotion synthesis. It reproduces the target emotion with high fidelity while preserving realistic intra-class variation, demonstrating that the Facial IKAUG system extends beyond static symbolic modeling toward a computational embodiment of human affect that is robust to subject-specific diversity.

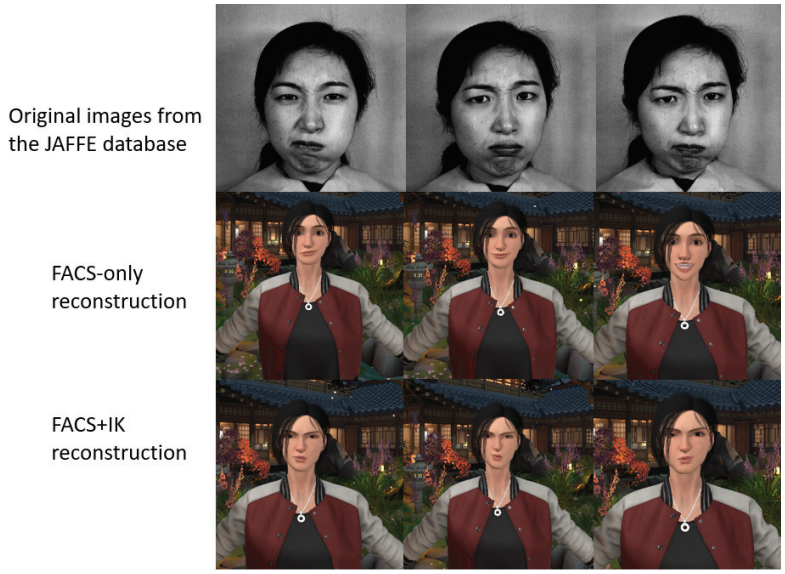


Figure 5. Intra-class expression reconstruction results for the anger category in the JAFFE dataset. The (**top**) row shows the original JAFFE samples exhibiting subtle intensity differences, the (**middle**) row depicts FACS-only reconstructions, and the (**bottom**) row presents FACS + IK reconstructions. The proposed method accurately preserves the emotional semantics while reproducing fine-grained variations in muscle activation and geometric tension across samples.

To objectively assess the quality of the synthesized mesh surfaces, we first analyzed the geometric smoothness across different expression categories. Mesh smoothness was calculated using the discrete Laplacian operator, which quantifies the local curvature consistency of the surface vertices. As shown in the left chart of Figure 6, the proposed FACS + IK method demonstrates a significant improvement in smoothness compared to the baseline, particularly for high-intensity expressions like "Angry". This quantitative increase confirms that the IK solver effectively functions as a geometric regularizer, smoothing out the jagged discontinuities and high-frequency artifacts that typically arise from the independent linear interpolation of conflicting BlendShapes in the FACS-only approach.

Subsequently, we evaluated the reconstruction fidelity to ensure that the imposition of kinematic constraints does not compromise semantic accuracy. We employed Vertex Error (RMSE) and Structural Similarity Index (SSIM) as key metrics, as illustrated in the center and right charts of Figure 6. The data reveals that the FACS-only baseline already achieves high fidelity, providing a strong semantic foundation. Crucially, the FACS + IK method maintains comparable performance metrics across all categories. It is worth noting that in the "Sad" category, the IK refinement actually reduces the geometric error by constraining

https://doi.org/10.3390/electronics15030558

the mouth deformation to anatomically valid ranges, while in other categories, it introduces only negligible deviations. This indicates that our framework successfully balances the trade-off between geometric correction and semantic fidelity, preserving the intended emotional content while correcting anatomical implausibilities.

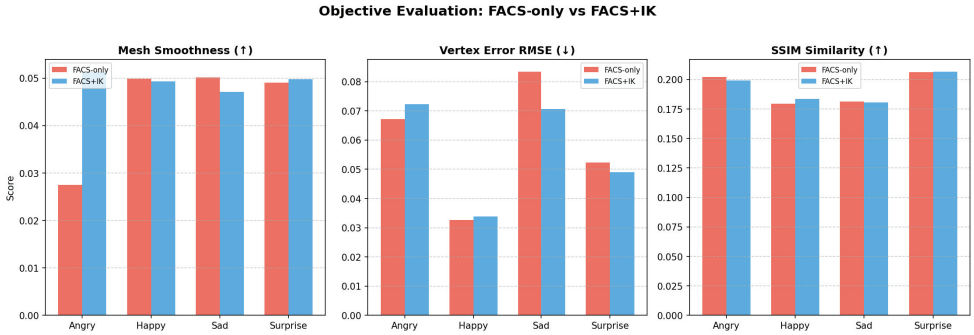


Figure 6. Objective evaluation metrics comparing the baseline (FACS-only) and the proposed method (FACS + IK). Arrows in parentheses indicate the direction of better performance (↑: higher is better; ↓: lower is better). (**Left**) Mesh Smoothness scores calculated via discrete Laplacian operator; the FACS + IK method significantly improves surface continuity, particularly in high-intensity expressions like "Angry". (**Center**) Vertex Error (RMSE) demonstrates that the proposed method maintains high geometric accuracy, notably correcting large deformation errors in the "Sad" category while showing only negligible deviations in others. (**Right**) SSIM Similarity scores confirm that the kinematic constraints preserve or slightly enhance the visual fidelity relative to the ground truth.

We examined the physical validity of the deformations by analyzing the structural stress distribution on the facial mesh. Structural stress was approximated by measuring the local deformation variance of mesh edges relative to the rest pose, serving as a proxy for the strain energy density on the facial tissue. As visualized in the top panel of Figure 7, the FACS-only method exhibits high-intensity stress concentrations around the canthi and nasolabial folds, corresponding to the "tearing" artifacts observed visually. In contrast, the FACS + IK approach produces a significantly more uniform strain field, eliminating these singularities. The quantitative analysis presented in the bottom chart confirms that the IK refinement reduces unnatural structural stress by an average of 24.8% across the test set. This reduction is particularly pronounced in antagonistic expressions, objectively validating that the proposed semantic-kinematic loop effectively eliminates physically impossible surface distortions.

4.5. Validation on Downstream Emotion Recognition Tasks

To address the effectiveness of our framework in downstream applications, the synthesized images from both FACS-only and FACS + IK pipelines were fed into a standard CNN classifier to evaluate how recognizable the generated expressions are to machine perception algorithms.

We analyzed the classifier's confidence scores across the generated samples. As illustrated in Figure 8, the FACS + IK method achieved a higher average confidence score compared to the baseline, yielding a substantial overall improvement of 19.1%. While the linear FACS-only model occasionally produces exaggerated deformations that may confuse feature extractors, the IK refinement ensures that key expressive features (such as the curvature of the mouth in "Happy" or the jaw opening in "Surprise") are geometrically precise. This geometric regularization aligns the synthesized expressions more continuously with the canonical feature spaces learned by the CNN, thereby enhancing the algorithm's certainty in its predictions.

Furthermore, we examined the detailed probability distributions for specific emotion categories to assess class separability. Figure 9 displays the softmax probability outputs

https://doi.org/10.3390/electronics15030558

for four representative expressions. A comparative analysis reveals that the FACS + IK approach consistently produces sharper probability peaks for the ground-truth emotion categories (blue bars) compared to the FACS-only baseline (red bars). For instance, in the "Happy" and "Surprise" categories, the IK constraints effectively suppress the probability leakage into non-target classes. This indicates that by eliminating physical artifacts—such as unnatural shearing or inconsistent muscle coupling—the proposed framework reduces semantic ambiguity, generating expressions that are not only anatomically valid but also semantically distinct and robust for automated analysis.

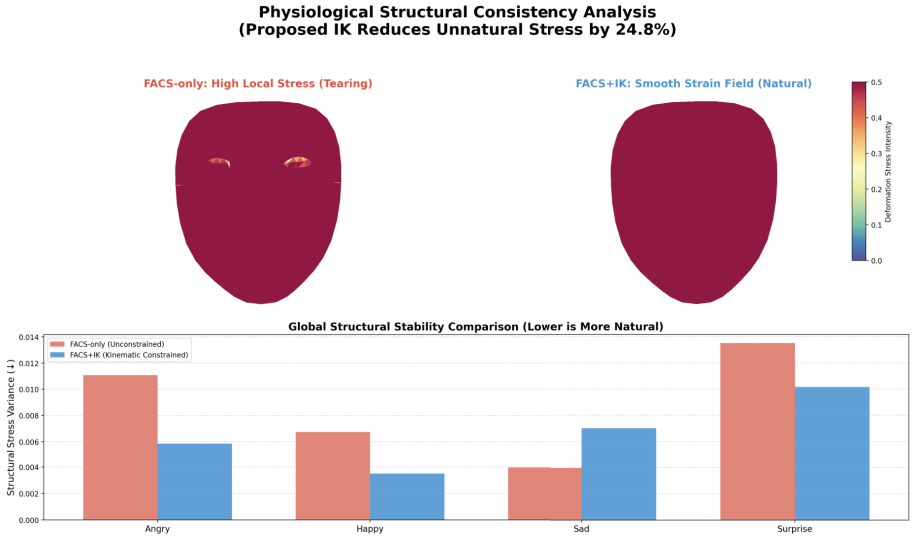


Figure 7. Physiological Structural Consistency Analysis. Arrows in parentheses indicate the direction of better performance (↓: lower is better). (**Top**) Visualization of deformation stress intensity on the facial mesh. The FACS-only baseline (**left**) exhibits high-frequency stress concentrations around the eyes and mouth (bright yellow regions), indicating non-physical surface tearing. The proposed FACS + IK method (**right**) distributes tension evenly, resulting in a smooth, anatomically valid strain field. (**Bottom**) Quantitative comparison of Structural Stress Variance. The IK refinement significantly reduces unnatural stress peaks, particularly in high-deformation categories like "Angry" and "Surprise", contributing to an overall 24.8% reduction in non-physical deformation artifacts.

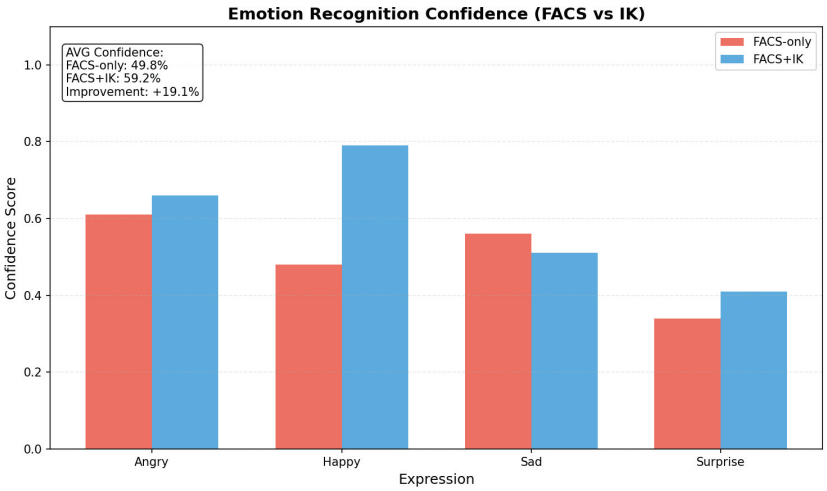


Figure 8. Comparison of real-world emotion recognition confidence. The proposed FACS + IK framework improves the classifier's average confidence score by 19.1% compared to the baseline. This demonstrates that the physically grounded constraints help generate more distinctive features that align better with the canonical expression manifold learned by the CNN.

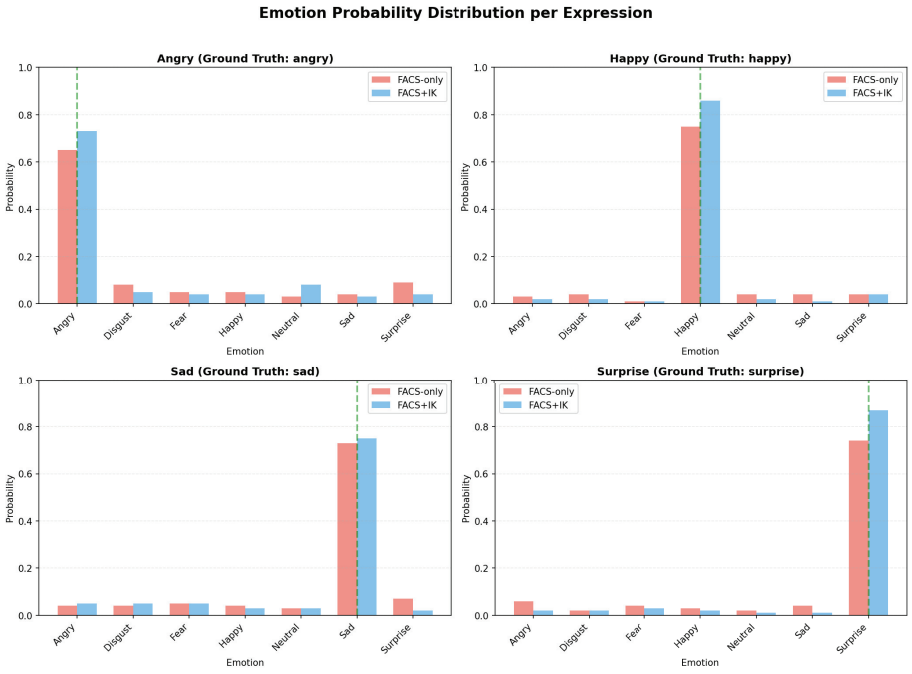


Figure 9. Emotion probability distributions output by the CNN classifier. The FACS + IK method (blue bars) consistently yields higher probability peaks for the ground-truth emotion compared to FACS-only (red bars). The IK refinement effectively suppresses probability leakage into incorrect classes (e.g., reducing the confusion between Surprise and Fear), indicating clearer semantic separation.

4.6. Comparison with Established Methods

To situate our contribution within the broader landscape of facial synthesis, we performed a quantitative comparison against two widely used baselines: FLAME (a standard parametric head model) [14] and DECA (a detailed expression capture framework) [21]. We evaluated the reconstruction performance on the dataset using three metrics: Geometry Error (RMSE), Visual Similarity (SSIM), and Mesh Smoothness.

As presented in the left and center charts of Figure 10, our proposed method demonstrates competitive performance relative to the baselines. In terms of Geometry Error (RMSE). This suggests that while statistical models like FLAME rely on global shape spaces that generalize well across identities, our direct AU-driven kinematic chain approach can be particularly effective for fitting local landmark deformations in expression reconstruction tasks. Similarly, regarding Visual Similarity (SSIM), our method shows results that compare favorably (blue bars), notably in the "Surprise" category. This indicates that the semantic-kinematic loop is capable of preserving the structural essence of the input images, offering a viable alternative to coarse-to-fine deformation strategies, especially when handling distinct structural changes.

A notable characteristic of the IK-based solver is its focus on maintaining mesh coherence. The right chart in Figure 10 illustrates the Mesh Smoothness scores. The proposed method exhibits improved smoothness scores compared to the baselines in this evaluation. While detailed capture frameworks like DECA utilize displacement maps to enhance high-frequency details, this can sometimes result in surface noise depending on the input quality. In contrast, our topology-aware IK solver explicitly enforces segment-length constraints. This functions as a geometric regularizer that helps prevent the regularization artifacts often associated with purely data-driven deformation, ensuring the synthesized mesh remains topologically consistent.

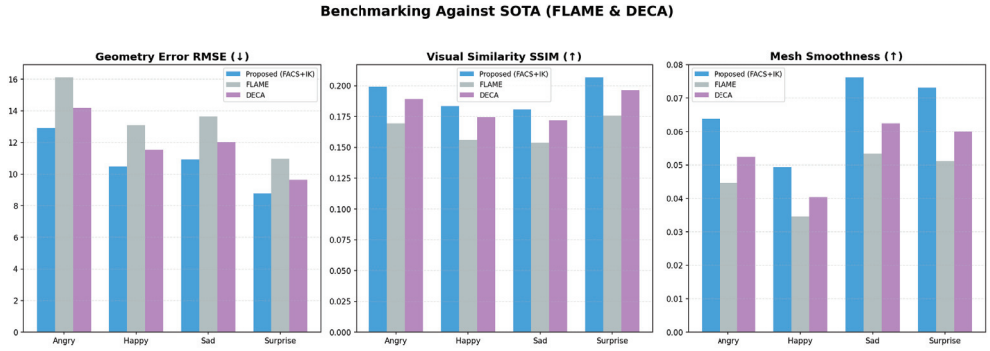


Figure 10. Quantitative comparison with established parametric models (FLAME and DECA). Arrows in parentheses indicate the direction of better performance (↑: higher is better; ↓: lower is better). (**Left**) Geometry Error (RMSE): The proposed method (blue) shows reduced landmark alignment error in this dataset compared to statistical baselines. (**Center**) Visual Similarity (SSIM): The structural preservation scores are competitive, particularly in expressions with large deformations like "Surprise". (**Right**) Mesh Smoothness: The higher smoothness scores suggest that the IK-based geometric regularization effectively mitigates high-frequency surface noise relative to displacement-based approaches.

4.7. Discussion

The experimental validation on the JAFFE dataset elucidates a fundamental dichotomy in facial modeling: the tension between semantic editability and biomechanical fidelity. To explicitly contextualize our contribution within this landscape, Table 2 contrasts the proposed framework with established paradigms. While traditional FACS-based control successfully isolates high-level affective intents (e.g., specific AU activations), relying solely on linear BlendShape interpolation inherently fails to capture the non-linear tissue dynamics of the human face. Conversely, physics-based simulations offer realism but lack real-time capability. Our Semantic-Kinematic approach effectively bridges this gap.

To substantiate the comparative ratings presented in Table 2, we define the evaluation criteria as follows:

- Physical Plausibility is evaluated based on the adherence to anatomical laws. Physics-based methods (e.g., FEA) achieve a "Very High" rating by modeling tissue mass, elasticity, and fluid dynamics. Our kinematic approach warrants a "High" rating because, unlike "Low" plausibility linear models that ignore structure entirely, it enforces strict geometric constraints, specifically, segment lengths and joint limits effectively eliminating surface tearing artifacts without the computational overhead of dense tissue simulation.
- Editability is measured by the semantic disentanglement of the control space. Our framework achieves "High" editability through explicit, localized Action Unit parameterization, allowing users to precisely modulate specific muscles. This contrasts with "Low" editability in neural renderers, which often rely on entangled latent representations ("black boxes") where targeted anatomical corrections are difficult.

Geometric Regularization via Inverse Kinematics. Our results indicate that the introduction of the Inverse Kinematics (IK) layer functions effectively as a non-linear geometric regularizer. By imposing segment-length constraints and inter-regional coupling on the mesh deformation, the IK solver restricts the solution space to anatomically plausible manifolds. This mitigates the "mask-like" discontinuities typical of vertex-level interpolation, particularly in high-curvature regions like the nasolabial folds and the brow-eye interface.

Complementarity to Data-Driven Approaches. Crucially, our framework is not positioned to replace modern neural rendering pipelines, but rather to provide a physically grounded "anatomical skeleton" that data-driven methods often lack. While end-to-end

https://doi.org/10.3390/electronics15030558

neural models excel at photorealism, they frequently operate as "black boxes." Our comparative experiments further empiricalize this distinction. Unlike statistical models such as FLAME or DECA, which are constrained by the variance of their training data (often leading to higher RMSE in outlier expressions), our IK-driven approach dynamically adapts to the target geometry. By explicitly parameterizing facial biomechanics, Facial IK-AUG offers a transparent control layer. This structure ensures that synthesized expressions remain structurally valid even when extrapolating beyond the training data, addressing the robustness issues common in pure learning-based approaches.

Limitations and Future Work. While our framework demonstrates improved geometric coherence, we acknowledge certain limitations. First, our evaluation primarily relies on the JAFFE dataset, which consists of static imagery. While sufficient for validating geometric constraints and topological consistency, it does not fully capture high-frequency temporal dynamics found in continuous video. Second, the current quantitative evaluation is constrained by the lack of ground-truth kinematic data for facial micro-expressions. Future work will focus on integrating this kinematic rig with dynamic 3D datasets (e.g., BU-4DFE) and exploring learnable parameters for the IK solver to automate the calibration process.

Table 2. Comparative positioning of Facial IK-AUG against standard facial synthesis paradigms. Our method offers a balanced solution for embodied AI, ensuring both interpretability and physical plausibility at low computational cost.

Method	Interpretability	Physical Plausibility	Comp. Cost	Editability
Pure FACS/BlendShape	High (Semantic)	Low (Linear Interpolation)	Very Low	High
Physics-based (FEA)	High	Very High (Tissue Dynamics)	Very High	Low
Neural/Black-box	Low (Latent)	Variable	High (GPU)	Low
Ours (Facial IK-AUG)	**High**	**High (Kinematic Constraints)**	**Low**	**High**

5. Conclusions

This study addresses the structural limitations of traditional parametric facial models by proposing the Facial IK-AUG framework, a unified architecture that embeds biomechanical constraints directly into the semantic control loop. The core premise of this work is that expressive realism in digital avatars requires more than just accurate texture or shape matching; it demands a synthesis pipeline that respects the anatomical connectivity of facial features. By integrating a specialized Inverse Kinematics solver with standard FACS-driven BlendShapes, the proposed "Semantic-Kinematic Loop" effectively bridges the gap between abstract affective intent and physical surface realization.

The experimental results confirm that modeling facial regions as coupled kinematic chains significantly enhances geometric continuity compared to standard linear interpolation methods. The IK refinement process automatically corrects surface tearing and volume loss artifacts in complex expressions—such as the antagonistic muscle actions found in anger or surprise—without requiring computationally prohibitive physics simulations. This approach ensures that the synthesized mesh deformations remain anatomically plausible while strictly adhering to the input Action Unit parameters. Furthermore, the framework extends beyond mere reconstruction by enabling physically grounded data augmentation. Through the systematic perturbation of solver parameters, such as damping factors and chain stiffness, the system generates a manifold of valid expression variants, thereby providing a robust mechanism for enriching training datasets used in downstream affective computing tasks.

Ultimately, this work establishes a scalable methodology for "physically grounded embodiment." By shifting the focus from purely visual reproduction to structurally con-

https://doi.org/10.3390/electronics15030558

strained synthesis, the Facial IK-AUG framework offers a resilient foundation for the next generation of embodied agents, capable of delivering high-fidelity, interpretable, and adaptable facial interactions in diverse virtual environments.

Author Contributions: Conceptualization, X.L.; methodology, B.W.; software, B.W.; validation, B.W., X.L., L.J. and J.S.; formal analysis, B.W.; investigation, B.W.; resources, L.J.; data curation, B.W.; writing original draft preparation, B.W.; writing review and editing, B.W. and J.S.; visualization, B.W.; supervision, X.L.; project administration, X.L.; funding acquisition, L.J. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JKA and its promotion funds from KEIRIN RACE.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here [30–32].

Acknowledgments: The authors would like to thank the reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Duan, J.; Yu, S.; Tan, H.L.; Zhu, H.; Tan, C. A survey of embodied AI: From simulators to research tasks. *IEEE Trans. Emerg. Top. Comput. Intell.* **2022**, *6*, 230–244. [CrossRef]
2. Li, S.; Deng, W. Deep facial expression recognition: A survey. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1195–1215. [CrossRef]
3. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
4. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
5. Tewari, A.; Fried, O.; Thies, J.; Sitzmann, V.; Lombardi, S.; Sunkavalli, K.; Martin-Brualla, R.; Simon, T.; Saragih, J.; Nießner, M.; et al. State of the art on neural rendering. *Comput. Graph. Forum* **2020**, *39*, 701–728. [CrossRef]
6. Daněček, R.; Black, M.J.; Bolkart, T. EMOCA: Emotion driven monocular face capture and animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, CO, USA, 2–6 June 2022*; IEEE: New York, NY, USA, 2022; pp. 20311–20322.
7. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
8. Ekman, P.; Friesen, W.V.; Hager, J.C. *Facial Action Coding System: The Manual*, 2nd ed.; Research Nexus: Salt Lake City, UT, USA, 2002.
9. Weise, T.; Bouaziz, S.; Li, H.; Pauly, M. Realtime performance-based facial animation. *ACM Trans. Graph. (TOG)* **2011**, *30*, 1–10. [CrossRef]
10. Cao, C.; Weng, Y.; Lin, S.; Zhou, K. FaceWarehouse: A 3D facial expression database for visual computing. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 413–425. [PubMed]
11. Cui, Z.; Kuang, C.; Gao, T.; Talamadupula, K.; Ji, Q. Biomechanics-guided facial action unit detection through force modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023*; IEEE: New York, NY, USA, 2023; pp. 8694–8703.
12. Martinez, B.; Valstar, M.F.; Jiang, B.; Pantic, M. Automatic analysis of facial actions: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2113–2133. [CrossRef]
13. Ponton, J.L.; Yun, H.; Aristidou, A.; Andujar, C.; Pelechano, N. Sparseposer: Real-time full-body motion reconstruction from sparse data. *ACM Trans. Graph. (TOG)* **2023**, *43*, 1–14. [CrossRef]
14. Li, T.; Bolkart, T.; Black, M.J.; Li, H.; Romero, J. Learning a model of facial shape and expression from 4D scans. *ACM Trans. Graph. (ToG)* **2017**, *36*, 1–17. [CrossRef]
15. Racković, S.; Li, C.; Sinisalo, O. Refined Inverse Rigging: A Balanced Approach to High-fidelity Blendshape Animation. *arXiv* **2024**, arXiv:2401.16496. [CrossRef]

16. Zhuang, Y.; Zhu, H.; Sun, X.; Cao, X. MoFaNeRF: Morphable facial neural radiance field. In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 268–285.
17. Galanakis, S.; Gecer, B.; Deng, J.; Zafeiriou, S. 3DMM-RF: Convolutional radiance fields for 3d face modeling. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–6 January 2023; pp. 3536–3547.
18. Ichim, A.E.; Kadleček, P.; Kavan, L.; Pauly, M. Phace: Physics-based face modeling and animation. *ACM Trans. Graph. (ToG)* **2017**, *36*, 1–14. [CrossRef]
19. Sumner, R.W.; Popović, J. Deformation transfer for triangle meshes. *ACM Trans. Graph. (TOG)* **2004**, *23*, 399–405. [CrossRef]
20. Garrido, P.; Valgaerts, L.; Sarmadi, H.; Steiner, I.; Varanasi, K.; Perez, P.; Theobalt, C. Reconstruction of personalized 3D face rigs from monocular video. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–15. [CrossRef]
21. Feng, Y.; Feng, H.; Black, M.J.; Bolkart, T. Learning an animatable detailed 3D face model from in-the-wild images. *ACM Trans. Graph. (TOG)* **2021**, *40*, 1–13. [CrossRef]
22. Kim, H.; Garrido, P.; Tewari, A.; Xu, W.; Thies, J.; Niessner, M.; Pérez, P.; Richardt, C.; Theobalt, C. Deep video portraits. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–14. [CrossRef]
23. Thies, J.; Zollhofer, M.; Stamminger, M.; Theobalt, C.; Nießner, M. Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2387–2395.
24. Wood, E.; Baltrušaitis, T.; Hewitt, C.; Dziadzio, S.; Cashman, T.J.; Shotton, J. Fake it till you make it: Face analysis in the wild using synthetic data alone. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3681–3691.
25. Qiu, H.; Yu, B.; Gong, D.; Li, Z.; Liu, W.; Tao, D. Synface: Face recognition with synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10880–10890.
26. Bertiche, H.; Madadi, M.; Escalera, S. Cloth3d: Clothed 3d humans. In *Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 344–359.
27. Chandran, P.; Bradley, D.; Gross, M.; Beeler, T. Semantic deep face models. In *Proceedings of the 2020 International Conference on 3D Vision (3DV), Virtual Event, Japan, 25–28 November 2020*; IEEE: New York, NY, USA, 2020; pp. 345–354.
28. Kollias, D.; Schulc, A.; Hajiyev, E.; Zafeiriou, S. Analysing affective behavior in the first abaw 2020 competition. In *Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020*; IEEE: New York, NY, USA, 2020; pp. 637–643.
29. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5203–5212.
30. Lyons, M.J.; Kamachi, M.; Gyoba, J. *The Japanese Female Facial Expression (JAFFE) Database*; Department of Psychology, Kyoto University: Kyoto, Japan, 1998. Available online: https://zenodo.org/records/14974867 (accessed on 24 January 2026).
31. Lyons, M.J.; Kamachi, M.; Gyoba, J. Coding Facial Expressions with Gabor Wavelets (IVC Special Issue). *arXiv* **2020**, arXiv:2009.05938. [CrossRef]
32. Lyons, M.J. "Excavating AI" Re-excavated: Debunking a Fallacious Account of the JAFFE Dataset. *arXiv* **2021**, arXiv:2107.13998. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Beyond "One-Size-Fits-All": Estimating Driver Attention with Physiological Clustering and LSTM Models

Juan Camilo Peña [1], Evelyn Vásquez [1], Guiselle A. Feo-Cediel [1], Alanis Negroni [2] and Juan Felipe Medina-Lee [2,*]

[1] Department of Electrical and Computer Engineering, Faculty of Engineering, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR 00680, USA; juan.pena3@upr.edu (J.C.P.); evelyn.vasquez@upr.edu (E.V.); guisellealejandra.feo@upr.edu (G.A.F.-C.)

[2] Department of Computer Science and Engineering, Faculty of Engineering, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR 00680, USA; alanis.negroni@upr.edu

* Correspondence: juan.medina26@upr.edu

Abstract

In the dynamic and complex environment of highly automated vehicles, ensuring driver safety is the most critical task. While automation promises to reduce human error, the driver's role is shifting to that of a teammate who must remain vigilant and ready to intervene, making it essential to monitor their attention level. However, a significant challenge in this domain is the considerable inter-individual variability in how people physiologically respond to cognitive states, such as distraction. This study addresses this by developing a methodology that first groups drivers into distinct physiology-based clusters before training a predictive model. The study was conducted in a high-fidelity driving simulator, where multimodal data streams, including heart rate variability and electrodermal activity, were collected from 30 participants during conditional-automated driving experiments. Using a time-series k-means clustering algorithm, the researchers successfully partitioned the drivers into clusters based on their physiological and behavioral patterns, which did not correlate with demographic factors. Then, a Long Short-Term Memory model was trained for each cluster, which achieved similar predictive performance compared to a single, generalized model. This finding demonstrates that a personalized, cluster-based approach is feasible for physiology-based driver monitoring, providing a robust and replicable solution for developing accurate and reliable attention estimation systems.

Keywords: autonomous driving; attention level; time series clustering; k-means; LSTM

1. Introduction

Safety in autonomous driving scenarios has become a constantly growing field of research in recent years [1,2], reflecting a growing interest in understanding the human factors that influence this activity. An important element is the reaction time of the drivers, which largely depends on maintaining an adequate level of attention [1,3–5]. This relationship between attention and reaction time is particularly relevant in situations where drivers temporarily divert their attention from driving [6], since a reduction in attention increases response time and increases the risk of incidents when regaining control of the vehicle [6,7], making it essential to have methods that allow monitoring and anticipating the driver's attention level [1,7–9].

Methods proposed in the literature for determining attention levels still pose key challenges. Image-based approaches struggle to interpret data when dynamic conditions, such as weather, lighting, or ambient motion, are present. In contrast, physiological

data allow direct assessment of the driver's internal state without relying on external factors. However, these types of signals present high individual variability [1,9–11], since the human body's reactions to mental stress are highly subjective and depend on the characteristics of each person [8,9]. This variability makes it challenging to build universal models and raises the need for approaches capable of handling such heterogeneity.

1.1. Similar Studies

The use of recurrent neural networks, particularly Long Short-Term Memory (LSTM), has established itself as a robust method for analyzing and predicting time series in driver monitoring. Several studies have used LSTM-based models to estimate driver states such as distraction or attentional engagement from physiological data [12]. For example, ref. [13] evaluated distraction and drowsiness in a simulator with 45 participants, using pulse blood volume, breathing, conductance, and cutaneous temperature signals. They compared traditional classifiers with an LSTM network and found that deep spacetime representations outperformed classical methods. Likewise, ref. [14] proposed an architecture with attention and Bidirectional LSTM (BILSTM) to predict performance in takeover situations during automated driving; this approach, which integrated indicators of the driver's state, the traffic environment, and personal attributes, surpassed other comparative algorithms, showing the effectiveness of explicitly modeling the temporal dependence on these predictions. However, these works demonstrate the potential of LSTM-based architectures, but they generally train a single model on grouped data from all participants, implicitly assuming homogeneity across subjects.

Physiological sensing in driving and related human–machine interaction scenarios increasingly relies on sequential models to capture the dynamics of the subject's electrocardiogram (ECG) and electrodermal activity (EDA). LSTM architectures are a natural fit for these data due to their ability to model long-range temporal dependencies and nonstationarity in autonomic responses. Recent studies show that Bayesian optimization can efficiently tune LSTM hyperparameters and thereby improve accuracy and robustness on physiological tasks; for example, Bayesian optimization has been used to adjust BILSTM learning rates and hidden sizes for ECG classification, and to optimize LSTM-based lead transformation models intended for wearable or reduced-lead settings [15,16]. Bayesian optimization has also proven valuable in other fields beyond biosignals, where spatiotemporal LSTM variants optimized by BO have advanced driver-state and control prediction problems such as steering-angle estimation, further underscoring the interplay between sequence models and sample-efficient hyperparameter search [17].

There are reported studies that have emphasized not only the usefulness of physiological data, but also their substantial variability between subjects, as individuals may react differently to the same stimulus [1,9–11]. In the context of driving, ref. [1] developed a multisensor system that simultaneously recorded EEG, ECG, and EDA-SPR signals during manual and autonomous driving scenarios in a simulator device. From these signals, they extracted blink frequency and beta-band power from EEG, heart rate from ECG, and RMS measures from EDA-SPR. Their results showed that manual driving elicited greater attentional engagement than autonomous driving, demonstrating that these physiological indicators reliably reflect the driver's perceived level of attention. In addition to physiological signals, behavioral indicators such as the hands on wheel (HOW) can be a valuable indicator of the driver's level of attention and willingness to regain control of the vehicle [18,19]. Research on partial collaborative systems has shown that keeping hands on the wheel improves monitoring behavior and recovery performance during automated command transitions [19].

In other application domains, hybrid approaches combining clustering techniques with LSTMs have been explored to improve model efficiency and accuracy. For example, in ref. [20], k-means clustering was applied to air quality time series to predict the average daily concentration of particles with a diameter less than 10 μm in Malaysia, reducing the number of models and the total training time. Similarly, ref. [21] proposed a scheme based on time-series clustering and sequence-to-sequence LSTM models for forecasting electrical load in homes, achieving notable improvements in efficiency and accuracy.

1.1.1. Research Gap and Contributions

While recent advanced data-driven methods have shown promising results in processing complex physiological time series for driver-state estimation, a challenge persists: inter-individual variability in physiological responses to cognitive states. Drivers exhibit unique patterns in their heart rate, electrodermal activity, and other physiological signals even under similar levels of distraction or vigilance. Most deep-learning-based studies train these models on aggregated data from diverse participants, often overlooking explicit segmentation or adaptation to individual differences. This typically results in either per-person model training (which is sometimes impractical due to a lack of enough data and cold-start problems [22]) or reliance on demographic factors that may not truly capture underlying psychophysiological similarities. Consequently, a significant gap remains in developing scalable, effective methodologies that explicitly account for inherent physiological differences among drivers to group them. This approach would enable the training of more specialized, accurate predictive models. Specifically, the literature lacks a comprehensive strategy that systematically integrates physiological time-series clustering as a preliminary, adaptive step before training deep learning models for personalized, precise attention level prediction in autonomous driving contexts.

The main contribution of this study is the introduction of a novel methodology that systematically groups drivers into different clusters based on their physiological and behavioral time-series patterns, moving beyond generic or demographically driven classifications. By aggregating data from multiple individuals who exhibit similar underlying psychophysiological response patterns, the trained models can learn more generalizable and robust representations of attention within each cluster. This approach mitigates data scarcity for any individual, reduces the computational burden of managing countless unique models, and provides a scalable solution for adding new drivers by assigning them to an appropriate pre-trained cluster. Furthermore, we empirically discuss the capabilities of customized models applied to the identified physiological clusters, representing a notable advancement over conventional global models. This work provides practical evidence that intrinsic physiological patterns, rather than broad demographic factors, are more relevant for determining how individuals exhibit and process varying levels of attention, offering insight for the design of adaptive and personalized driver-monitoring systems.

1.1.2. Approach

Our approach integrates two key steps: (i) using a time-series k-means algorithm to group participants according to their unique physiological and behavioral patterns; and (ii) training cluster-specific LSTM models to capture more representative dynamics of attention. This alternative methodology, which uses signals such as ECG, EDA, and hands-on-wheel data as features, effectively complements global and personalized approaches by providing a practical, replicable solution that accounts for inter-individual variability. To evaluate the strategy's effectiveness, its performance was compared with three alternative scenarios: a global model trained on all participants, models trained on

principal component analysis (PCA)-based clusters, and models trained on clusters defined by demographic characteristics.

1.1.3. Research Questions and Hypotheses

Having analyzed the methods used for monitoring a driver's attention level, it was observed that the grouping of physiological signals has rarely been considered a preparatory step for predictive attention models. Therefore, this study is based on the following research questions:

- R1: Does clustering of physiological time series using time-series k-means offer an advantage in predicting the level of attention compared to global or random clustering?
- R2: Can it be shown that membership in the obtained clusters does not directly depend on demographic variables such as age, gender, or years of experience, but instead on the observed physiological and behavioral dynamics?

To answer these questions, this study proposes an experimental design in which physiological signals (ECG and EDA) are collected, a clustering scheme is employed, and cluster-specific LSTM models are trained. The results are then compared against two reference conditions: (i) a global model trained on all participants and (ii) models trained on randomly generated clusters.

Consequently, this study investigates the following hypotheses:

- H1: Cluster-specialized models, defined using time-series k-means, achieve higher predictive performance than global models and models trained with random clustering.
- H2: Participant assignment to clusters is not explained by demographic characteristics, but by physiological and behavioral patterns.

The structure of this work is organized as follows: Section 2 describes in detail the implemented methods used in this study, including a description of the driving simulation experimental environment and the different scenarios designed, as well as the structure of the testing protocol, the technologies and equipment used, and the characterization of the participants. Next, Section 3 details the preprocessing methodology, which covers the initial data processing, the preprocessing of physiological characteristics, attention levels, and the hands-on-wheel metric, as well as the application of time-series k-means clustering, the design of the LSTM architecture, and Bayesian hyperparameter optimization. The results are presented in Section 4, including the findings from the clustering process, Bayesian optimization, and LSTM models. Subsequently, Section 5 discusses the performance of the LSTM models and the identified clusters. Finally, Section 6 summarizes the main findings and proposes future directions to strengthen the applicability of physiology-based clustering in real-life driving scenarios.

2. Methodology

The experimental study was conducted in a controlled driving simulation environment that allowed the incorporation of various scenarios, enabling engagement with different automated modes and attention levels. The aim was to simulate realistic urban and semi-urban driving scenarios with sudden events that required the driver to regain control, while systematically capturing multimodal data to assess reaction time, distraction level, and behavioral performance.

Throughout the test sessions, a structured experimental protocol was applied to a group of 30 participants, with their physiological signals and attentional states continuously recorded as they engaged in various driving situations. During the experiment, unexpected events were introduced to trigger takeover requests (TORs), including unsafe pedestrian crossings, traffic-impeding vehicles, and the sudden appearance of lane-blocking obstacles,

enabling real-time evaluation of drivers' reactions and physiological responses. To capture individual variability, the collected physiological data were used to generate participant clusters through unsupervised learning techniques. Based on these clusters, two machine learning training strategies were implemented and compared: (i) a global model, trained on the entire dataset as if the population were homogeneous, and (ii) segmented models, trained independently for each cluster.

Subsequently, the clusters were analyzed based on similarities in physiological parameter patterns to identify distinct behavioral trends and improve the interpretability of driver responses. Finally, the performance of both modeling strategies was compared based on their classification accuracy and their ability to adapt to the participants' physiological diversity.

The remainder of this section outlines the test protocol's structure, the experimental setup, the technologies and equipment employed, the data collected, and the characteristics of the participant group involved in the study.

2.1. Structure of the Test Protocol

To guarantee the quality and integrity of the physiological and attentional data collected during simulated driving tests, a structured experimental protocol was implemented. A total of $N_S = 30$ participants were subjected to this procedure, which was organized into four distinct phases, as shown in Figure 1.

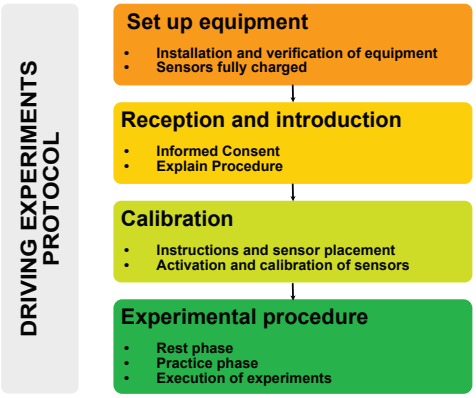


Figure 1. General diagram of the driving experiment protocol.

In the first phase, the functionality and connectivity of all critical systems were verified. This includes the devices listed in Section 2.2, which are responsible for data acquisition and logging.

The second phase focused on introducing the participants to the experiments. Individuals were presented with the informed consent form and received detailed instructions regarding the experiment. After consent was granted, the research team proceeded with the calibration and placement of physiological sensors.

The fourth phase of the protocol was dedicated to executing the main experiment. This phase begins with 5 min of rest to establish baseline data for each subject, followed by a 5 min experimental stage of manual driving practice in the environment. The primary objective of this practice is to allow participants to familiarize themselves with the sensitivity of the simulator's controls—brake, accelerator, and steering wheel—and to understand the interaction with the human–machine interface.

Following this familiarization period, participants undertook experimental tests to perform takeover requests during autonomous driving segments. When a takeover maneuver was required, an audible and visible alert on a tablet positioned next to the steering wheel prompted the participant to assume control. Once the critical situation was avoided,

the human had to return the vehicle to autonomous driving mode. The experiments are comprehensively detailed in Table 1, which summarizes the experimental setup used in the study and provides an overview of the key characteristics of each evaluated scenario. The experiments were conducted on two different maps in the CARLA (Car Learning to Act) version 0.9.15 simulator to analyze participant behavior under increasingly complex conditions. The table includes information on the urgency levels of TORs (low, medium, and critical) and the concentration levels evaluated for each. The scenarios included sudden lane changes, pedestrian crossings, road obstructions, traffic accidents, and dynamic moving obstacles, designed to induce varying degrees of cognitive load and attention demands. Some of these scenarios are illustrated in Figure 2. The hazardous objects in the scenes are highlighted with green boxes: a sudden lane change, a fire truck in the middle of the road, and a sudden pedestrian crossing. Additionally, the table provides the approximate time of each TOR activation and the total duration of each experimental session, offering a comprehensive overview of the test design, sequence, and complexity.

Table 1. Summary of experiments by topology.

Exp.	CARLA Town	Environment	TOR Urgency Level	Concentration Level	Scenarios	TOR Time	EXP. Time
1	4	Semi-urban	Low/high	Focused/distracted	Abrupt lane change, pedestrian crossing, obstacle in the road/stopped vehicle	1′21″ 3′28″ 4′37″	5′16″
2	5	Urban	Low/high/critical	Focused/distracted	Obstruction by stopped vehicle, traffic accident, dynamic obstacle in lane	1′43″ 3′40″ 6′00″	7′14″

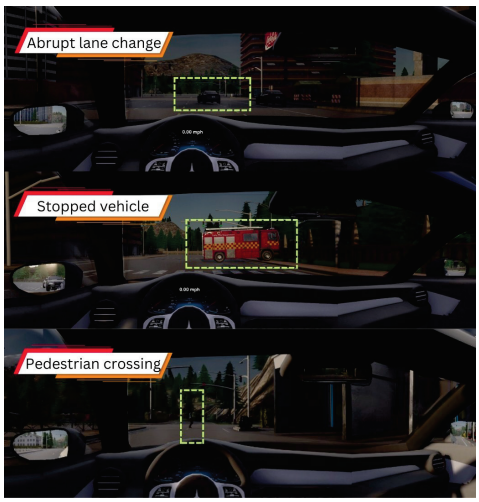


Figure 2. Examples of adverse driving events used to trigger takeover requests (TORs).

To systematically evaluate the participants' attention levels during driving experiments, a dual-task experimental paradigm was implemented, combining the primary driving simulation with secondary tasks. To induce varying levels of cognitive load, participants interacted with a mobile word search application at predefined stages of the experiment. This task simulated realistic driver distraction and provided a richer dataset for characterizing fluctuations in attention levels, complementing the primary performance metrics collected.

In parallel, a visual attention assessment was integrated into the driving environment. Geometric figures—circles, squares, triangles, and stars—were projected onto the side screens in distinct colors (red, green, blue, yellow, violet, and cyan) at 20 s intervals, each displayed for 2.5 s. Participants were instructed to respond by manipulating a switch lever to indicate whether the star stimulus appeared on the upper (Figure 3a) or lower region (Figure 3b) of the respective screen. This action evaluated their level of attention and

generated a detailed record in the Attention Logs, which will be described in Section 3.4. This procedure allowed for continuous, time-resolved assessment of visual detection performance and attentional engagement. Similar approaches have been successfully applied in driver workload studies, where secondary visual tasks combined with psychophysiological markers provide reliable indicators of attention allocation and cognitive state [23].



(**a**) (**b**)

Figure 3. Attention level assessment through stimulus detection and gear-shift response. (**a**) Shifter up. (**b**) Shifter down.

Additionally, takeover requests were introduced to evaluate participants' readiness to regain control during automation disengagement events. TORs were triggered seven seconds before a critical scenario, communicated via a tablet-based human–machine interface (HMI). Participants were required to press the "Take Control" button on the touchscreen to manually navigate around the obstacle and subsequently reactivate autonomous mode after clearing the hazard. These TOR events, combined with concurrent visual and mobile tasks, generated a multimodal dataset that linked attentional performance and physiological indicators.

2.2. Technologies and Equipment Used

This section describes the complete experimental setup used in this work, integrating software, wearable sensors, and computational systems. Figure 4 provides an overview of the entire architecture, illustrating the interaction between simulation tools, physiological monitoring devices, and computing platforms, and summarizes the multi-stage experimental framework designed to capture the complexity of the interaction between humans and vehicles in autonomous driving. Each module, from vehicle control to physiological detection, was developed to interact systematically, enabling analysis of the user's attention dynamics and interaction with the experiment.

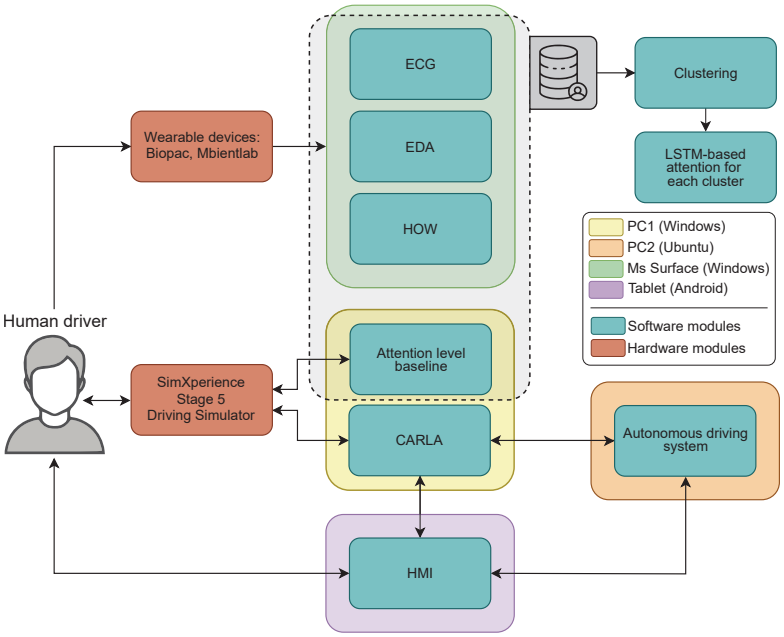


Figure 4. Simulation setup general diagram.

The experiments were conducted using the SimXperience Stage 5 simulation platform (SimXperience, Akron, OH, USA), which incorporates a multi-axis motion system, surround sound, and feedback from the steering wheel and seat. This setup aims to increase drivers' immersion in the physical sensations of real-life driving, allowing them to experience acceleration, braking, and cornering with high fidelity. Using an environment with these characteristics is essential for ensuring that the physiological responses recorded in the ECG and EDA accurately reflect driving conditions similar to those in a real-life environment [24,25]. Figure 5 illustrates the whole experimental setup, including the driving simulator, the human–machine interface for control interaction, the MetaWear IMU sensors for motion tracking, and the Biopac Bionomadix system (BIOPAC Systems, Inc., Goleta, CA, USA) for physiological signal acquisition (a demonstration video of the system is available at https://www.youtube.com/watch?v=Xa7oBMYvrkg (accessed on 19 October 2025).). This integrated platform enabled synchronized collection of multimodal data streams necessary for evaluating driver behavior and physiological responses under semi-automated driving conditions.



Figure 5. Simulation platform and data acquisition setup used in the experiments.

The experiments were conducted in CARLA, an open-source simulator designed to support the development, training, and validation of autonomous systems in highly configurable virtual environments. CARLA is powered by the Unreal Engine 4.26.2 graphics engine, which enables high-fidelity rendering and realistic physics. It supports a wide range of sensor modalities (e.g., cameras, LiDAR, IMU, GPS) and environmental conditions (e.g., lighting, weather), making it a widely adopted tool in the autonomous vehicle research community. More details about CARLA and its documentation can be found in [26].

The experimental environment was built using CARLA towns 4 and 5, which provide diverse urban and semi-urban settings. These maps were digitized in the lanelet2 format [27], allowing the autonomous driving system to navigate them with precise definitions of lane geometries, regulatory elements, and routing logic.

While the combination of the SimXperience Stage 5 motion platform and the CARLA simulator provides a high-fidelity, immersive, and highly controllable environment, essential for the safe and repeatable investigation of human-automation teaming scenarios, it is important to acknowledge certain limitations that may impact the validity of our findings. The simulated environment cannot fully replicate the unpredictable complexities of real-world traffic, the detailed sensory input from genuine vehicle dynamics, or the full spectrum of environmental conditions (e.g., unexpected glare, complex weather) that influence a driver's behavior and physiological state. Moreover, participants' conscious

awareness of being in a simulation, however realistic, might subtly alter their risk perception, engagement levels, or propensity to intervene compared to real-world driving. Nevertheless, the ability of our setup to precisely and safely induce hazardous conditions, along with its capacity for multimodal physiological data collection in a controlled setting, provides a valuable, ethically sound platform for systematically exploring the fundamental dynamics of human–AI interaction under critical scenarios.

Physiological monitoring was conducted using the Biopac BioNomadix system (Goleta, CA, USA), a wireless wearable device capable of capturing physiological signals at a sampling rate of 2000 Hz, enabling high-resolution continuous data acquisition. This setup included:

- Electrocardiogram (ECG): Data were acquired using Biopac Bionomadix wireless device RSP&ECG (Model BN-RSPEC-TGED-T) sensors, placed on the chest in a Lead I configuration. These sensors were connected via BN-EL45-LEAD3 cables to electrodes pre-gelled with conductive gel. The red, black, and white electrodes corresponded to the positive (E^+), negative (E^-), and ground (GND) inputs, respectively (Figure 6 chest placement).
- Electrodermal Activity (EDA): Measurements were taken using a Bionomadix PPG&EDA (Model BN-PPGED-T) sensor. This sensor, connected via a BN-EDA25-LEAD2 cable, utilized electrodes affixed to the index and middle fingers of the left hand (Figure 6 hand placement).

To detect if the driver has their hands on the steering wheel, two inertial motion tracking MetaWear IMU sensors were positioned on both wrists to collect 3-axis accelerometer data, 3-axis gyroscope data, and Quaternion orientation values (Figure 6 hand placement).

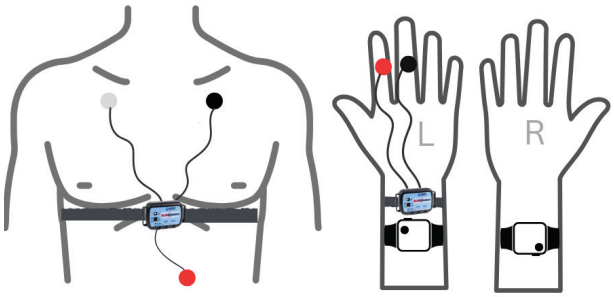


Figure 6. Location of devices on the user's body.

2.3. Description of Data Acquisition and Computational Equipment

A Microsoft Surface tablet was used to acquire physiological data using AcqKnowledge (v6.0), a BIOPAC data acquisition and visualization software. To replicate realistic self-driving behavior during the experiment, CARLA was integrated with a state-of-the-art autonomous driving system [28], which controlled the vehicle's trajectory, speed, and reaction to traffic scenarios in the virtual environment. This autonomous driving system was executed on a 15.6″ Dell Precision 5570 mobile (Dell Technologies Inc., Round Rock, TX, USA) workstation equipped with an Intel Core i7-12800H vPro processor, 32 GB DDR5 4800 MHz RAM, and an NVIDIA RTX A1000 GPU with 4 GB GDDR6 dedicated memory. On the other hand, the CARLA simulator was running on an Alienware Aurora R16 desktop computer (Dell Technologies Inc., Round Rock, TX, USA) with an Intel Core i9-14900F processor (24 cores, 32 threads, base 2.0 GHz), 32 GB DDR5 RAM, 2 TB SSD storage, and an NVIDIA GeForce RTX 4070 SUPER GPU with 12 GB GDDR6 dedicated memory. Additionally, a separate touchscreen Google Pixel tablet (Google LLC, Mountain View, CA, USA) (10.95″ LCD, 2560 × 1600 resolution, 267 ppi) served as the HMI device.

A multimodal dataset was collected during each experimental session to comprehensively assess driver behavior, physiological state, and attentional performance during TOR scenarios. The data acquisition system was designed to synchronize and log high-resolution information from various sources, enabling robust training and evaluation of machine learning models.

The core data streams include the following:

- **Physiological Signals (ECG and EDA):** The signals were acquired in real time. Data were streamed continuously during the experiment and stored in raw .csv format for offline analysis. A custom-designed graphical interface was used for live monitoring, buffering, and data export.
  Signal processing and feature extraction were performed using NeuroKit2 (v0.2.11) (Python) and supplementary MATLAB (R2020b, Update 7) routines. The preprocessing pipeline included noise reduction, signal normalization, and event detection—specifically R-peak identification for ECG and skin conductance response (SCR) peak detection for EDA. Signals were segmented into fixed-length windows (1 min), and features were extracted to quantify physiological states associated with stress, arousal, and cognitive workload.
  The extracted features include widely used psychophysiological metrics. For example, RMSSD (Root Mean Square of Successive Differences) and SDNN (Standard Deviation of NN intervals) are time-domain indicators of heart rate variability (HRV), commonly associated with parasympathetic nervous system activity and emotional regulation. Similarly, the LF/HF ratio (Low-Frequency to High-Frequency power ratio) is a frequency-domain HRV metric frequently used as a marker of sympathetic–vagal balance. In electrodermal activity, SCL (skin conductance level) reflects tonic arousal, while SCR (skin conductance response) quantifies phasic responses to discrete stimuli or cognitive shifts. These features serve as core inputs for downstream modeling of driver attention and state classification. For a more detailed overview of the physiological metrics and their interpretation, see [29–31].
- **Inertial Motion Data (IMU):** MetaWear sensors (MbientLab Inc., San Jose, CA, USA) on both wrists recorded accelerometer, gyroscope, and quaternion values from both wrists at 10 Hz. This data enabled hands-on-wheel detection, processed via LSTM-based deep learning models. The IMU logs were labeled with binary contact states and exported as structured CSV files. More details on this model can be found in [32].
- **Attention Logs:** As illustrated in Figure 3, participants responded to visual stimuli through a gear-shift lever, with each interaction automatically recorded in .json files including both response time and accuracy. Each recorded response was subsequently scored and categorized into discrete attention levels—focused, semi-distracted, or distracted. These attention labels were then resampled and temporally synchronized with the physiological signals at 1s intervals to facilitate data fusion and training processes.
- **Simulator and Vehicle State Logs:** The simulator environment recorded continuous data on vehicle trajectory, speed, TOR events, traffic objects, and user control inputs. These logs were structured via a Lightweight Communications and Marshaling (LCM) framework and timestamped for alignment.

All data files were time-synchronized using UNIX timestamps and exported in standardized CSV and JSON formats. The final dataset comprises both raw and processed versions of each modality, enabling flexible downstream analysis and integration. This rich dataset supports supervised model training, attention-state classification, and the development of a real-time driver-monitoring system. Table 2 details the characteristics extracted from each process.

Table 2. Features extracted by signal type.

Signal Type	Feature Category	Extracted Features
ECG	HRV–Time Domain	MeanNN, SDNN, RMSSD, MedianNN, MadNN, MinNN, MaxNN, pNN50
	HRV–Frequency Domain	LF, HF, LF/HF ratio
	HRV–Nonlinear	SD1, SD2, SD1/SD2 ratio, Approximate Entropy (ApEn)
EDA	Tonic (SCL)	Mean SCL, Max SCL, Min SCL, SCL Slope, SCL Std, Variance, Energy
	Phasic (SCR)	SCR Count, SCR Rate, Mean Amplitude, Max Amplitude, Amplitude Std
	Statistical	Mean Diff, Max Diff between successive values
IMU (MetaWear)	Raw Motion Data	Linear Acceleration (X, Y, Z), Angular Velocity (X, Y, Z), Quaternion (W, X, Y, Z)
	Behavioral State Label	Hands-on-Wheel binary label (1 = contact, 0 = no contact)
Attention Logs	Attentional Score	Coded values based on visual stimulus (0 = distracted, [0.6–1.0] = semi-distracted, 2 = focused)
Simulator Logs	Vehicle and Scenario State	Speed, Control Mode, Planner State, Obstacle Proximity, TOR trigger timestamps

2.4. Participants

The experimental protocol was conducted with $N_S = 30$ volunteer participants, following a public call for participation distributed among the university community. Interested individuals contacted the research team and completed the informed consent process in accordance with institutional ethical protocols. The sample consisted of a balanced mix of individuals, representing a diverse range of ages, genders, and driving experiences. Specifically, 63.3% of participants identified as male and 36.7% as female. Their ages ranged from 22 to 71 years, with a mean of 36 and a standard deviation of 12.7, reflecting a diverse adult population.

Driving experience varied considerably across participants, ranging from 0.6 to 55 years, with a mean of 14.95 years and a median of 8.5 years. This variability enabled the analysis to capture the behavioral and physiological responses of both less- and more-experienced drivers. Only 13.3% of participants reported prior experience with autonomous vehicle systems, making the findings particularly relevant for those facing real-time takeover situations for the first time.

Participants were divided equally into two experimental groups (GR1 and GR2), each comprising 50% of the total sample. Both groups presented similar demographic characteristics, ensuring a balanced distribution of variables such as gender, age, driving experience, and prior experience with autonomous vehicles. Table 3 provides detailed descriptive statistics.

Each group was exposed to a different attention level configuration, as described in Table 4. This attention-state configuration was chosen for each group and experiment to simulate realistic and varied situations that could arise in conditioned autonomous driving. This configuration enables analysis of how changes in attention level, whether gradual or sudden, affect the driver's physiological response to critical events, such as a traffic accident. This intentional variability in the data helps assess the physiological responses of different profiles to cognitive changes.

Table 3. Demographic summary statistics by group (General, GR1, GR2).

Variable	Category	General (100%)			GR1			GR2		
		FR#	%	Stats	FR#	%	Stats	FR#	%	Stats
Gender	Male (M)	19	63.3	–	10	66.7	–	9	60	–
	Female (F)	11	36.7	–	5	33.3	–	6	40	–
Age (years)	Min	-	-	22	-	-	23	-	-	22
	Max	-	-	71	-	-	71	-	-	70
	Mean	-	-	36	-	-	36.33	-	-	35.87
	Median	-	-	33.5	-	-	33	-	-	34
	SD	-	-	12.7	-	-	12.73	-	-	12.62
Driving Experience (Years)	Min	-	-	0.6	-	-	1	-	-	0.6
	Max	-	-	55	-	-	50	-	-	55
	Mean	-	-	14.95	-	-	14.13	-	-	15.77
	Median	-	-	8.5	-	-	7	-	-	13
	SD	-	-	14.33	-	-	13.81	-	-	14.78
Previous Autonomous Driving Experience	Yes (Y)	4	13.3	–	2	13.3	–	2	13.3	–
	No (N)	26	86.7	–	13	86.7	–	13	86.7	–

FR# = frequency count; Stats = statistical measures (mean, median, min, max, SD).

Table 4. Distribution of attention configuration by group and experiment.

Group	Experiment	Attention Configuration
1	1	FOCUSED → DISTRACTED → DISTRACTED (FDD)
	2	DISTRACTED → FOCUSED → FOCUSED (DFF)
2	1	DISTRACTED → DISTRACTED → FOCUSED (DDF)
	2	FOCUSED → FOCUSED → DISTRACTED (FFD)

3. Data Processing and Modeling Framework

This section details the end-to-end modeling pipeline. First, the dataset is organized and standardized by curating features and applying global min–max normalization to the physiological channels (HOW remains binary). Second, this section explains how the attention target was defined for participants during the experiment. Third, subject grouping is derived with time-series k-means based on physiological parameters to form coherent clusters. Finally, an LSTM sequence-to-label architecture is specified and its hyperparameters are tuned via Bayesian optimization.

3.1. Physiological Feature Preprocessing

A multimodal dataset was created combining peripheral physiological features extracted from raw electrocardiogram and electrodermal activity with a binary driver interaction signal indicating hands on wheel. For each subject $s_i \in S,\ i = \{1, \ldots, N_s\}$, continuous ECG and EDA streams were segmented into non-overlapping windows of 1 s to derive a set of 29 physiological features in a time interval $t = [0, T_i]$, with T_i being the duration of all the driving experiments for subject i combined. In parallel, the HOW channel was sampled at 1 Hz and encoded as $h_t \in \{0, 1\}$ (0: no hands on the wheel; 1: hands on the wheel). Let $N_F = 30$ be the feature dimensionality (29 physiological features plus HOW). Each row at time t is therefore an N_F-dimensional feature vector

$$\mathbf{z}_{it} \;=\; \begin{bmatrix} \mathbf{x}_{it} & h_{it} \end{bmatrix}, \quad \mathbf{x}_{it} \in \mathbb{R}^{1\times 29},\ h_{it} \in \{0,1\},\ \mathbf{z}_{it} \in \mathbb{R}^{1\times N_F}.$$

where $\mathbf{x}_{it} \in \mathbb{R}^{29}$ are the physiological features and h_{it} the HOW indicator for the subjects s_i.

To mitigate inter-subject amplitude variability while keeping feature scales comparable, a global min–max normalization was applied to the 29 physiological features and mapped them to $[-1, 1]$:

$$\tilde{x}_j \;=\; 2\,\frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)} \;-\; 1, \quad j = \{1, \ldots, 29\}, \tag{1}$$

leaving the binary HOW feature unscaled. This $[-1, 1]$ scaling is standard in physiological affect/stress modeling because it bounds the dynamic range and speeds up optimization [33,34].

3.2. ANOVA

To identify the physiological features most strongly associated with variations in driver attention, a one-way ANOVA was conducted between the two attention classes. This statistical test quantifies the extent to which each feature exhibits significant differences between the two conditions, providing an interpretable measure of discriminative power. Features with lower p-values indicate greater statistical relevance to the classification task, guiding subsequent comparisons and the selection of machine learning models. The ranked features according to their p-values are summarized in Table 5.

Table 5. ANOVA results showing the most discriminative physiological features ranked by p-value.

Feature	p-Value
prediction	1.88×10^{-69}
eda_energy	9.39×10^{-62}
ecg_HRV_MinNN	2.60×10^{-61}
eda_min_scl	4.12×10^{-52}
eda_max_diff	1.00×10^{-51}
eda_mean_scl	7.41×10^{-47}
ecg_HRV_MedianNN	1.54×10^{-41}
eda_max_scl	5.63×10^{-32}
ecg_HRV_RMSSD	5.75×10^{-23}
ecg_HRV_SD1	9.79×10^{-23}
ecg_HRV_HF	2.67×10^{-21}
ecg_HRV_SD1SD2	8.28×10^{-21}
ecg_HRV_SDNN	1.17×10^{-19}
eda_mean_diff	2.94×10^{-18}
ecg_HRV_SD2	6.26×10^{-17}
eda_scr_rate	2.20×10^{-15}
eda_scr_count	2.20×10^{-15}
ecg_HRV_MeanNN	7.06×10^{-15}
ecg_HRV_MaxNN	4.81×10^{-13}
ecg_HRV_LF	2.80×10^{-12}
eda_scl_slope	1.09×10^{-11}
ecg_HRV_ApEn	6.18×10^{-6}
eda_scl_std	3.46×10^{-5}
eda_scr_amplitude_mean	3.97×10^{-2}
eda_variance	5.18×10^{-2}
ecg_HRV_MadNN	1.18×10^{-1}
ecg_HRV_LFHF	2.11×10^{-1}
eda_scr_amplitude_max	2.28×10^{-1}
eda_scr_amplitude_std	4.34×10^{-1}
ecg_HRV_pNN50	8.54×10^{-1}

3.3. Feature Selection

In the dataset selection stage, feature dimensionality was determined through the clustering strategy described in Section 3.5 and by training the models detailed in Section 3.6.2. Based on the comparative performance across different configurations, it was concluded that using 30 features yields the most consistent and robust results, as summarized in Table 6.

Table 6. Accuracy (%) by feature count.

LSTM	24 Features	27 Features	30 Features
Global	95.34%	94.50%	95.15%
Cluster 1	75.58%	78.47%	94.08%
Cluster 2	68.17%	75.05%	89.22%

3.4. Attention Level Preprocessing

A data preprocessing pipeline was designed using the responses recorded to each attention stimulus during the experiments. It should be noted that each stimulus was presented at 20-s intervals, giving the participant a maximum of 4 s to select the correct option. The results obtained for Subject 6 are shown in Figure 7, where the red dots reflect the stimuli without response or incorrect responses, and the green dots represent correct responses. Additionally, the average correct response time for this subject was 2.26s. According to this group and the experiment (Group 1, experiment 2), the suggested attention setting for this driver corresponded to distracted–focused–focused (see Table 4).

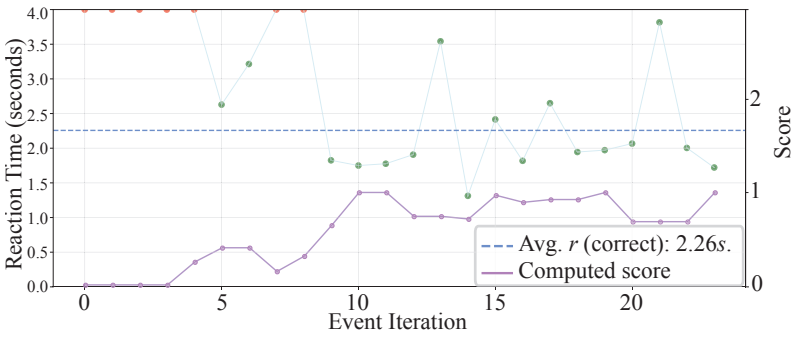


Figure 7. Driver S_6 reaction times per event and average response of experiment 2.

To transform the raw responses into labels suitable for the model, a preprocessing pipeline was designed with four main steps: (1) calculate the characteristic reaction time of each subject from the correct responses, (2) determine, in each trial, if the response occurred within the individual reference time, (3) apply a penalty when the selection was incorrect, or there was a late response with respect to its individual reference, and (4) define the final attention labels at two levels: focused and distracted.

First, each reaction time $r_{n,i}$ (bounded to the interval $[0, RT_{\max}]$ seconds, with $RT_{\max} = 4$) is normalized with respect to the individual average $\bar{r}$, ensuring that the score reflects the relative performance of the driver rather than the overall speed of the group, as detailed in Equation (2):

$$\bar{r}_i = \frac{1}{N_A} \sum_{n=1}^{N_A} r_{ni} \qquad \text{score}(r_{ni}, \bar{r}) = \frac{RT_{\max} - r_{ni}}{RT_{\max} - \bar{r}_i}. \tag{2}$$

where N_A is the number of attention level evaluations requested during the experiment.

Once the individual yield is calculated, each attempt is labeled by a variable Y_{in}, where i denotes the subject and n denotes the attempt. This variable can take three values: correct

where the star was selected in adequate time, incorrect where a different figure was chosen, or failed where no response was recorded within 4 s, as shown in Equation (3).

$$y_{in} = \begin{cases} 0 & = missed, \\ \text{score}(r_{in}, \bar{r}) & = correct, \\ \alpha \cdot \text{score}(r_{in}, \bar{r}) & = incorrect, \end{cases} \tag{3}$$

where $\alpha = 0.5$ is a weighting coefficient introduced to penalize incorrect responses. This choice ensures that errors, even if produced with fast reaction times, contribute less to the overall score than correct answers, but are not entirely disregarded; this computed score can be seen reflected in Figure 7 , where each response is compared with its score for each stimulus.

From y_{in}, two attention levels are defined in Equation (4):

$$a_{in} = \begin{cases} y_{in} < \tau & = distracted, \\ y_{in} \geq \tau & = focused, \end{cases} \tag{4}$$

where the threshold $\tau = 0.6$ was established by analyzing the distribution of scores obtained only from correct answers, which is illustrated in Figure 8. This distribution is visualized by the light blue histogram bars, which indicate the frequency of each score. Fitting a Gaussian function (orange line with $\mu = 0.91$ $\sigma = 0.20$) to showed that 93.5% of the values were above the threshold, which is represented by the dashed blue vertical line at $\tau = 0.6$ and highlighted by the orange shaded area. This value enabled distinguishing between correct answers associated with high concentration (higher scores) and those slower than the average response time, which could indicate a lower level of sustained attention. It also balanced the proportions of instances labeled as focused and distracted to avoid bias towards a single class.

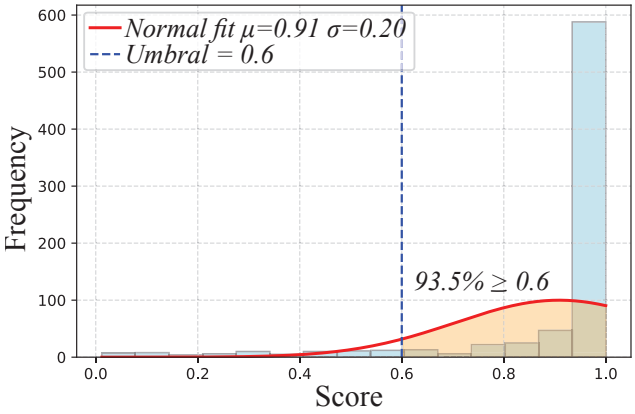


Figure 8. Attention score distribution.

In this way, each attempt is categorized at one of two levels of attention: focused when the answer was correct and in adequate time, or distracted when the answer was late, incorrect, or nonexistent. This scheme ensures that the data are consistent, equivalent, and aligned with the cognitive relevance of each moment, as illustrated in Figure 9.

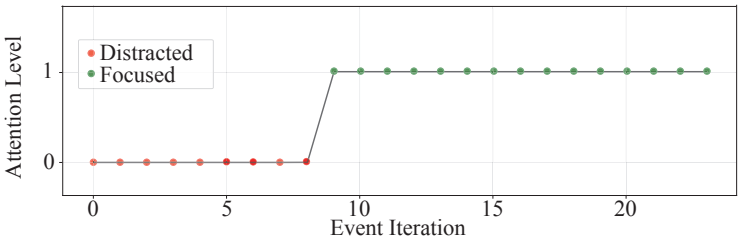


Figure 9. Driver S_6 reaction times per event and average response of experiment 2.

Hands-on-Wheel Prediction Feature Preprocessing

The hands-on-wheel (HOW) signal was incorporated as an auxiliary feature derived from a previously developed LSTM-based detector [32].

Let $b = 0, 1, 2, \ldots$ index samples with $t_b = b \cdot 0.1$ s. Define the HOW label $h_b \in \{0, 1\}$ as

$$h_b = \begin{cases} 1, & \text{if hands are on the wheel at } t_b, \\ 0, & \text{otherwise.} \end{cases}$$

Due to the LSTM's 10 Hz output rate, the HOW stream was downsampled to 1 Hz to match the 1 s resolution of the physiological dataset: each 1 s window averages its ten samples and thresholds the mean at 0.5 to yield a binary label (1 = hands on wheel, 0 = otherwise).

The resulting data was aligned by timestamp, and the HOW series was appended as an additional column to the feature matrix used for model training.

3.5. Time-Series k-Means Clustering

To analyze each driver's behavior, a time-series clustering strategy was employed, as it leverages the sequential structure of the data and yields more representative groupings than strategies based on single values, such as classic k-means [35]. In particular, the time-series k-means algorithm was used, which transforms the sequences into a set of clusters defined by centroids that represent similar temporal patterns [20].

Before applying the clustering algorithm, each series was normalized to the range $[-1, 1]$. Then, the data from experiments 1 and 2 for each subject were combined to form a time series, as shown in Equation (5).

$$\hat{x}_{tix} = 2 \cdot \frac{x_{tix} - \min\limits_{i \in S}(x_{tix})}{\max\limits_{i \in S}(x_{tix}) - \min\limits_{i \in S}(x_{tix})} - 1 \tag{5}$$

where x_{tix} is the value of the feature (x) for subject (i) at time (t), and $\hat{x}_{tix}$ represents the normalized value. The operators $\min(x_{tix})$ and $\max(x_{tix})$ represent the global minimum and maximum found for the feature, respectively. These were calculated considering all subjects $i \in S$ and all times t. Comparability between characteristics with different magnitude ranges is ensured this way.

In this study, drivers completed the experiments at different time intervals, generating sequences of varying lengths. For this reason, instead of using a Euclidean distance metric, dynamic time warping (DTW) was chosen, a measure that allows aligning time series of different durations or with possible distortions on the time axis, although its computational cost is higher.

The k-means clustering algorithm seeks to assign data to clusters and position centroids such that the sum of all squared distances is as small as possible [36], as shown in its objective equation:

$$J_i = \min_{\{u_{ik}\}} \left[\sum_{i=1}^{N_S} \sum_{x=1}^{N_F-1} \sum_{k=1}^{N_K} u_{ik} \left\| \hat{x}_{tix} - c_{tkx} \right\|^2 \right], \tag{6}$$

where $\hat{x}_{tix}$ is the normalized value of the physiological parameter x for subject i at time t, c_{tkx} is the corresponding value of the centroid of cluster $\varsigma \in \mathcal{K}$, with $\mathcal{K}$ denoting the set of N_k clusters. Finally, u_{ik} is the membership indicator, which takes the value 1 if subject i belongs to cluster ς_k , $k \in [1, \ldots, N_k]$ and 0 otherwise. In this context, the objective function measures the total cost of clustering. It is obtained by minimizing the sum of the squared distances between the subjects' series and the centroids of their respective clusters.

One of the most commonly used strategies for determining the optimal number of clusters is the elbow plot, which identifies the point at which the inertia (the sum of squared intra-cluster distances) stops decreasing significantly. On the other hand, the silhouette score, which measures how compact and separated the clusters are, reaches higher values when the data are well grouped and differentiated. In this work, both metrics are employed in a complementary manner: the elbow plot is used to determine the number of clusters, and the maximum value of the silhouette index confirms the final choice.

To validate this choice of N_k, a strategy was implemented that simultaneously evaluates stability and silhouette in time-series clustering. The analysis is based on the clustering stability principle, which states that a partition is stable if it produces consistent results in the face of slight data perturbations [37]. This principle, widely demonstrated for the K-means algorithm, can be extended to TimeSeriesKMeans, given that both share a centroid-based nature [37]. Unlike supervised classification methods, where there is a "ground truth" that guides the evaluation, in unsupervised clustering, stability is considered a meta-principle that allows estimating the reliability and reproducibility of the structure found.

To generate perturbed versions of the original dataset N_S, the bootstrap method is used, taking 80% ($\lambda = 0.80$) of the data in $B = 50$ replicates. In each replicate b, a cluster C_b is obtained.

The observed stability is calculated as the average distance between all pairs of obtained clusters, using the Adjusted Rand Index (ARI) based on the intersection of the data points that coincide in the two subsamples. The observed average instability is theoretically defined in Equation (7) [37].

$$\text{Instab}(k, n) = \frac{1}{b_{\max}^2} \sum_{b,b'=1}^{b_{\max}} d(C_b, C_{b'}), \tag{7}$$

where a higher ARI value indicates greater stability and reproducibility between the obtained clusters. It is essential to normalize stability, as instability tends to increase systematically with the number of clusters (k). To correct this bias, a null baseline was used, obtained by repeating the stability calculation, comparing partitions with randomly shuffled labels but preserving the original cluster size.

From this adjustment, the final selection of K is based on two key metrics that assess the significance of the observed stability compared to this null model: the Stability Gap and the Z-score Stability. Thus, the Stability Gap is defined as the difference between the observed stability and the expected stability under the null model, as seen in Equation (8), while the Z-score Stability evaluates whether the actual stability is significantly greater than the null Equation (9). Finally, the value of K that produces the highest Z-score Stability is considered to be the most statistically significant number of clusters [37].

$$\text{Stability Gap} = \text{mean}_{\text{obs}} - \text{mean}_{\text{null}}, \tag{8}$$

$$\text{Z-score Stability} = \frac{(\text{mean}_{\text{obs}} - \text{mean}_{\text{null}})}{\text{std}_{\text{null}}}. \quad (9)$$

In this way, to define whether a new subject belongs to one of the defined clusters, the same processing flow applied during training is followed. First, physiological and behavioral signals are recorded during the experiments. Subsequently, the series are preprocessed and normalized to the range $[-1, 1]$ to ensure compatibility with the original model's scale. Finally, the normalized sequences are evaluated using the pre-trained TimeSeriesKMeans model, assigning each subject to the cluster whose centroid has the lowest dynamic time warping (DTW) distance. This procedure allows new individuals to be classified without retraining the model, preserving consistency with the original segmentation.

In addition to the time-series k-means-based strategy, an alternative approach combining principal component analysis and K-means was implemented. In this case, the physiological datasets were transformed using principal components analysis. This multivariate technique seeks to reduce the dimensionality of the data by identifying a set of uncorrelated components that account for most of the original variance [38]. This transformation allows the data to be represented in a lower-dimensional space, facilitating comparisons between subjects and the detection of overall similarity patterns. K-means was applied to this data reduction to perform clustering, using the principal components rather than the entire sequences. However, this strategy provides a general view of the data structure.

3.6. Model Architecture and Hyperparameter Optimization

This section describes the Long Short-Term Memory (LSTM) model architecture and the Bayesian optimization implemented in this study.

3.6.1. Random Forest and Support Vector Machine

Two classical machine learning models were evaluated for attention level classification: Random Forest (RF) and Support Vector Machine (SVM). The best configuration for the RF model was Splits = 25, Cycles = 150, and Learning rate = 0.01, which provided a stable balance between model complexity and performance. For the SVM model, the optimal configuration used a Radial Basis Function (RBF) kernel with BoxConstraint = 0.10 and KernelScale = 0.10, achieving consistent results across subjects.

3.6.2. Long Short-Term Memory Architecture

The attention level was modeled from physiological time series using an LSTM network, a recurrent architecture designed to capture short- and mid-term temporal dependencies in sequential data [39]. The suitability of LSTM-based pipelines for driver vigilance and drowsiness using peripheral physiology (e.g., ECG/respiration) has been demonstrated in recent work, supporting the choice of a temporal model for this task [40].

Figure 10 summarizes the end-to-end model. First, the physiological features are min–max normalized to $[-1, 1]$ as detailed in Section 3.1. The normalized sequences are then processed by an LSTM-based sequence-to-label classifier with a last-step readout and a compact feedforward head, as depicted in the figure.

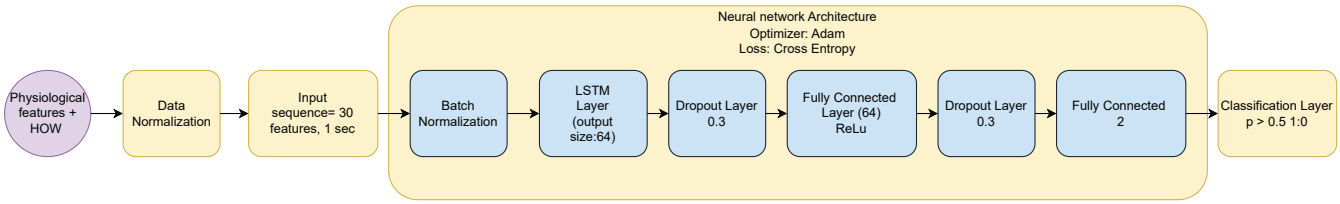


Figure 10. LSTM architecture model.

Windows were built at 1 Hz according to Welch PSD analysis, which showed that over 96% of the physiological features' spectral power lies below 0.1 Hz, confirming that 1 Hz sampling effectively captures the relevant dynamics without oversampling. Accordingly, sequences of length $L = 5$ s with a stride of $r = 1$ s were used to capture short-term temporal dependencies while preserving sufficient temporal resolution.

For each subject (s_i) with $(t = 1, \ldots, T_i)$ one-second samples and feature vectors $\mathbf{z}_{it} \in \mathbb{R}^{N_F}$, the set of valid window start indices is

$$\mathcal{Q}_i \;=\; \{\, 1,\, 1{+}r,\, 1{+}2r,\, \ldots,\, T_i - L + 1 \,\}. \tag{10}$$

Given any $q \in \mathcal{Q}_i$, the length-L window is defined as the concatenation of L consecutive time steps,

$$\mathbf{W}_{iq} \;=\; \left[\, \mathbf{z}_{iq} \,,\, \mathbf{z}_{i(q+1)} \,,\, \ldots \,,\, \mathbf{z}_{i(q+L-1)} \right] \;\in\; \mathbb{R}^{N_F \times L}. \tag{11}$$

Windows are constructed for the combined experiments of each subject (s_i) (two experiments per subject), so no window crosses subject or experiment boundaries.

The resulting sequences are processed by a unidirectional , causal LSTM with a last-step readout (Figure 10). A compact feedforward head then maps the sequence summary to two logits, which are converted to posterior class probabilities via a softmax. Batch normalization is applied at the network input to stabilize feature scales across subjects and sessions, and dropout is used around the recurrent and dense blocks to mitigate overfitting without inflating model size. Optimization was performed using categorical cross-entropy and the Adam algorithm over 30 epochs, with mini-batches of 64 and shuffling at every epoch.

To assess subject-independent performance, i.e., generalization across individuals, Leave-One-Subject-Out (LOSO) cross-validation was employed. This protocol is widely recommended in physiology-based affect and stress recognition because it mitigates optimistic bias associated with subject overlap and better reflects deployment to unseen users. Recent studies explicitly advocate for LOSO for fair evaluation in wearable physiology and emotion/stress modeling [34,41,42]. In particular, [43] demonstrated the necessity of LOSO for EEG-based disease diagnosis using a dataset of approximately 30 participants, showing that conventional k-fold cross-validation can substantially overestimate performance due to data leakage across subjects. Similarly, [44] employed a multi-input CNN-LSTM architecture for fear-level classification from EEG and peripheral physiological signals using a dataset of comparable size, emphasizing the importance of LOSO evaluation to ensure realistic cross-subject generalization. Following these recommendations, LOSO validation was applied in the present study to maximize the use of the available data while strictly preventing overlap between training and testing subjects. Accordingly, the reported results were aggregated to the arithmetic mean across all LOSO folds, with one fold per held-out subject.

After evaluating the three supervised models (Random Forest, SVM, and LSTM) with accuracies of 84.93%, 62.25%, and 95.12%, respectively, the LSTM model was selected for further analysis. Its superior performance indicates a better alignment with the temporal characteristics of the physiological data, making it the most appropriate approach for this study.

3.6.3. Bayesian Optimization

Deep learning models for time series, including LSTM architectures, are sensitive to hyperparameter choices (e.g., sequence length, hidden width, learning rate, dropout rate, and optimizer). Manual trial-and-error tuning across a large, multidimensional space is laborious and error-prone, and it often fails to identify strong configurations under realistic compute budgets. Evidence from time-series forecasting, as well as from driving

and driver-state monitoring, indicates that automated hyperparameter search enhances performance and efficiency compared to manual heuristics, thereby mitigating degradation due to suboptimal settings [17,45,46].

In advanced driving assistance and autonomous driving contexts, training deep temporal models is computationally expensive. Prior studies report that Bayesian optimization can improve spatiotemporal LSTM models for steering prediction and driver monitoring by locating better hyperparameters with fewer trials than manual search [17,46]. These characteristics motivate a sample-efficient optimizer capable of handling noisy, non-convex objectives under limited evaluation budgets.

Building on this rationale, the end-to-end pipeline (windowing, normalization, and subject-aware validation) is formalized as a black-box objective. Bayesian optimization with an expected improvement family acquisition guides the selection of each costly training run, and the resulting hyperparameters are reported both at the global level and per cluster. This procedure replaces ad hoc heuristics with a reproducible search designed to target cross-subject generalization in physiological driver-state modeling [17,45–48].

Bayesian optimization was used to automatically select the hyperparameters of the LSTM classifier shown in Figure 10. The set of hyperparameters is

$$\boldsymbol{\theta} = \{\text{window size},\ \text{LSTM layer units},\ \text{fully connected units},\ \text{dropout},\ \text{learning rate}\}.$$

Window size defines the temporal length of each input sequence, LSTM layer units determine the number of hidden units in the recurrent layer, fully connected units specify the width of the projection layer, dropout controls the fraction of activations dropped during training, and learning rate sets the step size of the Adam optimizer.

The training and validation pipeline is treated as a black-box function $O(\boldsymbol{\theta})$. For each candidate $\boldsymbol{\theta}$, the model is trained with reduced epochs and validated using subject-wise K-fold cross-validation. Let $\mathrm{Acc}_m(\boldsymbol{\theta})$ be the accuracy on fold $m = [1, \ldots, M]$. The objective is defined as the negative of the average accuracy across folds:

$$O(\boldsymbol{\theta}) = -\frac{1}{M}\sum_{m=1}^{M} \mathrm{Acc}_m(\boldsymbol{\theta}).$$

A Gaussian process surrogate is fitted to past evaluations $\{(\boldsymbol{\theta}_j, O(\boldsymbol{\theta}_j))\}$. Based on this surrogate, the expected improvement plus acquisition function selects the next hyperparameter configuration by balancing exploration and exploitation. This iterative process continues until the evaluation budget is exhausted. The best observed configuration $\boldsymbol{\theta}^\star$ is then retrained with an enlarged epoch budget, and its accuracy is reported as the final result. Once the optimal configuration is found, the model is retrained with a larger epoch budget for reporting purposes. Random seed initialization was fixed when assigning subjects to folds to enhance reproducibility.

This protocol aligns with standard practice in physiological modeling and driver-state monitoring, where inter-subject variability is substantial [48,49].

4. Results

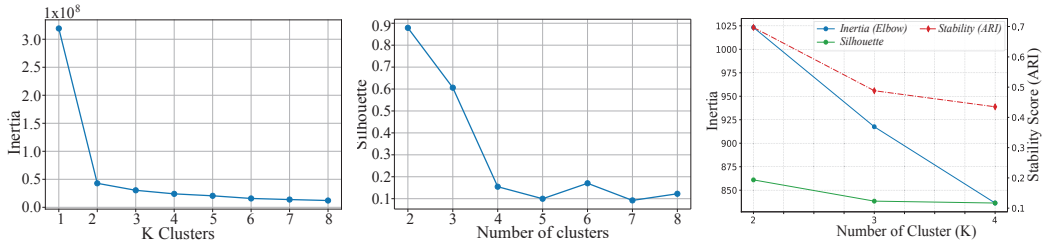
First, several machine learning models, RF, SVM, and LSTM, were compared to determine the most suitable architecture for the classification task. Once the optimal model was selected, the set of subjects was clustered based on their physiological profiles using the strategy described in Section 3.5 to obtain the optimal number of clusters and analyze the features of each of them. Then, the LSTM model is applied to each cluster. The performance of these cluster-specific models is evaluated by comparing them with that of a global model that includes all drivers in the dataset. First, the global model was trained using the original

hyperparameters; second, Bayesian optimization was applied to the global model to obtain the global reference. Finally, the physiology-based clusters are also compared against non-physiological partitions; four additional cluster pairs were defined: sex (female, male), driving experience (experienced, novice), and two random pairs.

4.1. Clustering

Once the dataset with the normalized series was prepared, the optimal number of clusters was determined. To accomplish this, the inertia obtained by varying the number of clusters from $N_k = 1$ to $N_k = 9$ was evaluated, as illustrated in Figure 11a. This upper bound was chosen because the dataset comprised 30 participants, and allowing more than 9 clusters would result in very small groups with limited representativeness. The objective of this analysis was to identify the point at which the inertia stops decreasing significantly, thus avoiding excessive oversegmentation of the data and ensuring an appropriate balance between simplicity and representativeness of the clustering model.

Figure 11a shows that, for $N_k = 2$ and $N_k = 3$, the inertia reaches relatively low values, while from $N_k > 3$, the reduction is no longer significant. For its part, the silhouette score (Figure 11b) shows a pronounced drop between $N_k = 2$ and $N_k = 3$, remaining at considerably lower values for $N_k > 3$. However, the most determining criterion comes from the stability analysis in Table 7, which shows that $N_k = 2$ presents the greatest consistency between executions (Figure 11c), indicating that the clustering structure is robust against perturbations of the dataset. Considering stability, silhouette, and inertia together, it is concluded that the optimal number of clusters for this dataset is $N_k = 2$, where the best balance between separation, compactness, and clustering reproducibility is achieved. Similarly, for the stability tests used for the dataset with noise in the physiological signals, an average stability was obtained for $k = 2$ of 0.88, for $k = 3$ of 0.504, and for $k = 4$ of 0.529.



(**a**) Inertia (**b**) Silhouette (**c**) Stability

Figure 11. Evaluation of metrics to determine the optimal number of clusters. (**a**) Inertia for different values of N_k. (**b**) Silhouette score for different values of N_k. (**c**) Stability-based evaluation.

Table 7. Clustering stability and internal validation metrics for different values of K.

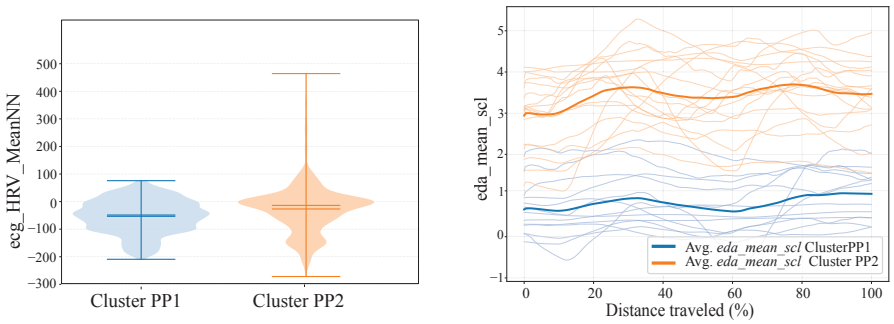
K	Silhouette	Inertia	Mean ARI	Null Mean	Gap	Z-Score
2	0.193	1023.26	0.697	0.0049	0.692	8.66
3	0.123	917.55	0.488	0.0004	0.488	6.62
4	0.117	836.34	0.435	0.0042	0.431	5.52

Once the optimal number of clusters was determined, the time-series k-means algorithm with DTW distance was applied, configured to divide the set into two groups ($N_k = 2$). The model described in Section 3.5 was trained for a maximum of 50 iterations, with a random state of 42 points to ensure stable, reproducible results. This procedure allowed each time series to be assigned a cluster label and, at the same time, to calculate the representative centroids of each group, from which Table 8 was obtained.

The clustering results identified two physiological parameter (PP) clusters (Cluster PP1 and Cluster PP2), allowing recognition of specific patterns associated with each group. An example of the physiological differences among clusters is depicted in Figure 12a. It shows how Cluster PP1 included subjects who, for the most part, presented negative mean heart rate variability values and low electrodermal conductance levels. This may indicate lower autonomic activation during the driving task for most of the subjects. In contrast, in Figure 12b, Cluster PP2 (orange line group) included participants with greater heterogeneity in their ECG values, some with elevated positive averages, and consistently higher EDA levels, reflecting a profile of greater physiological activation. Regarding demographic characteristics, both clusters included a mix of young men and women, indicating that age and driving experience were not directly associated with cluster membership. This suggests that the observed physiological dynamics primarily explain the cluster separation.

Table 8. Demographic and physiological experiment 1 and experiment 2 mean summary by cluster.

Cluster	Subjects	Demographic			ecg_HRV_MeanNN		eda_mean_scl	
		Age	Gender	Exp	Mean	Std	Mean	Std
Cluster PP 1	S03	44	M	4	−51.61	±41.51	0.14	±0.27
	S05	35	F	20	−20.05	±25.29	1.99	±0.3
	S07	42	M	24	−87.08	±30.49	0.63	±0.25
	S10	30	M	4	−136.74	±24.98	1.15	±0.15
	S12	24	M	7	−38.74	±13.67	0.14	±0.17
	S13	33	M	1	−17.34	±27.96	1.52	±0.24
	S15	31	F	3	−91.11	±37.76	3.04	±0.36
	S16	36	M	19	−63.42	±44.58	0.69	±0.13
	S18	22	M	0.6	31.16	±29.16	0.71	±0.67
	S20	40	M	15	0.62	±26.97	0.39	±0.03
	S25	70	M	55	−36.85	±59.78	0.08	±0.19
	S19	23	M	7	−9.24	±28.52	1.03	±0.29
	S29	24	M	7	24.79	±29.37	0.73	±0.26
Cluster PP 2	S01	24	M	8	−9.75	±27.6	4.0	±0.64
	S02	23	M	7	−20.24	±28.5	3.57	±0.2
	S04	49	F	33	77.95	±15.42	4.31	±0.31
	S06	25	F	1	1.44	±31.64	3.84	±0.5
	S08	31	M	11	−58.55	±18.56	2.45	±0.25
	S11	35	M	5	−67.06	±23.52	4.24	±0.11
	S14	29	M	13	16.04	±44.02	3.02	±0.21
	S17	25	M	9	−13.65	±23.63	3.3	±0.3
	S21	35	F	15	106.31	±74.8	3.29	±0.28
	S22	56	M	40	20.45	±12.05	3.16	±0.47
	S23	28	M	7	−100.08	±36.36	1.89	±0.25
	S24	30	F	3	−109.87	±65.33	1.23	±0.26
	S26	39	F	23	−8.4	±27.52	2.09	±0.28
	S27	34	F	1	−2.96	±20.54	3.32	±0.21
	S28	71	F	50	−149.21	±29.44	3.18	±0.9
	S29	51	F	30	1.85	±14.07	2.75	±0.99
	S30	44	F	26	111.02	±292.66	4.53	±0.56



(**a**) (**b**)

Figure 12. Comparison of physiological responses between clusters. (**a**) Distribution of ecg_HRV_MeanNN by cluster. (**b**) Evolution of eda_mean_scl along the cluster path.

4.2. Results of Bayesian Optimization

The Bayesian optimization described in Section 3.6.3 was applied consistently at both the global and cluster levels. For all cases, the hyperparameters shown in Figure 10 served as the base configuration, and the procedure refined these settings to accommodate between-subject and subpopulation variability.

The hyperparameters are tuned as depicted in Table 9: (i) *Window size* (sequence length/temporal receptive field), (ii) *LSTM units* (recurrent hidden width), (iii) *Dropout*, (iv) *Fully Connected Layers* (projection width), and (v) *Learning rate* (searched on a logarithmic scale due to multiplicative effects). This design keeps the search expressive yet tractable for expensive sequence models.

Table 9. Hyperparameter configurations per model.

Parameter	Global	Global BO	Cluster PP 1	Cluster PP 2	Young	Old	Male	Female	PCA 1.1	PCA 1.2	Novice	Experienced
Window size	5	15	13	11	10	14	5	9	11	10	5	14
LSTM units	64	80	117	127	120	48	102	114	37	120	101	115
Dropout	0.3	0.31278	0.33888	0.22343	0.49752	0.3815	0.22298	0.23995	0.20032	0.49752	0.28891	0.38473
Fully Connected Layers	64	114	84	24	99	123	116	33	93	99	49	86
Learning rate	1.0×10^{-3}	5.3×10^{-3}	1.0×10^{-2}	9.3×10^{-3}	4.2×10^{-3}	9.8×10^{-3}	9.3×10^{-3}	9.6×10^{-3}	9.1×10^{-3}	4.2×10^{-3}	8.1×10^{-3}	6.5×10^{-3}

4.3. LSTM Results

This subsection reports the LSTM results for the global model and for the cluster-specific models under a standard LOSO evaluation protocol. For each cluster, a separate Bayesian optimization was conducted within the same search space (LSTM layer units, fully connected units, dropout, learning rate) and with the same training protocol (Table 9).

Results are presented at two levels. First, the confusion matrices and derived metrics for the global model are given, as well as clusters PP1 and PP2. Secondly, the percentage differences relative to the global baseline are described to quantify subject-wise changes attributable to cluster specialization.

Figure 13a shows the confusion matrix of the global LSTM before Bayesian optimization, yielding an overall accuracy of 88.83%. From these counts, the per-class F_1 scores are 90.93% for Attention (class 1) and 85.46% for No-attention (class 0), with a macro F_1 of 88.20%. Qualitatively, the model is stronger on class 1, with some class 0 samples misclassified as class 1.

Figure 13b shows the confusion matrix of the LSTM after Bayesian optimization. The pooled accuracy is 94.00%. From these counts, the per-class F_1 scores are 95.8% for Attention (class 1) and 85.7% for No-attention (class 0), yielding a macro F_1 of 90.8%. Compared with the pre-optimization model (accuracy 88.83%, macro $F_1 = 88.20\%$), both overall accuracy and macro F_1 show clear improvements, while the total misclassifications

($FP + FN$) slightly decrease from 1076 to 1015, confirming that the Bayesian-optimized configuration enhances class discrimination and overall generalization.

Figure 13c reports PP1 under pooled LOSO aggregation. The model achieves a global accuracy of 94.17%. Class-wise F_1 scores are 92.72% for class 0 (No-attention) and 95.14% for class 1 (Attention), yielding a macro F_1 of 93.93%. The small gap between classes suggests a well-balanced behavior.

Figure 13d summarizes PP2 in the same pooled setting. The model attains 89.53% global accuracy and records F_1 scores of 85.97% for class 0 and 91.64% for class 1, for a macro F_1 of 88.81%. Performance remains solid, with a clearer advantage on the Attention class relative to No-attention.

Figure 14 presents the mean subject-wise accuracy with 95% confidence intervals obtained under the LOSO validation scheme for the global model and the two cluster-specific models, together with the subject-matched comparison of each cluster relative to the global configuration. The global model (blue) reached an overall accuracy of 93.33% with a narrow confidence interval of ±2.08 percentage points, showing stable performance across all subjects. Cluster 1 (green) achieved 94.08% with a confidence interval of ±2.40 percentage points, while Cluster 2 (orange) reached 89.22% with a wider interval of ±4.44 percentage points. Although the global model exhibits the most stable average behavior, the cluster-based configurations capture subject-specific dynamics that can enhance performance within more homogeneous participant groups.

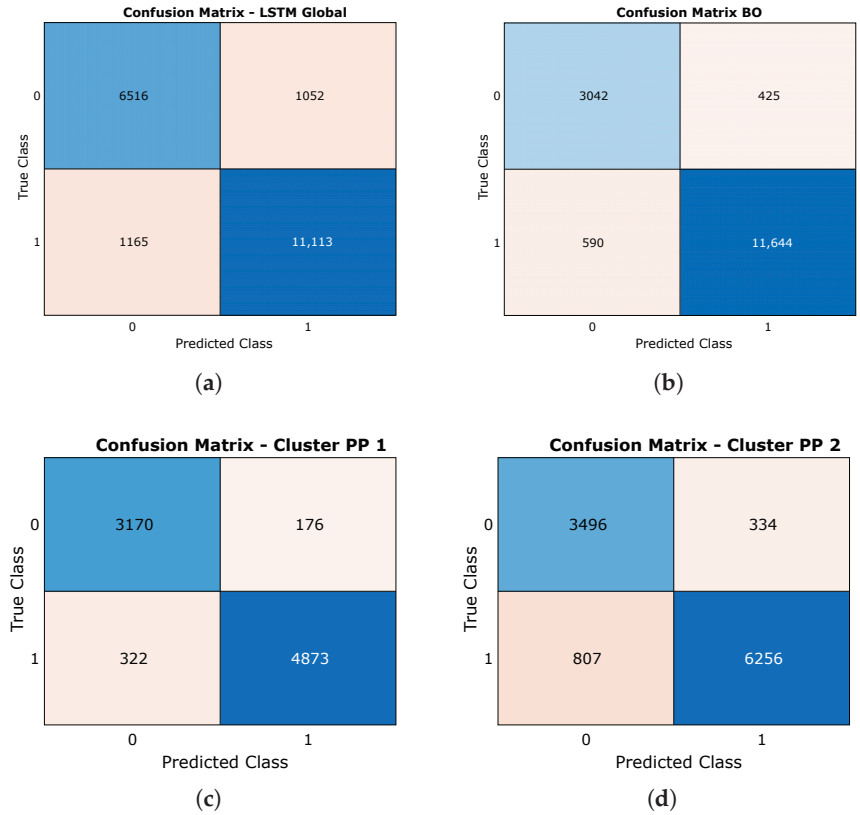


Figure 13. Confusion matrix. (**a**) LSTM global before Bayesian optimization. (**b**) LSTM global after Bayesian optimization. (**c**) Cluster PP 1 LSTM. (**d**) Cluster PP 2 LSTM.

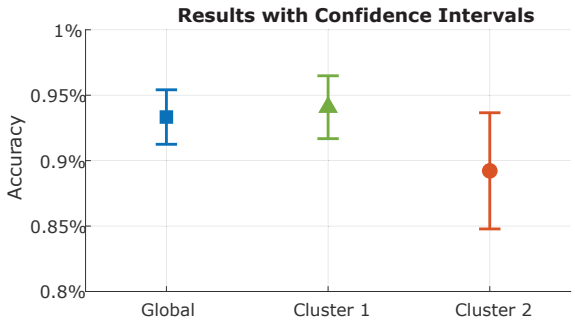


Figure 14. Confidence intervals for global model, Cluster 1 and Cluster 2.

Figure 15 presents an analysis that compares each cluster against a subject-matched global reference. Each bar shows the difference in average LOSO accuracy for the subjects in that cluster when evaluated with their cluster model versus the same subjects evaluated with the global model. Positive values indicate that the cluster outperforms the subject-matched global reference; negative values indicate the opposite.

Colors denote partition families: blue corresponds to physiology-driven clusters, yellow to gender-based clusters, green to driving-experience clusters, purple to PCA-based clusters, and red to age partitions. Within each color, the two bars represent the two clusters that form the pair. The dashed blue rectangle highlights the physiology-driven pair, which achieves the best performance among the other partition families shown. The blue dashed square highlights the pair of clusters identified using the TimeSeriesKMeans method.

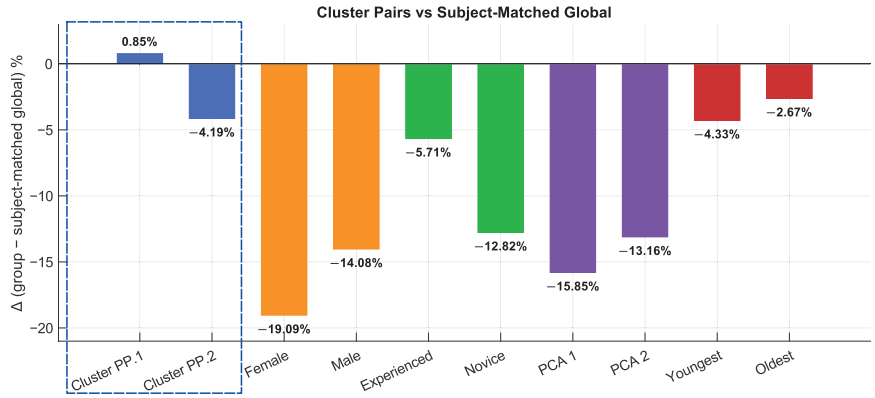


Figure 15. Cluster-based model performance vs. subject-matched global model performance.

To evaluate the robustness of the LSTM model to signal noise, Additive White Gaussian Noise (AWGN) was systematically introduced to the ECG and EDA signals in the dataset. Based on the literature, a fixed Signal-to-Noise Ratio (SNR) of 12 dB was applied to the ECG signals. For the EDA signals, a fixed SNR of 15 dB was used, which falls within the range of low- to high-SNR conditions explored in previous studies [50–52]. After augmenting the dataset with this controlled noise, the performance of the LSTM model was evaluated using the same architecture as with the original, clean signals. The evaluation yielded nearly identical results as is shown in Figure 16, with accuracies of 93.24% for Cluster 1, 85.51% for Cluster 2, and a global accuracy of 91.36%.

Table 10 shows that the cluster-specific models present a lower computational load per physiological window compared with the global configuration. Cluster 1 and Cluster 2 reduce the total number of operations by approximately 25–27%, mainly in the recurrent (LSTM) component, which accounts for more than 98% of the total FLOPs in all models. Despite this reduction, both clusters maintain sub-millisecond inference times ($\approx$0.10 ms per sample) and comparable throughput ($\sim$9.6–9.7k samples/s) to the global model while

preserving an almost identical memory footprint (∼0.3 MB). These results indicate that the cluster architectures provide a more compact and computationally efficient representation of physiological dynamics within groups of subjects sharing similar ECG and EDA characteristics.

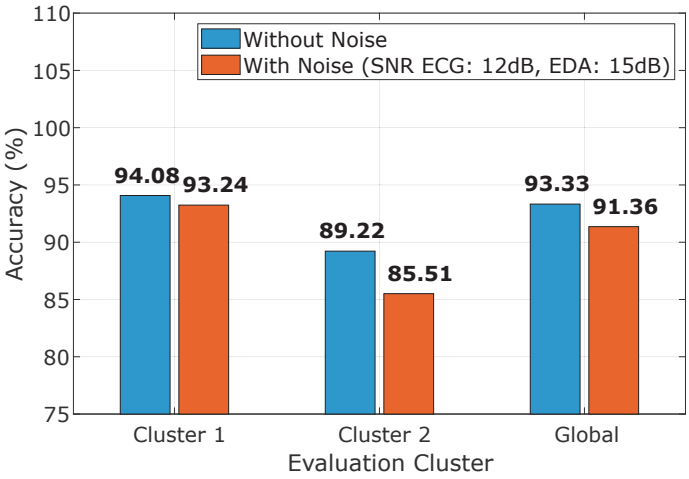


Figure 16. Model performance comparison with a higher noise level.

Table 10. Computational performance comparison between global and cluster-specific LSTM models.

Metric	Global	Cluster 1	Cluster 2
Total parameters	83,446	79,406	83,446
Model size (KB)	325.96	310.18	325.96
FLOPs per sample	2,415,012	1,821,636	1,772,660
Avg. inference time per sample (ms)	0.099	0.104	0.103
Throughput (samples/s)	10,148.1	9621.5	9703.3

5. Discussion

The decision to retain two clusters for the analysis was based on a balance between validating internal metrics and ensuring the stability of the clustering model. Joint analysis of the elbow graph (Figure 11a) and the silhouette index (Figure 11b) showed that $N_K = 2$ simultaneously achieved a significant reduction in inertia and the maximum silhouette value, demonstrating a well-defined structure and a clear separation between physiological groups. This consistency is complemented by statistical criteria that ensure the stability and generalization capacity of the predictive model. In this configuration, the clusters maintain sufficient size and adequate internal variability.

Furthermore, the stability observed for $k = 2$ was maintained even when the physiological ensembles were generated with different levels of random noise. This result indicates that the separation between groups is not due to minor signal fluctuations or instrumental perturbations, but rather, it reflects a consistent physiological organization. In theoretical terms, this resistance to noise suggests that the centroids of both groups are sufficiently distant in feature space to maintain their statistical identity, even in the face of local variations in signal amplitude or temporal dynamics. Additionally, this stability analysis was repeated on the same dataset after introducing a controlled level of noise into the physiological signals.

It is important to analyze whether the dataset size also influenced the definition of the number of clusters. With a sample of this size, increasing k results in groups that are too small, with few subjects per category, reducing internal diversity and the likelihood of identifying consistent physiological patterns. Conversely, maintaining two clusters allows for a balance between group separation and the representativeness of the signals within

each cluster. In the future, as the dataset expands and a more diverse sample becomes available, it will be necessary to re-evaluate the clustering structure. A larger number of participants could reveal additional physiological subgroups, allowing for a more detailed characterization of attentional responses and increasing the predictive capacity of the models. However, in the current context, empirical and statistical evidence clearly supports that $K = 2$ represents the most coherent, robust, and generalizable solution.

It is essential to highlight that the groups obtained do not match the demographic variables considered in the experiments, such as age, sex, or driving experience. This reinforces the idea that the separation is primarily due to physiological patterns. In this sense, the two clusters show apparent differences in stimulus reception; Cluster PP1 shows negative values for feature `ecg_HRV_MeanNN` and low levels of `eda_mean_scl`, suggesting a more stable profile but lower electrodermal activation. On the other hand, Cluster PP2 exhibits greater dispersion in `ecg_HRV_MeanNN`, with subjects reaching very high or very low values, accompanied by consistently higher `eda_mean_scl` levels. This pattern could indicate that participants in Cluster PP2 were more receptive to the experimental stimuli or experienced greater physiological activation, an aspect worth exploring in future studies. On the other hand, applying dimensionality reduction using PCA before clustering resulted in more clusters, but with lower physiological coherence. This is explained because PCA summarizes the signal information into values that represent its overall variance, without fully preserving the temporal structure or dynamic transitions.

The global LSTM optimized model establishes a firm reference, and the clustering experiments indicate that specialization yields selective benefits. One cluster consistently attains the highest subject-wise performance among the partitions considered. In contrast, its companion cluster averages below the global baseline, mainly because a small subset of participants within that group performs poorly. Demographic splits and random pairs do not match the best cluster.

The Bayesian optimization stages further indicate that each cluster benefits from different regions of the hyperparameter space, suggesting that model capacity, regularization, and sequence length interact with the temporal statistics of the cluster. This interaction helps explain why a single global setting cannot be optimal for all subjects and why cluster-specific tuning attains selective gains without relying on cluster size. The Bayesian optimization budget can be deepened by expanding the hyperparameter search space and increasing the iteration count, enabling a more thorough exploration of sequence length, hidden width, depth, dropout, learning rate, and fusion choices while keeping the evaluation procedure unchanged.

A limitation of this study is the number of participants, which may constrain the full representation of physiological diversity in the dataset. While the current results demonstrate consistent performance and meaningful clustering behavior, a larger sample size could further enhance cluster robustness and model generalization. Increasing the participant pool in future studies would allow a broader range of physiological variability, leading to more stable and representative clusters. In parallel, exploring architectures tailored to each cluster and conducting more extensive hyperparameter searches across sequence length, network depth and width, regularization, and fusion strategies may further enhance cluster performance without altering the evaluation protocol.

6. Conclusions

With respect to H1 (cluster models can optimize and improve the global baseline), the evidence supports the use of cluster specialization to raise subject-wise performance when the partition captures consistent physiological structure. Gains are not universal across all

clusters, indicating that improvement is conditional on the quality of the grouping rather than guaranteed by mere partitioning.

Regarding H2 (physiology-based grouping can be superior to demographic or random partitions), the results favor clusters derived from physiological features over demographic or random splits. The advantage appears to arise from the structure in the ECG/EDA dynamics and the hands-on-wheel signal, rather than from group size or coverage, which explains why non-physiological partitions tend to reproduce the global behavior.

As a practical takeaway, clustering is a viable approach to improving attention estimation when subjects share compatible physiological dynamics, while the global model remains a safe fallback in other cases.

Author Contributions: Conceptualization, J.F.M.-L., J.C.P., and E.V.; methodology, J.F.M.-L., J.C.P., E.V., and G.A.F.-C.; software, J.C.P., E.V., and A.N.; validation, J.C.P., E.V., G.A.F.-C., A.N., and J.F.M.-L.; formal analysis, J.F.M.-L., J.C.P., E.V., and A.N.; investigation, J.F.M.-L., J.C.P., E.V., A.N., and G.A.F.-C.; resources, J.F.M.-L., J.C.P., E.V., and G.A.F.-C.; data curation, J.C.P., E.V., and A.N.; writing—original draft preparation, J.C.P., E.V., G.A.F.-C., and J.F.M.-L.; writing—review and editing, J.F.M.-L., J.C.P., E.V., A.N., and G.A.F.-C.; visualization, J.C.P., E.V., G.A.F.-C., A.N., and J.F.M.-L.; supervision, J.F.M.-L.; project administration, J.F.M.-L.; funding acquisition, J.F.M.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received financial support from the NSF EPSCoR Center for the Advancement of Wearable Technologies (CAWT) under Grant No. OIA-1849243. It was also supported in part by the CAHSI–Google Institutional Research Program.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Puerto Rico (protocol code 2024100041, 25 October 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the privacy terms agreed with the Institutional Review Board.

Acknowledgments: During the preparation of this study, the authors used Python Neurokit2 for data processing, Matlab 2020 for training learning models, Python 3 for clustering and data visualization, and CARLA 0.9.15 for driving simulation. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BILSTM	Bidirectional LSTM
BO	Bayesian optimization
CARLA	Car Learning to Act
DTW	Dynamic time warping
ECG	Electrocardiogram
EDA	Electrodermal activity
FR	Frequency
HF	High frequency
HMI	Human–machine interface
HOW	Hands on wheel

HRV	Heart rate variability
IMU	Inertial motion data
LCM	Lightweight Communications and Marshaling
LF	Low frequency
LSTM	Long Short-Term Memory
LOSO	Leave-One-Subject-Out
PCA	Principal component analysis
PP	Physiological parameter
RMSSD	Root Mean Square of Successive Differences
SCL	Skin conductance level
SCR	Skin conductance response
SDNN	Standard deviation of NN intervals
TOR	Takeover requests

References

1. Aminosharieh Najafi, T.; Affanni, A.; Rinaldo, R.; Zontone, P. Driver attention assessment using physiological measures from EEG, ECG, and EDA signals. *Sensors* **2023**, *23*, 2039. [CrossRef] [PubMed]
2. Veluchamy, S.; Michael Mahesh, K.; Muthukrishnan, R.; Karthi, S. HY-LSTM: A new time series deep learning architecture for estimation of pedestrian time to cross in advanced driver assistance system. *J. Vis. Commun. Image Represent.* **2023**, *97*, 103982. [CrossRef]
3. Kashevnik, A.; Lashkov, I.; Ponomarev, A.; Teslya, N.; Gurtov, A. Cloud-based driver monitoring system using a smartphone. *IEEE Sens. J.* **2020**, *20*, 6701–6715. [CrossRef]
4. Schwarz, C.; Gaspar, J.; Miller, T.; Yousefian, R. The detection of drowsiness using a driver monitoring system. *Traffic Inj. Prev.* **2019**, *20*, S157–S161. [CrossRef]
5. Daza, I.G.; Bergasa, L.M.; Bronte, S.; Yebes, J.J.; Almazán, J.; Arroyo, R. Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. *Sensors* **2014**, *14*, 1106–1131. [CrossRef]
6. SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (Surface Vehicle Recommended Practice: Superseding J3016 Sep 2016)*; Technical report; SAE International: Warrendale PA, USA, 2016.
7. Gold, C.; Körber, M.; Lechner, D.; Bengler, K. Taking Over Control from Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Hum. Factors* **2016**, *58*, 642–652. [CrossRef]
8. Kumar, M.; Weippert, M.; Vilbrandt, R.; Kreuzfeld, S.; Stoll, R. Fuzzy evaluation of heart rate signals for mental stress assessment. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 791–808. [CrossRef]
9. Noh, Y.; Kim, S.; Jang, Y.J.; Yoon, Y. Modeling individual differences in driver workload inference using physiological data. *Int. J. Automot. Technol.* **2021**, *22*, 201–212. [CrossRef]
10. Nezamabadi, K.; Sardaripour, N.; Haghi, B.; Forouzanfar, M. Unsupervised ECG Analysis: A Review. *IEEE Rev. Biomed. Eng.* **2023**, *16*, 208–224. [CrossRef]
11. Collet, C.; Clarion, A.; Morel, M.; Chapon, A.; Petit, C. Physiological and behavioural changes associated to the management of secondary tasks while driving. *Appl. Ergon.* **2009**, *40*, 1041–1046. [CrossRef]
12. Shajari, A.; Asadi, H.; Alsanwy, S.; Nahavandi, S.; Lim, C.P. Leveraging Motion Platform Simulator for Detecting Driver Distraction: A CNN-LSTM Approach Integrating Physiological Signal and Head Motion Analysis. In Proceedings of the Neural Information Processing. ICONIP 2024, Auckland, New Zealand, 2–6 December 2024; Mahmud, M., Doborjeh, M., Wong, K., Leung, A.C.S., Doborjeh, Z., Tanveer, M., Eds.; Communications in Computer and Information Science; Springer Nature: Singapore, 2025; Volume 2284. [CrossRef]
13. Papakostas, M.; Das, K.; Abouelenien, M.; Mihalcea, R.; Burzo, M. Distracted and Drowsy Driving Modeling Using Deep Physiological Representations and Multitask Learning. *Appl. Sci.* **2021**, *11*, 88. [CrossRef]
14. Chen, L.; Li, D.; Wang, T.; Chen, J.; Yuan, Q. Driver Takeover Performance Prediction Based on LSTM-BiLSTM-ATTENTION Model. *Systems* **2025**, *13*, 46. [CrossRef]
15. Li, H.; Lin, Z.; An, Z.; Zuo, S.; Zhu, W.; Zhang, Z.; Mu, Y.; Cao, L.; Prades García, J.D. Automatic electrocardiogram detection and classification using bidirectional long short-term memory network improved by Bayesian optimization. *Biomed. Signal Process. Control* **2022**, *73*, 103424. [CrossRef]
16. Kumar, P.S.; Ramasamy, M.; Kallur, K.R.; Rai, P.; Varadan, V.K. Personalized LSTM Models for ECG Lead Transformations Led to Fewer Diagnostic Errors Than Generalized Models: Deriving 12-Lead ECG from Lead II, V2, and V6. *Sensors* **2023**, *23*, 1389. [CrossRef]

17. Riboni, A.; Ghioldi, N.; Candelieri, A.; Borrotti, M. Bayesian Optimization and Deep Learning for Steering Wheel Angle Prediction. *Sci. Rep.* **2022**, *12*, 8739. [CrossRef]
18. Amre, S.M.; Steelman, K.S. Keep your hands on the wheel: The effect of driver supervision strategy on change detection, mind wandering, and gaze behavior. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Washington, DC, USA, 23–27 October 2023; Volume 67, pp. 1214–1220. [CrossRef]
19. Wu, Y.; Hasegawa, K.; Kihara, K. How to request drivers to prepare for takeovers during automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2025**, *109*, 938–950. [CrossRef]
20. Ariff, N.M.; Bakar, M.A.A.; Lim, H.Y. Prediction of PM10 concentration in Malaysia using k-means clustering and LSTM hybrid model. *Atmosphere* **2023**, *14*, 853. [CrossRef]
21. Masood, Z.; Gantassi, R.; Ardiansyah; Choi, Y. A Multi-Step Time-Series Clustering-Based Seq2Seq LSTM Learning for a Single Household Electricity Load Forecasting. *Energies* **2022**, *15*, 2623. [CrossRef]
22. Kubota, K.; Togo, R.; Maeda, K.; Ogawa, T.; Haseyama, M. Balancing generalization and personalization by sharing layers in clustered federated learning. In Proceedings of the International Workshop on Advanced Imaging Technology (IWAIT), Douliu City, Taiwan, 5 February 2025; SPIE: Douliu City, Taiwan, 2025; Volume 13510, pp. 112–116.
23. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Re, A.; Montanari, R. Driver workload and eye blink duration. *TRansportation Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 199–208. [CrossRef]
24. Kazemi, M.; Rezaei, M.; Azarmi, M. Evaluating Driver Readiness in Conditionally Automated Vehicles From Eye-Tracking Data and Head Pose. *IET Intell. Transp. Syst.* **2025**, *19*, e70006. [CrossRef]
25. Sahayadhas, A.; Sundaraj, K.; Murugappan, P. Driver inattention detection methods: A review. In Proceedings of the 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT), Kuala Lumpur, Malaysia, 6–9 October 2012; Volume 10, pp. 1–6. [CrossRef]
26. CARLA Team. CARLA Simulator. 2025. Available online: https://carla.org/ (accessed on 5 August 2025).
27. Poggenhans, F.; Pauls, J.H.; Janosovits, J.; Orf, S.; Naumann, M.; Kuhnt, F.; Mayr, M. Lanelet2: A high-definition map framework for the future of automated driving. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018 ; pp. 1672–1679. [CrossRef]
28. Medina-Lee, J.; Artuñedo, A.; Godoy, J.; Villagra, J. Merit-based motion planning for autonomous vehicles in urban scenarios. *Sensors* **2021**, *21*, 3755. [CrossRef] [PubMed]
29. Benedek, M.; Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **2010**, *190*, 80–91. [CrossRef] [PubMed]
30. Shaffer, F.; Ginsberg, J.P. An overview of heart rate variability metrics and norms. *Front. Public Health* **2017**, *5*, 258. [CrossRef] [PubMed]
31. Makowski, D.; Pham, T.; Lau, Z.J.; Brammer, J.C.; Lespinasse, F.; Pham, H.; Martinez, A.; Chen, S. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* **2021**, *53*, 1689–1696. [CrossRef]
32. Peña, J.C.; Negroni Santiago, A.; Martínez García, D.; Vásquez, E.; Medina Lee, J.F. Comparative Study of Hands-on-Wheel Detection Using Wearable LSTM and Camera-Based Vision Models for Driver Monitoring. *IEEE Int. Conf. Veh. Electron. Saf.* 2025, *accepted*.
33. Ayata, D.; Yaslan, Y.; Kamasak, M.E. Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems. *J. Med. Biol. Eng.* **2020**, *40*, 149–157. [CrossRef]
34. Mishra, V.; Sen, S.; Chen, G.; Hao, T.; Rogers, J.; Chen, C.H.; Kotz, D. Evaluating the Reproducibility of Physiological Stress Detection Models. *NPJ Digit. Med.* **2020**, *3*, 125. [CrossRef]
35. Huang, X.; Ye, Y.; Xiong, L.; Lau, R.Y.; Jiang, N.; Wang, S. Time series k-means: A new k-means type smooth subspace clustering for time series data. *Inf. Sci.* **2016**, *367–368*, 1–13. [CrossRef]
36. Maharaj, E.A.; D'Urso, P.; Caiado, J. *Time Series Clustering and Classification*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2019.
37. Von Luxburg, U. Clustering stability: An overview. *Found. Trends® Mach. Learn.* **2010**, *2*, 235–274.
38. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
39. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
40. Ebrahimian, S.; Nahvi, A.; Tashakori, M.; Salmanzadeh, H.; Mohseni, O.; Leppänen, T. Multi-Level Classification of Driver Drowsiness by Simultaneous Analysis of ECG and Respiration Signals Using Deep Neural Networks. *Int. J. Environ. Res. Public Health* **2022**, *19*, 10736. [CrossRef]
41. Mattern, E.; Jackson, R.R.; Doshmanziari, R.; Dewitte, M.; Varagnolo, D.; Knorn, S. Emotion Recognition from Physiological Signals Collected with a Wrist Device and Emotional Recall. *Bioengineering* **2023**, *10*, 1308. [CrossRef]
42. Gholamiangonabadi, D.; Kiselov, N.; Grolinger, K. Deep Neural Networks for Human Activity Recognition With Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection. *IEEE Access* **2020**, *8*, 133982–133994. [CrossRef]

43. Kunjan, S.; Grummett, T.S.; Pope, K.J.; Powers, D.M.W.; Fitzgibbon, S.P.; Bastiampillai, T.; Battersby, M.; Lewis, T.W. The Necessity of Leave One Subject Out (LOSO) Cross Validation for EEG Disease Diagnosis. In Proceedings of the Brain Informatics, Virtual Event, 17–19 September 2021; Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 558–567.
44. Masuda, N.; Yairi, I.E. Multi-Input CNN-LSTM deep learning model for fear level classification based on EEG and peripheral physiological signals. *Front. Psychol.* **2023**, *14*, 1141801. [CrossRef] [PubMed]
45. Dhake, H.; Kashyap, Y.; Kosmopoulos, P. Algorithms for Hyperparameter Tuning of LSTMs for Time Series Forecasting. *Remote Sens.* **2023**, *15*, 2076. [CrossRef]
46. Vo, H.T.; Ngoc, H.T.; Quach, L.D. An Approach to Hyperparameter Tuning in Transfer Learning for Driver Drowsiness Detection Based on Bayesian Optimization and Random Search. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 828–837. [CrossRef]
47. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [CrossRef]
48. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175. [CrossRef]
49. MathWorks. *Bayesian Optimization Algorithm—Statistics and Machine Learning Toolbox*; MathWorks, Inc.: Natick, MA, USA,2025.
50. Greco, A.; Valenza, G.; Scilingo, E. *Advances in Electrodermal Activity Processing with Applications for Mental Health*; Springer: Cham, Switzerland, 2016. [CrossRef]
51. Mohd Apandi, Z.F.; Ikeura, R.; Hayakawa, S.; Tsutsumi, S. An Analysis of the Effects of Noisy Electrocardiogram Signal on Heartbeat Detection Performance. *Bioengineering* **2020**, *7*, 53. [CrossRef]
52. Moody, G.B.; Muldrow, W.; Mark, R.G. A noise stress test for arrhythmia detectors. *Comput. Cardiol.* **1984**, *11*, 381–384.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Development and Validation of an Unobtrusive Automatic Sleep Quality Assessment Index (ASQI) for Elderly Individuals

Zunyi Tang [1,2,*], Yoshinobu Murayama [2,3], Linlin Jiang [4] and Wenxi Chen [5]

[1] Center for Medical Education and Career Development, Fukushima Medical University, Fukushima City 960-1295, Fukushima, Japan
[2] Department of Electrical and Electronics Engineering, College of Engineering, Nihon University, Koriyama City 963-8642, Fukushima, Japan
[3] Department of Hematological and Biophysical System Sciences Analysis, Graduate School of Medical and Dental Sciences, Institute of Science Tokyo, Tokyo 152-8550, Japan; murayama.yoshinobu@tmd.ac.jp
[4] School of Nursing, Fukushima Medical University, Fukushima City 960-1295, Fukushima, Japan
[5] School of Computer Science and Engineering, University of Aizu, Aizuwakamatsu City 965-8580, Fukushima, Japan
* Correspondence: tangzy@fmu.ac.jp; Tel.: +81-24-5471435

Abstract

This study presents the development and validation of an unobtrusive automatic sleep quality assessment index (ASQI) designed for elderly individuals. The proposed method utilizes features such as sleep duration, sleep latency, and sleep efficiency, calculated from physiological data—heart rate, respiratory rate, body movements, and bed-exit behavior—captured by a non-contact bed sensor system installed in home environments. Based on these parameters, a six-component sleep quality index was constructed to objectively evaluate nightly sleep. To assess the reliability and validity of ASQI, sleep data were collected from eleven elderly participants over a one-year period. Results showed strong test–retest reliability ($r = 0.91$, $p < 0.001$) and moderate correlation with the widely used Pittsburgh Sleep Quality Index (PSQI) ($r = 0.52$, $p < 0.05$). Furthermore, ASQI successfully differentiated between self-reported good and poor sleepers, achieving a classification accuracy of 85.7%, with a sensitivity of 66.7% and specificity of 93.3%. These findings demonstrate that the ASQI system is a practical and scalable tool for continuous, home-based sleep monitoring in older populations. Its non-intrusive design and objective scoring make it well-suited for personalized sleep management and early detection of sleep-related issues. This work contributes to the growing field of unobtrusive health monitoring and highlights the potential of sensor-based systems in elderly care.

Keywords: sleep quality assessment; elderly; unobtrusive monitoring; heart rate; respiratory rate; body movement; ubiquitous healthcare system; Pittsburgh sleep quality index

1. Introduction

Insufficient sleep has been linked to a range of chronic diseases and conditions, including diabetes, cardiovascular disease, obesity, and depression, all of which pose significant threats to public health. According to a survey [1], approximately 21.3% of the adult population in Japan reported difficulties such as trouble initiating sleep, maintaining sleep, or experiencing early morning awakenings. Advances in methods for assessing sleep quality and treating sleep disorders offer hope to the millions of individuals affected by insufficient sleep. Sleep quality is a critical component of overall health and well-being,

influencing cognitive function, emotional stability, and physical health. While sleep quantity (total sleep time) is often emphasized [1,2], sleep quality—defined by factors such as sleep continuity, depth, and restorative value—has gained increasing attention in both clinical and research settings. The development of reliable and valid tools to assess sleep quality has become a focal point in sleep research. Sleep quality assessment can be broadly categorized into subjective and objective methods [3,4]. Subjective measures, such as self-reported questionnaires and sleep diaries, provide insights into an individual's perceived sleep experience [5]. These tools are valuable because they capture the psychological and emotional aspects of sleep, which are often not fully reflected in objective measures. For example, the Pittsburgh Sleep Quality Index (PSQI) [6] is a widely used subjective tool that evaluates sleep quality over a one-month period, assessing components such as sleep latency, duration, and disturbances [7]. However, subjective measures have limitations, including recall bias and the inability to capture detailed physiological data. Objective measures, such as polysomnography (PSG) and Actigraphy, provide quantifiable data on sleep architecture and patterns. PSG, often considered the gold standard, records brain activity, eye movements, and muscle tone during sleep [8–10]. However, PSG is expensive, intrusive, and impractical for long-term monitoring. Actigraphy, which uses wrist-worn devices to measure movement and infer sleep–wake patterns, offers a more practical alternative [11]. Despite their advantages, objective measures may not always align with subjective perceptions of sleep quality, underscoring the need for a multimodal approach to sleep assessment [9,11,12].

The rapid development of wearable technology has revolutionized sleep monitoring, enabling continuous, unobtrusive, and real-time assessment of sleep quality [13,14]. Wearable devices, such as smartwatches and fitness trackers, leverage sensors like accelerometers, heart rate monitors, and photoplethysmography (PPG) to collect data on sleep stages, heart rate variability, and respiratory rate [13,14]. These devices provide a convenient and cost-effective alternative to traditional methods, making sleep monitoring accessible to a broader population [13,14]. One of the key features of wearable sleep monitoring systems is their ability to integrate subjective and objective data. For instance, some devices allow users to input their perceived sleep quality, which can then be compared with physiological data to identify discrepancies and provide personalized feedback [15]. Additionally, advancements in machine learning and artificial intelligence have enhanced the accuracy of wearable devices in detecting sleep disorders, such as insomnia and sleep apnea [13].

Unobtrusive sleep monitoring systems, such as bed sensors and ambient devices, represent another significant advancement [16]. These systems use non-contact methods, such as ballistocardiography (BCG) and radar technology, to monitor sleep without requiring the user to wear any devices [16]. For example, bed sensors can detect subtle movements and respiratory patterns, providing detailed insights into sleep architecture [17]. Ambient devices, which use environmental sensors to monitor factors like room temperature and light levels, further contribute to a comprehensive understanding of sleep quality [18,19]. Recently, many kinds of sleep monitoring system have been developed to track various vital signs during sleeping [20,21]. Zhu et al. developed a sleep sensor using a piezoelectricity (PZT) sensor [22,23] to monitor heart rate (HR) and respiratory rate (RR), but the ease of use still needs to be improved. Widasari et al. developed an automatic and non-intrusive method of evaluation of sleep quality for patients with obtrusive sleep apnea using only an easy to conduct and record ECG signal [24]. Kalintha et al. propose SleepAge, a method that assesses sleep quality by analyzing nocturnal sounds recorded in a home environment, using machine learning to estimate a person's "sleep age" as an indicator of sleep health [25]. Nakajima et al. used video recordings to detect respiratory movements and movement of

the body position to aid in the diagnosis of sleep disturbances or to assist in the evaluation of sleep quality [26]. Liao et al. proposed a video-based sleep monitoring system that uses texture-based background modeling for real bedroom environments [27]. However, video recording raises privacy concerns that limit its broad applicability. Another challenge in using video images for sleep assessment is the large data capacity required. Meanwhile, the proliferation of smartphones and high-speed internet has enabled alternative approaches. For instance, some methods [28] leverage smartphone-embedded accelerometers to detect physiological activities caused by heartbeat, respiration, and body motion, thereby estimating HR, RR, and sleep quality without intrusive monitoring. However, such methods are not very suitable for the elderly because it is cumbersome and difficult to use [29,30]. In addition, sheet-type and under-mattress non-wearable devices have gained increasing attention for unobtrusive sleep monitoring. For example, Beddit uses BCG to estimate heart rate, respiration, and sleep parameters without requiring body contact [31], while Nemuri SCAN, an under-mattress pressure sensor, detects sleep, wakefulness, and out-of-bed states with high temporal resolution [32]. Owing to their non-contact and user-friendly nature, these devices are particularly suitable for elderly individuals, offering promising potential for long-term home-based sleep assessment [31,32].

In this work, we proposed a novel method for assessing sleep quality using characteristics and patterns of sleep obtained from an unobtrusive sleep monitoring system. The purpose of this study was to design and evaluate an objective sleep quality index using an unobtrusive sleep monitoring system as an aid to the conventional subjective Pittsburgh sleep quality index to facilitate a comprehensive and subjective assessment of long-term sleep quality. This method analyzes data on heart rate, respiratory rate, body movement (BW), and leaving-bed activities recorded during sleep, and evaluates sleep quality using a predefined scale. Eleven elderly participants were recruited to collect nocturnal sleep data and to assess the consistency and validity of the proposed method. The standard PSQI questionnaire was administered to the participants as a reference measure. Experimental results demonstrate that the sleep quality scores obtained using the proposed method exhibit high consistency with the reference PSQI scores. This preliminary study demonstrates the promise of the proposed method for automated sleep assessment and user health management.

2. Materials and Methods

2.1. System Design

An unobtrusive sleep monitoring system was developed to record sleep-related physiological signals using a non-contact sleep sensor (Safety Sheep Sensor α, NJI Co., Ltd., Koriyama, Japan) [33]. A schematic diagram of the system is shown in Figure 1. The monitoring device consists primarily of twelve piezoelectric sensors and a control circuit equipped with LAN communication capability. The piezoelectric sensors were assembled into a matrix and housed within a soft, durable plastic case measuring 800 mm $\times$ 150 mm $\times$ 17 mm. When placed beneath the mattress, the sensor monitors HR, RR, and BW by detecting voltage fluctuations from the piezoelectric elements. It also determines whether the user is on or off the bed based on these signals. According to [34], the sensor achieves a high measurement accuracy, reported as 96.9 $\pm$ 0.1% for HR and 90.5 $\pm$ 0.7% for RR.

The unobtrusive sleep monitoring system consists of hardware that collects vital sign signals during sleep and a software client that processes the collected data in real time, which is deployed on a remote AWS (Amazon Web Services (AWS) is a cloud platform that provides online services for computing, storage, and databases.) server using Node-RED (Verison 0.13) (Node-RED is a tool for connecting and controlling devices using a simple

flow-based interface.) (See Figure 1). The obtained results is composed of four kinds of data, such as heart rate, respiration rate, body movement, state whether lying on the bed or leaving. Data is stored into PostgresSQL database. In the every morning, the four kinds of data is further processed to generate six kinds of indices that are used to score the sleep in terms of the pre-defined scale and for assessing sleep quality.

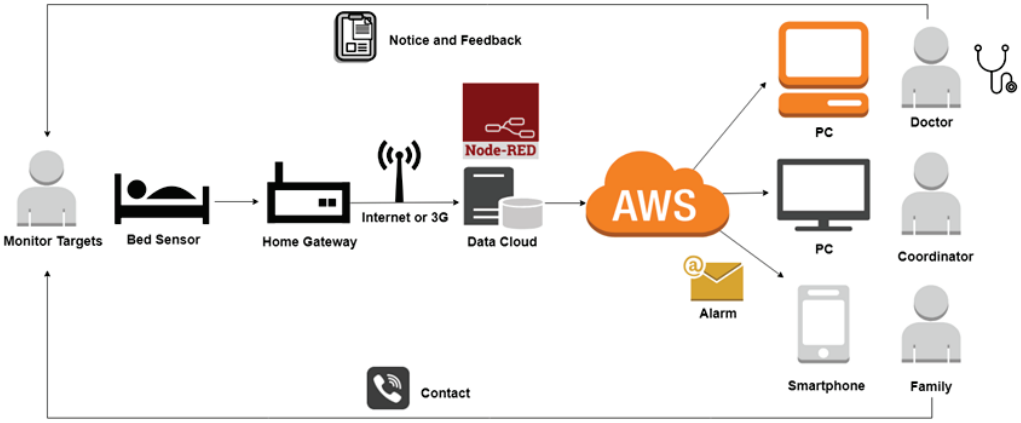


Figure 1. The schematic diagram of the unconstrained sleep health monitoring system on which the proposed automatic sleep quality index relies (reproduced from [34], with permission).

2.2. Scale Design of Sleep Quality Index

The proposed automatic sleep quality index (ASQI) was derived from two sources: a review of previous sleep quality assessment methods and data analysis with the data collected from the sleep monitoring system. The index includes six components and each component is set 0 to 3 in terms of corresponding pre-defined scale. The total score for all components ranges from 0 to 18. A total score of greater than "5" is defined as poor sleep quality. In fact, the three evaluation criteria presented in Table 1—sleep latency, sleep duration, and sleep efficiency—are derived from PSQI. We adopted the original scoring standards of these three PSQI components in this study. Regarding sleep start time, previous research has shown that a late bedtime is associated with poor sleep and adverse health outcomes, and that irregular sleep timing also leads to unfavorable effects [35,36]. Although most of these reviews did not specify precise hourly thresholds for example, defining 19:00–21:00 as "good" and >24:00 as "poor", considering that the present participants were older adults with a mean age of 77 years, and based on the actual distribution of their bedtimes, we established the evaluation criteria shown in Table 1. As for nighttime leaving of bed, previous studies have demonstrated that frequent nighttime bed-leaving in older adults-particularly those with dementia or cognitive impairment-is associated with sleep deprivation, increased risk of falls, and worse health outcomes [37,38], thus, such behavior has been identified as an indicator of poor sleep quality. Furthermore, studies examining body movements during sleep have reported that the number of movements is closely related to sleep depth and awakenings, sleep disturbances such as difficulty initiating sleep and frequent awakenings often result in increased turning and body movements [29,39,40]. Based on these prior findings, we summarized the criteria for evaluating sleep quality as presented in Table 1. Regarding the weighting strategy employed in the construction of the six ASQI components, we adopted the same methodology as that used in the PSQI, whereby each component is assigned equal weight.

Table 1. Assessment criteria of components judging sleep quality (reproduced from [34], with permission).

	Assessment Criteria			
Levels	**Good**	**Fair**	**Poor**	**Very Poor**
Sleep start time	19:00–21:00	21:00–22:30	22:30–24:00	>24:00
Sleep latency (minutes)	≤15	16–30	31–60	>60
Sleep duration (hours)	>7	6–7	5–6	<5
Sleep efficiency (%)	>85%	75–84%	65–74%	<65%
Body movement times	<1000	1000–3000	3001–6000	>6000
Leaving times	≤1	2	3–4	≥5

2.3. Signal Processing and Data Analysis

The overnight sleep data of each participant were processed in batches on the server to extract multiple sleep-related parameters, including sleep start time, sleep latency, sleep duration, sleep efficiency, body movement times during sleeping, and times of leaving bed during sleep. The sleep monitoring sensor detects pressure changes through PZT sensors and outputs a flag signal to determine whether the subject is on the bed or has left it. When the subject lies quietly on the bed without noticeable body movements, the sensor continuously outputs a signal of "0". When body movements occur, such as turning over or leg twitches, the sensor continuously outputs a signal of "1". When the subject leaves the bed, the sensor continuously outputs a signal of "2". Based on this flag signal, four parameters can be conveniently calculated: sleep start time, sleep duration, body movement times, and leaving times. Furthermore, the entire sleep data are segmented into 1-min epoches. According to the Actigraph method [41], the weighted body movement times for each epoch is calculated as follows,

$$S = 0.04c_{-2} + 0.2c_{-1} + c_0 + 0.2c_1 + 0.04c_2 \tag{1}$$

where c_i is the body movement times in the ith epoch relative to the current epoch. The epoch is judged as sleep or wake according to the following criteria,

$$\text{result} = \begin{cases} \text{sleep}, & S \leq \text{threshold} \\ \text{wake}, & S > \text{threshold} \end{cases} \quad \text{where threshold} = 5$$

where the threshold was set to 5 due to differences in the method for calculating body movements between this sleep sensor and the Actigraph [41]. Based on the above algorithm, the first epoch classified as sleep is identified, and the sleep start time is determined accordingly. The sleep latency is then derived by calculating the time difference between this point and the recorded sleep start time. Furthermore, the total wake time can be calculated based on the state of each epoch. Finally, sleep efficiency can be readily computed using the following equation,

$$\text{sleep efficiency} = \frac{\text{sleep duration - wake time}}{\text{sleep duration}} \times 100 \tag{2}$$

At this stage, all six parameters can be obtained. Based on these parameters and Table 1, sleep quality of each participant for each night was scored.

2.4. Participants

Eleven local elder volunteers (7 males and 4 females; mean age: 77.0 ± 7.2 years) were recruited for this study. Because the number of participants was limited, no strict

inclusion or exclusion criteria were applied, and data from all volunteers were included in the analysis. The participants were divided into two groups based on their self-reported sleep survey results (See Table 2). Group 1 consisted of "good" sleepers (8 participants: 5 males and 3 females; mean age: 80.6 $\pm$ 4.3 years), and Group 2 consisted of "poor" sleepers (3 participants: 2 males and 1 female; mean age: 67.3 $\pm$ 0.6 years). The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Southern TOHOKU Hospital, Fukushima, Japan.

Table 2. A summary of participant characteristics.

Select Factor	Number	Sex	Age (Year)
Good sleeper group	8	M: 5; F: 3	80.6 $\pm$ 4.3
Poor sleeper group	3	M: 2; F: 1	67.3 $\pm$ 0.6
Total	11	M: 7; F: 4	77.0 $\pm$ 7.2

2.5. Study Protocol

Following a detailed explanation of the investigation objective, all participants provided written informed consent before enrollment. After that, we visited each participant's house one by one to deploy the devices for sleep data collection. The sleep monitoring sensors (see Figure 2) were set under the mattress of every participants. Home gateway for data communication were also set up to the household of each participant. Physiological data, including HR, RR, body movements, and bed-exit events, were continuously monitored during sleep. As a reference for validation, the PSQI [42] was administered three times to each participant at staggered intervals. (The three measurement dates were as follows: 12 November 2016; 23 March 2017; and 17 November 2017). In addition, to ensure prompt monitoring of the participants' daily status, a dedicated coordinator was assigned to regularly check each participant's sleep and water usage data through a web-based GUI (see Figure 1). Whenever abnormal data or suspected device issues arose, the coordinator would promptly contact the participant or their family members.



Figure 2. The sleep monitoring sensor used in this proposed unobtrusive sleep monitoring system.

2.6. Statistics

The results analyzed from the proposed ASQI were compared with the reference PSQI. Test–retest reliability (consistency) was assessed with paired t tests and Pearson product-moment correlations for the proposed ASQI global score between any two adjacent months (e.g., July and August). This analysis was conducted across all participants. In addition, the intraclass correlation coefficient (ICC) was calculated to assess the reliability of the ASQI global score between consecutive months, and the results were interpreted according to the ICC classification [43]. To evaluate the validity of the proposed ASQI, we assessed its ability to discriminate between groups known to differ in sleep quality. To this end, we compared the component and global scores of "good" sleepers against those of "poor" sleepers.

3. Results

In this study, data collection was conducted over approximately 16 months, beginning in April 2016. In total, around 400 million records were accumulated on the data server. Figure 3 presents an example of an entire night's sleep data from one participant, illustrating variations in HR, RR, and BW, as well as instances of leaving the bed, which together reflect the participant's sleep rhythm throughout the night.

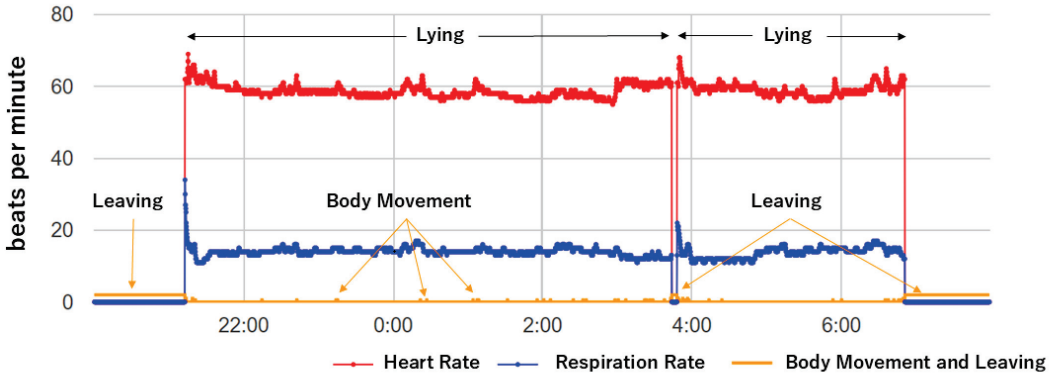


Figure 3. Representative sleep data obtained from one participant are shown, including variations in HR, RR, and BW. The red, blue, and orange lines represent HR, RR, and the states of lying on the bed or leaving the bed, respectively. The state, which includes body movement and leaving, is dimensionless. In this case, 0 represents lying, 1 represents body movement, and 2 represents leaving. (adapted from [34]).

3.1. Performance Consistency (Test–Retest Reliability)

All participants completed the analysis of automatic sleep quality assessment for one year. To analysis the performance of test–retest reliability, we convert the data to 144 pairs of sleep quality assessment results by tieing data of any two adjacent months. Paired t tests for the global ASQI score showed no significant differences between any two adjacent months. Pearson product-moment correlations demonstrated the stability of the proposed ASQI global scores, as shown in Figure 4, where the solid line represents perfect estimation ($y = x$). The corresponding correlation coefficient between the global ASQI scores of any two adjacent months was 0.91 ($p < 0.001$). The ICC analysis also revealed test–retest reliability for the proposed index scores. The test–retest reliability was assessed using ICC (3,1), which showed excellent reliability (ICC = 0.95) with a 95% confidence interval ranging from 0.93 to 0.96 ($p < 0.001$) according to the classification by Koo and Li [43].

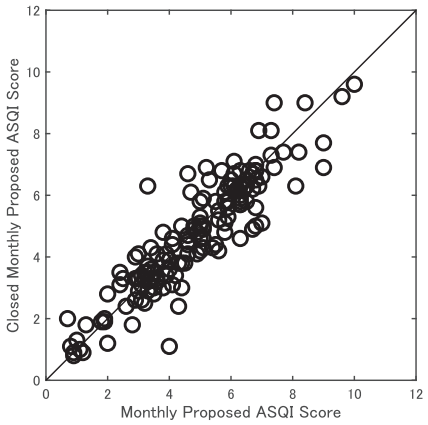


Figure 4. The pearson product-moment correlations between the monthly ASQI Score and the next month ASQI score are 0.91 ($p < 0.001$) (reproduced from [34], with permission).

3.2. Validity

The primary analysis of validity involved a comparison of global scores between the PSQI and the ASQI. Data from three PSQI surveys yielded 21 matched pairs of results. A visual comparison of the histograms in Figure 5 shows substantial agreement between the ASQI and PSQI outcomes; the red horizontal line denotes the sleep quality threshold. This convergent validity was further supported by a significant Pearson correlation coefficient of $r = 0.52$ ($p < 0.05$) for the global scores, presented in Figure 6.

As a secondary analysis of validity, the comparison between the two contrast groups revealed distinct component and global score profiles, as shown in Figure 7. We can see in Figure 7 that the most components and global scores of ASQI differed significantly between two contrast groups except the component of leaving times. The distribution of global ASQI scores also differed significantly between the two groups (Figure 7, right side). The application of a post hoc cutoff score of 5 for ASQI yielded an overall correct classification rate of 85.7% (18/21), corresponding to a sensitivity of 66.7% and a specificity of 93.3%, respectively.

Figure 8 illustrates the two-month trends in sleep quality for two participants, as evaluated by the proposed ASQI. The upper subfigure in Figure 8 indicates a generally better sleep quality compared with that shown in the lower subfigure.

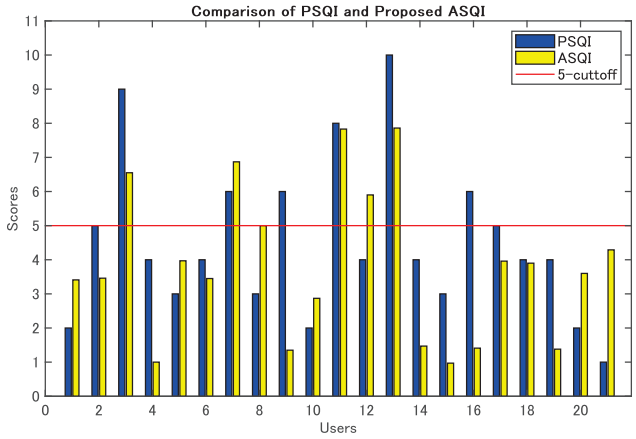


Figure 5. Comparison of the analysis results of the reference PSQI and the proposed ASQI from 11 participants (totaling 21 data entries) (reproduced from [34], with permission).

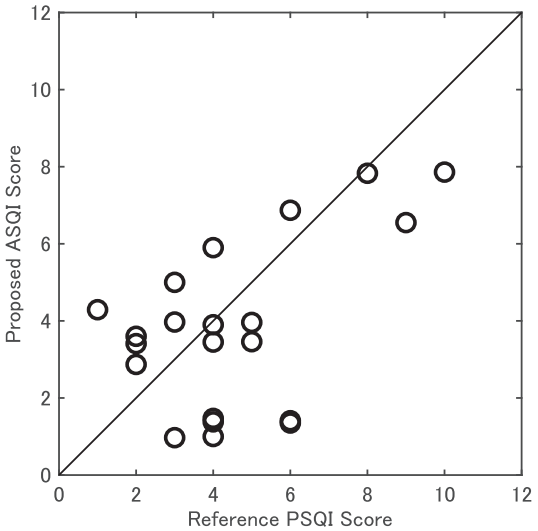


Figure 6. Correlation comparison between the reference PSQI and the proposed ASQI for 21 paired data ($r = 0.52$, $p < 0.05$) (reproduced from [34], with permission).

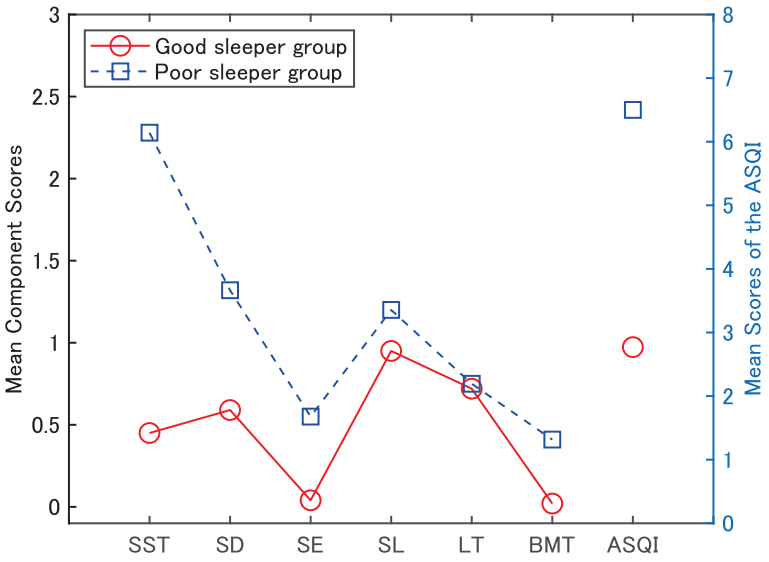


Figure 7. Proposed Automatic Sleep Quality Index (ASQI): Mean Component Scores and Total Score. In this figure, SST, SD, SE, SL, LT, and BMT denote Sleep Start Time, Sleep Duration, Sleep Efficiency, Sleep Latency, Leaving Times, and Body Movement Times, respectively.

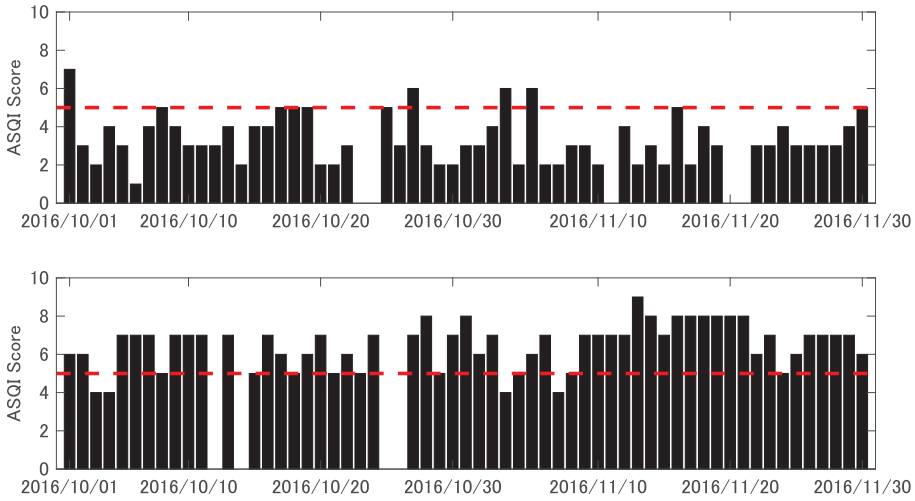


Figure 8. Daily trend of the proposed ASQI of a "good" sleeper and a "poor" sleeper for two months. The red dotted lines denote threshold (5) judging sleep quality to be good or poor.

4. Discussion

The precision of the estimation of sleep quality for this method is a crucial factor. Whether estimated sleep quality is accurate depends mainly on several main factors, including measurement of clear and accurate biological signals and the effective combination of components related to sleep features. Because three components of ASQI were inherited directly from PSQI, the results of ASQI were quite close to those of PSQI.

In this study, we proposed an unobtrusive ASQI specifically designed for elderly individuals, and validated it using long-term sleep data collected over one year. The results demonstrated that ASQI shows high consistency with the well-established PSQI, with a Pearson correlation coefficient of 0.52 ($p < 0.05$), and excellent test–retest reliability $r = 0.91$ ($p < 0.001$).

Compared to traditional objective methods such as PSG and Actigraphy, the proposed system offers significant advantages in terms of cost effectiveness, ease of use, and minimal burden on users. Although PSG remains the gold standard, its intrusiveness and complexity

limit its suitability for continuous long-term home monitoring. In contrast, the proposed bed sensor system is fully non-contact and suitable for elderly individuals who may have difficulty using wearable or smartphone-based systems. This makes it highly applicable for daily health management in home environments.

The ASQI index includes six components—bedtime, sleep latency, sleep duration, sleep efficiency, body movement count, and number of times leaving bed—all of which are objectively derived from physiological data. Incorporating parameters such as body movement and bed-exit events is particularly relevant to the elderly population, who often experience fragmented sleep due to nocturia and other age-related issues.

Validity analysis revealed that ASQI could effectively distinguish between good and poor sleepers, with statistically significant differences observed in all components except the "leaving times" category. The classification accuracy of ASQI was also promising, with a sensitivity of 66.7% and a specificity of 93.3%, correctly identifying 85.7% of participants based on a threshold score of 5. These findings indicate that the ASQI has potential not only for continuous sleep monitoring but also as a screening tool for sleep disturbances in the elderly.

Nevertheless, the proposed sleep quality index has certain limitations that present opportunities for further refinement in future studies. First, a key limitation of our work is the lack of a prospective sample size calculation. The cohort was constituted from elderly residents in a designated region, which fixed the number of participants. Therefore, while our analyses revealed statistically significant findings for the primary outcome, the precision of our estimates for secondary outcomes may be limited. We acknowledge that this affects the generalizability of our conclusions. Nonetheless, our findings provide a foundational platform for future research that employs a hypothesis-driven design with pre-specified power. To rigorously evaluate the validity of the proposed method, future studies should enroll larger and more diverse cohorts from wider populations. Moreover, while PSQI was used as a reference, it is a subjective measure, and some discrepancies between subjective and objective data are inevitable [12]. Future studies should involve larger, more diverse populations, and possibly include PSG as a clinical benchmark. Third, the proposed sleep quality index did not incorporate heart rate or respiration rate, even though these parameters can be measured by the sensor used in this study. Consequently, a key direction for future work is to integrate these features, or their derivatives, into the index. Fourth, the maximum score of our index (18) is lower than that of the PSQI (21), as it currently lacks one comparative component. This structural difference can lead to discrepancies between the two scores. To address this limitation and enhance comparability, we plan to add another component, such as one including the cardiorespiratory metrics mentioned above. Fifth, the bedside sensors employed in this project successfully passed a one-year long-term field test. While the majority of sensors performed reliably, a small number of units experienced failures—such as broken wiring due to poor soldering or sensor detachment—requiring replacement during the testing period. Some participants slept on traditional tatami mats, which necessitated repositioning the sensors each night. This repeated setup and removal likely contributed to internal wear and eventual malfunction. Additionally, the sensors required wired power and LAN connections, which posed usability challenges for tatami users. Replacing wired LAN with wireless communication in future versions could improve both convenience and system robustness.

Despite these limitations, our findings highlight the potential of noncontact, sensor-based sleep monitoring systems for elderly care. The proposed ASQI provides a practical and scalable solution for long-term sleep quality assessment, with implications for personalized health monitoring, early detection of sleep-related disorders, and better overall well-being among the aging population.

5. Conclusions

In conclusion, the proposed ASQI framework offers a viable and scalable solution for continuous sleep quality assessment, representing a meaningful advancement in unobtrusive health monitoring for the aging population.

Author Contributions: Conceptualization, Z.T. and L.J.; methodology, Z.T. and Y.M.; software, Z.T.; validation, Z.T. and L.J.; data curation, L.J.; writing—original draft preparation, Z.T.; writing—review and editing, Y.M., L.J. and W.C.; visualization, L.J.; supervision, Y.M. and W.C.; project administration, Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Southern TOHOKU Hospital, Fukushima, Japan. The approval was granted on 28 September 2015.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request.

Acknowledgments: We thank NJI Co., Ltd. for providing the Safety Sheep Sensor α for this study and for their assistance in data collection.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASQI	Automatic Sleep Quality Index
PSQI	Pittsburgh Sleep Quality Index
PSG	Polysomnography
PPG	Photoplethysmography
BCG	Ballistocardiography
HR	Heart Rate
RR	Respiratory Rate
BW	Body Movement
SST	Sleep Start Time
SD	Sleep Duration
SE	Sleep Efficiency
SL	Sleep Latency
LT	Leaving Times
BMT	Body Movement Times

References

1. Kim, K.; Uchiyama, M.; Okawa, M.; Liu, X.; Ogihara, R. An epidemiological study of insomnia among the Japanese general population. *Sleep* **2000**, *23*, 41–47. [CrossRef]
2. Itani, O.; Kaneita, Y.; Munezawa, T.; Mishima, K.; Jike, M.; Nakagome, S.; Tokiya, M.; Ohida, T. Nationwide epidemiological study of insomnia in Japan. *Sleep Med.* **2016**, *25*, 130–138. [CrossRef] [PubMed]
3. Penzel, T. Sleep Quality Assessment: Challenges and Opportunities. *IEEE Pulse* **2016**, *7*, 3. [CrossRef]
4. Fabbri, M.; Beracci, A.; Martoni, M.; Meneo, D.; Tonetti, L.; Natale, V. Measuring Subjective Sleep Quality: A Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1082. [CrossRef]
5. Chen, W.H.; Chen, C.T.; Cheng, K.H.; Tsao, Y.C.; Lin, Y.H.; Chen, J.Y. Associations Between Sleep Quality and Self-Reported Health Status in Middle-Aged and Older Adults: A Community-Based, Cross-Sectional Study in Northern Taiwan. *Healthcare* **2025**, *13*, 1272. [CrossRef]

6. Buysse, D.J.; Reynolds, C.F.; Monk, T.H.; Berman, S.R.; Kupfer, D.J. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **1989**, *28*, 193–213. [CrossRef] [PubMed]
7. Hinz, A.; Glaesmer, H.; Brähler, E.; Löffler, M.; Engel, C.; Enzenbach, C.; Hegerl, U.; Sander, C. Sleep quality in the general population: Psychometric properties of the Pittsburgh Sleep Quality Index, derived from a German community sample of 9284 people. *Sleep Med.* **2017**, *30*, 57–63. [CrossRef]
8. Berry, R.B.; Brooks, R.; Gamaldo, C.E.; Harding, S.M.; Marcus, C.; Vaughn, B.V. *The AASM Manual for the Scoring of Sleep and Associated Events. Rules, Terminology and Technical Specifications*; American Academy of Sleep Medicine: Darien, IL, USA, 2012; Volume 176, p. 7.
9. Kaplan, K.A.; Hirshman, J.; Hernandez, B.; Stefanick, M.L.; Hoffman, A.R.; Redline, S.; Ancoli-Israel, S.; Stone, K.; Friedman, L.; Zeitzer, J.M.; et al. When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography. *Biol. Psychol.* **2017**, *123*, 37–46. [CrossRef]
10. Chen, Y.; Zhou, E.; Wang, Y.; Wu, Y.; Xu, G.; Chen, L. The past, present, and future of sleep quality assessment and monitoring. *Brain Res.* **2023**, *1810*, 148333. [CrossRef] [PubMed]
11. Girschik, J.; Fritschi, L.; Heyworth, J.; Waters, F. Validation of self-reported sleep against actigraphy. *J. Epidemiol.* **2012**, *22*, 462–468. [CrossRef]
12. Masaki, M.; Tsumoto, S.; Tani, A.; Tominaga, M.; Seol, J.; Chiba, S.; Miyanishi, K.; Nishida, K.; Kawana, F.; Amemiya, T.; et al. Discrepancies between subjective and objective sleep assessments revealed by in-home electroencephalography during real-world sleep. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2412895121. [CrossRef] [PubMed]
13. Lechat, B.; Hirotsu, C.; Appleton, S.; Younes, M.; Adams, R.J.; Vakulin, A.; Hansen, K.; Zajamsek, B.; Wittert, G.; Catcheside, P.; et al. A novel EEG marker predicts perceived sleepiness and poor sleep quality. *Sleep* **2022**, *45*, zsac051. [CrossRef]
14. Li, Y.; Li, J.; Yan, C.; Dong, K.; Kang, Z.; Zhang, H.; Liu, C. Sleep Quality Evaluation Based on Single-Lead Wearable Cardiac Cycle Acquisition Device. *Sensors* **2023**, *23*, 328. [CrossRef] [PubMed]
15. Ibáñez, V.; Silva, J.; Cauli, O. A survey on sleep questionnaires and diaries. *Sleep Med.* **2018**, *42*, 90–96. [CrossRef]
16. Tamura, T.; Huang, M. Unobtrusive Bed Monitor State of the Art. *Sensors* **2025**, *25*, 1879. [CrossRef]
17. Yamakawa, M.; Kang, H.S.; Wang, H.; Konno, R. Sleep quality assessment of adults in care settings using non-wearable sleep trackers: Scoping review. *Int. J. Nurs. Pract.* **2024**, *30*, e13240. [CrossRef] [PubMed]
18. Baniassadi, A.; Manor, B.; Yu, W.; Travison, T.; Lipsitz, L. Nighttime ambient temperature and sleep in community-dwelling older adults. *Sci. Total Environ.* **2023**, *899*, 165623. [CrossRef]
19. Dashkevych, O.; Portnov, B.A. Using Low-Cost Technology Devices for Monitoring Sleep and Environmental Factors Affecting It: A Systematic Review of the Literature. *Appl. Sci.* **2025**, *15*, 1188. [CrossRef]
20. Jiang, L.; Tang, Z.; Liu, Z.; Chen, W.; Kitamura, K.-i.; Nemoto, T. Automatic sleep monitoring system for home healthcare. In Proceedings of the 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, Hong Kong, China, 5–7 January 2012; pp. 894–897.
21. Cay, G.; Ravichandran, V.; Sadhu, S.; Zisk, A.; Salisbury, A.; Solanki, D.; Mankodiya, K. Recent Advancement in Sleep Technologies: A Literature Review on Clinical Standards, Sensors, Apps, and AI Methods. *IEEE Access* **2022**, *10*, 104737–104756. [CrossRef]
22. Zhu, X.; Chen, W.; Nemoto, T.; Kanemitsu, Y.; Kitamura, K.i.; Yamakoshi, K.i.; Wei, D. Real-time monitoring of respiration rhythm and pulse rate during sleep. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2553–2563. [CrossRef]
23. Zhu, X.; Chen, W.; Tang, Z.; Nemoto, T.; Wei, D. Automatic home care system for monitoring HR/RR during sleep. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 21–24 August 2008; pp. 522–525.
24. Widasari, E.R.; Tanno, K.; Tamura, H. Automatic sleep quality assessment for obstructive sleep apnea patients based on HRV spectrum analysis. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 1187–1192.
25. Kalintha, W.; Kato, T.; Fukui, K. SleepAge: Sleep Quality Assessment from Nocturnal Sounds in Home Environment. *Procedia Comput. Sci.* **2020**, *176*, 898–907. [CrossRef]
26. Nakajima, K.; Matsumoto, Y.; Tamura, T. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiol. Meas.* **2001**, *22*, N21. [CrossRef]
27. Liao, W.H.; Kuo, J.H. Sleep monitoring system in real bedroom environment using texture-based background modeling approaches. *J. Ambient. Intell. Humaniz. Comput.* **2013**, *4*, 57–66. [CrossRef]
28. Chen, Z.; Lin, M.; Chen, F.; Lane, N.D.; Cardone, G.; Wang, R.; Li, T.; Chen, Y.; Choudhury, T.; Campbell, A.T. Unobtrusive sleep monitoring using smartphones. In Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), Venice, Italy, 5–8 May 2013; pp. 145–152.

29. Ghazi, S.N.; Behrens, A.; Berner, J.; Sanmartin Berglund, J.; Anderberg, P. Objective sleep monitoring at home in older adults: A scoping review. *J. Sleep Res.* **2025**, *34*, e14436. [CrossRef]
30. Lam, A.; Simonetti, S.; D'Rozario, A.; Ireland, D.; Bradford, D.; Fripp, J.; Naismith, S.L. Perceptions of the Use of Mobile Apps to Assess Sleep-Dependent Memory in Older Adults With Subjective and Objective Cognitive Impairment: Focus Group Approach. *JMIR Aging* **2025**, *8*, e68147. [CrossRef]
31. Tuominen, J.; Peltola, K.; Saaresranta, T.; Valli, K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph beddit sleep tracker: A validation study. *J. Clin. Sleep Med.* **2019**, *15*, 483–487. [CrossRef] [PubMed]
32. Nagatomo, K.; Masuyama, T.; Iizuka, Y.; Makino, J.; Shiotsuka, J.; Sanui, M. Validity of an under-mattress sensor for objective sleep measurement in critically ill patients: A prospective observational study. *J. Intensive Care* **2020**, *8*, 16. [CrossRef]
33. Ikuta, K.; Aishima, M.; Noguchi-Watanabe, M.; Fukui, S. Feasibility of Monitoring Heart and Respiratory Rates Using Non-wearable Devices and Consistency of the Measured Parameters: Pilot Feasibility Study. *JMIR Hum. Factors* **2024**, *11*, e56547. [CrossRef]
34. Tang, Z.; Jiang, L.; Sakatani, K. Development of Sleep Quality Assessment Method for Elderly People Based on Unobtrusive Sleep Monitoring System. In Proceedings of the Life Engineering Symposium 2018 (LE 2018), Aizu, Japan, 10–12 September 2018; pp. 1–4.
35. Chaput, J.P.; Dutil, C.; Featherstone, R.; Ross, R.; Giangregorio, L.; Saunders, T.J.; Janssen, I.; Poitras, V.J.; Kho, M.E.; Ross-White, A.; et al. Sleep timing, sleep consistency, and health in adults: A systematic review. *Appl. Physiol. Nutr. Metab.* **2020**, *45*, S232–S247. [CrossRef]
36. Sletten, T.L.; Weaver, M.D.; Foster, R.G.; Gozal, D.; Klerman, E.B.; Rajaratnam, S.M.; Roenneberg, T.; Takahashi, J.S.; Turek, F.W.; Vitiello, M.V.; et al. The importance of sleep regularity: A consensus statement of the National Sleep Foundation sleep timing and variability panel. *Sleep Health* **2023**, *9*, 801–820. [CrossRef] [PubMed]
37. Higami, Y.; Yamakawa, M.; Shigenobu, K.; Kamide, K.; Makimoto, K. High frequency of getting out of bed in patients with Alzheimer's disease monitored by non-wearable actigraphy. *Geriatr. Gerontol. Int.* **2019**, *19*, 130–134. [CrossRef]
38. Walsh, L.; Moloney, E.; McLoone, S. Identification of nocturnal movements during sleep using the non-contact under mattress bed sensor. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 1660–1663.
39. Huysmans, D.; Borzée, P.; Testelmans, D.; Buyse, B.; Willemen, T.; Van Huffel, S.; Varon, C. Evaluation of a commercial ballistocardiography sensor for sleep apnea screening and sleep monitoring. *Sensors* **2019**, *19*, 2133. [CrossRef]
40. Saleh, D.; Bertisch, S.M.; Reid, M.; Lim, A.; Purcell, S.; Redline, S. Actigraphy-derived sleep fragmentation index: Convergent validity and associations with clinical outcomes. *J. Clin. Sleep Med.* **2025**, *21*, 11754. [CrossRef]
41. Peng, Y.T.; Lin, C.Y.; Sun, M.T.; Landis, C.A. Multimodality sensor system for long-term sleep quality monitoring. *IEEE Trans. Biomed. Circuits Syst.* **2008**, *1*, 217–227. [CrossRef] [PubMed]
42. Doi, Y.; Minowa, M.; Uchiyama, M.; Okawa, M.; Kim, K.; Shibui, K.; Kamei, Y. Psychometric assessment of subjective sleep quality using the Japanese version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects. *Psychiatry Res.* **2000**, *97*, 165–172. [CrossRef] [PubMed]
43. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Model-Based Design and Sensitivity Optimization of Frequency-Output Pressure Sensors for Real-Time Monitoring in Intelligent Rowing Systems

Iaroslav Osadchuk [1], Oleksandr Osadchuk [1], Serhii Baraban [2,*], Andrii Semenov [1] and Mariia Baraban [3]

[1] Faculty of Information Electronic Systems, Vinnytsia National Technical University, Khmelnytske Highway 95, 21021 Vinnytsia, Ukraine; osadchuk.j93@vntu.edu.ua (I.O.); osadchuk.o.v@vntu.edu.ua (O.O.); semenov.a.o@vntu.edu.ua (A.S.)
[2] Faculty of Computing and Telecommunications, Poznan University of Technology, 60965 Poznan, Poland
[3] Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Khmelnytske Highway 95, 21021 Vinnytsia, Ukraine; baraban@vntu.edu.ua
* Correspondence: serhii.baraban@put.poznan.pl; Tel.: +48-5151-799-06

Abstract

This study presents a model-driven approach to the design, calibration, and application of frequency-output pressure sensors integrated within an intelligent system for real-time monitoring of rowing performance. The proposed system captures biomechanical parameters of the "boat–rower" complex across 50 parallel channels with a temporal resolution of 8–12 ms. At the core of the sensing architecture are parametric pressure transducers incorporating strain-gauge primary elements and microelectronic auto-generator circuits featuring negative differential resistance (NDR). These oscillating circuits convert mechanical stress into high-frequency output signals in the 1749.9–1751.9 MHz range, with pressure sensitivities from 0.365 kHz/kPa to 1.370 kHz/kPa. The sensor models are derived using physical energy conversion principles, enabling the formulation of analytical expressions for transformation and sensitivity functions. These models simplify sensitivity tuning and allow clear interpretation of how structural and electronic parameters influence output frequency. The system architecture eliminates the need for analog-to-digital converters and signal amplifiers, reducing cost and power consumption, while enabling wireless ultra high frequency (UHF) transmission of sensor data. Integrated algorithms analyze the influence of biomechanical variables on athlete performance, enabling real-time diagnostics. The proposed model-based methodology offers a scalable and accurate solution for intelligent sports instrumentation and beyond.

Keywords: negative differential resistance; frequency-output pressure sensor; self-oscillating circuits; sensitivity; measurement system; real-time biomechanical monitoring

1. Introduction

1.1. Background

The popularization of rowing, the possibility of good earnings, commercialization, increasing media attention to rowing, the development of training methods, and many other factors have led to the emergence of strong crews in many countries that until recently could not reach the C final. The wide variety of boat and oar manufacturers has eliminated the limiting factor of equipment quality. The increase in the number of strong and technical athletes, together with the lack of differences in equipment, has led to increased competition

at international competitions. The consequence of increased competition was a reduction in the gap between the outsiders and the race leaders (both in the final and qualifying races), respectively, which led to an increase in the speed of overcoming the distance [1–5].

The long-term training planning of successful rowing athletes is based on the principles of perspective, phased training, results programming, and an individual approach to training for each athlete. Successful performance in international competitions is mainly due to the effectiveness of the training process, which depends to a large extent on an individual approach to the selection of work modes. When training highly skilled athletes, the individualization of work modes is of particular importance. Multifactorial training involves the division of the load by the main direction and the distribution of multidirectional cycles, which is due to the individual residual training effects of each athlete [6–13].

Therefore, one of the most promising areas is the development of a radio-measuring system designed to improve the athlete's movements during the stroke and the entire distance in general in order to bring them to automaticity and further control them at the subconscious level.

Capacitive, piezoresistive, magnetoelastic, inductive, piezoelectric, and acoustoelectric sensors are used to measure pressure. Capacitive sensors have high accuracy and sensitivity and thermal stability, but their disadvantages are that they are not suitable for high pressures and are also expensive. Inductive sensors are characterized by temperature instability and high sensitivity to shock and vibration. Piezoelectric sensors have low output signals, low accuracy and sensitivity, and insufficient stability. Piezoresistive pressure sensors have low output signals and are sensitive to temperature changes, but they have high strength, low hysteresis, and a wide temperature range [14–16].

As can be seen from the analysis of the literature [14–18], one of the main disadvantages of existing pressure sensors is the low output signal, except for inductive pressure sensors, whose manufacture is incompatible with integrated microelectronic technology. This leads to significant measurement errors due to information losses in the channel between the sensor output and the input of the amplifier and converting equipment, which causes their low accuracy and sensitivity and low noise immunity and speed. It is possible to eliminate the above shortcomings on the basis of parametric self-oscillating sensors operating in the pressure–frequency conversion mode, which makes it possible to significantly improve their metrological performance [19–21]. The advantages and disadvantages of parametric oscillator sensors, taking into account their limits of applicability, are shown in Table 1.

Despite these advantages, sensor devices also have certain disadvantages that affect their practical application. These include sensitivity to external noise and interference, limited power resources, especially in portable or autonomous systems, and limited sensing capabilities [22,23]. In addition, sensors are often characterized by low computing and storage resources, making it difficult to process and store large amounts of data. All this can reduce the reliability and accuracy of the measurements and create difficulties in analyzing and comparing the data [22,23]. Thus, in addition to the advantages of sensor technologies, their limitations must also be taken into account, especially under difficult operating conditions.

1.2. Research Gap

Although frequency output pressure sensors have been explored in various engineering and industrial applications, their use in high-speed multichannel sports instrumentation, particularly for the real-time monitoring of rowing performance, remains underdeveloped. Existing studies on rowing biomechanics often employ analog output sensors combined with analog to digital converters (ADC) and amplifiers, which increase the cost, latency, and system complexity. Moreover, sensitivity calibration in such systems

is usually empirical and lacking robust analytical models that explicitly relate structural and electronic parameters to output characteristics.

Table 1. Advantages and drawbacks of parametric self-oscillating sensors with boundaries of applicability.

N	Criterion	Advantages	Drawbacks	Boundaries of Applicability
1.	**Principles of operation**	Simple construction; possibility of contactless measurement	Dependence on parasitic parameters and on the environment	Used when permissible errors are not critical
2.	**Sensitivity**	Very high (can detect very small changes in inductance/capacitance)	Strong susceptibility to external interference	Not suitable for precise measurements in high-noise environments
3.	**Measurement range**	Suitable for small deviations (where high sensitivity is required)	Narrow operating range	Used within a narrow range of physical quantities
4.	**Temperature stability**	Compensation by circuit techniques is possible	Often exhibits significant temperature drift	Not suitable under large temperature variations
5.	**Power consumption**	Active excitation enables a stable output signal path	Power consumption higher than in passive sensors	Limited in autonomous/battery-powered systems with strict energy budgets
6.	**Durability**	No mechanical contact (no wear of contacts)	Sensitive to aging of generator elements and material stability	Not suitable for long-term stable measurements without recalibration
7.	**Resistance to environmental influence**	Possibility of hermetically sealing the sensitive element	Susceptible to electromagnetic interference, humidity, and contamination	Not suitable in aggressive or unstable environmental conditions
8.	**Cost and simplicity**	Relatively inexpensive, easy to implement	Requires calibration, compensation, and stabilization	Effective where low cost and simplicity are priorities, provided accuracy and sensitivity are met

In the domain of frequency-output sensing, research has focused primarily on general-purpose transducers without addressing the integration of strain gauge primary elements into active self-oscillating circuits with negative differential resistance (NDR). The absence of such integration limits opportunities for compact low-power designs and real-time wireless data transmission. In addition, there is a lack of model-based methodologies that simultaneously address the sensor design, transformation function derivation, and sensitivity optimization in the context of intelligent sports systems.

This gap underscores the need for a unified physically grounded design approach that not only predicts the sensor performance but also simplifies the hardware architecture for scalable, high-resolution, and low-latency biomechanical monitoring.

1.3. Problem Statement

Recent trends in rowing show a significant increase in competition at both the international and national levels. According to BioRow, the average speed of the winners of the World Championships from 1992 to 2018 increased by about 0.4% annually, and at the Paris 2024 Olympic Games, the victory was decided by increases within tenths of a percent of relative speed [24,25]. Even the difference in technique can account for up to 2% of the result deviation (∼7–8 s at a distance of 2000 m), which is critical at the level of the final races [26]. At the same time, national rowing competitions, the results of which are systematically published on the rowing.su platform, demonstrate increased competition at the regional level: the participation of a large number of athletes in different age groups

confirms the narrowing of time gaps between participants [27]. It is important to note that academic rowing (water discipline) and rowing indoors (competition on ergometers) are different sports, but both areas reflect the same tendency to minimize the difference in results. This makes the development of real-time monitoring systems especially relevant, as even the slightest deviation in technique or power can determine the final placement of an athlete.

Accurate and rapid monitoring of biomechanical parameters in rowing is crucial to optimize technique and increase the efficiency of force transfer to boat motion. Key indicators include force curves, stroke and finish angles, length and power of the pulling stroke, as well as speed and acceleration of the boat, handles, and seat [28]. It is also important to analyze the coordination of leg, trunk, and arm movements, which determines the consistency of the stroke [28]. Assessment of these parameters allows for the timely detection of technical errors and adjustment of the training process.

Accurate and high-speed monitoring of biomechanical parameters in rowing is critical for evaluating athlete performance and optimizing training processes. Conventional pressure sensing solutions in sports instrumentation often rely on analog output signals, which require analog-to-digital conversion and amplification before processing. This increases the system cost, complexity, and power consumption while introducing latency and potential signal degradation.

Furthermore, the sensitivity of such systems to structural and environmental variations is often not adequately modeled in the design stage, limiting the accuracy and scalability of the calibration. Existing designs rarely integrate model-based methodologies for predicting and optimizing the sensitivity of frequency output pressure sensors, especially in multichannel real-time monitoring environments, such as intelligent rowing systems.

1.4. Objective

The primary objective of this research is to develop a model-based design and a sensitivity optimization methodology for pressure sensors with frequency output intended for integration into an intelligent rowing monitoring system. Specifically, the work aims to

1. Establish analytical models of parametric pressure transducers that incorporate primary strain gauge elements and microelectronic self-oscillating circuits with negative differential resistance.
2. Derive transformation and sensitivity functions based on physical energy conversion principles to enable straightforward sensitivity tuning and performance prediction.
3. Implement and validate a multichannel (50-channel) real-time monitoring system capable of acquiring biomechanical data from the "boat–rower" complex at a temporal resolution of 8–12 ms.
4. Minimize hardware complexity and cost by eliminating traditional ADCs and amplifiers while enabling wireless UHF data transmission for real-time analysis.

1.5. Main Contribution

This study advances the state of intelligent sports instrumentation by introducing the following.

- A unified model-driven framework for the design and calibration of frequency-output pressure sensors, allowing precise prediction of sensitivity and the influence of structural/electronic parameters;
- Novel integration of strain-gauge primary transducers as active elements within NDR-based auto-generator circuits, reducing the component count and enhancing the signal stability;

- Derivation of parametric transformation functions that explicitly relate mechanical pressure changes to equivalent capacitance and negative differential resistance variations in the oscillating system, enabling accurate frequency-domain performance optimization;
- A fully implemented 50-channel real-time monitoring system for rowing, with demonstrated sensor sensitivity ranging from 0.365 kHz/kPa to 1.370 kHz/kPa over a 0–2050 kPa pressure range and output frequencies between 1749.9 MHz and 1751.9 MHz;
- System-level benefits, including elimination of ADCs/amplifiers, reduced cost and power consumption, and seamless wireless data transmission, making the approach scalable for broader applications in sports analytics and industrial sensing.

1.6. Novelty Statement

The novelty of this work lies in the integration of primary strain gauge transducers directly as active elements within microelectronic self-oscillating circuits featuring negative differential resistance, enabling the development of a compact pressure sensor with a low output that does not require analog-to-digital converters or signal amplifiers. Unlike conventional designs, the proposed sensors are modeled from the ground up using the energy conversion method, allowing the analytical derivation of transformation and sensitivity functions that explicitly link changes in mechanical pressure with oscillator parameters and output frequency.

This approach enables

1. High-speed multi-channel (50-channel) real-time monitoring of rowing biomechanics with a temporal resolution of 8–12 ms.
2. Scalable sensitivity optimization through direct parametric adjustment informed by analytical models, rather than purely empirical calibration.
3. Seamless wireless UHF data transmission without bias, reducing the cost, size, and power consumption while maintaining measurement accuracy in the 1749.9–1751.9 MHz range.

To our knowledge, this is the first reported application of such model-based frequency output pressure sensors in intelligent rowing systems and one of the few in sports instrumentation to completely eliminate ADC while maintaining subkHz/kPa sensitivity resolution.

2. Materials and Methods

A self-oscillating pressure transducer with a frequency output signal for computer and information measurement systems and networks is based on a microelectronic transistor structure with negative differential resistance, in which a strain-sensitive element is a strain resistor. Figure 1 shows the electrical diagram of a microelectronic pressure transducer.

It is an integrated microelectronic circuit consisting of bipolar and field-effect transistors VT1 and VT2 with different types of conductivity, leading to the appearance of negative differential resistance at the collector–emitter of the transistors, which significantly reduces the hardware costs in computer and information and measurement systems (Figure 1). The self-oscillating system has three dynamic variables, two of which are voltages on the equivalent capacitances of the transistor structure with negative differential resistance (respectively, on the drain electrodes of the field-effect transistor VT2 and the collector of the bipolar transistor VT1), and the third is the inductance current L1. The dynamic processes occurring in the self-oscillator are determined by the reactive properties of the bipolar field-effect transistor structure with negative differential resistance and are similar in nature to quasi-sinusoidal processes, as in a generator based on an inductive triac circuit.

To change the dynamics of oscillations and expand the working volume of the phase space, a circuit R5C3 was added to the inductive branch of the basic generator circuit based on a bipolar field-effect transistor structure with negative differential resistance. A C1R3 circuit has been introduced to thermally stabilize the generator [29]. The ratings of the radioelectronic components were selected based on the following considerations: (1) to ensure the maximum length of the falling section of the static VAC; (2) to ensure a wide change in the equivalent capacitance of the transistor structure and, accordingly, the range of frequency change under the influence of pressure.

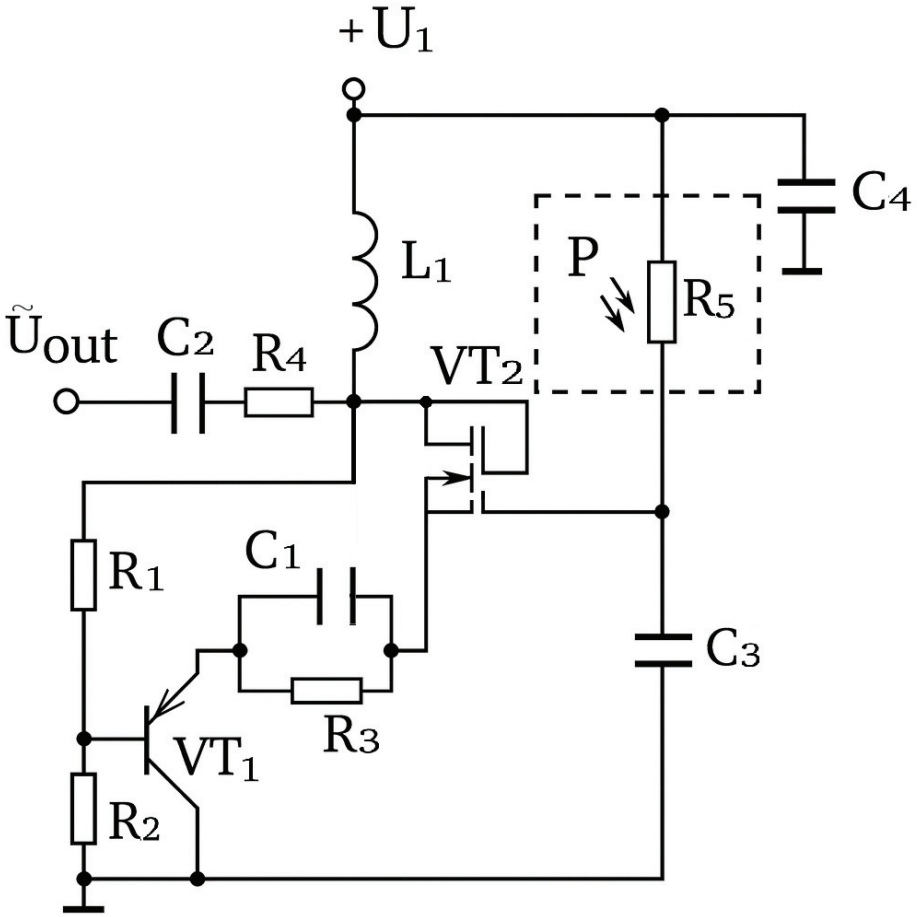


Figure 1. Electrical diagram of a self-oscillating pressure transducer based on a bipolar and field-effect transistor structure.

The strain-sensitive resistor R5 of the Interlink Electronics FSR-400 type is included in the positive feedback loop. There is total resistance in the collector drain electrodes VT1 and VT2, the active component of which has a negative value, and the reactive component has a capacitive character. Connecting the inductor L1 to the collector VT1 and to the upper terminal of the capacitance C4 creates an oscillating counter, the energy losses of which are compensated by the energy of the negative differential resistance. The capacitance C4 prevents the passage of alternating high-frequency current through the DC voltage source U1, which supplies the device circuit. The resonant frequency of the oscillating circuit depends on the pressure on the strain-sensitive resistor R5, since its current changes, which leads to a change in the equivalent capacitance and negative differential resistance in the oscillating system of the device's oscillator. The generated voltage is removed from the drain of the transistor VT2 (the main output of the oscillator). In this case, the selection of the initial generation frequency without pressure is controlled by changing the voltage of the power supplies V1 and V2. When connecting the load resistance, it is

necessary to take into account the active and reactive resistance introduced by it to the oscillating circuit of the generator. In this case, the generator operates in idle mode on the main output, which allows it to be widely used in measurement control and automation devices circuits. A mathematical model was developed to analyze the characteristics of the generated frequency as a function of pressure, and further calculations were performed in MATLAB R2018a (MathWorks, Natick, MA, USA).

In the course of analyzing the optimal design of the self-oscillating pressure transducer, it was concluded that it is advisable to use the frequency method of information conversion, which will increase the output power of the self-oscillating and expand the frequency range of the generated signal. In principle, oscillator transistor structures with negative differential resistance are a transistor analog of a negatron, whose waveform has a downward section corresponding to a negative differential resistance, which is provided by internal feedback and serves to compensate for energy losses on the active resistances of the circuit. The complex resistance of such a structure is capacitive, and its value depends on the voltage applied to its input. When such a structure is connected to an inductor, a resonant oscillating circuit is formed.

Figure 2 shows a diagram of a self-oscillating pressure transducer based on a bipolar-field-effect transistor structure with the introduction of a control voltage source for developing a mathematical model of this oscillator.

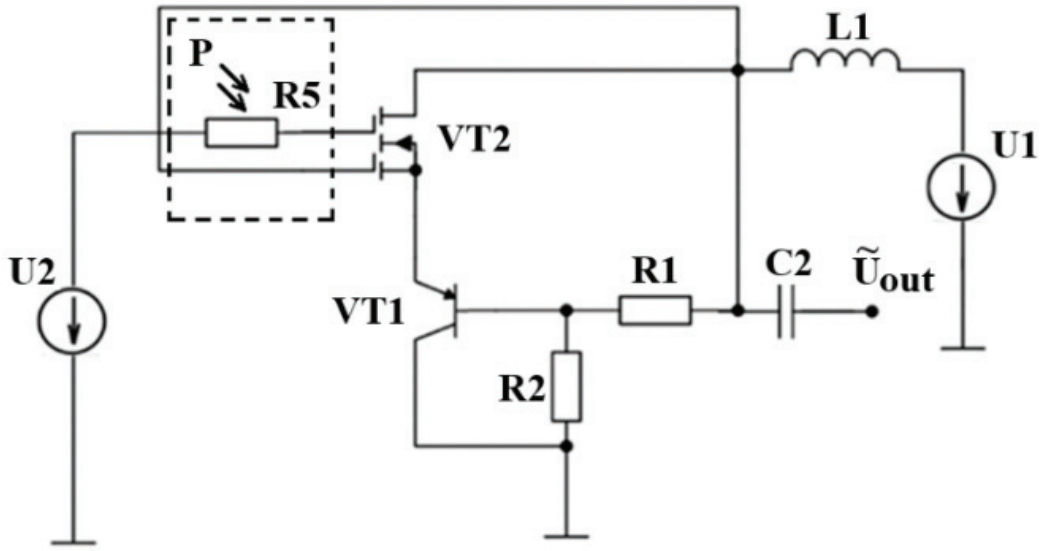


Figure 2. Electrical diagram of a self-oscillator pressure transmitter with two sources of constant voltage: supply voltage and control voltage.

Finding the optimal operating point of the power supply of such a circuit is carried out using two sources of constant voltage: the supply voltage and the control voltage (Figure 2).

The equivalent capacitance is part of the total capacitance of a transistor structure with a negative differential resistance arising at the electrodes of the drain of a double-gate MOSFET—collector of a bipolar transistor, which, together with the inductance L1, form a resonant oscillating circuit, which, in view of the above, has a resonant oscillation frequency dependent on the control voltage. To create a mathematical model of the self-oscillating pressure transducer, Figure 3 shows a nonlinear equivalent circuit.

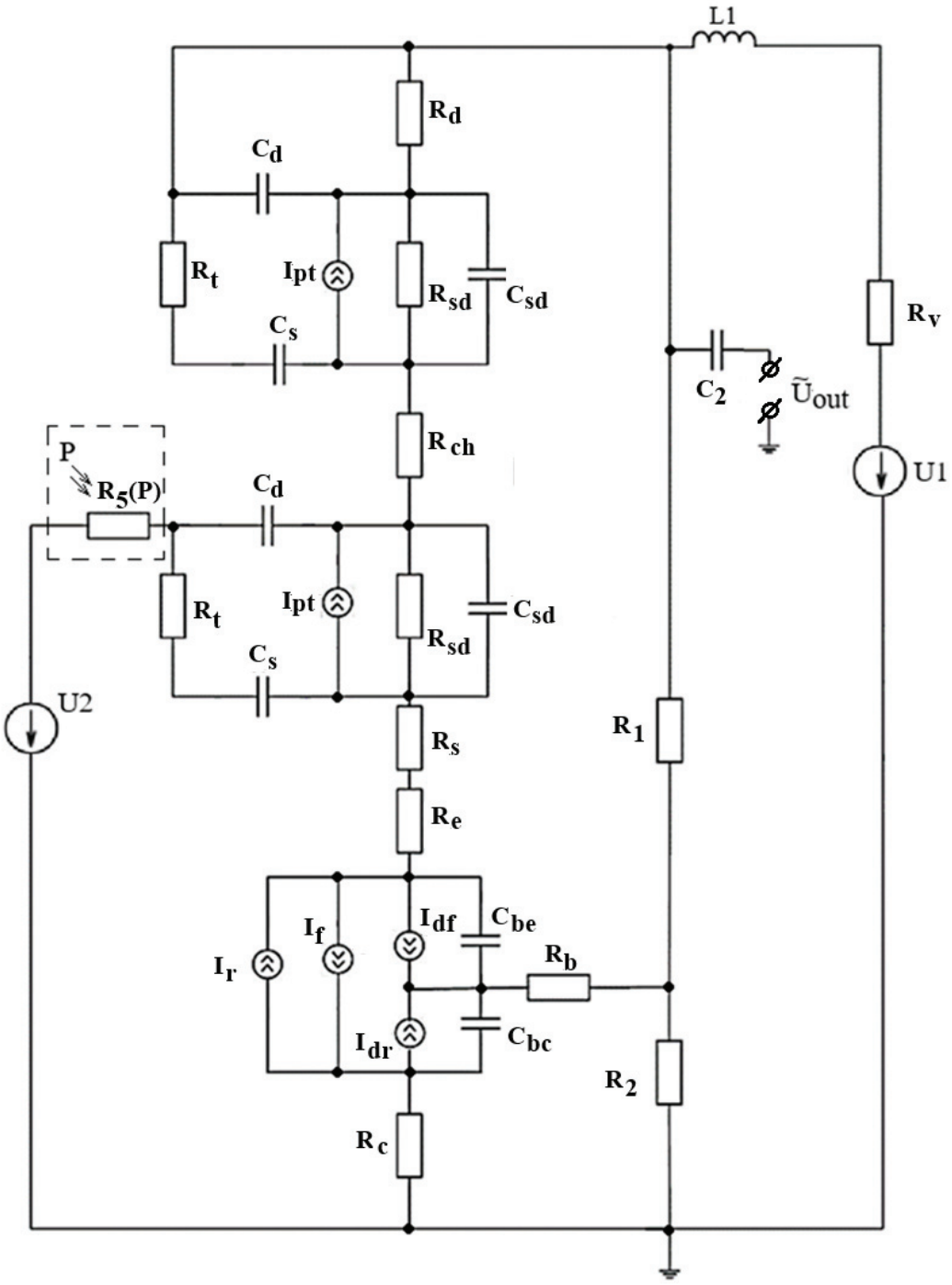


Figure 3. Nonlinear equivalent circuit of a self-oscillating pressure transducer based on a bipolar field-effect transistor structure.

The elements of the equivalent circuit are described as follows: $L1$—inductance; R_V—internal inductance resistance; R_1, R_2, and R_3—load resistances of drain–drain and base-collector transistors; $R_5(P)$—resistance of the strain-sensitive resistor; R_{ds}, R_d, and R_s—channel, drain, and leakage resistances of a double-gate MOSFET; R_e, R_c, and R_b—emitter, collector, and base resistances of a bipolar transistor; C_e and C_c—capacitances of the emitter and collector junctions of a bipolar transistor; C_s and C_d—gate-drain and gate-emitter capacitances; C_{sd}—capacitance of the MOSFET channel; I_f and I_r—forward and reverse currents of a bipolar transistor; I_{dr} and I_{df}—currents of the internal base-collector

and base-emitter junctions; I_{pt}—channel current of a double-gate MOSFET; U_1—power supply voltage source; U_2—control voltage source.

To build a mathematical model of a self-oscillating pressure transducer based on a bipolar-field-effect transistor structure, we apply the state variable method, which allows us to determine the values of voltages and currents of circuit elements at each moment of time and changes in pressure on a pressure-sensitive strain gauge. Therefore, using the state variable method, we present a transformed equivalent circuit (Figure 4), in which all capacitances are represented as independent voltage sources, and inductance is a current source. To consider the properties of the proposed circuit and create its mathematical model, we created a transformed equivalent DC circuit shown in Figure 4. The following simplification was made in the equivalent circuit: $I_{bt} = \frac{I_f - I_\gamma}{QB}$, where QB is the coefficient of the imperfection of the junction of the bipolar transistor [30].

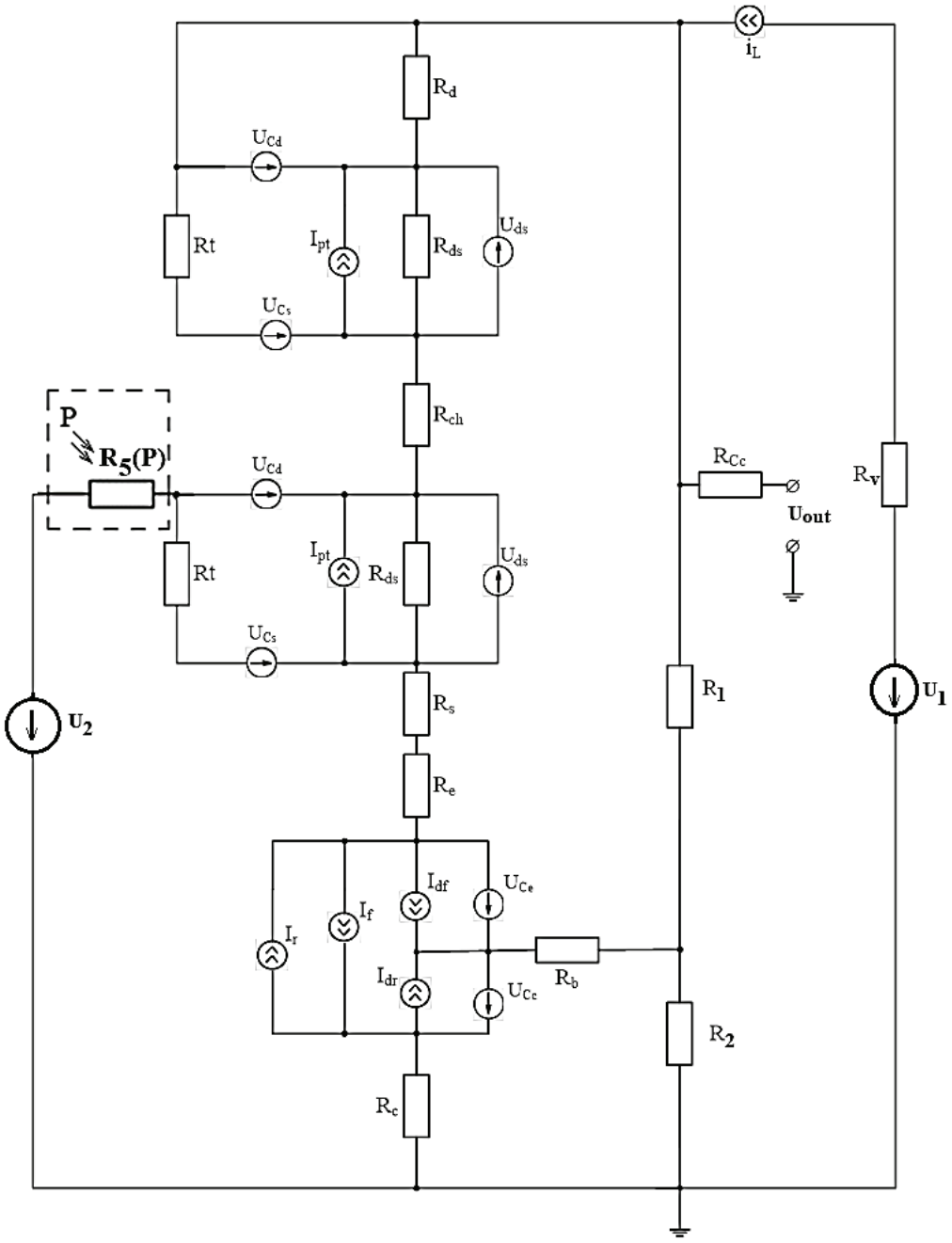


Figure 4. Transformed nonlinear equivalent circuit of a self-oscillating pressure transducer.

Having chosen the directions of the circuit bypass and designated the corresponding currents, we write down a system of equations according to Kirchhoff's laws (its variant, the method of circuit currents). The resulting system of equations will be used to analyze the circuit using the state variable method.

$$\begin{cases} U_L = U_1 - \left(i_{R_2}(R_1 + R_{cc}) + i_{R_5}R_2 - i_L R_V\right), \\ i_{R_3} = i_{R_2} - i_{R_5}, \\ i_{R_{d3}} = \dfrac{-U_{C_{SD}}}{R_{d3}}, \\ i_{c_a} = \dfrac{U_{c_s} - U_{C_{SD}} - U_{c_a}}{R_t}, \\ i_{R_{d4}} = \dfrac{U_{C_{SD}}}{R_{d4}}, \\ i_{R_2} = i_{c_a} + i_{R_{d4}} + I_L, \\ i_{R_5} = \dfrac{-U_{Cc} - i_{R_c}R_c - i_{R_3}R_3}{R_b}, \\ i_{R_c} = \dfrac{-R_s(P)(i_{c_a} + i_{c_s}) - i_{R_c}(R_e + R_s) - (U_2 + U_{c_s} + U_{Cc} + U_{c_e})}{R_c}, \\ i_{R_1} = i_{c_d} + i_{c_s}, \\ i_{c_s} = -I_{pt} - i_{R_{d3}} + i_{c_a} + i_{R_5}, \\ i_{c_a} = i_{R_{d4}} + I_{pt} - i_{c_d} + i_{R_{d3}}, \\ i_{c_d} = i_{R_1} + I_{bt} - I_{d\gamma}, \\ i_{R_1} = i_{R_c} + I_{bt} - I_{d\gamma}, \\ i_{R_3} = i_{R_2} - i_{R_5} \end{cases} \tag{1}$$

Assuming the corresponding values of the voltages across the capacitors and the current through the inductor, we solve the system of Equation (1) with respect to the following values: U_L, i_{c_d}, i_{c_s}, i_{c_d}s, i_{c_c}, i_{c_e}, by making the following substitutions:

$$\begin{cases} i_{c_d} = A_4\,; \\ i_c = \dfrac{A_{10} + A_{11}R_3A_{17}}{A_9\,R_c}\,; \\ i_s = \dfrac{A_{14} + A_{16}A_{17}}{A_5} + I_{bt} - I_{d\gamma}\,; \\ i_e = \dfrac{A_4 + A_{16}A_{17}}{A_5} + I_{pt} - A_4 - \dfrac{U_{c_a}}{R_*}\,; \\ i_{c_d} = A_4\,; \\ U_L = U_1 - \left(A_{12} + i_L(R_1 + R_2 + R_V)\right) + A_{17}, \end{cases} \tag{2}$$

where

$$A_1 = \frac{-U_{C_{SD}} + U_{C_s} - U_{C_d}}{R_t} + I_{pt} + \frac{U_{C_d}}{R_d} - \frac{U_{C_{SD}}}{R_{ds}};$$
$$A_2 = I_{bt} + I_{d\gamma}; \qquad A_3 = U_2 + U_{C_s} + U_{C_c} + U_{C_e};$$
$$A_4 = i'_{C_d}; \qquad A_5 = -R_s(P)A_4 + I_{pt} - A_1 - \frac{U_{C_S D}}{R_{ds}} - A_3; \qquad A_6 = R_e + R_s + R_5(P);$$
$$A_7 = 2U_{C_S D} + U_{C_s} + U_{Cd} + U_{out};$$
$$A_8 = \left(\frac{U_{C_s} - U_{C_{SD}} - U_{C_d}}{R_t} + \frac{U_{C_d}}{R_d} + i_L\right)(R_2 + R_3);$$
$$A_9 = 1 - A_6 - \frac{A_6}{R_b}(2R_2 + R_3);$$
$$A_{10} = A_5 + A_6A_4 + A_6A_8 + \frac{A_6U_{Cc}(2R_1 + R_2)}{R_b}; \qquad A_{11} = \frac{A_6(2R_1 + R_2)}{R_b};$$
$$A_{12} = \left(A_4 + U_{C_d}/R_d\right)(R_2 + R_3); \qquad A_{13} = U_{Ce}A_9 + A_{10};$$
$$A_{14} = -A_7A_9 - A_{10} + A_4A_9 - \frac{U_{C_d}}{R_d}A_9;$$
$$A_{15} = A_9(2R_s + R_e); \qquad A_{16} = A_{11}R_3 + A_9R_3;$$
$$A_{17} = \frac{-U_{Ce}A_9 - A_{10} - 2A_{14}R_b - A_{15}R_b(I_{bt} + I_{pt}) + A_5A_1R_b + A_{15}R_bI_{d\gamma} + \frac{A_{15}R_bU_{C_{SD}}}{R_{ds}}}{2A_{16}R_b + A_{11}R_3 + A_9R_3}.$$

Since the voltages and currents of the circuit of an oscillator pressure transducer based on a bipolar-field-effect transistor structure vary over time, the capacitance current and inductance voltage are described by the expressions: $i_C = C\frac{dU_C}{dt}$, $U_L = L\frac{di_L}{dt}$. If we take into account that each instantaneous value of voltage and current in the equivalent circuit is a function of time, then we can write the left-hand sides of the equations in the form of first-order differential equations, and system (2) can be rewritten as follows:

$$\begin{cases} C_d\dfrac{dU_{C_d}(t)}{dt} = \dfrac{U_{C_s}(t) - U_{C_{SD}}(t) - U_{C_d}(t)}{R_t}, \\ C_C\dfrac{dU_{C_c}(t)}{dt} = \dfrac{U_{C_s}(t) - U_{C_{SD}}(t) - U_{C_d}(t)}{R_t} + I_{bt} - I_{d\gamma}, \\ C_e\dfrac{dU_{C_e}(t)}{dt} = A_4 + I_{bt} - I_{d\gamma}, \\ C_s\dfrac{dU_{C_s}(t)}{dt} = -I_{pt} - \dfrac{U_{C_{SD}}(t)}{R_{d3}} + A_3, \\ L\dfrac{di_L(t)}{dt} = U_2 - i_L(t)R_4 - \dfrac{U_{C_s}(t) + U_{C_c}(t) + U_{C_e}(t)}{R_t + R_s}, \\ C_{SD}(t)\dfrac{dU_{C_{SD}}(t)}{dt} = \dfrac{U_{C_d}(t)}{R_{d4}} + I_{pt} + \dfrac{U_{C_s}(t) - U_{C_{SD}}(t) - U_{C_d}(t)}{R_t} + \dfrac{U_{C_{SD}}(t)}{R_{d3}}. \end{cases} \tag{3}$$

Given that the system of Equation (3) contains terms that describe nonlinear elements (current sources), the system is nonlinear. Nonlinear circuit elements describe the currents of nonlinear internal sources of circuit elements. The forward and reverse junction currents of a bipolar transistor are described by equations:

$$I_f = I_s\left(e^{\frac{U_{BE}}{U_T N_F}} - 1\right); \qquad I_r = I_s\left(e^{\frac{U_{BK}}{U_T N_R}} - 1\right). \tag{4}$$

The currents of the internal junctions of the base-emitter and base-collector bipolar transistors are described by equations:

$$I_{DF} = \frac{I_f}{B_F}; \qquad I_{DR} = \frac{I_r}{B_R}, \tag{5}$$

where B_F and B_R are the maximum current transfer coefficients in the common emitter switching scheme.

The channel current of a double-gate MOSFET is represented as a nonlinear piecewise approximation:

$$I_{pt} = \begin{cases} 0, & U_{GS} - U_{VTO} \le 0, \\ \beta\,(U_{GS} - U_{VTO})^2, & U_{GS} - U_{VTO} \le U_{DS}, \\ \beta\,U_{DS}\left[2\,(U_{GS} - U_{VTO}) - U_{DS}\right], & U_{GS} - U_{VTO} > U_{DS}, \end{cases} \tag{6}$$

where *NF* and *NR* are the MOSFET imperfection coefficients in the forward and inverse modes of operation; *BF* and *BR*—maximum current transfer coefficient; U_{VTO}—threshold voltage; $\beta = \frac{1}{2} S_C W_{pL}$—slope of the double-gate MOSFET characteristic, where *Sc* is the specific slope of the characteristic; W_{pL}—ratio of channel width to its length; $U_{GS} = U_G - U_S$—voltage at the internal gate-drain point; $U_{DS} = U_D - U_S$—voltage at the internal drain-drain point.

Any capacitance of a bipolar transistor has diffusion and barrier components in the linear mode of operation. The capacitance of the collector junction is described by the equation:

$$\begin{aligned} C_c &= C_{ceq} + C_{cs}, \\ C_{ceq} &= \frac{(I_r + I_s)\,TR}{U_T \cdot NR}, \end{aligned} \tag{7}$$

where

$$C_{cs} = \begin{cases} \dfrac{C_{JC}}{\left(1 - \dfrac{U_{BC}}{V_{JC}}\right)^{M_{JC}}}, & U_{BC} < 0, \\ C_{JC}, & U_{BC} \ge 0, \end{cases} \tag{8}$$

where U_t is the temperature potential of the *p-n* junction.

The capacitance of the emitter junction is described by the following expressions:

$$\begin{aligned} C_e &= C_{eeq} + C_{es}; \\ C_{eeq} &= \frac{(I_r + I_s)\,\tau_{in}}{U_T\,\tau_{nom}} \end{aligned} \tag{9}$$

where

$$C_{es} = \begin{cases} \dfrac{C_{JE}}{\left(1 - \dfrac{U_{BE}}{V_{JE}}\right)^{M_{JE}}}, & U_{BE} < 0, \\ C_{JE}, & U_{BE} \ge 0, \end{cases} \tag{10}$$

where C_{e_dif}, C_{e_b}, C_{c_dif}, and C_{c_b} are the differential and barrier capacitances of the emitter and collector junctions of the bipolar transistor; τ_{inv} and τ_{nor}—the time of charge transfer through the base region in the inverted and normal operating modes; U_{JE} and U_{JC}—contact potential difference of the emitter and collector junctions of a bipolar transistor; C_{JE} and

C_{JC}—capacitance of the emitter and collector junctions of a bipolar transistor; M_{JE} and C_{JC}—smoothness coefficients of the emitter and collector junctions of a bipolar transistor.

The dynamic model of the self-oscillating pressure transducer (system of Equation (3)) allows us to determine the value of the output signal frequency depending on the control voltage, power supply, and pressure action, as well as the parameters of the main elements of the self-oscillator at any given time. Using the MATLAB (MathWorks, Natick, MA, USA) software package, a computer circuit study of the parameters and characteristics of the generated electrical oscillations was performed.

The conversion function is described by the expression

$$F = \frac{1}{2\pi R_s(P) C_{\text{EKV}}(P)} \left(\frac{R_s^2(P) C_{\text{EKV}}(P)}{L_1} - 1 \right)^{\frac{1}{2}}, \tag{11}$$

where $R_g(P)$ is the dynamic differential negative impedance of the generator; $C_{EKV}(P)$ is the equivalent capacitance of the transducer; $L1$ is the inductance of the oscillating circuit.

The sensitivity of the pressure transducer is determined on the basis of expression (11) and is described by the following formula:

$$\begin{aligned} S_F^P = & \frac{\frac{R_s(P) C_{\text{EXT}}(P)}{L_1} \frac{dR_s(P)}{dP} + \frac{R_s^2(P)}{L_1} \frac{dC_{\text{EKV}}(P)}{dP}}{4\pi R_s(P) C_{\text{EKV}}(P) \sqrt{\frac{R_s^2(P) C_{\text{EKV}}(P)}{L_1} - 1}} \\ & - \frac{\sqrt{\frac{R_s^2(P) C_{\text{EKV}}(P)}{L_1} - 1}}{2\pi R_s^2(P) C_{\text{EKV}}(P)} \frac{dR_s(P)}{dP} - \frac{\sqrt{\frac{R_s^2(P) C_{\text{EKV}}(P)}{L_1} - 1}}{2\pi R_s(P) C_{\text{EKV}}^2(P)} \frac{dC_{\text{EKV}}(P)}{dP}. \end{aligned} \tag{12}$$

Approximate Conversion Function

$$y = \frac{a + c \cdot \ln(x) + e \cdot (\ln(x))^2}{1 + b \cdot \ln(x) + d \cdot (\ln(x))^2 + f \cdot (\ln(x))^3} \tag{13}$$

where y is the specified parameter (pressure); x is the generation frequency; a, b, c, d, e, and f are the coefficients.

The experimental studies were carried out using an Arinst SSA TG-LC spectrum analyzer (Arinst, Saint Petersburg, Russia).

The frequency range of autogenerated parametric pressure devices based on transistor microelectronic structures with negative differential resistance with primary strain-sensitive elements such as strain gauge and strain gauge is selected in the Long Term Evolution (LTE) Band 3 (DCS-1800) uplink: 1710–1785 MHz. The standard uses the 1800 MHz frequency band, which coincides with the range of Global System for Mobile Communications (GSM) networks. When creating a 4G network on the basis of an already built 2G infrastructure, operators usually retain the original network and allocate part of the band for the new 4G network. LTE Band 3 is found in both urban and rural areas [31]. This type of communication operates in Frequency Division Duplex (FDD) mode with frequency separation of the downlink and uplink. The Internet of Things (IoT) will redefine the way people interact with each other and with the objects around them, with the goal of creating a global network that connects everyone and everything. The IoT is expected to enable the use of mass machine-to-machine communication (mMTC), by which devices transmit and receive small amounts of data, to improve industrial and measurement processes [32]. Low power radio communication (LPWA) technologies have emerged as a solution to developing mMTC that meets the requirements of low power consumption, low data rate, scalability,

and long range. The use of Narrowband Internet of Things (NB-IoT), as a completely new cellular technology, will allow a step forward in the realization of 5G-based IoT, and NB-IoT will be compatible with the 4G LTE system and with the 2G GSM [31–33]. This advantage, coupled with a higher spectrum utilization efficiency, puts the technology in a privileged position for the implementation of LPWA networks [32].

The radiomeasuring information system (RIS) for monitoring the parameters of an athlete rower in academic rowing is designed to analyze and determine the main characteristics of the "boat–rower" system in real time. Figures 5–8 show the structural blocks of the RIS system.

Measurements are made every 8–12 ms in parallel on 50 channels. The developed system consists of measuring units on the right and left oars, a measuring unit on the sliding seat, and the main measuring unit with a transmitter on the boat. All units are connected via a radio channel, and the main unit, which collects and processes the measured information, is connected to a personal computer located on the shore with a more powerful transmitter in duplex mode.

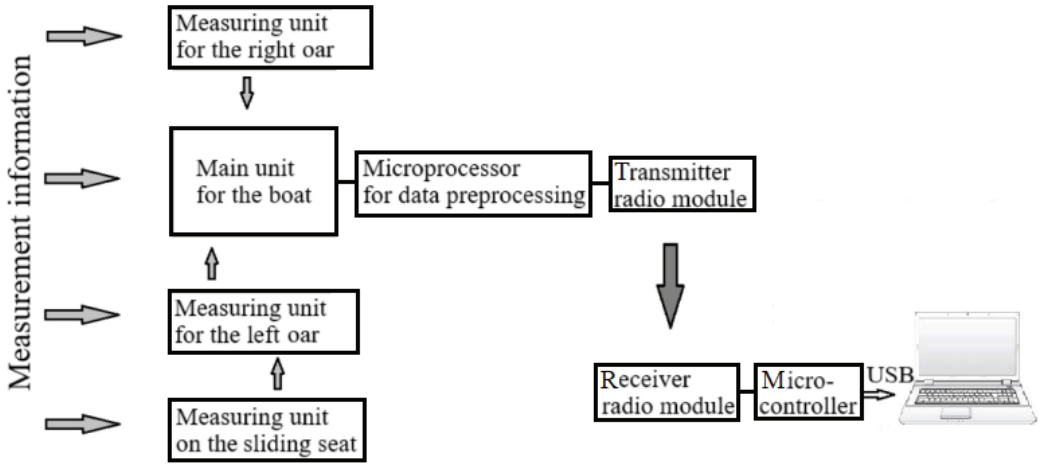


Figure 5. Block diagram of the VLSI for monitoring the parameters of an athlete-rower in rowing.

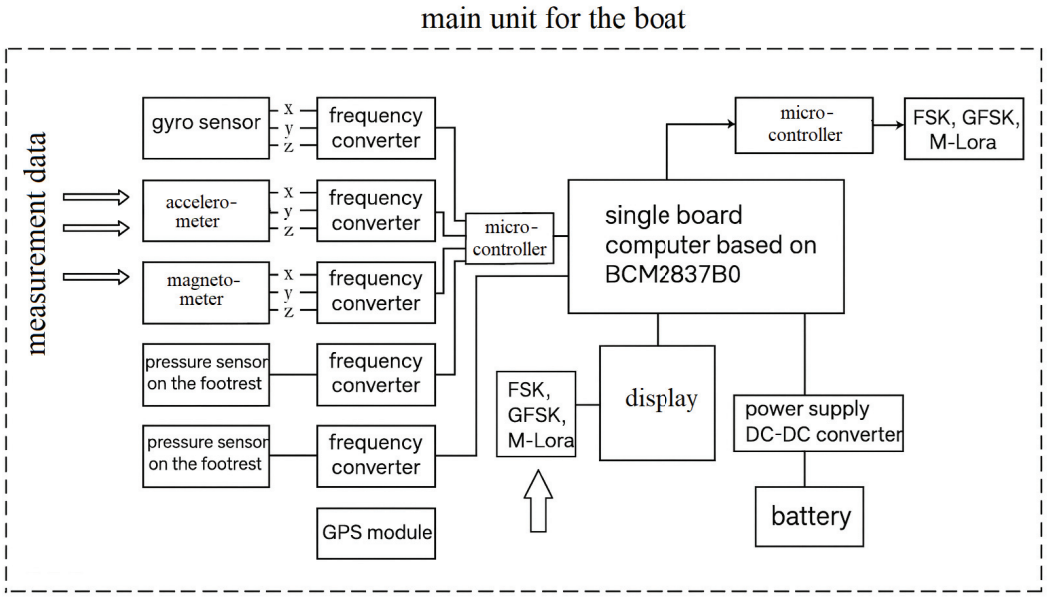


Figure 6. Block diagram of the main unit of the VLSI.

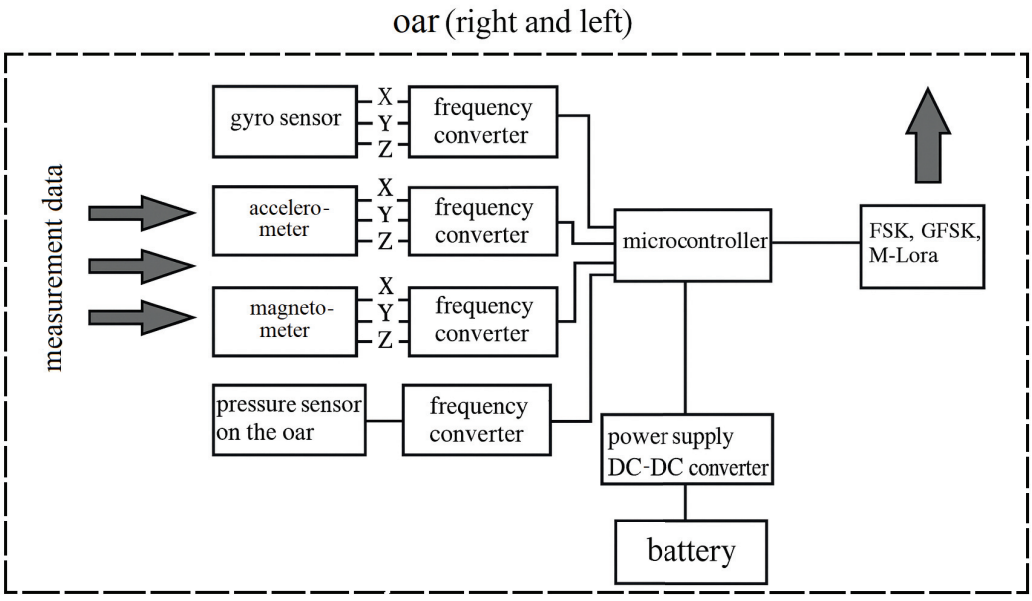


Figure 7. Block diagram of the oar block of the RVIS.

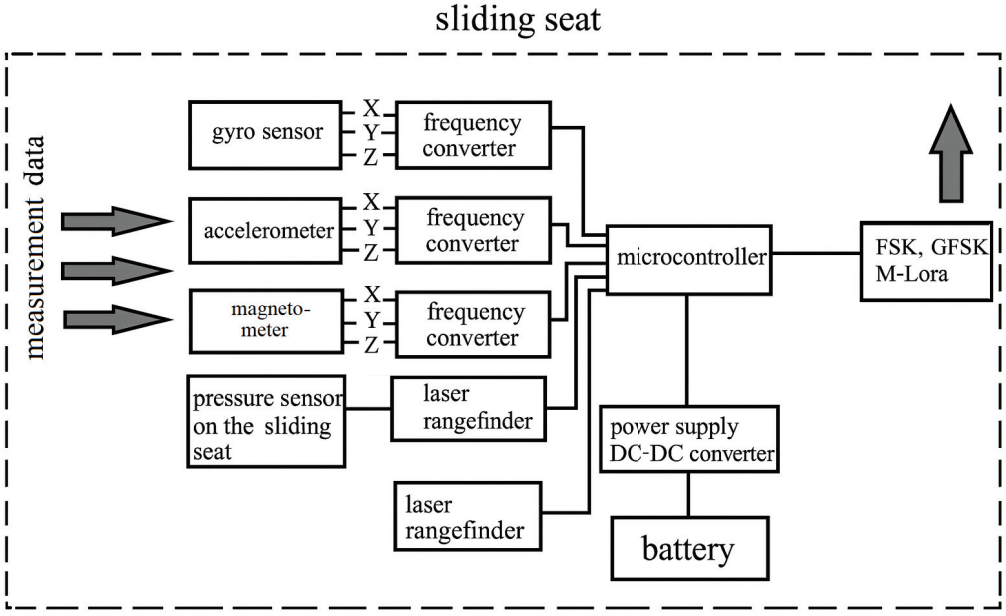


Figure 8. Block diagram of the RVIS sliding seat unit.

3. Results

The dynamic model of the self-generating pressure transducer (system of Equation (3)) allows determining the value of the output signal frequency depending on the control voltage, power supply, and pressure action, as well as the parameters of the main elements of the self-generator at any time. Using the MATLAB (MathWorks, Natick, MA, USA) software package, a computer circuit design study of the parameters and characteristics of the generated electrical oscillations was performed. Figure 9 shows the simulated output signal of the self-generating pressure transducer, and Figure 10 shows the simulated output signal of the self-oscillating pressure transducer under pressure 18 kg/cm^2.

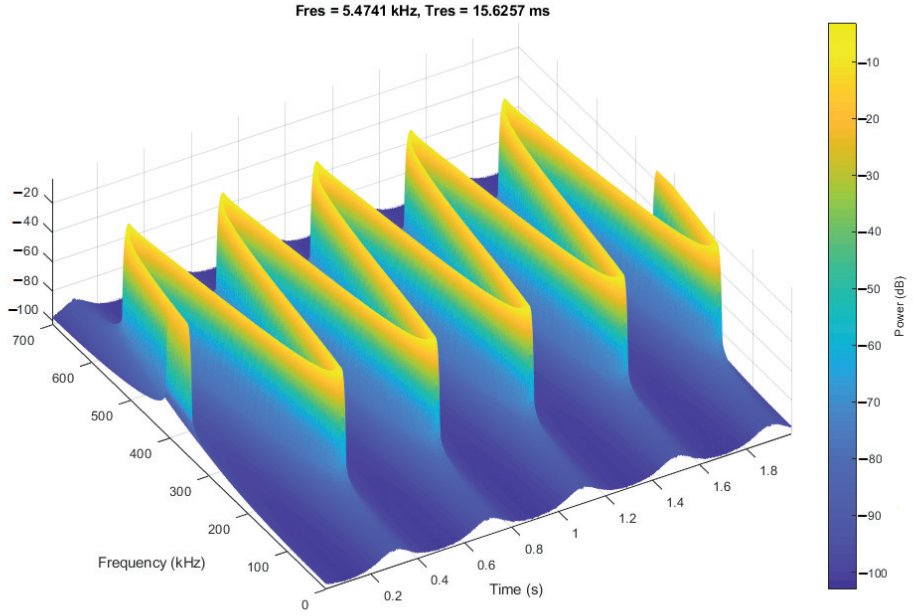


Figure 9. Simulated output signal of an autogenic pressure transducer.

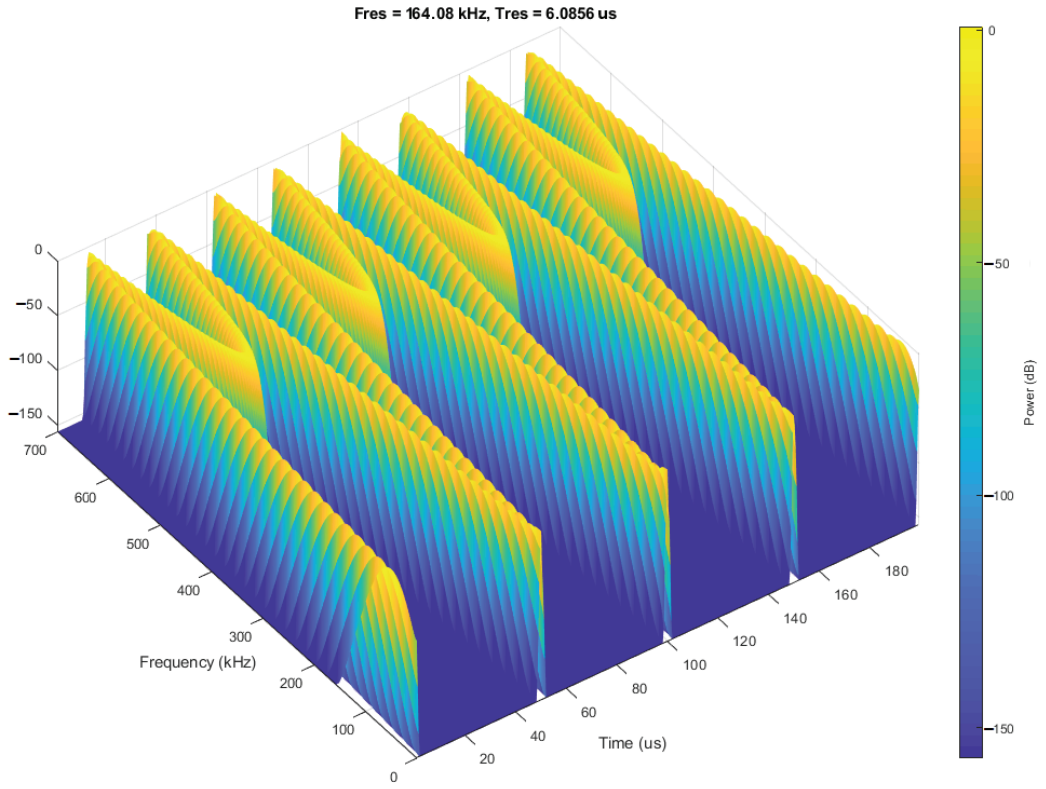


Figure 10. Simulated output signal of an autogenous pressure transducer under pressure (18 kg/cm^2).

The conversion function graph is shown in Figure 11, and the sensitivity graph is shown in Figure 12. As can be seen from Figure 12, the sensitivity of the transducer varies from 42.5 Hz/g/cm^2 to 78.0 Hz/g/cm^2 in the range of 0 to 26 kg/cm^2.

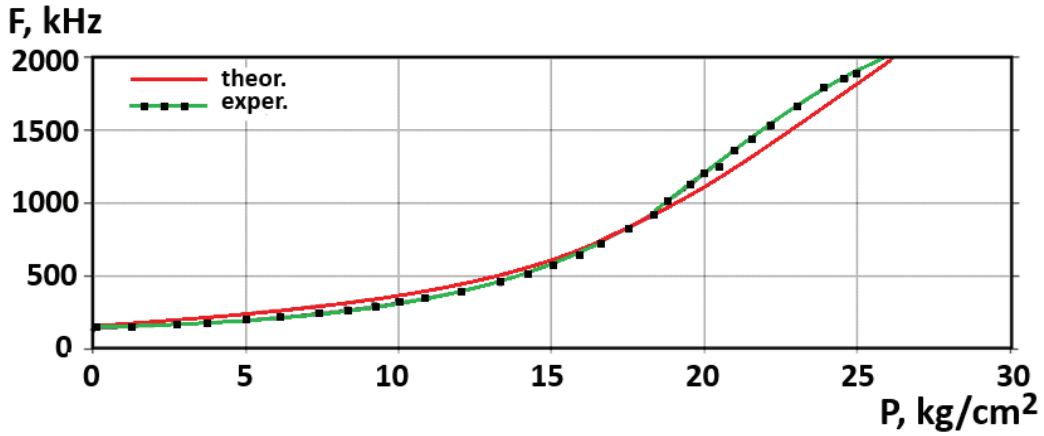


Figure 11. Theoretical and experimental dependences of generation frequency on pressure changes.

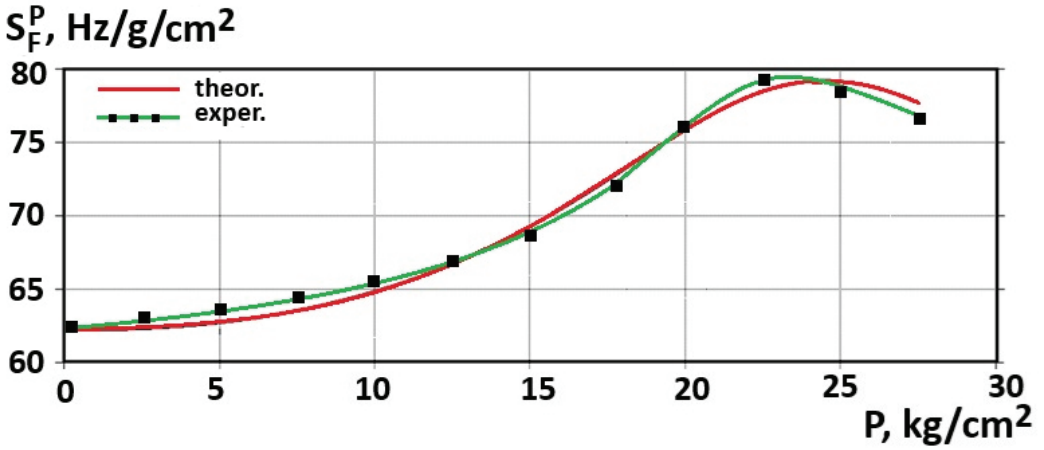


Figure 12. Dependence of transducer sensitivity on pressure changes.

As can be seen in the graph in Figure 11, the highest sensitivity of the device lies in the range of 15 kg/cm^2 to 25 kg/cm^2 and is 62.5–79.5 Hz/g/cm^2. The discrepancy between the theoretical and experimental studies is ±0.5%.

In Figure 13, the approximated transformation function is shown.

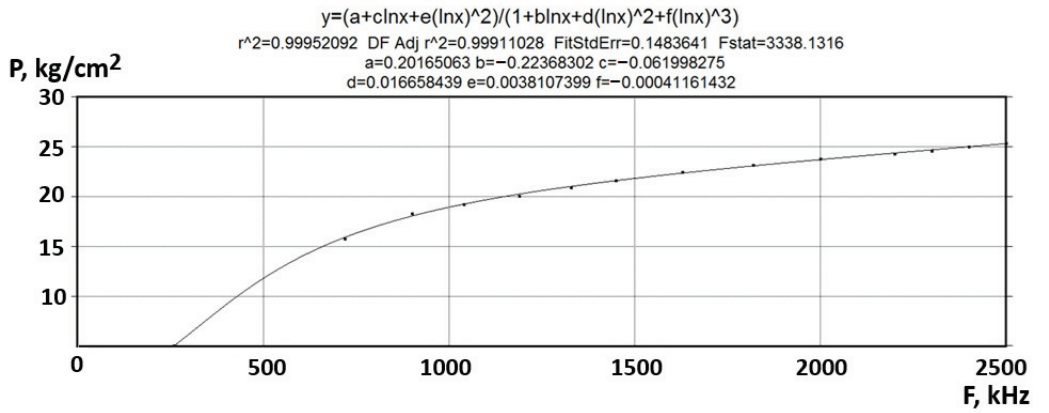


Figure 13. Approximate transformation function.

In Figure 13, the experimental dependence of pressure P (kg·cm^{-2}) on excitation frequency F (kHz) for the sensor is shown, together with a rational–logarithmic fit $y = (a + c\ln x + e(\ln x)^2)/(1 + b\ln x + d(\ln x)^2 + f(\ln x)^3)$. The response is monotonic and sublinear: P increases rapidly to ∼0.5–1.0 MHz and then gradually approaches

a quasi-saturation above ~1.5–2.0 MHz. The fitted parameters are $a = 0.20165603$, $b = -0.22368302$, $c = -0.061998275$, $d = 0.016685439$, $e = 0.0038107399$, and $f = -0.00041161432$. The quality of fit metrics indicate an excellent agreement with the data ($R^2 = 0.99952092$, adjusted $R^2 = 0.99911028$, standard fit error = 0.1483641, $F = 3338.1316$).

Figure 14 shows the signal spectrum of a self-oscillating parametric pressure device based on a transistor microelectronic structure in the DCS-1800 uplink band; the transmission frequency is 1750.0 MHz.

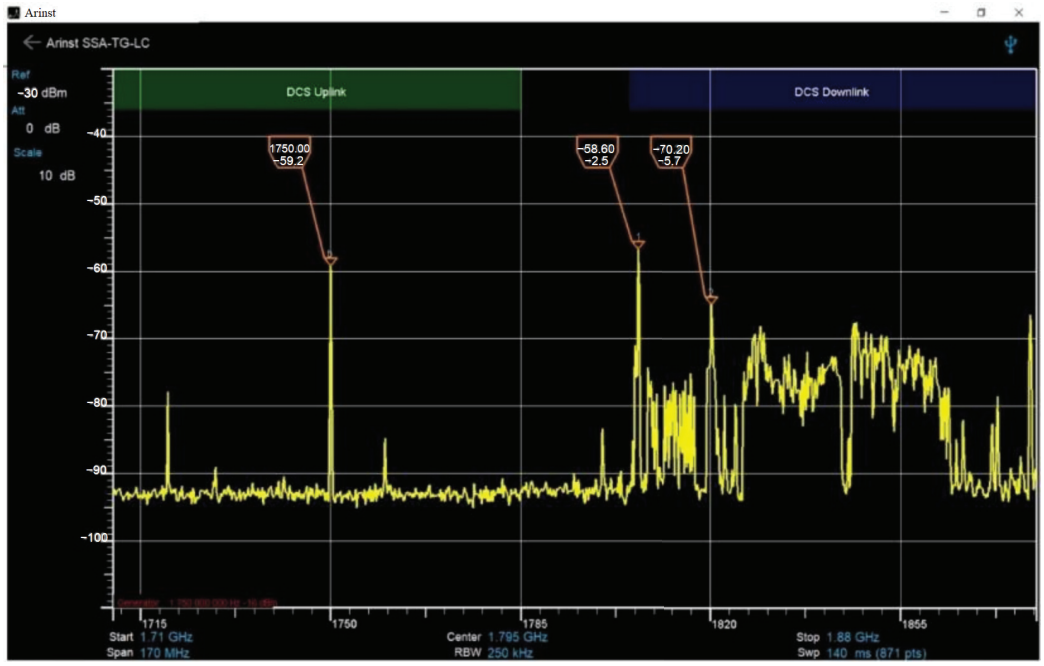


Figure 14. Signal spectrum of a self-oscillated parametric pressure device based on a transistor microelectronic structure.

Figures 15 and 16 present photos of the device. Figure 16 shows a view of the system with the data display, and Figure 17 shows the interface of the data processing program. The developed system provides data packet transmission at the level of 60–70 FPS. The data received from each sensor are displayed in digital form and in the form of graphs.

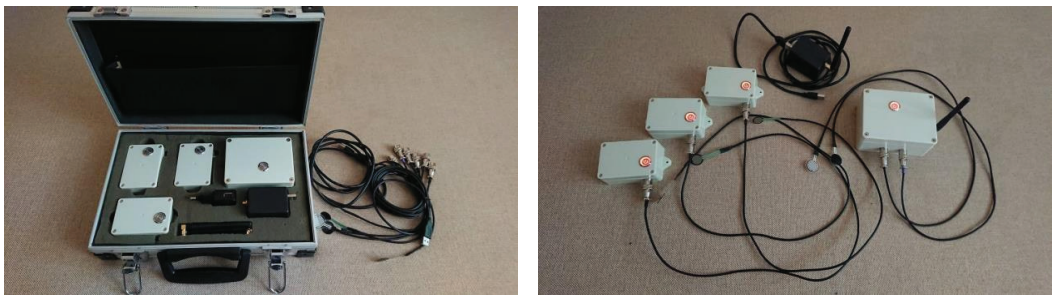


Figure 15. Exterior view of the RIS.

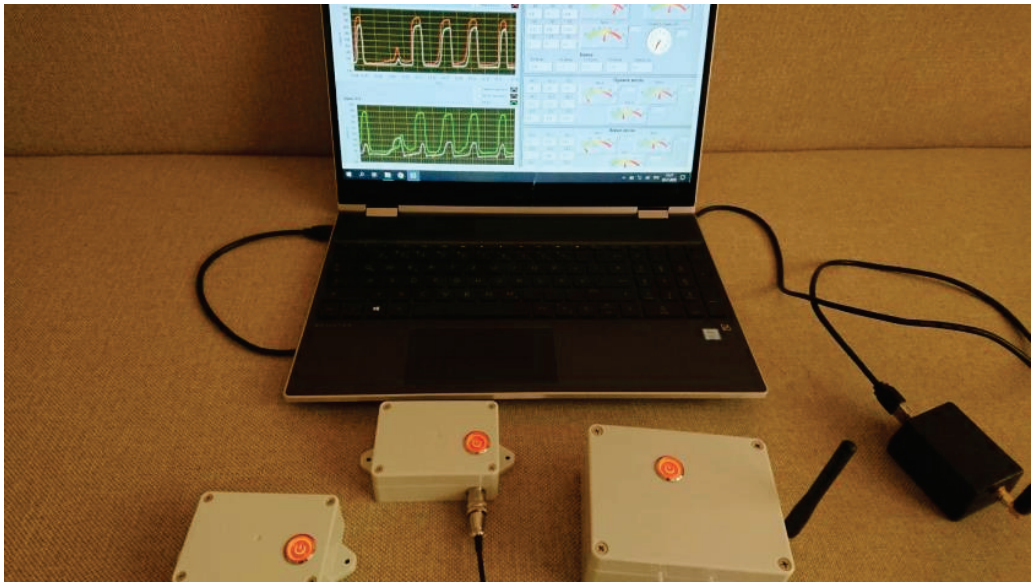


Figure 16. System view with data display on a personal computer.

The general data of the RVIS monitoring the parameters of an athlete-rower in rowing are as follows:

- Weight of the entire system mounted on the boat—0.75 kg;
- Water resistance—IP65 protection standard;
- Working hours—up to 24 h;
- Operation in the temperature range—$-10\,^{\circ}\mathrm{C}\ \ldots\ +60\,^{\circ}\mathrm{C}$;
- Installation without compromising the integrity of the rowing equipment;
- High-speed telemetry data transmission in real time to the PC—data packet transmission at the level of 60–70 FPS;
- Transmission range without an additional amplifier—up to 2500 m (with an additional amplifier—up to 15 km);
- Real-time data recording to a PC in a .txt file, which ensures minimal file size and maximum possibility of further data processing;
- Data packet recording resolution up to 100 Hz;
- Visualization of large amounts of data in real time.

In Figure 17, the LabVIEW-based acquisition/visualization interface used for on-water rowing experiments is presented. The left panel shows two synchronized time series of force F (N) versus time: the upper plot displays the forces of the right/left shoulder (color coded traces), and the lower plot shows the forces of the bench/footrests. Periodic stroke cycles with peaks approaching $\sim$800 N and near-zero recovery phases are evident. The right panel provides real-time telemetry and controls the following: serial port status and time/date, channel selectors, and instrument clusters for the boat, bench, and each oar. For each module, raw IMU channels (AX, AY, AZ accelerations, and GX, GY, GZ angular rates) are displayed alongside gauge indicators of yaw/roll/pitch; a dial reports the boat speed in m/s. Together, the interface enables the synchronous monitoring of propulsive forces and 6-DOF kinematics for the boat, bench, and both oars during rowing trials. The quantitative parameters and calculated characteristics obtained from these measurements are summarized in Table 2.



Figure 17. Interface of the data processing and visualization program. Non-English terms correspond to: Сила = Force; Лодка = Boat; Банка = Bench; Правое весло = Right oar; Левое весло = Left oar; Правая подножка = Right footrest; Левая подножка = Left footrest; Рыскание/Крен/Дифферент = Yaw / Roll/Pitch; Скорость лодки, м/с = Boat speed, m/s.

Table 2. Parameters and characteristics of the biomechanical system "boat–rower" measured and calculated by the RVIS.

№	Measured Parameter	Measured Value Range	Measurement Error	Real-Time Measurement	Calculation from the Database
1x	Coordinates in space—latitude	49.233903	2 m	+	
2x	Coordinates in space—longitude	28.415983	2 m	+	
3x	Boat speed (instantaneous) per 100 ms	0–25 m/s	0.05 m/s	+	
4x	Footrest pressure is instantaneous, for the right foot	0–1500 H	0.5 H	+	
5x	Pressure on the footrest is instantaneous, for the left foot	0–1500 H	0.5 H	+	
6x	Yaw, boat heading (Yaw)	360°	0.1°	+	
7x	Boat roll (Roll)	90°	0.1°	+	
8x	Boat differential (Pitch)	90°	0.1°		+
	Boat speed (average per stroke resulting in (x, y, z))	0–25 m/s	0.05 m/s		+
	Boat movement (average per stroke and for the supported and unsupported part of the stroke (x, y, z))	0–25 m	0.01 m		+
	Pulse of pressure force on the right footrest	0–1500 N·s	1 N·s		+
	Pulse of pressure force on the left footrest	0–1500 N·s	1 N·s		+
	Pressure gradient on the footrest for the right foot	0–70,000 N/s	1 N/s		+
	Pressure gradient on the footrest for the left foot	0–70,000 N/s	1 N/s		+
9x	Angular acceleration of the boat along the x-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
10x	Angular acceleration of the boat along the y-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
11x	Angular acceleration of the boat along the z-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
12x	Boat angular velocity x-axis	0–300 °/s	0.5 °/s	+	
13x	Boat angular velocity y-axis	0–300 °/s	0.5 °/s	+	
14x	Boat angular speed z-axis	0–300 °/s	0.5 °/s	+	
15x	Linear acceleration of the boat along the x-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
16x	Linear acceleration of the boat along the y-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
17x	Linear acceleration of the boat along the z-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
18x	Pressure on the right oar blade	0–1500 H	0.5 H	+	
	Force impulse on the right oar blade	0–1500 N·s	0.5 N·s		+
	Force gradient on the right oar blade	0–50,000 N/s	1 N/s		+
	Work of the right oar blade per stroke	0–5000 J	1 J		+
	Power of the right oar blade per stroke	0–10,000 W	1 W		+
19x	Angular movement of the right oar along the x-axis	0–360°	0.1°	+	
20x	Angular movement of the right oar along the y-axis	0–180°	0.1°		
21x	Angular movement of the right oar along the z-axis	0–180°	0.1°		
22x	Angular acceleration of the right oar along the x-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
23x	Angular acceleration of the right oar along the y-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
24x	Angular acceleration of the right oar along the z-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
25x	Angular velocity of the right oar x-axis	0–300°/s	0.5°/s	+	
26x	Angular velocity of the right oar y-axis	0–300°/s	0.5°/s	+	
27x	Angular velocity of the right oar z-axis	0–300°/s	0.5°/s	+	
28x	Linear acceleration of the right oar along the x-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
29x	Linear acceleration of the right oar along the y-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
30x	Linear acceleration of the right oar along the z-axis	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
	Movement and length of the trajectory of the right oar blade	0...5 m	0.01 m		+
	Angular movement of the right oar around the longitudinal axis	0–180°	0.1°		+
	Linear speed of the right oar	0–50 m/s	0.05 m/s		+

Table 2. *Cont.*

№	Measured Parameter	Measured Value Range	Measurement Error	Real-Time Measurement	Calculation from the Database
31x	Pressure on the left oar blade	0–1500 H	0.5 H	+	
	Force impulse on the left oar blade	0–1500 N·s	0.5 N·s		+
	Force gradient on the left oar blade	0–50,000 N/s	1 N/s		+
	Work of the left oar blade per stroke	0–5000 J	1 J		+
	Power of the left oar blade per stroke	0–10,000 W	1 W		+
32x	Angular movement of the left oar along the x-axis	0–360°	0.1°	+	
33x	Angular movement of the left oar along the y-axis	0–180°	0.1°	+	
34x	Angular movement of the left oar along the z-axis	0–180°	0.1°	+	
35x	Angular acceleration of the left oar along the x-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
36x	Angular acceleration of the left oar along the y-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
37x	Angular acceleration of the left oar along the z-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
38x	Left oar angular velocity x-axis	0–300°/s	0.5°/s	+	
39x	Angular velocity of the left oar y-axis	0–300°/s	0.5°/s	+	
40x	Angular velocity of the left oar z-axis	0–300°/s	0.5°/s	+	
41x	Linear acceleration of the left oar along x	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
42x	Linear acceleration of the left oar along y	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
43x	Linear acceleration of the left oar along z	$-250 \ldots +250$ m/s^2	0.01 m/s^2	+	
	Displacement and length of the trajectory of the left oar blade (resultant and in two planes)	0...5 m	0.01 m		+
	Angular movement of the left oar around the longitudinal axis	0–180°	0.1°		+
	Linear velocity of the left oar (average and instantaneous in two planes)	0–50 m/s	0.05 m/s		+
	Left oar angular velocity around the longitudinal axis (average and instantaneous)	0–300°/s	0.5°/s		+
	Duration of the whole stroke and the supported and unsupported parts of the stroke	0–20 s	0.01 s		+
	Rowing pace	0–60 g/min	0.1 g/min	+	
	The rhythm of rowing			+	
44x	Pressure on the sliding seat	0–1500 H	0.5 H	+	
45x	Angular acceleration of the sliding seat along the x-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
46x	Angular acceleration of the sliding seat in the y-axis	$-500 \ldots +500$ m/s^2	0.01 m/s^2	+	
47x	Angular velocity of the sliding seat along the x-axis	0–300°/s	0.5°/s	+	
48x	Angular velocity of the left oar y-axis	0–300°/s	0.5°/s	+	
49x	Moving the sliding seat	0–0.9 cm	0.001 m	+	
	The speed of the sliding seat	0–50 m/s	0.05 m/s	+	
	Impulse of pressure force on the sliding seat	0–1500 N·s	1 N·s		+
	Pressure gradient on the sliding seat	0–50,000 N/s	5 N/s	+	

Note: The "+" symbol indicates parameters included in real-time measurement or derived from the database.

4. Discussion

The high speed of the course imposes special requirements on the rowing power supply mechanisms. Maximum efficiency was found to be achieved when the power supply mechanism reaches steady state (uniform fuel consumption for uniformly high power output), that is, speed fluctuations of no more than 2%. Thus, a successful performance in international competitions involves reaching a "cruising" speed, maintaining it throughout the race, and maximizing mobilization in the finish zone [8–11].

The tactic of evenly covering the distance requires a high level of development of the aerobic and glycolytic mechanisms of energy supply. However, these mechanisms cannot develop synchronously, only sequentially [8–13].

It is reliable that all components of sportsmanship (power and capacity of energy processes, level of strength abilities, speed of movements and locomotion, technical skills, special endurance and performance) are highly developed in top-class athletes. To create a developmental training environment for such athletes, high-intensity specific influences are necessary on each of the components of their special training. At the same time, adaptive reactions of the organism can prevent the high efficiency of such loads. In case of exceeding the physiological bar of its capabilities, the body levels the special character of loads that belong to high intensity. In this case, one should not expect a high training effect; the training will have the character of general physical training, with low efficiency and large unbalanced energy expenditures [2,4,11]. In order to eliminate these contradictions or minimize them, the following methodological foundations for organizing the training process were formulated when developing the block system [1,2,4]:

- advanced use of special physical training loads in relation to other areas of training;
- concentration of special physical training loads at the beginning of a large adaptation cycle as a condition that ensures the creation of a functional and energy basis for the intensification of subsequent loads;

- time separation of training loads of different predominant orientation and their inclusion in training in accordance with the objective logic of the development of the process of long-term adaptation;
- utilization of the long-term delayed training effect of concentrated loads of special physical training.

Due to the increased level of the body's functional state, favorable conditions are created for the performance of subsequent high-intensity loads.

A radio-measuring information system for monitoring the parameters of an academic rowing athlete has been developed, which is designed to improve the athlete's movements during the stroke and, in general, the entire distance in order to bring them to automaticity and further control them at the subconscious level. One of the promising scientific directions in the development of radio-measuring microelectronic transducers proposed in this work is the use of the dependence of the reactive properties and negative differential resistance of semiconductor devices on the influence of external physical quantities and the creation of a new class of microelectronic frequency transducers for pressure, angular position, and acceleration on this basis. In devices of this type, the above external influences are converted into a frequency signal, which allows increased accuracy, sensitivity, speed, and stability.

Measurements were made at intervals of 8–12 ms (≈80–120 measurements/s) and averaged over 1-second windows ($n \approx 100$), ensuring a statistically reliable estimate of the mean and standard error. Switching to a shorter interval (e.g., 5 ms → 200 measurements/s) would theoretically reduce the standard error of the mean by approximately $\sqrt{2}$ (1.41 times) under the condition of independent samples, but the actual benefit depends on the sensor bandwidth and correlation between samples. In addition, shorter time intervals require more power consumption and microcontroller and storage load and may introduce more high-frequency interference. Therefore, the chosen interval of 8–12 ms is a compromise between the temporal resolution, statistical reliability, and the practical limitations of the equipment. The question of the feasibility of switching to shorter time intervals (e.g., 5 ms) requires additional tests to determine the sensor bandwidth and assess the overall system performance.

The authors found that the discrepancy between the theoretical and experimental results was no more than ±0.5%. The experimental measurements were performed at a frequency of 80–120 samples/s, with a sample of more than 10^3 values for each point, followed by averaging. The total reduced error, which takes into account the instrumental error, reproducibility, and error of the reference measuring instrument, did not exceed ±0.5% of the upper limit of the range. The average experimental values obtained are consistent with the results of numerical modeling within this error.

To evaluate the quality and compare various physical quantity transducers, we will use the concept of efficiency, which means the suitability of a device to perform a task. When conducting such an analysis, it is necessary to have a measure that is suitable for this purpose, i.e., the efficiency and quality criteria. Efficiency and quality criteria characterize certain aspects of the operation of physical quantity transducers. A device that is optimal for one of the individual criteria may not be optimal for the other. Therefore, when designing and optimizing physical quantity transducers, one strives not for the extremum of any single characteristic but for the overall optimality of the device, the measure of which is the generalized efficiency criterion. The basis for the formation of generalized criteria are individual criteria. They can be divided into three groups: metrological, informational, and economic. Individual metrological criteria coincide with the main normalized metrological characteristics of measuring devices, which are determined by state standards, and allow evaluating certain properties of measuring devices that affect the measurement results.

When evaluating self-oscillating transducers of physical quantities, it is necessary to start with the selection of a performance evaluation criterion. In this case, it is necessary to choose a criterion that would allow the synthesis of optimal devices taking into account the most important performance indicators. These indicators include

- Indicators of constructive perfection;
- Indicators of technical characteristics;
- Indicators of device complexity;
- Indicators of reliability and durability;
- Indicators of manufacturability.

Calculations of the complex technical level indicator are performed using dimensionless relative indicators. The normalization of the indicators P_{ij}, that is, the calculations of their relative values Q_{ij}, are performed using the indicators of the base product $P_{i(basic)}$. The "ideal" device is used as the base device, which contains a set of the highest achievements in all indicators analyzed for the studied group of devices without exception. Calculations of normalized indicators for devices are performed by the following expressions:

$$\begin{aligned} Q_i &= \frac{P_i}{P_{i(\text{basic})}} \quad (\text{if } P_i < P_{i(\text{basic})}) \\ \text{or} \quad Q_i &= \frac{P_{i(\text{basic})}}{P_i} \quad (\text{if } P_i > P_{i(\text{basic})}), \end{aligned} \tag{14}$$

where the index i= 1...6 is the column number of the type of strain gauge (Table 3).

For homogeneous devices, the complex index is determined based on the expression

$$KP_j = \sum_{i=1}^{n} \lg(Q_{ij}), \tag{15}$$

where n is the number of informative indicators, and j= 1...11 is the line number of Table 3 of the technical data from the pressure transmitter.

For heterogeneous devices, the complex index is determined based on the expression

$$KP_j = \sqrt{\sum_{i=1}^{n} (1 - Q_{ij})^2}. \tag{16}$$

The calculation of the normalized indicators of the devices is carried out according to the expression Q_{ij}, according to which the growth of the normalized indicator corresponds to the growth of the quality of the device. In this case, the smaller the absolute value of the complex indicator KP_j, the higher the level of the j-number of the device.

Table 3. Pressure transducers.

№	Indicators of Pressure Transducers	Types of Pressure Transducers						
		Strain-Sensitive Resistors M1	Strain-Sensitive Diodes M2	Strain-Sensitive Bipolar Transistors M3	Strain-Sensitive FETs with p–n Junction M4	MIS Transistors M5	Self-Oscillating Pressure Transducers M6	Ideal Pressure Transducer
1	Supply voltage, V	5	3.3–9	0.8–5	3.3–12	3.3–5	0.8–5	0.8–5
2	Sensitivity, mV/bar	9	12	15	17	22	50	50
3	Nonlinearity of conversion function, %	0.15	0.3	0.5	0.55	0.45	0.3	0.25
4	Repeatability of measurement results, ±%	0.25	0.5	0.3	0.3	0.4	0.25	0.25
5	Temperature sensitivity coefficient, %/K	0.20	0.3	0.5	0.5	0.3	0.4	0.2
6	Power consumption, mW	1.5	0.5	15	5	3	2	1.5

Table 3. *Cont.*

№	Indicators of Pressure Transducers	Types of Pressure Transducers						
		Strain-Sensitive Resistors M1	Strain-Sensitive Diodes M2	Strain-Sensitive Bipolar Transistors M3	Strain-Sensitive FETs with p–n Junction M4	MIS Transistors M5	Self-Oscillating Pressure Transducers M6	Ideal Pressure Transducer
7	Operating frequency range, Hz	500	10^3	10^6	10^6	10^6	10^9	10^9
8	Output signal level, V	0.3	0.1	0.3	0.3	0.4	5	5
9	Manufacturability	1	1	1	1	1	1	1
10	Temperature range, °C	−40 to +85	−40 to +100	−40 to +80	−40 to +80	−40 to +100	−40 to +100	−40 to +100
11	Operating pressure range, bar	0–100	0–100	0–100	0–100	0–100	0–100	0–100
	Summary	**M1 = 1.35**	**M2 = 1.76**	**M3 = 1.55**	**M4 = 1.68**	**M5 = 1.53**	**M6 = 0.78**	

5. Conclusions

The principle of the operation, design, and block diagram of a radio-measuring information system for monitoring the parameters of an athlete-rower in academic rowing, which is designed to analyze and determine the main characteristics of the "boat–rower" system in real time, have been developed. Measurements are made every 8–12 milliseconds in parallel on 50 channels. Based on multivariate analysis, data processing determines the significance of the influence of each factor on the athlete's physical fitness.

Parametric pressure devices are proposed for self-oscillators based on transistor microelectronic structures with negative differential resistance with a primary strain gauge element based on strain gauge, and the primary strain gauge transducers are active elements of self-oscillating circuits, which simplifies the design of pressure devices.

Based on the consideration of physical processes in primary strain gauges and self-oscillators based on the energy conversion method, mathematical models of pressure devices are developed, on the basis of which parametric dependencies of the conversion and sensitivity functions are obtained. The main contribution to the variables of the transformation function is shown to be the change in pressure. This causes a change in the equivalent capacitance and negative differential resistance of the oscillating system of the sensor oscillators, which in turn changes the output frequency of the devices.

The parametric dependences obtained from the transformation functions of pressure devices show the possibility of calculating the main characteristics of sensors more easily and clearly show the influence of each parameter of the primary transducers and the parameters of the oscillators on the output frequency of devices in comparison with the calculations of transformation functions from equivalent device circuits based on Kirchhoff equations.

Pressure sensors with frequency output do not require analog-to-digital converters and amplifiers for further processing of information signals, which reduces the cost of information and measuring equipment, and it is possible to transmit information over a distance when the devices operate at ultrahigh frequencies.

The use of frequency as an output signal provides a number of advantages: high noise immunity, nonzero drift problem, and the ability to transmit a signal over long distances without losing accuracy. For microcontrollers, frequency is a convenient digital parameter that can be measured by counters and timers without additional conversions. Thus, the implementation of pressure–frequency sensors is optimal for modern measurement systems.

6. Patents

The following Ukrainian patents are associated with the work reported in this paper [34,35].

Author Contributions: Conceptualization, I.O., O.O. and S.B.; methodology, I.O., O.O. and A.S.; software, S.B. and M.B.; validation, I.O., O.O. and A.S.; formal analysis, I.O. and O.O.; investigation, I.O.; resources, O.O.; data curation, S.B. and M.B.; writing—original draft preparation, I.O., O.O. and S.B.; writing—review and editing, I.O., O.O. and A.S.; visualization, S.B.; supervision, O.O.; project administration, M.B.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors are grateful to all colleagues and institutions that contributed to the research and made it possible to publish its results.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NDR	Negative Differential Resistance
RIS	Radio-measuring Information System
IoT	Internet of Things
GSM	Global System for Mobile Communications

References

1. Platonov, V.N.; Danko, G.V. Theory of adaptation and reserves for improving the training system for athletes. *Nauka Olimp. Sport.* **2007**, *1*, 3–16.
2. D'iachenko, A.I. Modern concept of improving special endurance of high-class athletes in rowing. *Nauka Olimp. Sport.* **2007**, *1*, 54–61.
3. Platonov, V.N. *Sistema Podgotovki Sportsmenov v Olimpiiskom Sporte [System of Training Athletes in Olympic Sports]*; Olympic Literature: Kyiv, Ukraine, 2004.
4. Mishchenko, V.; D'iachenko, A.; Tomiak, T. Individual characteristics of anaerobic capacity as a component of special endurance of athletes. *Nauka Olimp. Sport.* **2003**, *1*, 57–62.
5. Shantarovich, V.V.; Kallaur, E.G.; Medvedev, P.Y. *Theoretical and Practical Aspects of Improving the Strength Training of Kayak and Canoe Rowers*; Mozyr State Pedagogical University Named After I.P. Shamyakin: Mozyr, Belarus, 2018; pp. 1–111.
6. Castañeda-Babarro, A.; León-Guereño, P.; Viribay, A.; Gutiérrez-Santamaría, B.; López, I.; Mielgo-Ayuso, J. The influence of anthropometric variables on the performance of elite traditional rowers. *Sports* **2024**, *12*, 185. [CrossRef]
7. Gavala-González, J.; Porras-García, M.E.; Fernández-García, J.C.; Real-Pérez, M. Effects of specific training using a rowing ergometer on sport performance in adolescents. *Appl. Sci.* **2024**, *14*, 3180. [CrossRef]
8. Ding, Z.; Li, C.-P.; Lin, H.-H.; Hung, S.-T.; Tseng, C.-H.; Hsu, C.-H. Exploring the flow experience and re-experience intention of students participating in water sports from the perspective of regional tourism and leisure environment suitability. *Sustainability* **2023**, *15*, 14614. [CrossRef]
9. Chen, C.-C.; Lin, C.-S.; Chen, Y.-T.; Chen, W.-H.; Chen, C.-H.; Chen, I.-C. Intelligent performance evaluation in rowing sport using a graph-matching network. *J. Imaging* **2023**, *9*, 181. [CrossRef] [PubMed]
10. Mpimis, T.; Gikas, V.; Gourgoulis, V. A rigorous and integrated on-water monitoring system for performance and technique improvement in rowing. *Sensors* **2023**, *23*, 6150. [CrossRef]
11. Agius, T.P.; Cerasola, D.; Gauci, M.; Sciriha, A.; Sillato, D.; Formosa, C.; Gatt, A.; Xerri de Caro, J.; Needham, R.; Chockalingam, N.; et al. The Kinematics of Fixed-Seat Rowing: A Structured Synthesis. *Bioengineering* **2023**, *10*, 774. [CrossRef]
12. Puce, L.; Biz, C.; Trompetto, C.; Marinelli, L.; Currà, A.; Cavaggioni, L.; Formica, M.; Vecchi, V.; Cerchiaro, M.C.; Trabelsi, K.; et al. A Scoping Review with Bibliometric Analysis of Para-Rowing: State of the Art and Future Directions. *Healthcare* **2023**, *11*, 849. [CrossRef]
13. Penichet-Tomás, A.; Jimenez-Olmedo, J.M.; Pueo, B.; Olaya-Cuartero, J. Physiological and Mechanical Responses to a Graded Exercise Test in Traditional Rowing. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3664. [CrossRef]

14. Guo, Y.; Li, B.; Zhang, Q.; He, X.T.; Sun, J.Y. A Further Theoretical Study of Capacitive Pressure Sensors Based on Thin Film Elastic Deflection and Parallel Plate Capacitor: Refined Closed-Form Solution and Numerical Calibration. *Sensors* **2022**, *22*, 2848. [CrossRef]
15. Tai, G.; Wei, D.; Su, M.; Li, P.; Xie, L.; Yang, J. Force-Sensitive Interface Engineering in Flexible Pressure Sensors: A Review. *Sensors* **2022**, *22*, 2652. [CrossRef] [PubMed]
16. Czakó, Z.; Surdea-Blaga, T.; Sebestyén, G.; Hangan, A.; Dumitraşcu, D.L.; David, L.; Chiarioni, G.; Savarino, E.; Popa, S.L. Integrated Relaxation Pressure Classification and Probe Positioning Failure Detection in High-Resolution Esophageal Manometry Using Machine Learning. *Sensors* **2022**, *22*, 253. [CrossRef] [PubMed]
17. Jang, J.; Jun, Y.S.; Seo, H.; Kim, M.; Park, J.U. Motion Detection Using Tactile Sensors Based on Pressure-Sensitive Transistor Arrays. *Sensors* **2020**, *20*, 3624. [CrossRef]
18. Deng, F.; He, Y.; Li, B.; Zuo, L.; Wu, X.; Fu, Z. A CMOS Pressure Sensor Tag Chip for Passive Wireless Applications. *Sensors* **2015**, *15*, 6872–6884. [CrossRef]
19. Osadchuk, I.A.; Osadchuk, A.V.; Osadchuk, V.S.; Semenov, A.O. Nanoelectronic Pressure Transducer with a Frequency Output Based on a Resonance Tunnel Diode. In Proceedings of the 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, 25–29 February 2020; pp. 452–457. [CrossRef]
20. Osadchuk, O.V.; Osadchuk, V.S.; Osadchuk, I.A. Mathematical Model of a Frequency Pressure Transducer Based on a Resonant Tunneling Diode. *Phys. Chem. Solid State* **2022**, *23*, 277–284. [CrossRef]
21. Osadchuk, O.V.; Osadchuk, I.O.; Suleimenov, B.; Zyska, T.; Arman, A.; Tleshova, A.; Grądz, Ż. Frequency Pressure Transducer with a Sensitivity of Mem Capacitor on the Basis of Transistor Structure with Negative Resistance. *Proc. SPIE* **2017**, *10445*, 1044559. [CrossRef]
22. Kenyeres, M.; Kenyeres, J.; Hassankhani Dolatabadi, S. Distributed Consensus Gossip-Based Data Fusion for Suppressing Incorrect Sensor Readings in Wireless Sensor Networks. *J. Low Power Electron. Appl.* **2025**, *15*, 6. [CrossRef]
23. Bulowski, W.; Knura, R.; Socha, R.P.; Basiura, M.; Skibińska, K.; Wojnicki, M. Thin Film Semiconductor Metal Oxide Oxygen Sensors: Limitations, Challenges, and Future Progress. *Electronics* **2024**, *13*, 3409. [CrossRef]
24. Kleshnev, V. Analysis of the Results of the 2018 World Rowing Championships. BioRow. 2018. Available online: https://old.biorow.com/index.php?news_id=38&route=information/news/news (accessed on 5 October 2025).
25. Kleshnev, V. Paris 2024 Olympic Games—Analysis of Rowing Results. BioRow. 2024. Available online: https://old.biorow.com/index.php?news_id=118&route=information/news/news (accessed on 5 October 2025).
26. Kleshnev, V. Rowing Science: Case Study on Technique Efficiency. row2k.com. 2023. Available online: https://www.row2k.com/features/6563/rowing-science-case-study-on-technique-efficiency/ (accessed on 5 October 2025).
27. Rowing.su. Results of the St. Petersburg Indoor Rowing Championships (2020). 2020. Available online: https://www.rowing.su/result2020.html (accessed on 5 October 2025).
28. Kleshnev, V. BioRow. 2025. Available online: https://biorow.com/ (accessed on 5 October 2025).
29. Neamen, D.A. *Semiconductor Physics and Devices: Basic Principles*, 3rd ed.; McGraw–Hill: New York, NY, USA, 2003.
30. Sun, Y.; Thompson, S.; Nishida, T. *Strain Effect in Semiconductors: Theory and Device Applications*; Springer: New York, NY, USA, 2010. [CrossRef]
31. Dorave, J.; Sadiwala, R. Uplink and Downlink Variation in Drone Technology for Cloud, Edge, Fog and Smart Dust Integrated IoT Architecture: Demonstrated Over WSNs. *J. Phys. Conf. Ser.* **2021**, *2089*, 012023. [CrossRef]
32. Korowajczuk, L. *LTE, WiMAX and WLAN Network Design, Optimization and Performance Analysis*; John Wiley & Sons: Chichester, UK, 2011.
33. Di Benedetto, M.-G.; Serra i Lidón, P. Analysis of the NB-IoT Technology Towards Massive Machine Type Communication. Bachelor Thesis, Università degli Studi di Roma "La Sapienza", Rome, Italy, 2018.
34. Osadchuk, O.V.; Osadchuk, V.S.; Krylyk, L.V.; Osadchuk, Y.O. Microelectronic Device for Pressure Measurement with Frequency Output. Ukrainian Utility Model. UA 155932 U, 17 April 2024. Available online: https://patents.google.com/patent/UA155932U (accessed on 11 August 2025).
35. Osadchuk, V.S.; Osadchuk, O.V.; Krylyk, L.V.; Osadchuk, Y.O.; Chervak, O.P. Self-Oscillating Pressure Transducer. Ukrainian Patent UA 124907 C2, 8 December 2021. Available online: https://patents.google.com/patent/UA124907C2 (accessed on 11 August 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

A Distributed Wearable Computing Framework for Human Activity Classification

Jhonathan L. Rivas-Caicedo [1,*], Kevin Niño-Tejada [1], Laura Saldaña-Aristizabal [1] and Juan F. Patarroyo-Montenegro [2,*]

[1] Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA; kevin.nino@upr.edu (K.N.-T.); laura.saldana@upr.edu (L.S.-A.)
[2] Department of Computer Science and Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA
* Correspondence: jhonathanleonardo.rivas@upr.edu (J.L.R.-C.); juan.patarroyo@upr.edu (J.F.P.-M.)

Abstract

Human Activity Recognition (HAR) using wearable sensors plays a critical role in applications such as healthcare, sports monitoring, and rehabilitation. Traditional approaches typically rely on centralized models that aggregate and process data from multiple sensors simultaneously. However, such architecture often suffers from high latency, increased communication overhead, limited scalability, and reduced robustness, particularly in dynamic environments where wearable systems operate under resource constraints. This paper proposes a distributed neural network framework for HAR, where each wearable sensor independently processes its data using a lightweight neural model and transmits high-level features or predictions to a central neural network for final classification. This strategy alleviates the computational load on the central node, reduces data transmission across the network, and enhances user privacy. We evaluated the proposed distributed framework using our publicly available multi-sensor HAR dataset and compared its performance against a centralized neural network trained on the same data. The results demonstrate that the distributed approach achieves comparable or superior classification accuracy while significantly lowering inference latency and energy consumption. These findings underscore the promise of distributed intelligence in wearable systems for real-time and energy-efficient human activity monitoring.

Keywords: decentralized activity classification; deep learning; distributed neural networks; edge computing; human activity recognition (HAR); multi-sensor systems; wearable sensors

1. Introduction

Human Activity Recognition (HAR) plays a pivotal role in a wide range of applications, including healthcare, assisted living, sports performance analysis, and rehabilitation monitoring. Fundamentally, HAR entails classifying physical activities, such as walking, running, sitting, or more complex movements, based on sensor data collected from the human body. The growing availability and miniaturization of wearable sensors, including accelerometers, gyroscopes, magnetometers, and physiological monitors, have significantly improved the feasibility and granularity of HAR systems. These technological advancements have made it possible to perform continuous, real-time monitoring in naturalistic environments, thereby enabling personalized and proactive interventions [1].

Despite significant advancements in sensing technologies, achieving accurate and robust HAR under real-world conditions remains a considerable challenge. Traditional

HAR systems often employ a centralized architecture, where raw or pre-processed data from multiple wearable devices is transmitted to a central processor, typically a smartphone, cloud service, or edge hub, for classification. However, this centralized design introduces several technical limitations, particularly in dynamic, resource-constrained environments. These limitations include increased communication overhead, elevated energy consumption, data synchronization complexities, vulnerability to node disconnections, and concerns related to user data privacy [2]. Furthermore, centralized inference requires the central node to possess substantial processing power to handle multi-sensor data streams, which may be impractical for low-power, real-time applications such as continuous patient monitoring or remote fitness tracking.

1.1. Wearable Systems and Their Challenges

The adoption of wearable technologies in healthcare and activity monitoring has skyrocketed, particularly with the emergence of smartwatches, fitness trackers, and wearable medical devices. However, these systems are often limited by their hardware capabilities, especially in terms of computation, storage, and power [3]. This makes it impractical to run complex machine learning models locally on each device, pushing most processing tasks to the central node. Furthermore, wearable systems often operate under variable conditions: sensors may lose contact, network availability may fluctuate, and the physical placement of the device may introduce variability in signal quality. These challenges make robustness and fault tolerance essential design goals in modern HAR systems.

From a system architecture perspective, current wearable HAR solutions primarily focus on data capture and transmission, with minimal onboard processing [4,5]. The captured data is typically streamed to a mobile phone or cloud service, where inference is conducted using centralized neural network models. These models, while powerful, depend heavily on uninterrupted data flow and fail gracefully only in rare circumstances. A disconnection or failure in any part of the sensor network can significantly degrade performance or halt monitoring altogether [6].

1.2. Neural Networks in HAR and Edge Constraints

Recent advances in deep learning have dramatically improved HAR accuracy. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models have been applied successfully to multi-modal sensor data [4,5]. These models are capable of learning complex temporal and spatial features directly from raw signals, reducing the need for handcrafted features. However, their deployment remains challenging in wearable contexts due to computational and memory limitations.

Most deep learning-based HAR frameworks are still implemented in centralized architectures where data from all sensors is required simultaneously [1]. Edge deployment, which runs models directly on wearable devices, offers potential benefits in terms of latency reduction and data privacy, but is constrained by the limited processing capabilities of individual devices. Moreover, scaling such systems to accommodate multiple sensors or users exacerbates these limitations. Edge–cloud hybrid solutions have been proposed, but they still rely heavily on constant connectivity and central control, which may not be sustainable in all use cases.

1.3. Problem Statement: The Limits of Centralized and Distributed HAR

As shown in Figure 1a, a centralized Machine Learning (ML) model computes all measurements and extracted features together to obtain a series of outputs, which may be classifications or regressions. In a distributed sequential ML approach (Figure 1b), the model is partitioned by layers such that all measurements and features are received by the first core, which contains the first layer. The subsequent cores contain one layer at

a time, operating in a pipeline to reduce computation time. However, a single core will receive all measurements, which requires a direct connection to all sensors in the network. Finally, suppose each sensor contains a portion of the original network, unaware of the other signals. In that case, it will lose observability because some measurements will not be accounted for, resulting in an inaccurate inference (Figure 1c).

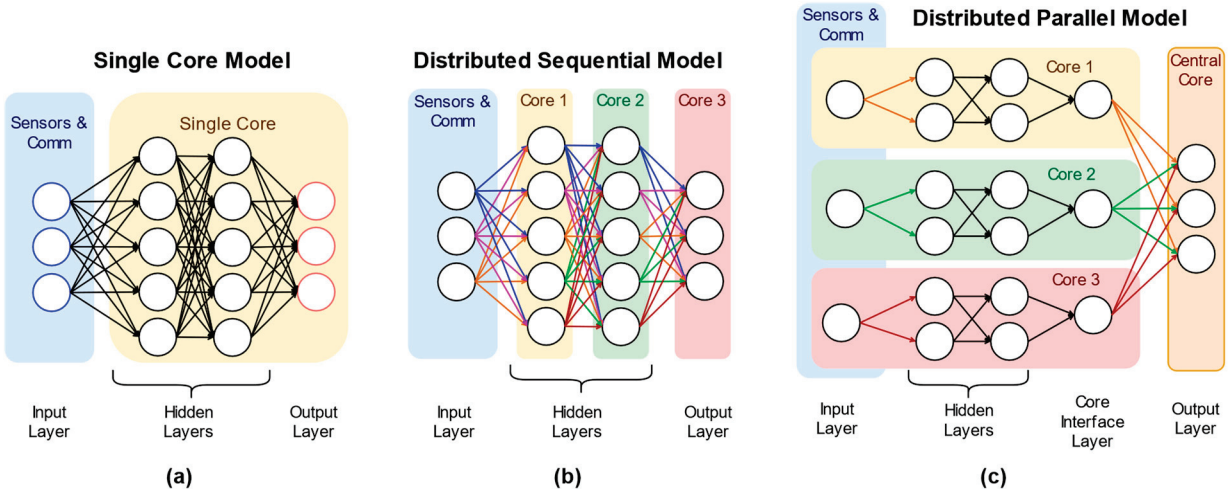


Figure 1. ML model distribution using sequential and parallel approaches: (**a**) example of a regular ML model running in a single mathematical core; (**b**) distributed model using a sequential approach. The arrows with the same color between layers indicate that they represent the same signal communicated between layers, and they are separated for illustrative purposes; (**c**) distributed parallel model. Each sensor signal is computed by its core, which is isolated from the cores of other sensors.

The fundamental problem with centralized HAR models lies in their structural dependency on full-sensor data integration at a single location. This dependency introduces significant bottlenecks in terms of latency, scalability, and fault tolerance. In scenarios where multiple sensors are deployed across the body, maintaining synchronization and uninterrupted data streams becomes increasingly difficult. Centralized systems are also more vulnerable to performance degradation due to packet loss, sensor dropout, or delayed data transmission.

In addition to these technical limitations, centralized systems pose privacy risks. Raw sensor data often contains sensitive information about users' behaviors and routines. Transmitting this data across networks increases the risk of potential security breaches. Similarly, energy consumption becomes a critical concern when high volumes of data must be wirelessly transmitted, particularly in battery-constrained devices like smartwatches or wearable patches [7].

1.4. Proposal: A Distributed Framework for HAR

To address these challenges, we propose a *distributed neural network architecture* for human activity classification using wearable sensors. In this architecture, each wearable device processes its local data stream using a lightweight neural model trained to produce high-level feature representations or preliminary predictions. These outputs, rather than the raw data, are transmitted to a central classifier, which integrates information from all devices to perform final activity recognition. This distributed framework offers several compelling advantages:

- **Reduced Communication Overhead:** only concise representations are transmitted, minimizing bandwidth use and energy consumption.

- **Improved Fault Tolerance:** the system remains functional even if one or more devices disconnect or fail temporarily.
- **Enhanced Scalability:** new sensors can be added with minimal impact on the central node's computational load.
- **Greater Privacy:** since raw data is not transmitted, the system inherently protects sensitive user information.

This architectural shift allows us to reimagine HAR systems not as monolithic, centrally controlled pipelines but as collaborative, adaptive, and modular networks of intelligent agents. Each sensor node contributes to the overall classification task without requiring complete access to all data, thereby enabling deployment in real-time, low-power, and privacy-sensitive scenarios.

1.5. Related Work

Recent research in HAR has increasingly shifted from traditional centralized processing toward edge-based and distributed approaches that better accommodate real-world deployment constraints. Centralized models such as those employing deep convolutional and recurrent networks have demonstrated strong performance across various benchmark datasets [8,9]. These models often aggregate raw multi-sensor data to a cloud or local server for classification, achieving high accuracy but suffering from high communication costs, latency, and vulnerability to sensor failures or dropouts [10]. To address these issues, edge intelligence techniques have been developed, allowing devices to perform localized inference. For example, Haque et al. introduced LightHART, a transformer-based HAR model optimized for execution on mobile processors without sacrificing accuracy [11]. Similarly, the work of Agarwal and Alam proposed an ultra-lightweight deep learning model tailored to wearable hardware constraints [7].

Federated learning (FL) has emerged as a promising solution to reduce data transmission and enhance privacy. Frameworks such as FedHAR [12], FedHealth [13], and FedOpenHAR [14] enable multiple wearable devices to collaboratively train models while retaining data locally. These approaches mitigate privacy risks but still depend on stable connectivity and frequent model synchronization. Spiking neural networks (SNNs), like those explored in [15], offer biologically inspired, event-driven computation with low energy consumption, making them suitable for on-body deployment, though training remains challenging. Meanwhile, hybrid distributed architectures such as adARC [16] and Sannara et al.'s distributed sensor fusion scheme [17] combine local lightweight inference with central aggregation of intermediate representations or decisions, offering robustness to disconnections and improved scalability.

In addition to these frameworks, recent studies have explored enhancements to traditional FL through communication-efficient and heterogeneous model approaches. Sozinov et al. [18] evaluated FL for HAR using both softmax regression and deep neural networks, showing that although FL achieves slightly lower accuracy than centralized models, it significantly reduces privacy risks and data transmission costs. Gad et al. [19] proposed FedAKD, an FL strategy based on augmented knowledge distillation, which supports heterogeneous client models and improves communication efficiency by exchanging soft labels rather than gradients. Their results show superior performance under non-IID conditions, addressing one of the key limitations of traditional FL algorithms.

Beyond FL, split learning (SL) and federated split learning (FSL) have been introduced to reduce the computational load on wearable devices further while preserving privacy. Ndeko et al. [20] presented an FSL framework with differential privacy that partitions the model between the client and the server. Their method achieved improved accuracy,

reduced latency, and better privacy preservation compared to conventional FL, particularly in edge computing scenarios.

Building on these advancements in FL and SL, recent studies have emphasized the importance of energy awareness and heterogeneous device capabilities in real-world HAR deployments. Nguyen et al. proposed an energy-aware FL framework that dynamically selects participating clients based on energy status and data quality, achieving a better trade-off between accuracy, fairness, and device longevity [21]. Similarly, Thakur et al. introduced FedMeta, a meta-learning-based approach that enhances personalization and adaptation to non-iid data distributions commonly encountered in wearable HAR [22]. Saylam and Incel [23] provided a comprehensive review of FL techniques for edge devices, highlighting the challenges posed by device heterogeneity, communication bottlenecks, and the need for lightweight aggregation methods tailored to wearable systems.

In parallel, research has also focused on distributed training and inference paradigms that relax the server-centric assumptions of traditional FL. Khan et al. developed a multi-frequency federated learning (MFFL) algorithm for HAR using head-worn sensors, enabling asynchronous updates from devices operating at different sampling and computation rates while maintaining both accuracy and energy efficiency [24]. Furthermore, decentralized training approaches, such as ring-based and peer-to-peer topologies, have been explored to eliminate single points of failure, allowing HAR systems to maintain progress even under intermittent connectivity. These emerging strategies emphasize adaptability, privacy preservation, and scalability, making them particularly relevant for next-generation distributed HAR frameworks.

Complementary to learning strategies, energy efficiency remains a central concern for wearable HAR systems. Ding et al. [25] proposed dynamic inference schemes and sensor sampling strategies to reduce power consumption without sacrificing performance. Similarly, Rezaie et al. [26] introduced an adaptive activity-aware algorithm that adjusts sensing and computation rates based on user context, achieving substantial energy savings. These techniques focus on prolonging battery life, though they often assume single-device processing without inter-sensor coordination.

Despite these advances, many existing solutions either compromise accuracy for efficiency or require architectural assumptions (e.g., synchronized updates or homogeneous sensors) that are difficult to maintain in real-world deployments. In contrast, the framework proposed in this work adopts a modular, distributed architecture in which each sensor node runs an independent neural model to generate compact, informative outputs for fusion. This reduces communication overhead, enhances fault tolerance, and supports asynchronous operation, making it well-suited for scalable, privacy-conscious, and resource-constrained HAR deployments.

1.6. Contributions and Structure of the Paper

In this work, we attempt the following:

- Propose a **novel distributed neural network framework** for multi-sensor HAR using wearable devices, as shown in Figure 2.
- Implement and optimize lightweight local models for individual sensor nodes and a central fusion model for final classification.
- Validate our approach on a publicly available multi-sensor HAR dataset that we created and compare its performance against a centralized model.
- Demonstrate that our distributed model achieves **comparable or even better accuracy** while significantly reducing communication overhead and improving robustness to sensor disconnections.

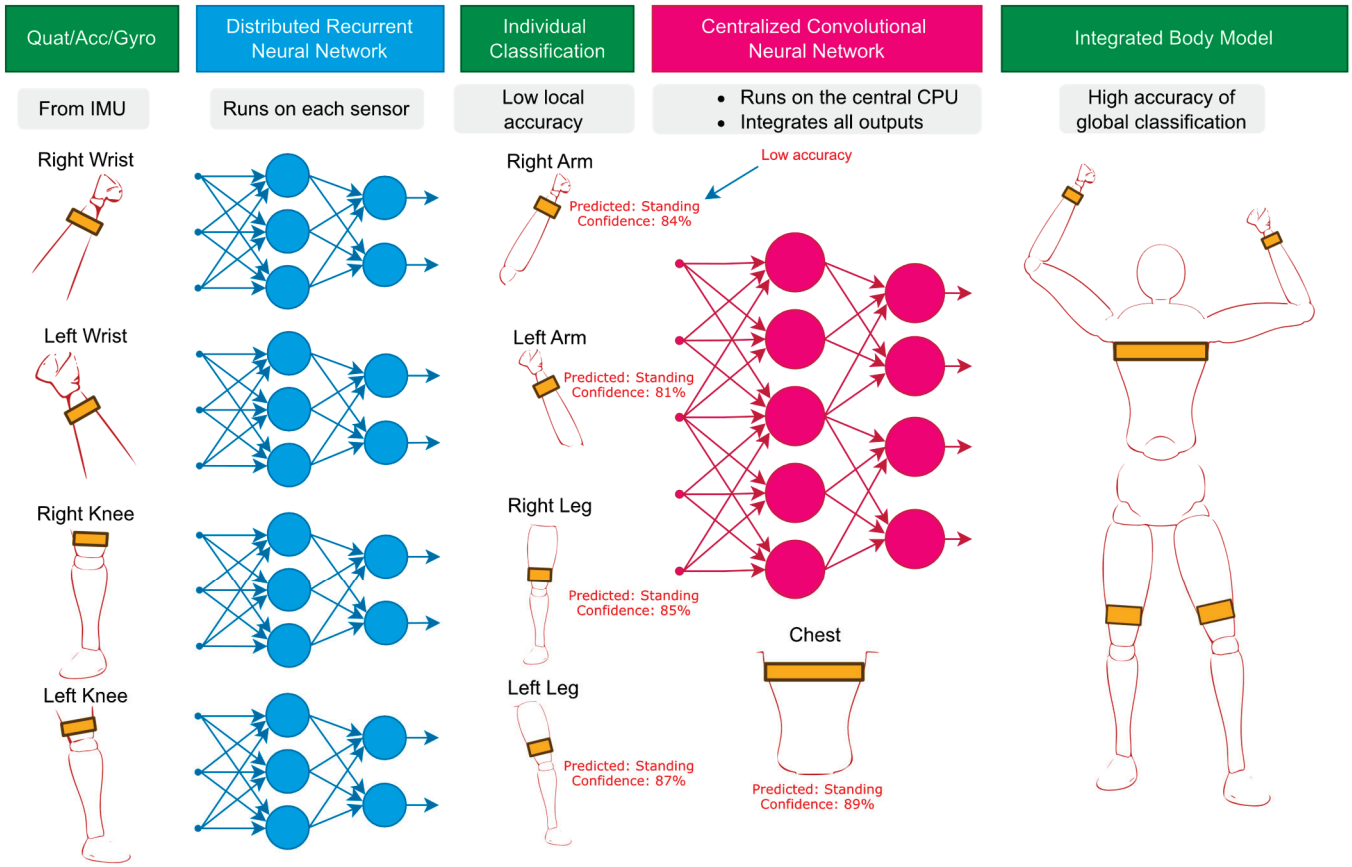


Figure 2. General scheme of the proposed distributed computing wearable sensor network to be used as an application case. As shown, the Individual Classification stage is characterized by low accuracy, as each limb processes limited data. However, this accuracy is increased in the central core when all the outputs from each limb are integrated into a single classification verdict.

The remainder of the paper is organized as follows: Section 2 presents the dataset, experimental setup, and implementation details. Section 3 outlines the proposed distributed framework, detailing the design of both local and central neural network components. Section 4 reports the results, including accuracy comparisons, robust analysis, and energy consumption metrics. Finally, Section 5 discusses implications, limitations, and directions for future research.

2. Materials and Methods

2.1. Participants

A total of 67 participants (40 males and 27 females), aged between 20 and 60 years, were recruited for this study from the local university population and the surrounding community. All individuals voluntarily agreed to participate and provided informed consent before data collection. Participants self-reported being in good general health and physical condition at the time of the study.

The primary inclusion criterion was the absence of any known cardiovascular conditions or medical history that could compromise the participant's ability to perform the physical activities safely as outlined in the experimental protocol. Individuals with mobility impairments, recent injuries, or conditions affecting balance or coordination were excluded to ensure consistency and safety during data acquisition.

This study was approved by the University of Puerto Rico's Institutional Review Board (IRB) in compliance with the ethical standards established by the Collaborative Institutional Training Initiative (CITI Program) for research involving human subjects. All

procedures were reviewed and authorized before data collection. Before participating, all individuals received a detailed explanation of the study's objectives and methods and provided written informed consent.

2.2. Dataset

We collected a dataset using wearable sensors strategically placed on the human body. The dataset comprises multiple sessions of physical activity performed by 67 participants, with each session consisting of several repetitions of predefined activities. These activities were recorded using three types of signals: quaternion, linear acceleration, and angular velocity. These signals provide a rich representation of body movement and orientation, which is essential for robust classification.

Each session was recorded at a fixed sampling rate of 50 Hz, meaning that one sample corresponds to one timestamp of recorded data. On average, each session lasted approximately 37 min, resulting in thousands of individual data samples per session. The "Duration per session" column in Table 1 indicates the length of each activity per participant.

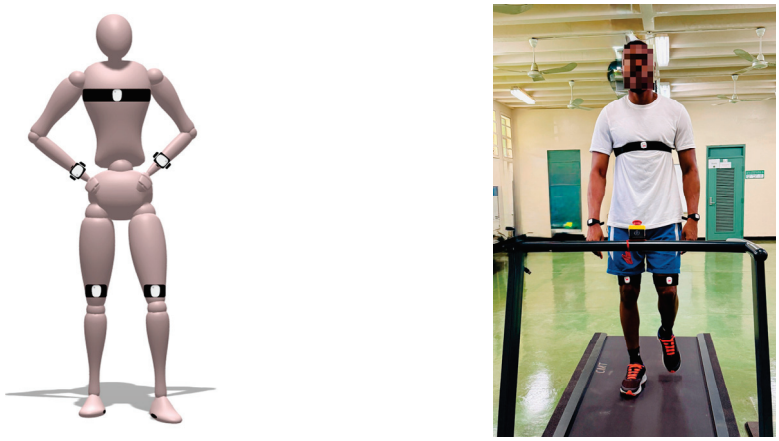
Table 1. List of classified activities with their corresponding label, description, and duration per session used in the training and evaluation of the model.

Activities	Label	Description	Duration per Session (s)
Sitting	0	Participant remains seated and mostly still	600
Folding Clothes	1	Participant simulates folding garments by hand	300
Sweeping	2	Participant mimics a sweeping motion with arms	300
Walking	3	Participant walks at a regular pace	300
Moving Boxes	4	Participant lifts and moves small to medium boxes	300
Riding a Bike	5	Participant performs a cycling motion while seated	300

Table 1 summarizes the list of activities selected for classification, their corresponding labels, and the average duration per subject for each activity. This provides a detailed overview of how the dataset was constructed for training and evaluating the classification models.

Many studies in the literature rely primarily on acceleration and gyroscope signals for activity recognition, as these modalities capture translational and rotational motion effectively [27]. However, in this work, we also incorporated quaternions because they represent the sensor's orientation in 3D space without suffering from the singularities and ambiguities [28]. By incorporating quaternions, the system achieves a more stable and comprehensive understanding of body posture and joint movement, which is particularly valuable when classifying activities that involve complex or continuous changes in orientation and speed.

To capture the full-body dynamics during the selected activities, wearable sensors were strategically placed on key locations of the body. In this study, five sensors were used and positioned as follows: on the chest to provide a reference for torso orientation and posture; on the right and left wrists to monitor upper limb movements; and on the left and right knees to capture lower limb dynamics. This sensor configuration enables the model to learn both global and local motion patterns across different body segments, improving its ability to distinguish between similar activities. A visual representation of the sensor placement is shown in Figure 3. The dataset collected is available in the resource [29].



(**a**) (**b**)

Figure 3. Body sensor placement configuration: (**a**) wearable sensing prototype; (**b**) human subject walking while instrumented with the proposed sensor setup.

Anatomical considerations guided the placement strategy and aligned with established practices in the human activity recognition literature. Sensors located on the chest and limbs offer a suitable trade-off between capturing relevant motion features and ensuring wearability. While this configuration proved effective for the activities included in this study, it is essential to note that no experiments were conducted to assess alternative sensor placements. Classification performance may vary with different sensor configurations, particularly when motion sensitivity or signal discriminability is reduced due to location-specific factors.

2.3. Experimental Design

To evaluate the proposed distributed framework for human activity recognition, a controlled experiment was conducted using five MetaMotionRL wearable sensors (version r0.5) from Mbientlab Inc. (San Jose, CA, USA). Each sensor was configured to stream specific inertial data types at a frequency of 50 Hz, utilizing Bluetooth Low Energy (BLE) for wireless communication with a MetaBase app running on an iOS device.

The chest sensor captured quaternions, 3-axis linear acceleration, and 3-axis angular velocity. The sensors placed on the right hand and left knee streamed only acceleration and gyroscope data, while the sensors on the left hand and right knee transmitted quaternion data exclusively. All data streams were timestamped at the source and subsequently synchronized offline using a standard reference clock. Synchronization was achieved through linear extrapolation based on timestamps to ensure temporal alignment across all sensor inputs.

Each of the 67 participants underwent a structured experimental session with a total duration of approximately 37 min. The session consisted of six predefined activities listed in Table 1: folding clothes, sweeping, walking, moving boxes, riding a bike, and sitting. The first five activities were performed continuously for 5 min each, while the remaining time was allocated to the sitting activity.

All sessions were conducted under supervised conditions in a controlled indoor environment, following a strict and consistent timeline across participants to ensure standardization and reproducibility of the dataset. Each session began immediately after sensor initialization and synchronization, and data acquisition was conducted in real-time over approximately 37 min. per participant. However, model training and classification were performed offline after data collection. The sensor signals remained stable throughout each session, with no noticeable drift or degradation in data quality. This ensured consistent labeling and reliable input for offline model development and evaluation. Future work

will investigate longer-term deployments to assess the impact of continuous use, including possible effects of sensor drift or synchronization loss on classification performance.

Transitions between activities were excluded from the final dataset to avoid introducing ambiguity in the activity labels. The sitting activity occurred at the beginning of the session and after every two subsequent activities, resulting in a total of four sitting periods. The short intervals during which the participant transitioned from one activity to another were automatically identified by software using the predefined time intervals for each task and systematically discarded, ensuring that each data segment corresponded strictly to a single, well-defined activity. While this design decision promotes label consistency and simplifies the learning process, it does not account for the transitional or overlapping motions that often occur in real-world deployments. Therefore, in future work, we aim to extend the framework toward real-time implementation, where detecting and classifying transitions will be critical for enabling continuous and adaptive activity recognition in dynamic environments.

2.4. Data Processing

To ensure temporal consistency, all recordings were synchronized post hoc using the timestamp information provided by the sensors. A linear extrapolation method was applied to align the signals across devices with a uniform sampling grid. Each recording session consisted of multiple predefined activities, performed in a controlled environment. Activity labels were assigned based on the sequence and timing of each task, which were recorded during acquisition.

To ensure the validity of the labeling process, all data acquisition sessions were supervised in real time by a member of the research team, who followed a predefined script indicating the sequence and duration of each activity. The timing of each task was logged concurrently with the sensor data to provide accurate alignment between activity execution and label assignment. After acquisition, the recorded sensor signals were visually inspected using custom plotting tools to verify that signal transitions (e.g., changes in acceleration or orientation) were consistent with the expected activity changes. Sessions where inconsistencies or ambiguities in timing or execution were detected were either repeated or excluded from the dataset. Although inter-observer agreement was not employed, this manual supervision and verification process served to ensure the reliability of the assigned labels and mitigate the effects of possible temporal drift or misalignment.

During preprocessing, any data entries with invalid or missing labels were excluded. In addition, samples containing missing values in any of the sensor features were removed to ensure the integrity of the dataset. Feature values were then normalized using z-score normalization (zero mean and unit variance), computed based on the training set statistics. This normalization was applied using the utilities from the scikit-learn library and saved for consistent scaling of future data.

The resulting normalized time series were segmented into overlapping windows of 50 samples (equivalent to 1 s) using a sliding window approach with a stride of 25 samples (50% overlap). Each window was assigned the most representative label from the segment to generate the final dataset used for model training and evaluation.

3. Proposed Distributed Framework

To address the limitations of centralized architectures in wearable human activity recognition systems, we propose a distributed neural network framework designed for deployment across multiple edge devices. Unlike conventional models that rely on aggregating all sensor data at a central processing unit, our approach distributes the inference workload by assigning individual neural networks to each sensor node. These local models

extract relevant motion features independently and transmit only compact, high-level representations to a central model, which is responsible for final decision-making.

This distributed architecture is designed to minimize communication overhead and energy consumption while enhancing scalability and robustness against common issues, such as sensor dropout or data transmission delays. In this initial stage, the framework was evaluated offline, simulating the behavior of distributed inference by independently processing each sensor's data through local models and aggregating their outputs in a central model. This approach allows for the validation of the distributed logic and its effectiveness in multi-sensor fusion before deploying the system on actual edge devices. The real-time deployment of this framework is outside the scope of this paper.

3.1. Centralized Model

As an initial approach to human activity recognition, a centralized model was developed and trained using the complete set of synchronized sensor data. This model served as a baseline to evaluate the classification performance when a single neural network processes all information. A centralized architecture receives as input a concatenated vector that integrates the signals from all five sensors, combining features such as quaternions, linear acceleration, and angular velocity. This design assumes full availability of the data streams during inference, representing an ideal scenario without communication delays or disconnections. The purpose of this centralized model is twofold: to establish a performance benchmark under optimal conditions and to provide a direct comparison point for the subsequent implementation of a distributed architecture.

Figure 4 illustrates the structure of the input data and the operation of a single 1D convolutional layer applied to the raw sensor window. Each input sample consists of a sliding window of 50 timesteps and 30 features, where the 30 features correspond to the concatenation of sensor measurements (acceleration, angular velocity, and the four orientation components) collected from five body-worn devices. The vertical axis represents the feature dimension, while the horizontal axis spans the temporal sequence. The figure also shows how 1D convolutional filters with a kernel size of 3 operate along the time axis, integrating information across all 30 features at each step. This setup enables the network to extract local spatial-temporal patterns by sliding the filters across the sequence and generating feature maps that encode short-term dependencies. Although the figure depicts the first convolutional layer, the subsequent convolutional layers process the output of the preceding ones, similarly building progressively deeper hierarchical representations of the input signal.

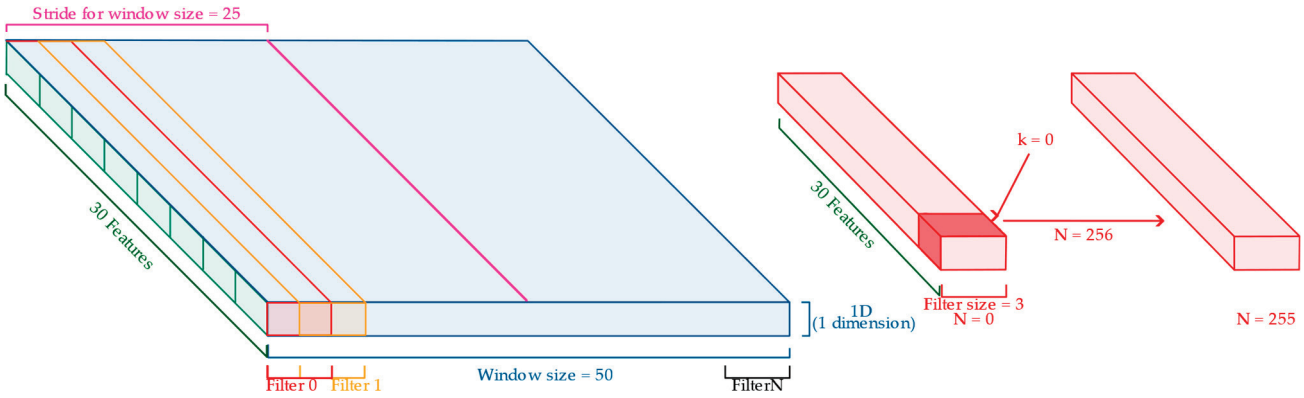


Figure 4. Illustration of the input tensor structure and the operation of a 1D convolutional layer in the centralized model.

Figure 5 presents the architecture of the centralized model. The first part of the network consists of three consecutive 1D convolutional layers with a kernel size of 3 and appropriate padding to preserve the temporal resolution. Each convolutional layer is followed by batch normalization and a ReLU activation function. A max pooling layer with a kernel size of 2 is applied afterward to reduce the sequence length and emphasize the most relevant features. This is followed by a two-layer LSTM (Long-Short Term Memory) with a hidden size of 256, which encodes the temporal evolution of the signal across the reduced time dimension. The LSTM output is then passed through a series of three fully connected layers. The first two linear layers include ReLU activation and dropout for regularization, while the final layer applies a SoftMax function to generate the activity class prediction. This architecture was designed to strike a balance between expressiveness and generalization in a centralized setting, serving as a strong baseline for comparison against the distributed framework.

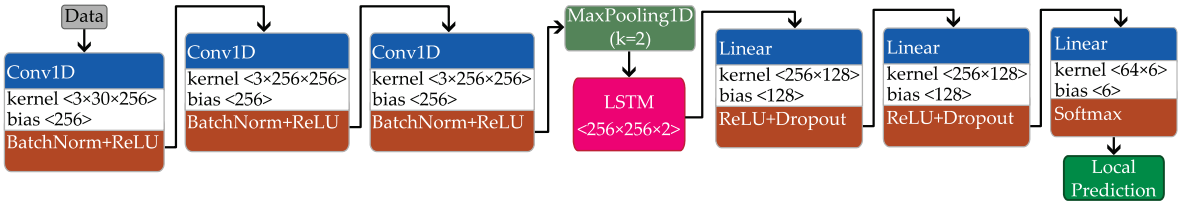


Figure 5. Implemented architecture for the Centralized Model.

To ensure a robust architecture for the centralized model, several configurations of the CNN–LSTM (Convolutional Neural Network and Long-Short Term Memory) structure were explored, varying the number of convolutional layers and LSTM units. These configurations were evaluated on a held-out test set to assess generalization. A detailed comparison of these configurations and their corresponding performance is presented in Section 4. The best-performing configuration identified was selected for a more in-depth evaluation. To better understand its behavior, a confusion matrix was computed using validation data, along with precision, recall, and F1-score for each activity class. These metrics provide a detailed view of the model's performance on activities observed during training, helping to identify both well-classified categories and classes that may be confused due to similar motion patterns.

3.2. Distributed Model

To address the challenges of scalability, latency, and robustness in real-world deployments, a distributed neural network framework was implemented. Unlike the centralized model, which processes all sensor data jointly, the distributed approach decouples computation across multiple local models, each associated with a specific sensor and a central model that integrates their outputs. This design enables independent inference at the sensor level while allowing fusion at a higher level, reducing the computational burden on a single node and enhancing fault tolerance in case of sensor disconnections or transmission delays, an aspect that was addressed in a previous study focused on mitigating sensor disconnections [6].

3.2.1. Local Model

Each local model is designed to operate independently on the data collected from a single wearable sensor. These models process sensor-specific signals (such as quaternions, accelerometer, and gyroscope data) and output a probability distribution over the predefined set of activity classes. Figure 6 illustrates the structure of the input data and the operation of a single 1D convolutional layer applied to the raw sensor window.

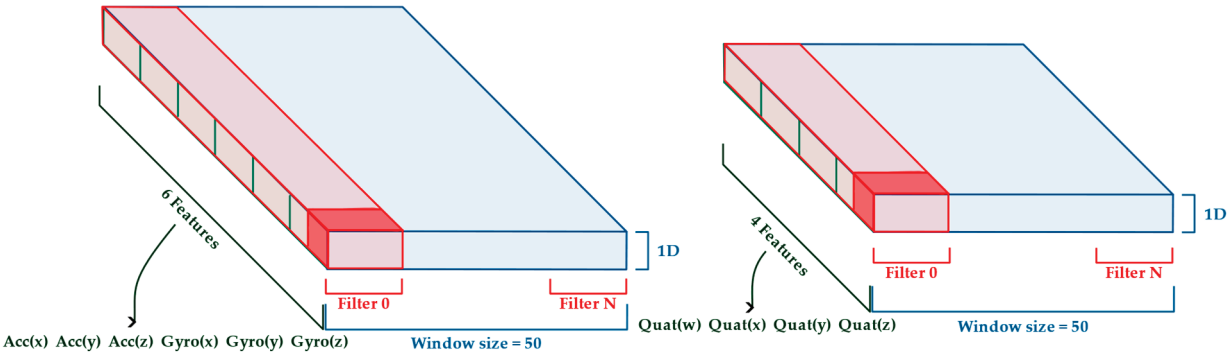


Figure 6. Input tensor structure and the operation of a 1D convolutional layer in the local model.

Each model receives input windows of 50 timesteps and either six features (for sensors with accelerometer and gyroscope data) or four features (for sensors with quaternion data), resulting in tensors of shape 50×6 or 50×4. The first stage consists of two 1D convolutional layers with a kernel size of 3 and increasing output channels (from 16 to 32), both of which are followed by ReLU activations. A max pooling layer with a kernel size of 1 is applied to preserve the temporal resolution. This is followed by a single-layer LSTM with a hidden size of 32, which captures temporal dynamics in the sensor data. The final representation is passed through two fully connected layers: the first with ReLU activation, and the second with a SoftMax output to produce class probabilities.

The architecture follows a lightweight CNN–LSTM structure optimized for execution on embedded devices, as illustrated in Figure 7.

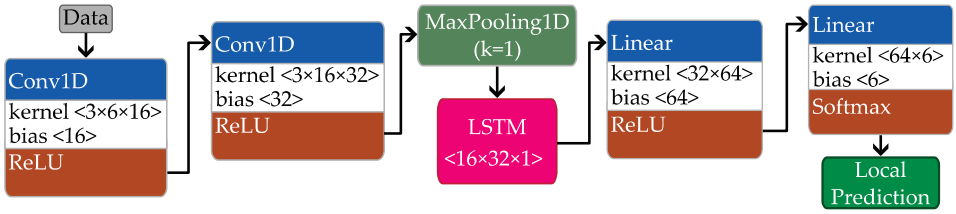


Figure 7. Implemented architecture for the local core on each sensor for the Distributed Model.

The lightweight design of each local model was intentionally chosen to facilitate real-time deployment on embedded platforms with limited computational resources. By minimizing the number of parameters and reducing architectural complexity, the models achieve low latency and energy efficiency while maintaining competitive classification performance. This design choice enables distributed inference directly at the sensor level, eliminating the need for constant communication with a centralized processor.

3.2.2. Central Model of the Distributed Network

The central model receives as input a single concatenated output vector formed by stacking the class probability distributions generated by each local model. Each local model, corresponding to a different body-worn sensor placed explicitly on the chest, left and right wrists, and left and right knees, outputs a probability vector of length **C**, where **C** is the number of activity classes. These vectors are concatenated vertically to form a unified input vector of dimension $\mathbf{N} \cdot \mathbf{C}$, where **N** is the number of sensors. Figure 8 illustrates how this representation encapsulates the local decisions of all sensors and enables the central multi-layer perceptron (MLP) to integrate them into a global activity classification.

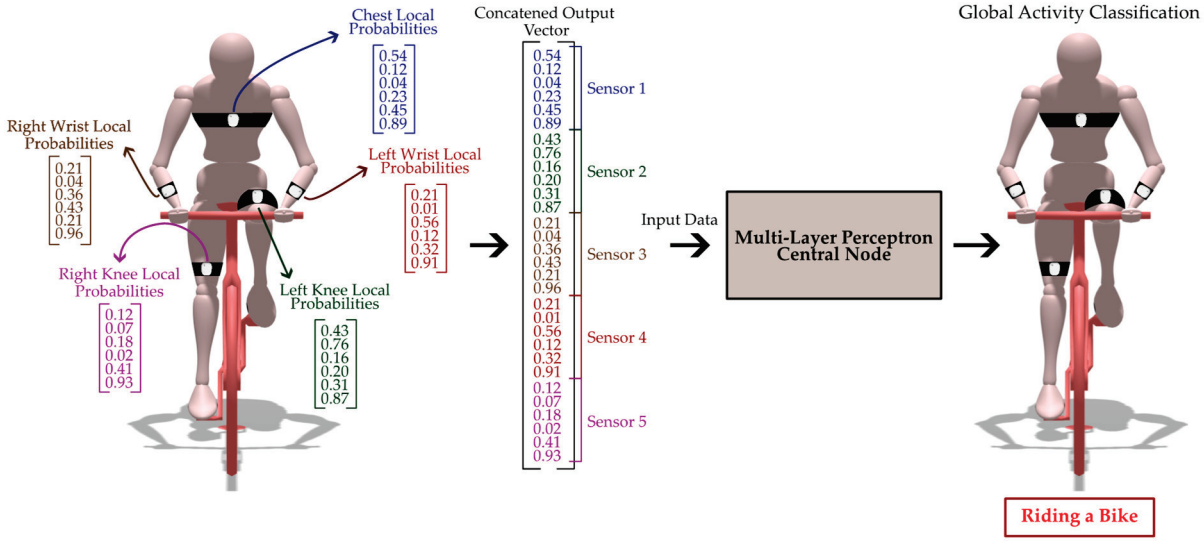


Figure 8. Overview of the centralized decision fusion process in the proposed distributed framework. Each local model processes sensor-specific data and outputs a probability vector over the set of activity classes.

The architecture consists of an MLP composed of two fully connected layers, as shown in Figure 9. The first layer maps the 30-dimensional input to a 64-dimensional hidden representation, followed by a ReLU activation and dropout to promote generalization. The second layer applies a SoftMax activation over the final 6-dimensional output, producing the global activity prediction. This architecture enables the integration of individual local inferences while maintaining a compact and efficient structure, making it suitable for real-time decision fusion.

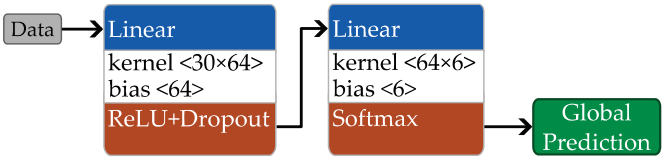


Figure 9. Implemented architecture for the central core on the Distributed Model.

This central model plays a key role in the distributed framework by acting as a decision-level fusion layer. While each local model is capable of performing standalone inference based on its respective sensor data, the central model enhances overall system performance by leveraging complementary information across sensors. This hierarchical design offers both robustness, allowing for partial predictions in the event of sensor failures, and scalability, as additional local models can be integrated seamlessly without requiring retraining of the entire system.

To assess the computational suitability of the proposed architecture for embedded and resource-constrained environments, we analyzed the complexity of both the local CNN–LSTM models and the central MLP. Table 2 presents the number of trainable parameters, floating point operations (FLOPs), and estimated memory usage for each model. These results demonstrate that the local models remain lightweight enough for on-device inference. At the same time, the central node incurs minimal computational overhead, making the entire framework suitable for deployment on wearable systems and low-power edge devices.

Table 2. Number of parameters, computational complexity (in FLOPs), and memory usage of the local CNN–LSTM models and the central MLP, highlighting the feasibility of real-time execution in embedded systems.

Model	Parameters		FLOPs	Memory Usage (MB)
	Layer	Trainable Parameters		
Local Node	Input Layer	0	229.83 KFLOPs	0.06
	Conv1D (1 × 3)	304		
	Activation (ReLU)	0		
	Conv1D (1 × 3)	784		
	Activation (ReLU)	0		
	MaxPooling (1 × 2)	0		
	LSTM (1 layer, hidden = 32)	6400		
	Dense (64)	2112		
	Activation (ReLU)	0		
	Predictions (Dense)	390		
	Total Trainable Parameters	9990		
Central Node	Input Layer	0	2.5 KFLOPs	0.01
	Dense (64)	1984		
	Activation (ReLU)	0		
	Dropout	0		
	Predictions (Dense)	390		
	Total Trainable Parameters	2374		

3.3. Training Configuration and Reproducibility

The local models, each based on a CNN–LSTM architecture, were trained independently for each sensor. Each model received as input only the data from its corresponding sensor and was trained as a standalone classifier. The central model, a lightweight MLP, was trained separately using the SoftMax outputs from the already trained local models as input features.

All models were trained using the categorical cross-entropy loss function, standard for multi-class classification problems. The Adam optimizer was selected due to its adaptive learning rate capabilities and efficient convergence, which have proven effective in human activity recognition (HAR) tasks [30]. A fixed learning rate of 0.001 was used, a value commonly adopted for stable convergence without excessive oscillations [31]. The batch size was set to 64, offering a good balance between generalization and training speed while remaining compatible with the memory constraints of embedded devices [32].

Training was limited to a maximum of 50 epochs. Early stopping was applied by monitoring validation loss, halting training if no improvement was observed for 10 consecutive epochs (patience = 10). While no additional regularization techniques such as dropout or L2 penalties were applied, the CNN–LSTM architecture incorporates max-pooling and ReLU activations, which naturally contribute to overfitting mitigation and feature sparsity [33].

The best-performing model on the validation set was saved and used for final evaluation. All training experiments were implemented in PyTorch v2.5.1 with CUDA 12.1 acceleration. Random seeds were fixed across all training and evaluation steps to ensure full reproducibility.

4. Results

This section presents a comprehensive evaluation of both the centralized and distributed models using a test set composed of unseen data. The results include accuracy,

confusion matrices, and class-wise performance metrics, enabling a comparative analysis of the two approaches.

4.1. Centralized Model Performance

To optimize the classification performance and assess the robustness of the centralized model, multiple architectural configurations were tested. Each configuration varied the number of convolutional and LSTM units, while maintaining the overall CNN–LSTM structure. The evaluation was conducted using a test set composed of data that was not used during training or validation, ensuring an unbiased assessment of generalization capability.

Figure 10 presents the accuracy obtained for each configuration. The models are grouped into three categories: C2L1, C3L2, and C4L3, where **C** denotes the number of convolutional layers and **L** the number of LSTM layers. Each configuration also varies in terms of convolutional filters (**F**) and the number of hidden units in the LSTM layers (**N**). Among all tested variants, the configuration with 256 convolutional filters and a two-layer LSTM with 256 hidden units (F = 256, N = 256, C = 3, L = 2) achieved the highest accuracy (93.68%), demonstrating the benefit of deeper representations when processing complex multi-sensor data.

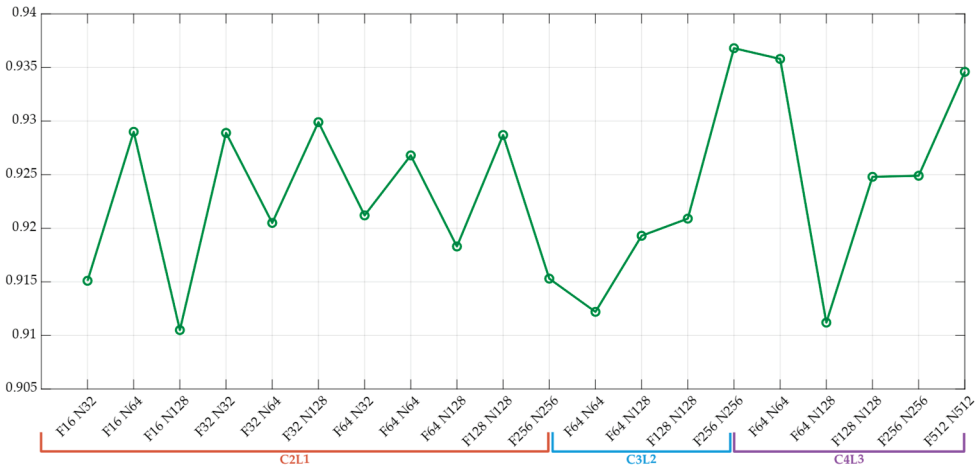


Figure 10. Accuracy obtained during training under various architectural configurations of the centralized model. Each configuration explores a different combination of the number of filters (F), fully connected neurons (N), convolutional layers (C), and LSTM layers (L), aiming to identify the most effective setup for maximizing classification performance.

These results establish a robust and well-performing centralized baseline, which serves as a strong reference point for evaluating the distributed framework. The effectiveness of this configuration not only validates the design choices of the centralized approach but also provides a meaningful benchmark for subsequent comparisons.

4.2. Distributed Model Performance

Building upon the centralized baseline, the distributed framework was evaluated under the same test conditions to enable a direct comparison. In this approach, each wearable sensor is associated with an independent local model responsible for processing its own data stream. These local models were trained separately using their respective sensor modalities and produce intermediate predictions in the form of SoftMax probability vectors.

To integrate the outputs from the local models, a lightweight central model based on an MLP was employed. This central node aggregates the class scores output by each sensor to generate the final activity classification.

Figure 11 presents the accuracy obtained for each configuration of the distributed framework. The models are grouped under the category C2L1, representing a lightweight and straightforward architecture tailored for embedded deployment. Among all tested variants, the configuration with 16 convolutional filters and a single-layer LSTM with 32 hidden units (F = 16, N = 32, C = 2, L = 1) achieved the highest accuracy (95.99%).

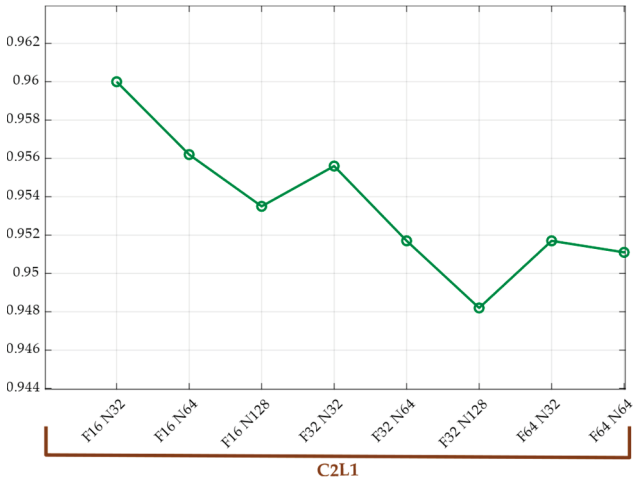


Figure 11. Accuracy achieved during training for various architectural configurations of the local models within the distributed framework. Each configuration represents a different combination of convolutional layers (C) and LSTM layers (L), aiming to identify the most effective structure for maximizing classification performance while maintaining a lightweight design.

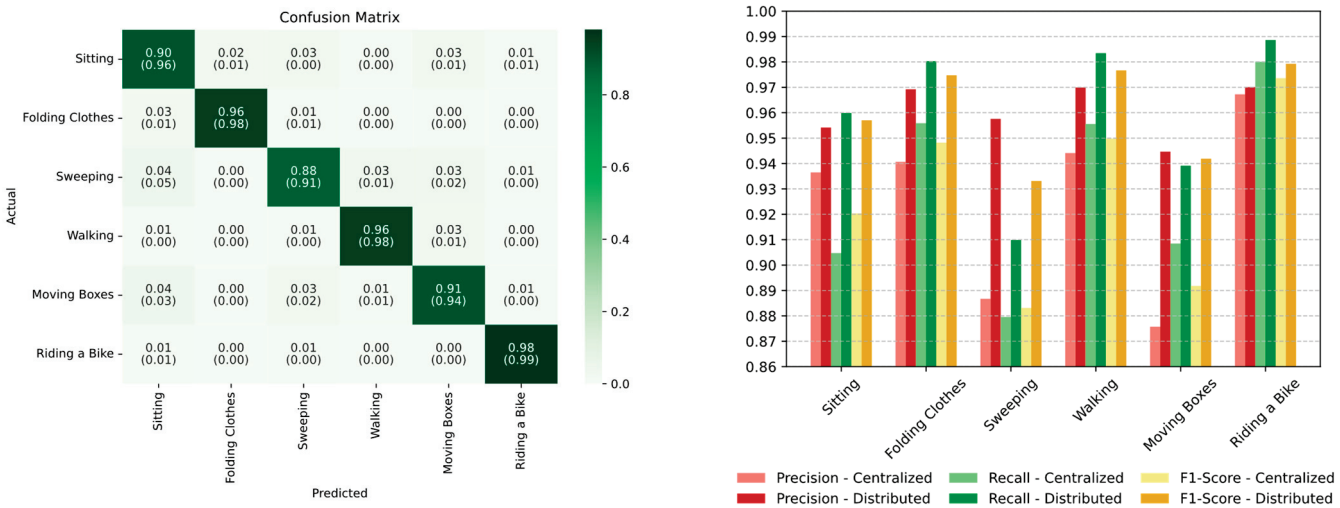
These results confirm that the distributed framework, despite relying on simpler and independently trained local models, can achieve competitive and better classification performance. The lightweight configuration with minimal computational complexity proved sufficient for capturing discriminative features when combined across sensors. This balance between efficiency and accuracy highlights the potential of distributed processing in wearable systems. A comparative analysis between the centralized and distributed approaches is presented in the following subsection.

4.3. Comparison

To enable a direct evaluation of both approaches, the centralized and distributed models were tested under identical conditions using the same dataset. Figure 12 provides a detailed comparison of the class-wise performance for both approaches. It includes confusion matrices and the per-class precision, recall, and F1-scores. It can be appreciated that the distributed model achieved higher precision than the centralized approach, indicating a reduced rate of false positives in its predictions. In terms of overall accuracy and recall, both models performed comparably, with slight variations depending on the activity class. These results highlight the effectiveness of the distributed strategy, particularly considering its modularity and scalability, and set the stage for a more in-depth discussion on the trade-offs between the two approaches.

The comparative evaluation between the centralized and distributed models reveals clear and consistent performance differences in favor of the distributed approach. As shown in Figure 12, which presents the normalized confusion matrix, each value from the centralized model is accompanied by its distributed counterpart in parentheses. This format enables a direct, class-by-class comparison. Notably, the distributed model demonstrates improved classification accuracy in several key activities. For instance, the class Sitting shows an increase in true positive rate from 0.90 to 0.96, while Walking improves from 0.96 to 0.98. In more complex activities such as Sweeping and Moving Boxes, the distributed

model also shows notable gains, increasing from 0.88 to 0.91 and from 0.91 to 0.94, respectively, while simultaneously reducing off-diagonal misclassifications. These improvements suggest that sensor-level processing enables better preservation of temporal patterns and motion nuances that may be lost during early fusion in centralized architectures.



(**a**) (**b**)

Figure 12. Comparison of classification performance between the centralized and distributed models on the test set: (**a**) normalized confusion matrix of the centralized model, where each centralized value is accompanied by its corresponding value in parentheses representing the distributed model. This enables direct comparison of prediction accuracy per class across both approaches; (**b**) class-wise bar plots of precision, recall, and F1-score for both models. Lighter bars represent the centralized model, while darker bars represent the distributed model, grouped by metric for each activity class.

Figure 12 further reinforces these findings by displaying per-class precision, recall, and F1-score for both models. The distributed model, represented by darker bars, consistently achieves higher recall across all classes. This trend is particularly evident in Sweeping, Walking, and Moving Boxes, where the recall gains are substantial. Since recall is critical in activity recognition systems to avoid missing actual events, these results highlight the distributed model's strength in capturing true positives. In terms of F1-score, which balances both precision and recall, the distributed model outperforms the centralized one in nearly every class, confirming its superior classification reliability.

While the centralized model performs slightly better in precision for a few activities, such as Sitting and Folding Clothes, the overall advantage shifts toward the distributed model due to its balanced improvement in both recall and F1-score. This trade-off is acceptable, especially in real-time systems where false negatives (missed detections) are more detrimental than occasional false positives.

In summary, the distributed model not only matches but often surpasses the centralized model in key evaluation metrics. Its superior performance in complex, dynamic activities demonstrates that localized processing with late fusion enhances robustness and generalization. These results support the design of distributed architectures in wearable activity recognition systems, where sensor-specific information and reduced latency are crucial for reliable real-time performance.

4.4. Simulation of BLE Packet Loss and Sensor Disconnection

To evaluate the resilience of the proposed distributed framework under BLE communication failures, we conducted a series of experiments simulating sensor disconnections and packet loss during inference. These tests were designed to reflect real-world challenges

in wearable systems, including dropped connections, sensor malfunctions, and partial data loss.

The distributed system includes five unique sensor sources: the Chest, Left Knee, Right Hand, Left Hand, and Right Knee. The local models associated with each sensor were independently trained and fixed during evaluation. A centralized MLP model receives a 30-dimensional concatenated input vector formed by the class probability distributions of all five sensors. To simulate sensor failures or transmission loss, we manipulated the input vectors at inference time without retraining the central model, reflecting deployment conditions where real-time adaptation is not feasible.

We evaluated the following fault scenarios:

- Baseline (No loss): all sensors are active and transmitting valid data. This serves as the reference for maximum expected performance.
- BLE packet loss (10%): randomly zeroes out 10% of the predicted outputs for each sensor to simulate moderate transmission dropouts. The central model continues to receive data from all sensors.
- BLE packet loss (30%): a more severe version, where 30% of each sensor's outputs are zeroed out, simulating high BLE packet loss or interference.
- Single-sensor disconnection: Each sensor is individually removed by replacing its full output vector with zeros. This represents a total communication failure or power loss in one node.
- Two-sensor disconnection: All 10 possible combinations of two sensors being disconnected simultaneously were tested. This simulates more critical failures or simultaneous signal dropout in two channels, allowing for the analysis of sensor pair importance in global inference.

Each scenario was evaluated using identical test subjects and consistent inference parameters to ensure reproducibility and comparability across conditions. Table 3 summarizes the impact of BLE packet loss and specific disconnection scenarios on classification performance, measured as accuracy and F1-score averaged across 10 held-out test subjects.

Table 3. Performance of the distributed model under BLE packet loss and critical sensor disconnections. (Sensor disconnection was modeled by zeroing out the corresponding probability vectors without retraining the central model.)

Scenario	Accuracy Mean (%)	F1-Score Mean (%)
Baseline (No loss)	95.924	95.83
BLE Packet Loss 10%	95.023	94.92
BLE Packet Loss 30%	91.913	91.78
1 Sensor Disconnected (Chest)	92.907	92.78
2 Sensors Disconnected (Chest and Left Hand)	84.772	84.43

While all disconnection scenarios negatively impacted performance, results show that the distributed model remained robust under moderate packet loss (10%) and some single-sensor disconnections. However, removing the Chest sensor or combinations involving the Chest and Left Hand caused significant degradation, suggesting these sensors contribute highly informative features.

To further analyze this, we conducted a complete evaluation of all one-sensor and two-sensor disconnection combinations, shown in Table 4. The values represent average metrics across the 10 test subjects. Results show a clear trend where accuracy and F1-score vary depending on which specific sensors are dropped, confirming their relative importance to the central model.

Table 4. Classification performance of the distributed model under all combinations of one or two sensor disconnections. Highlighted rows indicate the worst-performing scenarios per group.

Scenario	Sensor	Accuracy Mean (%)	F1-Score Mean (%)
One-Sensor Disconnection	Chest	92.907	92.78
	Left Knee	94.386	94.11
	Right Hand	94.835	94.68
	Left Hand	92.554	92.36
	Right Knee	95.908	95.87
Two-Sensor Disconnection	Chest and Left Knee	88.339	88.02
	Chest and Right Hand	90.751	90.35
	Chest and Left Hand	84.772	84.43
	Chest and Right Knee	92.487	92.37
	Right Hand and Left Knee	93.327	92.97
	Right Hand and Left Hand	89.764	89.48
	Right Hand and Right Knee	95.165	95.12
	Left Hand and Left Knee	90.188	89.86
	Left Hand and Right Knee	93.148	93.08
	Right Knee and Left Knee	92.654	92.51

As can be observed, the Chest and Left-Hand combination yields the lowest overall performance, underscoring the collective influence of these two elements. In contrast, disconnections involving only peripheral joints (e.g., Right Knee or Left Knee) have a minimal effect on the central model's prediction, demonstrating fault tolerance in those cases.

These experiments were designed to be fully reproducible and can be replicated. The robust results emphasize the advantage of the distributed architecture in maintaining reliable inference even in the presence of communication failures.

4.5. Latency and Energy Estimation

Beyond accuracy, latency and energy consumption are critical factors when evaluating the feasibility of activity recognition systems in real-time wearable scenarios. In this subsection, we report and compare the inference latency, energy consumption, power usage, and estimated battery life of both the centralized and distributed approaches.

All measurements were performed on an NVIDIA Jetson AGX Xavier (NVIDIA Corporation, Santa Clara, CA, USA) using TensorRT for model optimization and deployment. The Jetson platform was selected as a high-performance embedded system capable of profiling inference workloads, allowing relative comparison between centralized and distributed architectures. However, these results serve as a proxy and do not directly represent the on-device performance of ultra-low-power sensor nodes such as the MetaMotionRL (MbientLab Inc., San Jose, CA, USA).

Inference latency was recorded as the average time required to process a single input window and generate a prediction. For the centralized model, latency includes the entire processing pipeline of the fused multi-sensor input. In contrast, the distributed model performs independent local inferences at each node, followed by a lightweight aggregation step at the central unit. This modular and parallel structure led to significantly lower latency in the distributed setup.

Energy consumption per inference was estimated using TensorRT profiling tools, which account for both computational operations (FLOPs) and memory access patterns. The centralized model, due to its more complex architecture and multi-sensor integration, exhibited substantially higher energy consumption. The distributed framework, composed

of lightweight CNN–LSTM local models and a compact MLP fusion module, consumed considerably less energy per inference, a key advantage for deployment on low-power embedded platforms.

To further quantify performance, we computed the power consumption (W) of each model using the following expression:

$$\text{Power} = \text{Energy per inference (J)} \times \text{Inference rate (Hz)} \tag{1}$$

Assuming an inference rate of 50 predictions per second, consistent with the 50 Hz sampling rate commonly used in human activity recognition. This assumption enables extrapolation of average power usage. It is essential to note that the energy consumption of BLE communication was not included in this estimation, as accurate BLE transmission profiling requires direct measurement at the hardware level, which was beyond the scope of this study.

To estimate battery life, we used the specifications of the MetaMotionRL sensor (MbientLab Inc., San Jose, CA, USA), which includes a 190 mAh battery operating at 3.7 V, yielding a total energy capacity of:

$$\text{E}_{\text{battery}} = 190\ \text{mAh} \times 3.7\ \text{V} = 703\ \text{mWh} = 2530.8\ \text{J} \tag{2}$$

Using this energy capacity and the estimated power consumption for each model, the expected battery life was computed as:

$$\text{Battery Life(s)} = \frac{\text{E}_{\text{battery}}}{\text{Power consumption}} \tag{3}$$

For the distributed model, the total energy per inference was computed as the sum of the energy usage from all five local models (processing quaternion, accelerometer, and gyroscope data), plus the energy usage of the central model. Total latency was calculated assuming parallel local inference, using:

$$\text{Latency\_total} = \max\left(\text{Latency}_{\text{localmodels}}\right) + \text{Latency}_{\text{centralmodel}} \tag{4}$$

Table 5 provides a comprehensive summary of these performance efficiency metrics. The distributed model achieved a total inference time of 0.643 ms and energy usage of 3.91 mJ, compared to 3.184 ms and 9.219 mJ for the centralized model. These improvements are primarily attributed to the distributed execution of lightweight local models and the minimal overhead of the central fusion module.

Table 5. Comparison of latency, energy consumption, power usage, and estimated battery life between the centralized and distributed models.

Attribute	Centralized	Distributed	
		Local Node	Central Node
Total Latency [ms]	3.184	0.6065	0.0366
Energy Per Inference [mJ]	9.219	0.6472	0.0279
Power Consumption [W]	0.46095	0.19555 (Total)	
Battery Life [approx.]	1.52 h	3.59 h	

This efficiency translates directly into extended operational autonomy. While the centralized model supports approximately 1.52 h of continuous operation per full battery charge, the distributed framework enables an estimated runtime of roughly 3.6 h per sensor. These values represent theoretical estimates based on Jetson-level profiling and

battery specifications and should be interpreted as relative indicators rather than absolute performance guarantees.

These results highlight the practical advantages of the distributed architecture in scenarios where energy efficiency, low-latency inference, and prolonged operation are critical. By distributing the computational load across local nodes and reducing the complexity at each sensor, the system enables extended usage while maintaining competitive classification performance.

The experimental evaluation confirms that the distributed approach consistently outperformed the centralized model across all evaluated metrics, including accuracy, precision, recall, and F1-score. While the centralized model benefits from fused multi-sensor input, the distributed framework demonstrated superior performance, particularly in activities involving subtle or overlapping motion patterns. This suggests that local models, when coordinated adequately through an efficient fusion strategy, can effectively capture discriminative features while also offering additional advantages in modularity and scalability.

Beyond predictive performance, the efficiency gains achieved by the distributed model are substantial. As shown in Table 5, total inference latency decreased from 3.18 ms in the centralized model to just 0.64 ms in the distributed setup. Similarly, energy per inference was reduced from 9.22 mJ to 0.68 mJ when combining local and central components. These reductions led to a significantly lower power consumption (0.195 W vs. 0.461 W), which in turn resulted in a more than twofold increase in estimated battery life, from approximately 1.5 h in the centralized model to nearly 3.6 h in the distributed case.

These findings confirm that the distributed architecture is better suited for real-time, embedded deployments where energy autonomy and responsiveness are critical constraints. The balance between performance and efficiency offered by this framework lays a solid foundation for future work on scalable, robust, and context-aware activity recognition systems in real-world wearable applications.

Finally, we acknowledge that the current energy and latency estimates are derived from a high-performance embedded proxy (Jetson AGX Xavier), and do not include the energy overhead of BLE communication. Future work will focus on empirical energy profiling directly on ultra-low-power sensor nodes, incorporating BLE transmission cost and inference latency at the microcontroller level for full-system evaluation.

4.6. Implementation Challenges on Embedded Platforms

While this study focused on evaluating the proposed framework through offline simulations, transitioning the system to real-time embedded platforms introduces several practical challenges. First, BLE communication latency and packet loss must be carefully considered, as real-world transmission delays and asynchronous sensor updates can lead to temporary misalignments between sensor inputs. In our framework, such effects were simulated using zero-vector substitution during the inference process. Still, real-time handling may require buffering strategies, synchronization protocols, or lightweight extrapolation methods to maintain consistency across sensor nodes.

Second, the computational capabilities of embedded processors, such as microcontroller units (MCUs) or edge AI devices (e.g., Jetson Nano), can limit the complexity of local neural networks. While our CNN–LSTM models are designed to be lightweight, optimizing them further via model quantization, pruning, or hardware-specific inference engines (e.g., TensorRT, ONNX Runtime) would be necessary to meet real-time constraints.

Additionally, power consumption trade-offs must be managed carefully. Frequent wireless communication and local processing increase energy usage, which is critical in wearable applications. Efficient scheduling of inference and transmission tasks, possibly

driven by event-based triggering or adaptive sampling, will be essential for long-term operation.

Finally, synchronization among sensors becomes a non-trivial issue in multi-node systems. Without centralized clocking, drift can occur between nodes, resulting in temporal alignment issues. Implementing timestamp correction or leveraging protocols such as BLE time synchronization would be required.

Although real-time deployment was outside the scope of this work, future research will address these aspects through implementation on embedded prototypes with actual BLE communication links.

5. Discussion

The results presented in the previous section highlight the strengths and trade-offs between the centralized and distributed approaches for activity recognition using wearable sensors. This discussion builds on those findings by examining their practical implications, acknowledging limitations, and outlining directions for future research.

We acknowledge that no external motion capture system was used to validate the absolute accuracy of the Inertial Measurement Unit (IMU) orientation estimates. Instead, we relied on the onboard sensor fusion algorithms provided by the MetaMotionRL sensors, which integrate accelerometer, gyroscope, and magnetometer readings to produce stable orientation outputs in short recording windows. Since the activities were performed under controlled conditions and were of limited duration (~5 min per activity), the accumulated drift was negligible for activity classification. Furthermore, the model's generalization capability was evaluated across a diverse group of 67 participants, yielding consistent performance results. While validation with an external motion capture system would provide higher-fidelity ground truth for biomechanical analysis, it was not required for the goals of this study, which focus on robust and efficient activity recognition in practical wearable settings.

The centralized model, which processes fused data from all sensors simultaneously, achieved strong classification results. This aligns with previous findings, which show that centralized architectures benefit from access to multi-sensor input, enabling the capture of inter-sensor dependencies. However, the distributed model exhibited consistently higher precision across all activity classes, reflecting a more conservative prediction strategy with fewer false positives. This behavior is particularly advantageous in contexts where misclassifications could lead to undesired actions or safety risks.

Beyond classification performance, the distributed framework offers significant advantages in computational efficiency. By assigning local models to individual sensors and transmitting only their SoftMax outputs, the architecture reduces per-node processing requirements and enables parallel inference with lower latency. When deployed on the NVIDIA Jetson AGX Xavier using TensorRT, the distributed model achieved an order-of-magnitude improvement in both inference time and energy consumption compared to the centralized configuration. These gains translate into extended operational autonomy, with an estimated runtime increase from just over 1.52 h in the centralized case to approximately 3.6 h per sensor in the distributed configuration. This improvement makes the system more suitable for continuous applications such as physical therapy monitoring, occupational safety, and long-term activity tracking.

Although battery life was estimated and compared between configurations, we did not conduct experiments to assess how sensor performance or data quality may degrade as battery levels drop. Recording sessions in this study lasted approximately 37 min, well within the sensors' full runtime, and no perceptible degradation in signal quality, stability, or transmission was observed during that period. However, we acknowledge that

low battery conditions in extended deployments could potentially impact sampling rates, sensor accuracy, or communication reliability. Future work will investigate these aspects in the context of long-term, real-time operation.

It is essential to note that the distributed inference framework was evaluated using offline simulations, where data were collected and synchronized in advance under controlled conditions. This decision allowed us to focus on validating the inference architecture itself, specifically, the ability to classify activities based on partial, independently processed sensor inputs, without introducing variability from wireless communication or real-time processing constraints.

Nonetheless, we recognize that real-time implementation aspects such as BLE transmission delays, packet loss, and power consumption are critical for practical deployment and were not evaluated in this study. To address this, we have outlined a concrete plan for real-time implementation. Specifically, we intend to integrate BLE communication between each wearable sensor (MCU) and a central embedded hub (e.g., NVIDIA Jetson Nano or Xavier), using concurrent Bluetooth interfaces to emulate realistic operating conditions. During this phase, we will measure key transmission parameters, including:

- Total system delay, from sensing to central decision output (end-to-end latency);
- Packet loss rate, both random and due to temporary disconnections;
- Transmission power and energy consumption, measured per sensor over extended sessions.

Additionally, we aim to characterize the temporal behavior of the system as battery levels decrease, to analyze possible degradation in sampling rate, latency, or BLE stability over time. This implementation phase will also include the integration of error-handling mechanisms, predictive smoothing, and dropout-tolerant inference to mitigate the effects of intermittent data streams.

We also aim to characterize system behavior as battery levels decline, and implement error-handling mechanisms, predictive smoothing, and dropout-tolerant inference to mitigate instability in the data stream. This phase will involve a new round of data collection with an expanded participant cohort under dynamic and unconstrained conditions. Together, these efforts will bridge the gap between offline validation and real-time deployment, enabling a more comprehensive assessment of the system's viability in embedded and mobile contexts.

The distributed design additionally provides flexibility and robustness. Each local model can operate independently, allowing for partial inference even when one or more sensors become disconnected. This capability improves fault tolerance and supports modular scalability, critical features in real-world environments with unpredictable conditions. These aspects, along with reduced communication overhead, further reinforce the suitability of distributed learning systems for wearable applications with resource constraints.

Despite these advantages, the distributed framework presents certain limitations. Local models may lack the global context available in centralized architectures, which can impact classification performance in activities that involve inter-limb coordination. Furthermore, managing asynchronous data streams and synchronizing predictions across multiple nodes introduces architectural complexity, particularly in lossy or bandwidth-limited communication environments.

Another important consideration is the absence of explicit spatial calibration between sensors. In this study, each local model leveraged only its motion data, which was sufficient for the selected activities. However, tasks that require precise relative positioning, such as gait symmetry analysis, sign language interpretation, or specific rehabilitation protocols, may benefit from sensor alignment in a standard spatial frame. As the focus here was to

demonstrate a minimally coordinated distributed system, alternative spatial configurations and calibration strategies were not explored.

Lastly, it is essential to distinguish the proposed HAR framework from full-body biomechanical systems such as Xsens. These systems aim to reconstruct body pose using anatomically anchored sensor placements and high-precision inertial data, typically requiring extensive spatial calibration and personalized models. In contrast, our system focuses on activity classification using independently captured local motion features, without reconstructing global posture. This design supports computational efficiency, modular deployment, and ease of use in practical applications where full pose estimation is unnecessary or impractical.

In summary, while the centralized model offers high overall accuracy and more straightforward integration, the distributed approach provides superior precision, energy efficiency, and robustness. These qualities make it a strong candidate for real-time embedded activity recognition. Table 6 summarizes the key trade-offs discussed in this section and supports informed system design decisions based on specific deployment constraints.

Table 6. Summary of trade-offs between centralized and distributed.

Criterion	Centralized Model	Distributed Model
Overall Accuracy	High	Higher
Precision	Lower (more false positives)	Higher across all classes
Recall	High	Higher in some classes
Latency (Inference)	High (sequential processing)	Low (parallel local inference)
Energy Consumption	High	Low
Estimated Battery Life	1.52 h	3.6 h
Computational Load	Centralized, resource-intensive	Distributed across nodes, lightweight
Scalability	Limited (single point of processing)	High (modular, sensor-level models)
Fault Tolerance	Low (dependent on all sensor streams)	High (can tolerate partial sensor loss)
Implementation	Simpler to manage globally	More complex synchronization, but modular

Future work will focus on exploring architectures that adaptively combine centralized and distributed inference based on contextual factors and energy constraints. Promising research directions include developing methods to dynamically weight sensor contributions, incorporating attention mechanisms at the fusion stage, and enabling online learning directly at the edge. While the current evaluation was conducted in a controlled environment with a predefined activity protocol to ensure synchronization and label consistency, one of the primary motivations behind the proposed distributed framework is its potential scalability to real-world, dynamic scenarios. Accordingly, future efforts will include validating the model under unscripted, noisy conditions and free-form activity sessions to assess its robustness and generalizability. Moreover, extending evaluation to multi-user scenarios, outdoor environments, and intermittent connectivity will be crucial for advancing the deployment readiness of these systems. Finally, implementing and validating the proposed framework in real-world, real-time conditions beyond offline simulation remains an essential step toward practical application.

Author Contributions: Software, J.L.R.-C., K.N.-T., L.S.-A. and J.F.P.-M.; Investigation, J.L.R.-C., K.N.-T., L.S.-A. and J.F.P.-M.; Writing—review & editing, J.L.R.-C., K.N.-T., L.S.-A. and J.F.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received financial support from the NSF EPSCoR Center for the Advancement of Wearable Technologies (CAWT) under Grant No. OIA-1849243.

Institutional Review Board Statement: The study was conducted under the Declaration of Helsinki and approved by the University of Puerto Rico's Institutional Review Board (IRB) under protocol number 2024120022 (approved on 27 January 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset used in this study is openly available in Zenodo at https://doi.org/10.5281/zenodo.15830858.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
ML	Machine Learning
FL	Federated Learning
FSL	Federated Split Learning
SL	Split Learning
MFFL	Multi-Frequency Federated Learning
SNN	Spiking Neural Network
BLE	Bluetooth Low Energy
LSTM	Long-Short Term Memory
MLP	Multi-Layer Perceptron
FLOPs	Floating Point Operations
MCUs	Microcontroller Units
IMU	Inertial Measurement Unit

References

1. Baghezza, R.; Bouchard, K.; Bouzouane, A.; Gouin-Vallerand, C. From Offline to Real-Time Distributed Activity Recognition in Wireless Sensor Networks for Healthcare: A Review. *Sensors* **2021**, *21*, 2786. [CrossRef]
2. El-Adawi, E.; Essa, E.; Handosa, M.; Elmougy, S. Wireless Body Area Sensor Networks Based Human Activity Recognition Using Deep Learning. *Sci. Rep.* **2024**, *14*, 2702. [CrossRef] [PubMed]
3. Agarwal, P.; Alam, M. A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. *Procedia Comput. Sci.* **2020**, *167*, 2364–2373. [CrossRef]
4. Zhang, S.; Li, Y.; Zhang, S.; Shahabi, F.; Xia, S.; Deng, Y.; Alshurafa, N. Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors* **2022**, *22*, 1476. [CrossRef] [PubMed]
5. Saldana-Aristizabal, L.; Nino-Tejada, K.; Rivas-Caicedo, J.L.; Patarroyo-Montenegro, J.F. Evaluating Quaternion-Based Representations for Human Activity Recognition Using Motion Capture. In Proceedings of the International Symposium on Intelligent Computing and Networking (ISICN 2025), San Juan, PR, USA, 17–19 March 2025; Springer: Berlin/Heidelberg, Germany, 2025.
6. Rivas-Caicedo, J.L.; Saldana-Aristizabal, L.; Nino-Tejada, K.; Patarroyo-Montenegro, J.F. Mitigating Communication Failures in Multi-Sensor Wearable Systems: Extrapolation Methods for LSTM-Based Posture Classification. In Proceedings of the International Symposium on Intelligent Computing and Networking (ISICN 2025), San Juan, PR, USA, 17–19 March 2025; Springer: Berlin/Heidelberg, Germany, 2025.
7. Agarwal, S.; Alam, M. A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. *arXiv* **2019**, arXiv:1909.12917. [CrossRef]
8. Chen, Z.; Zhu, Q.; Soh, Y.C.; Zhang, L. Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3070–3080. [CrossRef]
9. Ord'oñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]

10. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for Human Activity Recognition Using Mobile Sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services (MobiCASE), Austin, TX, USA, 6–7 November 2014; pp. 197–205.
11. Haque, S.T.; Ni, J.; Li, J.; Yan, Y.; Ngu, A.H.H. LightHART: Lightweight Human Activity Recognition Transformer. In *Lecture Notes in Computer Science, Pattern Recognition*; Springer Nature: Cham, Switzerland, 2024; pp. 425–441. [CrossRef]
12. Yu, H.; Chen, Z.; Zhang, X.; Chen, X.; Zhuang, F.; Hiong, H.; Cheng, X. FedHAR: Semi-Supervised Online Learning for Personalized Federated Human Activity Recognition. *IEEE Trans. Mob. Comput.* **2021**, *22*, 3318–3332. [CrossRef]
13. Chaddad, A.; Wu, Y.; Desrosiers, C. Federated Learning for Healthcare Applications. *IEEE Internet Things J.* **2023**, *11*, 7339–7358. [CrossRef]
14. Dey, A.; Chakraborty, S.; Ray, R. FedOpenHAR: A Federated Learning Benchmark for Wearable Human Activity Recognition. *IEEE Internet Things J.* **2023**, *10*, 2163–2174.
15. Neftci, E.O.; Mostafa, H.; Zenke, F. Surrogate Gradient Learning in Spiking Neural Networks. *IEEE Signal Process. Mag.* **2019**, *36*, 61–63. [CrossRef]
16. Kalantarian, H.; Alshurafa, N.; Sarrafzadeh, M. Adaptive Activity Recognition with Crowdsourcing and Online Learning (AdARC). *IEEE J. Biomed. Health Inf.* **2017**, *21*, 1193–1201.
17. Sannara, P.; Maji, S.; Amintoosi, H. Edge-Distributed Sensor Fusion for Real-Time Human Activity Recognition. In Proceedings of the 21st IEEE Conference on Sensors, Dallas, TX, USA, 30 October–2 November 2022; IEEE: New York, NY, USA, 2022; pp. 1212–1217.
18. Sozinov, K.; Vlassov, V.; Girdzijauskas, S. Human Activity Recognition Using Federated Learning. In Proceedings of the 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), Melbourne, Australia, 11–13 December 2018; IEEE: New York, NY, USA, 2018; pp. 1103–1111.
19. Gad, G.; Fadlullah, Z.M.; Rabie, K.; Fouda, M.M. Communication-Efficient Privacy-Preserving Federated Learning via Knowledge Distillation for Human Activity Recognition Systems. In Proceedings of the IEEE International Conference on Communications, Rome, Italy, 28 May–1 June 2023; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023; pp. 1572–1578. [CrossRef]
20. Ndeko, J.; Shaon, S.; Beal, A.; Sahoo, A.; Nguyen, D.C. Federated Split Learning for Human Activity Recognition with Differential Privacy. In Proceedings of the IEEE Consumer Communications and Networking Conference, CCNC, Las Vegas, NV, USA, 10–13 January 2025; Institute of Electrical and Electronics Engineers Inc.: New Jersey, NJ, USA, 2025.
21. Fenoglio, D.; Li, M.; Casnici, D.; Laporte, M.; Gashi, S.; Santini, S.; Gjoreski, M.; Langheinrich, M. Multi-Frequency Federated Learning for Human Activity Recognition Using Head-Worn Sensors. In Proceedings of the 2024 International Conference on Intelligent Environments (IE), Ljubljana, Slovenia, 17–20 June 2024; pp. 17–24. [CrossRef]
22. Thakur, D.; Guzzo, A.; Fortino, G. Energy Aware Federated Learning with Application of Activity Recognition. In Proceedings of the 2023 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Abu Dhabi, United Arab Emirates, 14–17 November 2023; IEEE: New York, NY, USA, 2023; pp. 0636–0642.
23. Saylam, B.; Durmaz, O. Highlights Federated Learning on Edge Sensing Devices: A Review. *arXiv* **2023**, arXiv:2311.01201.
24. Khan, Q.W.; Khan, A.N.; Rizwan, A.; Ahmad, R.; Khan, S.; Kim, D.H. Decentralized Machine Learning Training: A Survey on Synchronization, Consolidation, and Topologies. *IEEE Access* **2023**, *11*, 68031–68050. [CrossRef]
25. Ding, G.; Tian, J.; Wu, J.; Zhao, Q.; Xie, L. Energy efficient human activity recognition using wearable sensors. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Barcelona, Spain, 15–18 April 2018; IEEE: New York, NY, USA, 2018; ISBN 9781538611548.
26. Rezaie, H.; Ghassemian, M. An Adaptive Algorithm to Improve Energy Efficiency in Wearable Activity Recognition Systems. *IEEE Sens. J.* **2017**, *17*, 5315–5323. [CrossRef]
27. Yang, L.; Amin, O.; Shihada, B. Intelligent Wearable Systems: Opportunities and Challenges in Health and Sports. *ACM Comput. Surv.* **2024**, *56*, 1–42. [CrossRef]
28. Zmitri, M.; Fourati, H.; Vuillerme, N. Human Activities and Postures Recognition: From Inertial Measurements to Quaternion-Based Approaches. *Sensors* **2019**, *19*, 4058. [CrossRef] [PubMed]
29. Rivas, J.; Niño, K.; Saldaña, L.; Patarroyo, J. Wearable IMU Sensor Dataset for Human Activity Recognition (HAR). 2025. Available online: https://zenodo.org/records/15769102 (accessed on 29 July 2025).
30. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980 2014.
31. Bengio, Y. Practical Recommendations for Gradient-Based Training of Deep Architectures. In *Neural Networks: Tricks of the Trade*; Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 437–478.

32. Masters, D.; Luschi, C. Revisiting Small Batch Training for Deep Neural Networks. *arXiv* **2018**, arXiv:1804.07612.
33. Prechelt, L. Early Stopping—But When? In *Neural Networks: Tricks of the Trade*; Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Indirect AI-Based Estimation of Cardiorespiratory Fitness from Daily Activities Using Wearables

Laura Saldaña-Aristizábal [1,*], Jhonathan L. Rivas-Caicedo [1], Kevin Niño-Tejada [1] and Juan F. Patarroyo-Montenegro [2,*]

[1] Department of Electrical and Computer Engineering, University of Puerto Rico, Mayagüez, PR 00680, USA; jhonathanleonardo.rivas@upr.edu (J.L.R.-C.); kevin.nino@upr.edu (K.N.-T.)
[2] Department of Computer Science and Engineering, University of Puerto Rico, Mayagüez, PR 00680, USA
* Correspondence: laura.saldana@upr.edu (L.S.-A.); juan.patarroyo@upr.edu (J.F.P.-M.)

Abstract

Cardiorespiratory fitness is a predictor of long-term health, traditionally assessed through structured exercise protocols that require maximal effort and controlled laboratory conditions. These protocols, while clinically validated, are often inaccessible, physically demanding, and unsuitable for unsupervised monitoring. This study proposes a non-invasive, unsupervised alternative—predicting the heart rate a person would reach after completing the step test, using wearable data collected during natural daily activities. Ground truth post-exercise heart rate was obtained through the Queens College Step Test, which is a submaximal protocol widely used in fitness settings. Separately, wearable sensors recorded heart rate (HR), blood oxygen saturation, and motion data during a protocol of lifestyle tasks spanning a range of intensities. Two machine learning models were developed—a Human Activity Recognition (HAR) model that classified daily activities from inertial data with 96.93% accuracy, and a regression model that estimated post step test HR using motion features, physiological trends, and demographic context. The regression model achieved an average root mean squared error (RMSE) of 5.13 beats per minute (bpm) and a mean absolute error (MAE) of 4.37 bpm. These findings demonstrate the potential of test-free methods to estimate standardized test outcomes from daily activity data, offering an accessible pathway to infer cardiorespiratory fitness.

Keywords: deep learning; health monitoring; human activity recognition; wearable sensors

1. Introduction

Modern society confronts an increasing burden of chronic, lifestyle-related conditions: hypertension affects approximately 1.28 billion adults aged 30–79 and causes around 10 million deaths annually [1]. Type 2 diabetes now impacts nearly 589 million adults, with this figure expected to rise to 853 million by 2050; alarmingly, about 40% of those affected remain undiagnosed [2]. These largely preventable conditions require ongoing monitoring, something episodic, clinic-based care often fails to provide. The COVID-19 pandemic further underscored the urgency for scalable, remote, and non-invasive health tools that empower individuals to manage their own health from home.

Cardiorespiratory Fitness (CRF) is widely recognized as one of the most powerful predictors of long-term health outcomes. A higher CRF level is strongly associated with reduced risk of cardiovascular disease, metabolic disorders, and all-cause mortality [3]. The gold standard for assessing CRF is maximal oxygen uptake (VO_2max); however, its

assessment typically requires a cardiopulmonary exercise test (CPET), involving lab-grade equipment (shown in Figure 1, e.g., metabolic gas analyzers, ergometers, electrocardiogram), trained personnel, and maximal exertion from the participant.



Image from Wikimedia Commons contributed by COSMED and shared under a license permitting reuse, verified through Wikimedia's VRT system (Ticket #2013073010007597).

Figure 1. Standard CPET equipment: Blood pressure monitor, mask (volume sensor and gas analyzer tubing), electrocardiogram (ECG), ergometer, pulse oximeter, gas analyzer, display of breath-by-breath data and exercise ECG.

Submaximal alternatives such as the Queens College Step Test [4] offer practical alternatives. This standardized protocol involves stepping at a prescribed cadence for three minutes and then measuring the participant's heart rate during recovery. The resulting post-test heart rate is used in validated equations to estimate VO_2max with reasonable accuracy. As a lower-burden alternative to CPET, the step test maintains clinical relevance while being simpler to administer, making it a practical bridge between wearable measurements and established fitness indicators. In this study, the heart rate predicted as if the participant had completed a step test serves as an intermediate variable, enabling VO_2max estimation from unconstrained, daily activity data.

While CPET is highly accurate, its complexity, cost, and physical demand limit its use in preventive care and for populations such as older adults or high-risk patients.

Wearable technologies offer a promising opportunity to improve CRF assessment by continuously collecting heart rate, SpO_2, and motion data in a non-invasive, user-friendly way. Beyond capturing raw biomarkers, they also support real-time activity recognition, providing essential context for interpreting physiological changes. For example, an elevated heart rate during walking is expected, but the same value at rest may indicate a problem. By integrating this contextual layer, wearables enable more personalized and meaningful health insights.

Despite its clinical importance, CRF remains difficult to assess in real-world conditions, as traditional protocols require structured effort and specialized equipment, while passive monitoring often lacks the context needed to interpret physiological signals. This study investigates whether it is possible to estimate an individual's cardiorespiratory capacity (specifically, their post step test heart rate) based on physiological behavior during natural, unconstrained activities. By analyzing responses to everyday tasks such as walking or folding clothes, our approach aims to bypass formal exercise protocols and support more scalable, user-centered fitness assessment in free-living environments.

1.1. Related Work

1.1.1. Protocol-Guided or Submaximal Exercise-Based Estimation

Several studies have estimated VO_2max using structured submaximal protocols such as step tests, cycling, or treadmill walks, aiming to reduce the burden of maximal cardiopulmonary testing while preserving predictive accuracy. For instance, Sheridan et al. [5] used team sport simulations to estimate oxygen uptake from inertial and heart rate data, and Wiecha et al. [6] predicted VO_2max using submaximal effort and body composition in athletes. Akay and Abut [7] reviewed machine learning and statistical models applied to VO_2max prediction in structured environments. Though more accessible than maximal tests, these methods still require supervised physical effort and synchronization with protocols, often targeting specific populations, limiting their applicability in broader, unsupervised real-world settings.

1.1.2. Wearable-Based Estimation of Cardiorespiratory Fitness

Recent studies have used consumer-grade wearables to estimate CRF in semi-controlled or free-living environments, typically using heart rate, step count, or photoplethysmography (PPG) data. Neshitov et al. [8], for example, estimated CRF from heart rate and step count collected during free-living activity, but their method required at least 200 min of active data, limiting its use for real-time or short-duration assessments.

Similarly, Spathis et al. [9] introduced a longitudinal approach to VO_2max prediction using wearable data, emphasizing the value of repeated measures over time. While promising, this approach requires prolonged usage and assumes consistent device adherence, as the window size is one week of data. In a more controlled setting, Hsiao et al. [10] developed a custom wearable system incorporating multichannel PPG to estimate VO_2 during treadmill exercise, showcasing real-time estimation capabilities but still requiring structured physical effort and specialized hardware.

While wearable-based CRF models have advanced non-invasive fitness assessments, many still depend on long-term tracking, structured effort, or device-specific setups, limiting their use for rapid or flexible deployment. In contrast, our two-model framework provides a faster, context-aware estimation from just a few minutes of wearable data collected during natural activities, making it well suited for opportunistic use in real-world settings.

1.1.3. Modeling CRF with Indirect or Non-Exercise Features

Recent approaches have explored estimating CRF without requiring structured physical exertion. Instead of relying on exercise protocols, these models infer fitness from demographic data, short physiological snapshots, or latent health indicators. For instance, Lee and Park [11] proposed a deep learning model that predicts health status from ordered questionnaire responses, demonstrating the feasibility of low-effort CRF estimation.

A promising alternative involves two-step strategies that first predict an intermediate value, like post-exercise heart rate, then estimate CRF using validated formulas. This preserves alignment with VO_2max while avoiding maximal or submaximal testing, offering a practical balance between accuracy and accessibility in settings where traditional protocols are impractical.

1.1.4. Human Activity Recognition Using Inertial Sensors

HAR is a core application of wearable sensing, especially with inertial measurement units (IMUs) that capture motion in free-living conditions. Deep learning models such as Convolutional Neural Networks (CNNs) and Long-Short Term-Memory LSTMs have improved classification accuracy, moving from handcrafted features to end-to-end learning [12]. Qureshi et al. [13] reviewed HAR challenges, including sensor variability, class

imbalance, and real-time demands. While HAR is mainly used for context inference, its integration into health analytics is expanding. Recognizing physical activity enriches physiological signal interpretation and supports context-aware modeling, yet few HAR systems connect activity context with individualized physiological responses in CRF estimation.

1.2. Study Proposal: Context-Aware Step Test HR Estimation from Wearable Data

This study introduces a novel two-model framework for indirectly estimating cardiorespiratory fitness from wearable sensor data collected during daily life, eliminating the need for structured exercise testing or prolonged tracking. The key innovation lies in combining context-aware activity recognition with physiological modeling to approximate a step test-equivalent heart rate.

The first model employs a HAR model to classify physical activities based on inertial sensor data. In the second model, the classified activity, combined with physiological signals (heart rate and oxygen saturation), motion features, and demographic information, is used to predict the heart rate the individual would exhibit after completing a standardized step test.

Figure 2 illustrates the full pipeline of the proposed method. Wearable devices capture accelerations, angular velocities, quaternions, heart rate, and SpO_2 during daily activities, while demographic variables such as age, sex, and body composition are collected separately. Model 1 uses motion features to infer the current activity. Model 2 then uses the corresponding activity label, collected independently during the protocol, alongside physiological signals and user metadata to estimate the step test-equivalent heart rate. Although both models use activity context, they are trained independently.

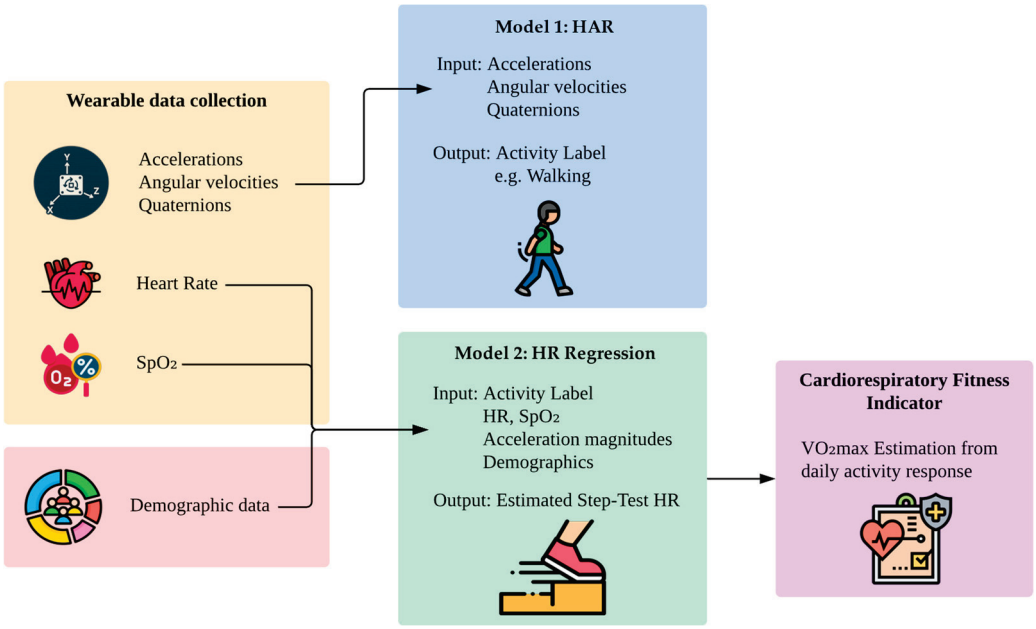


Figure 2. Overview of the proposed two-model framework for cardiorespiratory fitness estimation. Model 1 classifies physical activity using motion data from wearable sensors. Model 2 independently uses the ground-truth activity label along with physiological signals, movement features, and demographic metadata to estimate the heart rate a person would reach after a standardized step test.

By modeling natural physiological responses during free-living activities, this approach introduces a new path toward unsupervised, low-burden fitness assessment. Unlike prior CRF estimation methods that require structured exertion, dedicated testing, or long-term data aggregation, this framework allows opportunistic CRF estimation from

just a few minutes of wearable data, supporting real-world deployment in remote and preventive care.

1.3. Contributions and Paper Organization

The main contributions of this work include: (1) a unified framework that integrates activity recognition and physiological modeling to estimate step test-equivalent heart rate from daily wearable data, and (2) an indirect method for VO_2max estimation that does not require participants to follow formal exercise protocols. Future directions include expanding the dataset to incorporate broader demographic and health diversity, replacing activity classification with effort-based regression for more generalized modeling, and validating predicted fitness indicators against direct VO_2max measurements obtained via gas exchange analysis in clinical settings.

The rest of the paper is structured as follows: Section 2 details the study protocol, wearable setup, data processing, and models. Section 3 presents evaluation results for activity classification and heart rate prediction. Section 4 discusses the implications, limitations, and future directions. Section 5 concludes with a summary of contributions and the significance of this approach for health monitoring.

2. Materials and Methods

This section describes the experimental protocol, data collection procedures, wearable setup, preprocessing steps, and deep learning architectures used in the study. It outlines how participants were recruited and monitored using wearables during a structured protocol involving real-world physical activities. Two separate pipelines were established: one for HAR and another for estimating post-exercise heart rate through regression.

2.1. Participants and Protocol

This study received ethical approval from the Institutional Review Board (IRB) of the University of Puerto Rico. All research procedures complied with the ethical standards set by the Collaborative Institutional Training Initiative (CITI Program) for research involving human participants. Approval was granted prior to the start of data collection. Before participating, each subject received a full explanation of the study's objectives and procedures and provided written informed consent. To ensure participant's safety, all individuals completed the Physical Activity Readiness Questionnaire (PAR-Q) and disclosed basic health information prior to participation.

Data collection took place in a controlled indoor environment that adhered to ANSI/CTA recommendations for real-world testing [14]. The room was private, free from distractions, and maintained under consistent conditions, including a constant temperature of 21 °C, with lighting and humidity levels set within comfortable indoor ranges. These measures were intended to minimize external influences on physiological measurements and promote participant comfort throughout the experimental protocol.

A total of 67 participants were initially enrolled in the study. However, data from 7 participants were excluded from the final analysis due to issues affecting the quality or completeness of the recorded data. The resulting dataset includes 60 participants, consisting of 36 males and 24 females, with ages ranging from 19 to 64 years. For each participant, key demographic variables were collected, including age, sex, height, weight, body fat percentage, and body mass index (BMI). These variables were used to characterize the study population and were also included as features in the heart rate regression model.

Figure 3 illustrates the demographic distribution of the final participant pool, including sex proportions, age ranges, body fat percentage, and BMI categories. This visual summary

highlights the diversity in body composition and physical profiles across the sample, supporting the robustness of the physiological modeling approach.

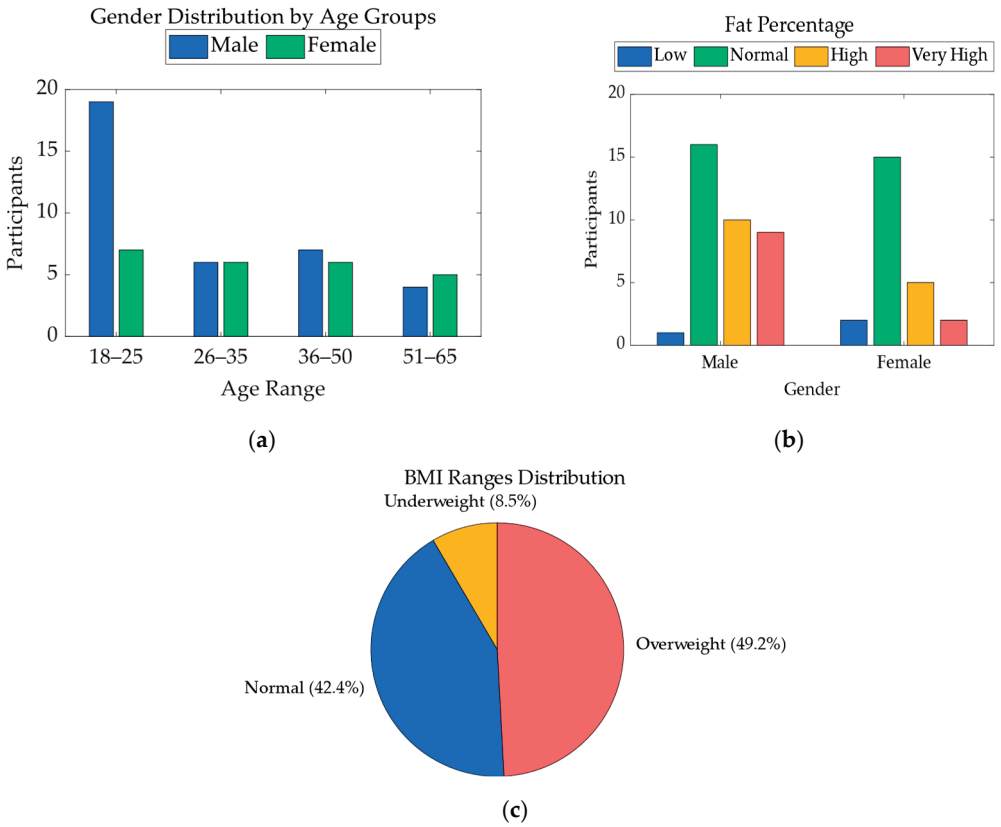


Figure 3. Demographic distribution of study participants. (**a**) Distribution of participants based on age and gender; (**b**) distribution of participants by gender and fat percentage ranges; (**c**) distribution of participants based on BMI ranges.

Each participant began the session by performing the Queens College Step Test [4], a standardized submaximal exercise protocol used to assess cardiorespiratory recovery. The test involves stepping up and down on a platform at a fixed cadence of 24 steps per minute for men and 22 steps per minute for women, over a period of three minutes. Maintaining this cadence is essential for producing consistent post-exercise heart rate responses, which are later used to estimate cardiorespiratory fitness. Participants were given verbal instructions, and a metronome was used to help them keep the correct rhythm. At the end of the test, the heart rate was measured manually at the wrist by a trained researcher.

Following the step test, wearable sensors were placed on the participants to monitor physiological and motion data during a sequence of activities. Details about the sensor setup, body placement, and signal types are provided in Section 2.2.

After the step test and sensor setup, each participant completed a structured activity protocol designed according to the ANSI/CTA-2108 Framework for Validation of Health Technology Derived Metrics under Naturalistic or Unconstrained Conditions [15] and the ANSI/CTA-2065.1 Standard for Physical Activity Monitoring for Heart Rate—Real World Analysis [14]. Activity intensity levels were categorized following CTA-2074 guidelines [16], which define physical effort using Metabolic Equivalent of Task (MET) values and real-world examples. The protocol included alternating periods of rest, light, moderate, and vigorous activity, simulating common daily tasks while eliciting varied cardiovascular responses.

The full activity protocol lasted 37 min and was consistent with all participants. Activities were selected to cover a representative spectrum of daily movements with increasing physiological demand. Table 1 summarizes the activity types, their assigned MET values, and their corresponding activity codes from the 2024 Adult Compendium of Physical Activities [17].

Table 1. Structured activity protocol with MET values.

Time [min]	Activity	Activity Code	MET
0–3	Sit	07021	1.0
3–8	Fold clothes while seated	05091	2.0
8–13	Sweep, light effort	05011	2.3
13–16	Sit	07021	1.0
16–21	Walk at 5.6 km/h, treadmill surface	17190	3.8
21–26	Move a 4.5 kg box from the table to the floor and vice versa	11615	4.5
26–29	Sit	07021	1.0
29–34	Static bike, 101–125 watts	01224	6.8
34–37	Sit	07021	1.0

This sequence allowed for the capture of wearable data across a realistic range of physical intensities, while incorporating recovery periods to observe heart rate dynamics. To ensure consistent task execution, an automated timer system provided visual and auditory cues to the participants at each activity transition, indicating the start and end of every segment.

2.2. Wearable Setup

To capture detailed biomechanical and physiological signals, participants were instrumented with a combination of motion sensors and wearable biomarker devices. The motion capture setup was designed to ensure symmetry and provide coverage of both the upper and lower limbs, as well as the torso. A total of five Inertial Measurement Units (IMUs) MetaMotionRL from MbientLab Inc. were used, positioned on the chest, left hand, right hand, left knee, and right knee. These IMUs operated in streaming mode during the activity protocol and transmitted their data via Bluetooth after each session for synchronization and storage. Each IMU captured a specific subset of motion features:

- Chest: Quaternions, tri-axial acceleration, and gyroscope (50 Hz);
- Left hand: Quaternions only (50 Hz);
- Right hand: Acceleration and gyroscope (50 Hz);
- Left knee: Acceleration and gyroscope (50 Hz);
- Right knee: Quaternions only (50 Hz).

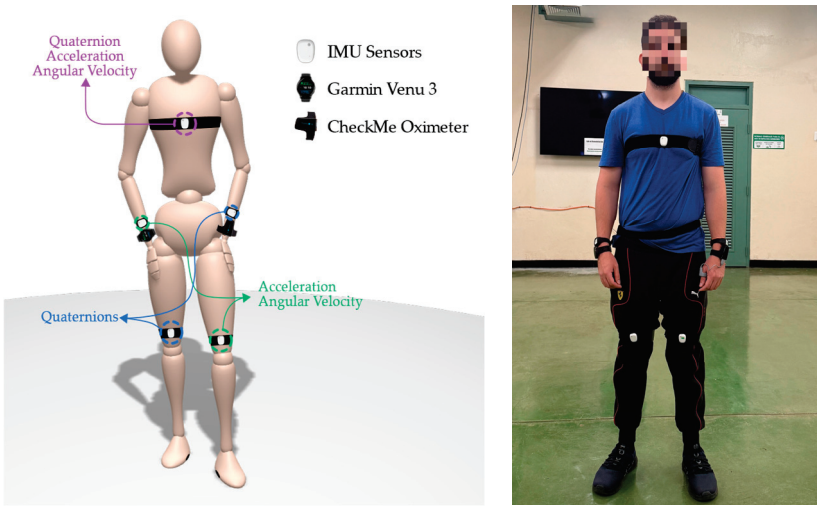
The asymmetrical configuration resulted from hardware limitations: the MetaMotionRL sensors could not record quaternions and raw acceleration/gyroscope data simultaneously. To balance signal diversity while minimizing device count, we adopted a complementary layout, assigning quaternions to one limb and acceleration/gyroscope to the opposite. This strategy leveraged the symmetric nature of the protocol's activities (e.g., walking, sweeping), allowing mirrored motion patterns to be inferred. Moreover, prior findings [18] showed that quaternion-based representations improve classification performance, and that full-body instrumentation can be reduced without significant loss of accuracy, supporting the idea that a reduced, strategically distributed configuration can

still capture the essential features for activity recognition. The chest sensor was assigned the most comprehensive sensing role, serving as a central anchor point to contextualize upper and lower limb movements with respect to the trunk.

IMU calibration was performed immediately following sensor placement on each participant. After all devices were secured, participants were instructed to sit upright with their hands resting on their knees and remain still. During this neutral and reproducible posture, the IMUs were initialized and calibrated to establish a reference frame for motion tracking. Since individual posture and body geometry vary across participants, the resulting orientation references naturally differed between subjects. This strategy promotes robustness and real-world applicability, enabling the model to tolerate small deviations in sensor alignment or anatomical differences, conditions that are common in everyday use.

To track orientation, quaternions were used instead of Euler angles, as they prevent the gimbal lock effect and support smooth, multi-axis rotational representation; as validated in our previous work [19]. This combination of data quality and minimal sensor burden supports both modeling accuracy and future real-world deployment.

For physiological monitoring, two wrist-worn devices were used, each capturing HR and SpO_2 at a sampling rate of 0.5 Hz. A Garmin Venu 3 smartwatch, worn on the right wrist, and a CheckMe™ oximeter, worn on the left wrist, provided redundant measurements for cross-validation and robustness during motion. During data review, the Garmin watch occasionally returned zero values for SpO_2 during movement periods, indicating signal dropout, in contrast, the CheckMe™ oximeter provided more stable and complete recordings. As a result, SpO_2 data from the CheckMe™ was prioritized during preprocessing to ensure signal continuity. Figure 4 provides a schematic overview of sensor placement, indicating the precise body locations and signal types associated with each device.



(**a**) (**b**)

Figure 4. Wearable sensor placement overview. (**a**) Wearable sensing prototype; (**b**) human subject instrumented with the proposed sensor setup.

To ensure temporal alignment across all sensor inputs, including IMUs and physiological devices, linear extrapolation based on timestamps was used for synchronization. This step enabled frame-level integration of biomechanical and cardiorespiratory data, resulting in a rich, unified dataset suitable for downstream analysis in both activity classification and physiological modeling. The collected dataset is available in the resource [20].

2.3. Data Processing

2.3.1. Preprocessing for Human Activity Recognition

Raw IMU data collected at 50 Hz, including acceleration, gyroscope, and quaternion signals, underwent a multi-step preprocessing pipeline prior to activity classification.

First, signals were smoothed using a fifth-order Butterworth low-pass filter with a cutoff frequency of 5 Hz, which preserves human movement frequencies while eliminating high-frequency noise. The filtered signals were then Z-score normalized feature-wise, with the mean and standard deviation computed per feature channel and per participant, to account for inter-individual differences in baseline physiology and motion patterns, ensuring that the model focuses on relative changes over time rather than being biased by absolute values that vary naturally across individuals.

To assign activity labels, each sample was timestamp-aligned and segmented into predefined time intervals according to the standardized activity protocol. A 5 s margin was trimmed after each activity transition to reduce transitional noise and avoid label ambiguity.

After labeling, the continuous signal was segmented into overlapping sliding windows to create fixed-length samples for input to the neural network. Figure 5 illustrates the segmentation and input structure used for HAR model training. Each input sample consists of a sliding window of 20 frames (0.4 s) and is defined by a set of individual features derived from the wearable sensors. These features include quaternions, linear accelerations, and angular velocities collected from the chest, knees, and hands, based on the specific sensor configuration described in Section 2.2. Each signal component (e.g., x, y, z of acceleration or w, x, y, z of quaternions) is treated as a separate input feature in the model. The windowing process uses a stride of 6 frames, resulting in a 70% overlap between windows. This approach maximizes the temporal resolution of the input while preserving the sequential structure of the movement data.

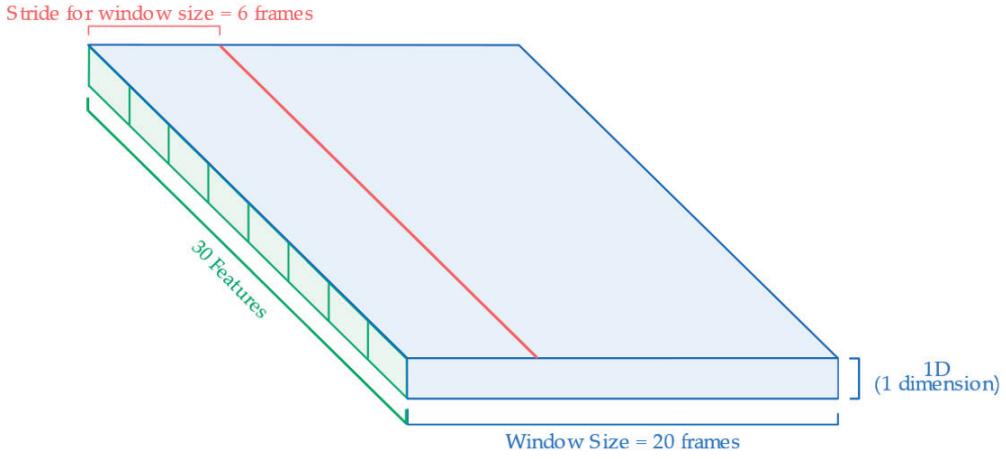


Figure 5. Input features and segmentation for the HAR model.

2.3.2. Preprocessing for Heart Rate Regression

For the HR regression task, the pipeline integrated synchronized physiological, biomechanical, and demographic data into a unified input structure suitable for deep learning. The goal was to model how each individual's body responded to natural activities to predict their step test-equivalent heart rate.

For the sequential data, time-series signals were collected at 0.5 Hz and formed the basis of the sequential input branch. These included:

- HR and SpO_2 recorded at 0.5 Hz;
- Acceleration magnitudes computed from tri-axial accelerometer data at the chest, left knee, and right hand. Each signal was first filtered with a 5th-order Butterworth low-pass filter (cutoff = 5 Hz) and then downsampled to match the 0.5 Hz sampling rate of the biomarkers.

This set of five features per time step formed the sequential portion of the input matrix, capturing the moment-to-moment physiological and motion response of the participant during the activity protocol.

To capture temporal patterns in physiological response, the sequential data was segmented into sliding windows of 30 frames, corresponding to 60 s of continuous activity. This window size was selected to reflect a typical adaptation timescale of HR and SpO_2 to physical effort. A stride of 8 frames was used between windows, enabling significant overlap and increasing the number of training samples. Additionally, the segmentation process was repeated using three different offsets (0, 4, and 8 frames), producing multiple views of the same signal and improving robustness against alignment variability.

In parallel with the sequential input, each window was enriched with a vector of contextual data and participant-specific variables, appended once per window:

- Demographic data: age, sex, height, weight, BMI, and body fat percentage;
- Baseline values: resting HR and SpO_2, computed from the initial resting segment of the protocol;
- Activity context: current activity label, previous activity label, duration (in seconds) of the current activity up to that point, and MET value representing the estimated physical intensity.

These metadata features provided the model with static context and physiological background for each subject, enabling more personalized and accurate predictions. The resulting multimodal input is illustrated in Figure 6, which shows the segmentation of the sequential features and the alignment of the contextual data with each window. The final input to the model consisted of a pair, i.e., a (30, 5) matrix representing the sequential features and a single 12-dimensional context data vector per window.

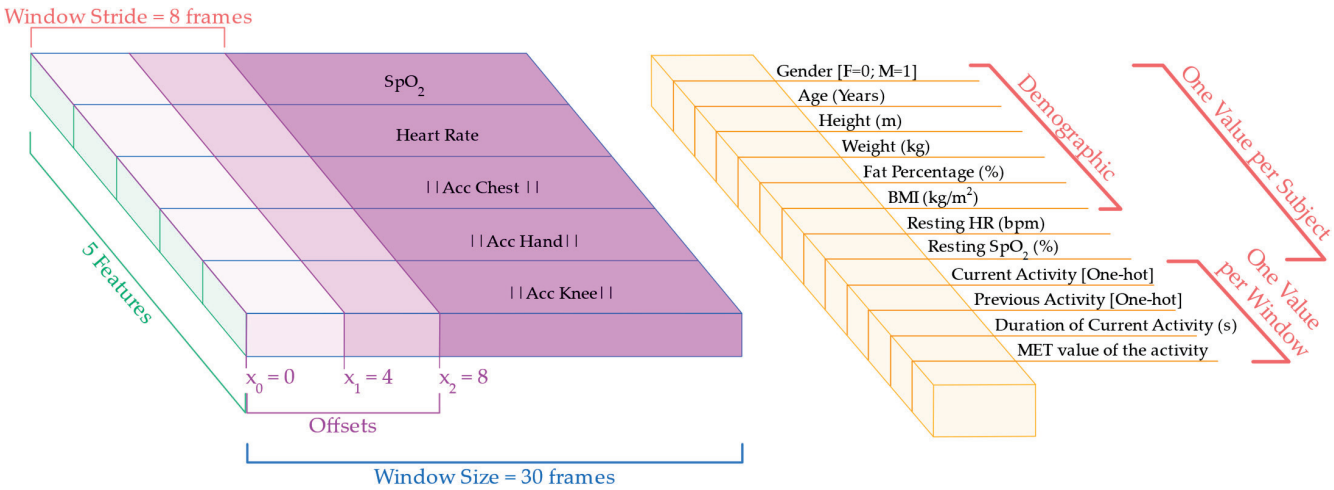


Figure 6. Structure of input data used for the heart rate regression model.

2.4. Model Architecture

2.4.1. Human Activity Recognition Classification Model

To classify physical activities from wearable motion data, we implemented a deep learning model that combines a one-dimensional CNN with an LSTM layer. The model

receives as input a time-series tensor of shape (batch size, time steps, features), where each frame contains multichannel IMU features.

Figure 7 illustrates the architecture of the model; it begins with a 1D convolutional layer that extracts short-term temporal patterns across input features using a kernel size of 3. This is followed by batch normalization and a ReLU activation function to stabilize training and introduce non-linearity. The resulting feature maps are then permuted and passed into an LSTM layer, which captures long-range temporal dependencies across the windowed sequence. A final fully connected layer maps the LSTM outputs to per-time-step activity class probabilities.

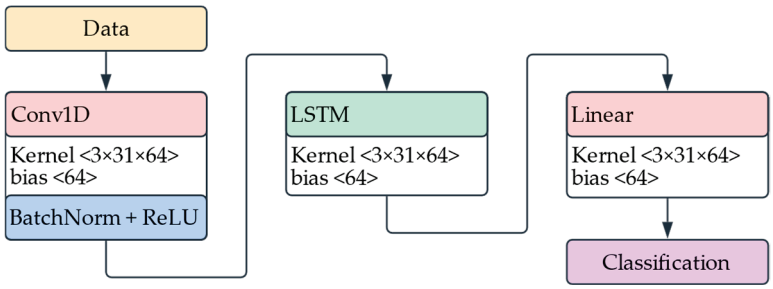


Figure 7. Architecture of the HAR model.

Hyperparameter tuning was conducted using Optuna [21], an open-source framework for automatic hyperparameter optimization based on Bayesian search strategies. Key parameters, including CNN output channels, LSTM hidden size, learning rate, and batch size were explored across predefined ranges. Trials were evaluated on validation accuracy after a few training epochs to identify optimal configurations.

2.4.2. Heart Rate Regression Model

To estimate the heart rate a participant would reach after completing a standardized step test, we developed a deep learning model that integrates sequential time-series data with contextual metadata. The model aims to learn cardiorespiratory responses to daily activities using fused physiological and biomechanical signals.

The first branch of the model processes sequential data: five time-series channels sampled at 0.5 Hz, including HR, SpO_2, and acceleration magnitudes from the chest, right hand, and left knee. These signals are passed through two stacked 1D convolutional layers with batch normalization, ReLU activation, max pooling, and dropout. The resulting temporal features are encoded by an LSTM layer into a fixed-length hidden representation.

In parallel, the context data branch encodes a 12-dimensional vector consisting of participant-specific features (age, sex, height, weight, BMI, body fat %), resting physiological baselines (HR, SpO_2), and contextual activity features (current/previous activity, duration, MET). This vector is processed through a fully connected layer with ReLU and dropout to produce a compact contextual embedding.

The two branches are concatenated and passed to a fusion module, a two-layer Multilayer Perceptron (MLP) that outputs a scalar prediction of the post-exercise heart rate. Figure 8 gives a visual representation of the architecture description.

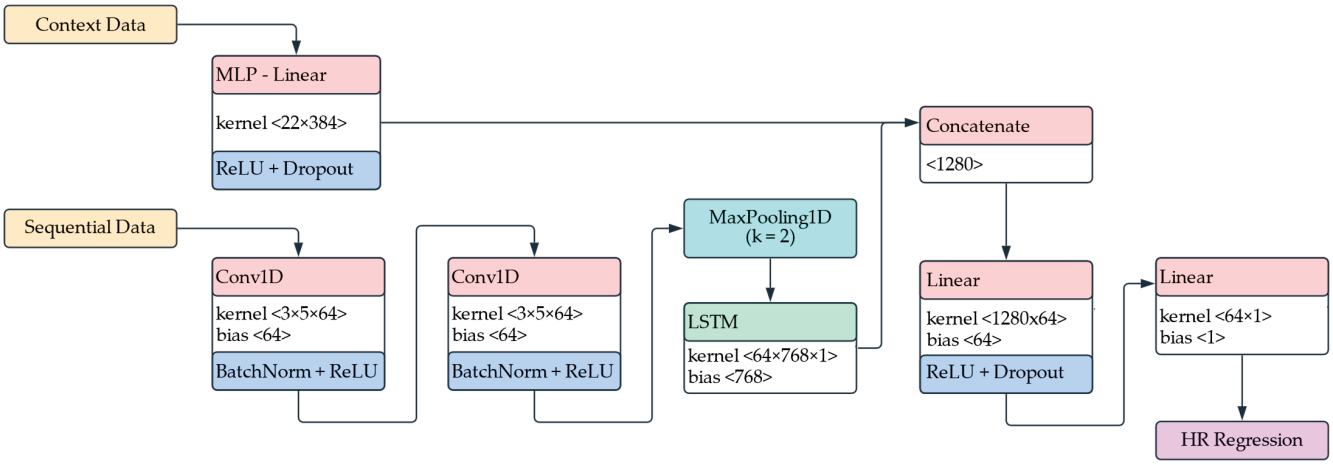


Figure 8. Architecture of the HR regression model.

3. Results

This section presents the performance outcomes of the two proposed models: the HAR classifier and the post step test HR regression model. Evaluation focused on their accuracy, generalizability across participants, and ability to capture both movement patterns and physiological responses. Metrics such as accuracy, F1-score, RMSE, and NRMSE were used to assess model effectiveness under real-world conditions.

3.1. Human Activity Recognition Classification Model's Performance

The final HAR model was evaluated on a held-out test set using the hyperparameters selected through Optuna tuning. The model showed strong performance in recognizing physical activities from wearable IMU data, with a weighted accuracy of 96.93% across all activity types.

To further assess model performance, we calculated precision, recall, and F1-score using a weighted average to account for potential class imbalance. All three metrics consistently exceeded 95%, suggesting both high sensitivity and precision in detecting a variety of everyday tasks. Figure 9 presents the confusion matrix of the HAR model, summarizing its classification performance across all six activity types.

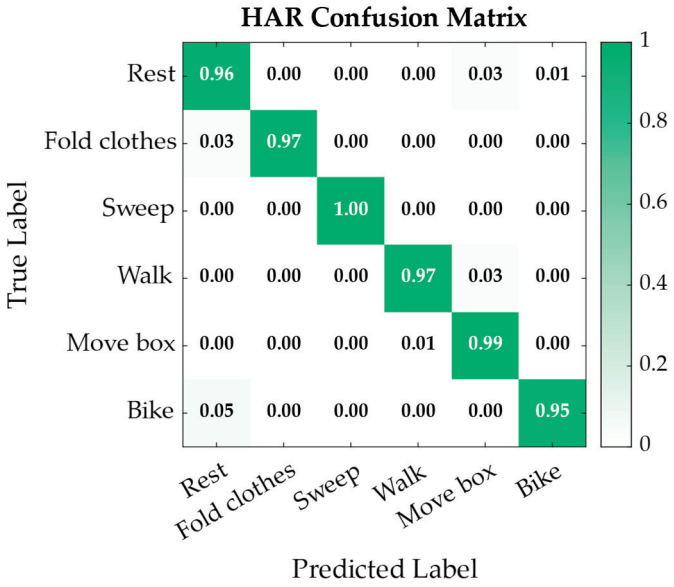


Figure 9. Confusion matrix for the HAR model predictions on the test set.

The matrix reveals high accuracy overall, with most predictions correctly aligned along the diagonal. This indicates that the model was highly effective at distinguishing between different physical activities based on wearable motion data.

While most samples were correctly classified, some minor misclassifications were observed. Interestingly, the greatest confusion occurred between the bike and rest activities, two tasks with clearly distinct physical intensities. This misclassification may be attributed to the similarity in upper body movement, as the hands remain relatively still in both activities. Additionally, the transition period at the beginning of the biking segment (when participants were seated and adjusting their position before pedaling) may have contributed to the overlap in motion patterns captured by the sensors. Despite these challenges, all activity classes maintained high recognition rates, indicating that the model generalizes effectively across the diverse set of movements in the protocol.

3.2. Heart Rate Regression Model's Performance

The heart rate regression model was evaluated using Leave-One-Subject-Out (LOSO) [22] cross-validation, ensuring generalizability by testing on each participant while training on all others. This approach was repeated for all 60 participants, ensuring robust and personalized performance evaluation across the population.

To identify the optimal model architecture, a grid search was performed over key hyperparameters: the number of output channels in the convolutional layers, the hidden size of the LSTM, and the dimensionality of the context branch. The context branch is a Multi-Layer Perceptron (MLP) responsible for processing non-sequential metadata, including demographic variables, resting biomarkers, and activity context. These features are projected into a compact representation and fused with the output of the sequential layers. Figure 10 displays the RMSE results across hyperparameter combinations: panel (a) shows results for models with 32 convolutional channels, and panel (b) for 64. Each line corresponds to a different number of neurons in the context MLP layer, while the x-axis indicates the LSTM hidden layer size.

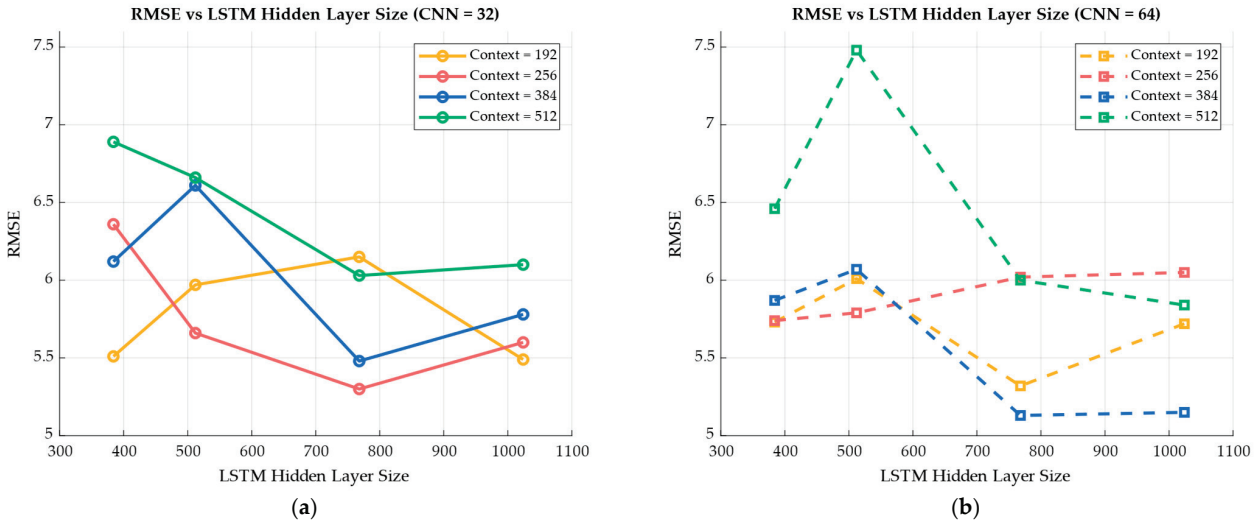


Figure 10. Grid search results show RMSE performance across LSTM, and context hidden sizes. (**a**) Results for CNN hidden size 32 and (**b**) Results for CNN hidden size 64. Each line represents a different context MLP layer size, illustrating how LSTM hidden layer size affects the results.

While Figure 10 highlights the best-performing configurations for clarity, the full hyperparameter search included additional LSTM sizes (up to 1024), as well as variations in dropout rates and learning rates. These extended configurations revealed that model

performance generally saturates at an LSTM hidden size of 768, with no consistent improvements beyond that point. In some cases, larger configurations led to slightly higher RMSE or training instability, particularly when the LSTM and context layers were both large. A consistent trend emerges across configurations: models perform better when the LSTM hidden layer size exceeds the context MLP layer size. This suggests that allocating more capacity to temporal modeling of physiological and motion signals is critical for accurate heart rate prediction, while a smaller context representation is sufficient to encode static metadata.

The best results were obtained with a model that used 64 convolutional output channels, an LSTM hidden layer size of 768, and 384 units in the context MLP branch. This configuration achieved an average RMSE of 5.13 beats per minute (bpm), a MAE of 4.37 bpm, and a NRMSE of 0.174. These results indicate that the model can accurately estimate the heart rate a participant would reach after performing the Queens College Step Test, based solely on their physiological and motion responses to free-living activities.

To assess the agreement between predicted and actual values, a Bland–Altman analysis [23] was performed. Figure 11 shows the mean bias between predicted and ground truth post-test heart rate values, which was just 0.13 bpm. Most differences fell within the 95% limits of agreement, suggesting strong alignment and minimal systematic bias in the model's predictions.

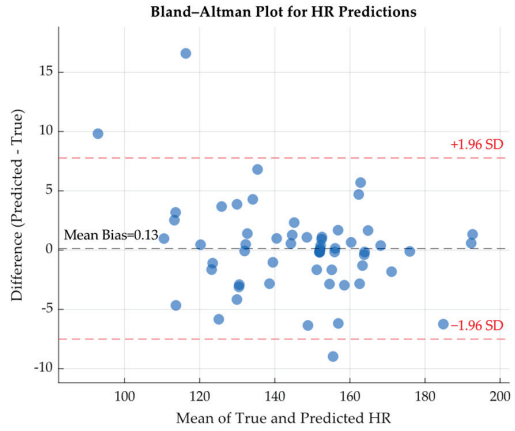


Figure 11. Bland–Altman plot comparing predicted and true HR values. Each point represents a participant's prediction. The solid black line shows the mean bias (0.13 bpm), while the dashed red lines indicate the 95% limits of agreement (±1.96 SD). Color intensity has no additional meaning, and all points reflect individual HR predictions.

Finally, model performance was also analyzed by activity type. As shown in Figure 12, the NRMSE was computed separately for each task.

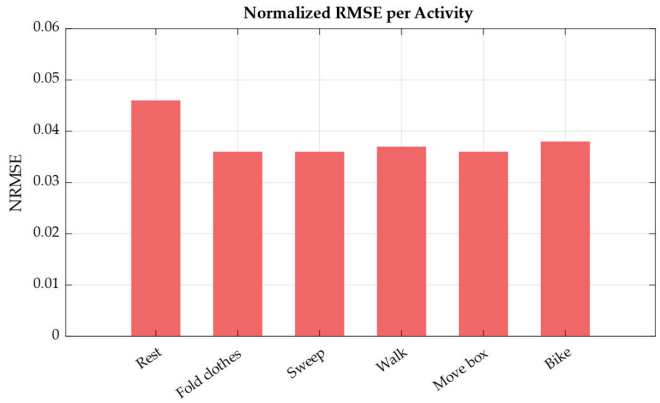


Figure 12. Normalized root mean squared error (NRMSE) of heart rate prediction across activity types.

Activities such as sweeping, walking, and moving a box led to the lowest prediction errors, with NRMSE values around 0.036–0.037. These activities typically elicit consistent physiological responses, making it easier to model across participants. On the other hand, resting periods showed the highest error (NRMSE = 0.046), likely due to higher inter-individual variability in baseline heart rate. These results reinforce the model's ability to learn exertion-driven trends from natural activity, while also highlighting the challenges of modeling personalized rest-state physiology.

4. Discussion and Future Work

This study presents a framework for estimating an individual's post step test heart rate using wearable data collected during natural daily activities. By integrating movement and physiological signals from wearable sensors, the proposed model enables an indirect estimation of VO_2max, the gold-standard indicator of CRF, without requiring structured exercise protocols. Unlike traditional assessments involving treadmills or cycle ergometers under supervision, our method allows users to be passively monitored through everyday movement.

The results demonstrate that this indirect VO_2max estimation is feasible, with the model achieving an average error of just 5.13 bpm across participants. The HAR model further reinforces this framework by providing reliable activity context information, with classification metrics consistently exceeding 96%. These components support the feasibility of context-aware, at-home monitoring of cardiorespiratory health, opening the door to more accessible and user-friendly fitness evaluation systems beyond clinical environments.

A key advantage of this approach lies in its efficiency. Unlike many existing methods in the literature, which rely on long-term data aggregation, ranging from hours to entire weeks of wearable recordings, our framework produces a fitness estimate from just a few minutes of daily activity. This substantial reduction in data requirements not only lowers the burden on users but also makes the method more compatible with opportunistic and on-demand assessment. It eliminates the need for continuous tracking or strict protocol adherence, offering greater flexibility for real-world deployment and broader population use.

Our approach emphasizes generalizability, as demonstrated through LOSO cross-validation, which confirmed the model's robustness across unseen individuals despite the moderate dataset size. While the activity protocol was structured, it was carefully designed to reflect free-living conditions by incorporating a variety of realistic daily tasks rather than isolated or treadmill-based exercises. Notably, the HAR model sustained high classification performance despite inter-individual differences in how movements were executed, suggesting that it reliably captures activity-specific patterns across diverse users. This design helps bridge the gap between controlled clinical protocols and real-world, unsupervised monitoring.

While the dataset used in this study includes participants with a range of ages and body compositions, it primarily represents individuals in average health. For safety reasons defined in the consent protocol, participants with known cardiovascular, pulmonary, or frailty-related conditions were excluded. As a result, the model has not yet been validated in clinical or high-risk populations, and its current findings should be interpreted within this context.

This model serves as a first step toward broader applications. Looking ahead, several improvements can be made to strengthen this framework. Future work should focus on expanding the dataset with a greater number of participants across a broader range of health conditions, ages, and fitness levels. Clinical validation in controlled environments (particularly with high-risk populations) will be essential to ensure safety and extend applicability to those who would benefit most from fitness monitoring. In these settings,

future studies should compare predicted values against gold-standard VO_2max measurements (e.g., via metabolic gas analysis) to formally establish the clinical accuracy of the proposed framework. Integrating additional biomarkers, such as respiration rate, ECG, or skin temperature, could enhance the model's understanding of internal physiological states, and it is also recommended to explore higher sampling rates for these signals to improve temporal resolution and capture finer physiological variations. In addition, analyzing performance across demographic groups could help ensure fairness and generalizability.

Validation under real-world free-living conditions is necessary to evaluate the model's robustness in practical, uncontrolled environments. Introducing variable intensity levels within the same activity class would support the development of more nuanced models capable of distinguishing effort gradients. Moreover, rather than relying on predefined activity classes, future models could benefit from regressing MET values directly, providing a continuous and more generalized representation of physical effort.

Finally, a key long-term goal is to move beyond offline estimation toward real-time feedback systems. With further refinement, this approach could power wearable-based fitness and recovery trackers that provide personalized health insights on the go, enabling early detection of decline, optimizing rehabilitation, and empowering users to engage in proactive health management from the comfort of their home. Future work should also consider optimizing model size and efficiency for on-device deployment, minimizing impact on battery life and processing.

5. Conclusions

This work presents a novel, data-driven approach to estimating cardiorespiratory fitness using wearable sensors in natural, unsupervised settings. By combining a deep learning-based HAR model with an HR regression model that integrates motion, physiological biomarkers, and contextual metadata, we demonstrate the feasibility of predicting the heart rate a person would reach after a standardized step test, without requiring them to perform the test. This predicted heart rate can then be used to estimate a cardiorespiratory fitness indicator.

A key strength of this framework lies in its ability to operate with just a few minutes of daily activity data, in contrast to many existing methods that require prolonged monitoring periods. This enables more opportunistic, user-friendly assessments compatible with real-world deployment. The model also showed strong generalizability across participants, achieving high accuracy in both activity classification and heart rate prediction, as demonstrated through LOSO cross-validation.

The ability to infer a CRF indicator from free-living activity data, without requiring treadmills, structured protocols, invasive equipment, or clinical supervision, represents a significant breakthrough. It bridges the gap between controlled laboratory assessments and real-world conditions, paving the way for more accessible, scalable, and personalized tools for cardiorespiratory fitness monitoring beyond clinical settings.

Author Contributions: Conceptualization, L.S.-A., J.L.R.-C., K.N.-T. and J.F.P.-M.; methodology, L.S.-A., J.L.R.-C. and J.F.P.-M.; software, L.S.-A.; validation, L.S.-A. and J.L.R.-C.; formal analysis, L.S.-A. and J.F.P.-M.; investigation, L.S.-A., J.L.R.-C. and J.F.P.-M.; resources, L.S.-A. and J.L.R.-C.; data curation, L.S.-A.; writing—original draft preparation, L.S.-A.; writing—review and editing, L.S.-A., J.L.R.-C., K.N.-T. and J.F.P.-M.; visualization, L.S.-A. and K.N.-T.; supervision, J.F.P.-M.; project administration, J.F.P.-M.; funding acquisition, J.F.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received financial support from the NSF EPSCoR Center for the Advancement of Wearable Technologies (CAWT) under Grant No. OIA-1849243 and the NSF CAREER Intelligent Biomarker Analysis based on Wearable Distributed Computing under Grant No. OAC-2439345.

Institutional Review Board Statement: The study was conducted under the Declaration of Helsinki and approved by the University of Puerto Rico's Institutional Review Board (IRB) (approved on 27 January 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.15857137.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HR	Heart Rate
HAR	Human Activity Recognition
RMSE	Root Mean Squared Error
bpm	beats per minute
MAE	Mean Absolute Value
CRF	Cardiorespiratory Fitness
VO_2max	Maximal Oxygen Uptake
CPET	Cardiopulmonary Exercise Test
ECG	Electrocardiogram
SpO_2	Blood Oxygen Saturation
PPG	Photoplethysmography
IMU	Inertial Measurement Unit
CNN	Convolutional Neural Network
LSTM	Long-Short Term-Memory
PAR-Q	Physical Activity Readiness Questionnaire
ANSI/CTA	American National Standards Institute/Consumer Technology Association
BMI	Body Mass Index
MET	Metabolic Equivalent of Task
LOSO	Leave-One-Subject-Out
MLP	Multi-Layer Perceptron

References

1. World Health Organization. Hypertension. Available online: https://www.who.int/news-room/fact-sheets/detail/hypertension (accessed on 30 July 2025).
2. International Diabetes Federation. IDF Diabetes Atlas, 10th ed.; 2021. Available online: https://diabetesatlas.org/atlas/tenth-edition/ (accessed on 30 July 2025).
3. Ross, R.; Blair, S.N.; Arena, R.; Church, T.S.; Despres, J.P.; Franklin, B.A.; Haskell, W.L.; Kaminsky, L.A.; Levine, B.D.; Lavie, C.J.; et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* **2016**, *134*, e653–e699. [CrossRef] [PubMed]
4. Rafiq, N.; Nabi, T.; Qayoom, O. Assessment of cardiovascular fitness [VO_{2max}] among medical students by Queens College step test. *Int. J. Biomed. Adv. Res.* **2015**, *6*, 418–421. [CrossRef]
5. Sheridan, D.; Jaspers, A.; Cuong, D.V.; De Beéck, T.O.; Moyna, N.M.; de Beukelaar, T.T.; Roantree, M.; Naseer, N. Estimating oxygen uptake in simulated team sports using machine learning models and wearable sensor data: A pilot study. *PLoS ONE* **2025**, *20*, e0319760. [CrossRef] [PubMed]
6. Wiecha, S.; Kasiak, P.S.; Szwed, P.; Kowalski, T.; Cieśliński, I.; Postuła, M.; Klusiewicz, A. VO_{2max} prediction based on submaximal cardiorespiratory relationships and body composition in male runners and cyclists: A population study. *eLife* **2023**, *12*, e86291. [CrossRef] [PubMed]
7. Akay, F.; Abut, F. Machine learning and statistical methods for the prediction of maximal oxygen uptake: Recent advances. *Med. Devices Evid. Res.* **2015**, *8*, 369–379. [CrossRef] [PubMed]
8. Neshitov, A.; Tyapochkin, K.; Kovaleva, M.; Dreneva, A.; Surkova, E.; Smorodnikova, E.; Pravdin, P. Estimation of cardiorespiratory fitness using heart rate and step count data. *Sci. Rep.* **2023**, *13*, 15808. [CrossRef] [PubMed]

9. Spathis, D.; Perez-Pozuelo, I.; Gonzales, T.I.; Wu, Y.; Brage, S.; Wareham, N.; Mascolo, C. Longitudinal cardio-respiratory fitness prediction through wearables in free-living environments. *NPJ Digit. Med.* **2022**, *5*, 176. [CrossRef] [PubMed]
10. Hsiao, C.-T.; Tong, C.; Coté, G.L. Machine Learning-Based VO_2 Estimation Using a Wearable Multiwavelength Photoplethysmography Device. *Biosensors* **2025**, *15*, 208. [CrossRef] [PubMed]
11. Lee, C.; Park, T. Deep learning health space model for ordered responses. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 191. [CrossRef] [PubMed]
12. Kaseris, M.; Kostavelis, I.; Malassiotis, S. A Comprehensive Survey on Deep Learning Methods in Human Activity Recognition. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 842–876. [CrossRef]
13. Qureshi, T.S.; Shahid, M.H.; Farhan, A.A.; Alamri, S. A systematic literature review on human activity recognition using smart devices: Advances, challenges, and future directions. *Artif. Intell. Rev.* **2025**, *58*, 276. [CrossRef]
14. *ANSI/CTA-2065.1*; Physical Activity Monitoring for Heart Rate-Real-World Analysis. Consumer Technology Association: Arlington, VA, USA, 2023. Available online: https://www.cta.tech/standards/ansicta-20651/ (accessed on 30 July 2025).
15. *ANSI/CTA-2108-2022*; Framework for Validation of Health Technology-Derived Metrics Under Naturalistic or Unconstrained Conditions. Consumer Technology Association: Arlington, VA, USA, 2022. Available online: https://www.cta.tech/standards/ansicta-2108/ (accessed on 30 July 2025).
16. *ANSI/CTA-2074 R-2025*; Intensity Metrics: Physical Activity Monitoring. Consumer Technology Association: Arlington, VA, USA, 2025.
17. Herrmann, S.D.; Willis, E.A.; Ainsworth, B.E.; Barreira, T.V.; Hastert, M.; Kracht, C.L.; Schuna, J.M.; Cai, Z.; Quan, M.; Tudor-Locke, C.; et al. 2024 Adult Compendium of Physical Activities: A third update of the energy costs of human activities. *J. Sport Health Sci.* **2024**, *13*, 6–12. [CrossRef] [PubMed]
18. Zmitri, M.; Fourati, H.; Vuillerme, N. Human Activities and Postures Recognition: From Inertial Measurements to Quaternion-Based Approaches. *Sensors* **2019**, *19*, 4058. [CrossRef] [PubMed]
19. Saldaña-Aristizábal, L.; Niño-Tejada, K.; Rivas-Caicedo, J.L.; Patarroyo-Montenegro, J.F. Evaluating Quaternion-Based Representations for Human Activity Recognition Using Motion Capture. In Proceedings of the International Symposium on Intelligent Computing Netowrking (ISICN2025), San Juan, PR, USA, 17–19 March 2025.
20. Saldaña-Aristizabal, L.; Rivas-Caicedo, J.L.; Niño-Tejada, K.; Patarroyo-Montenegro, J.F. Daily Activities Wearable Dataset for Cardiorespiratory Fitness Estimation. 2025. Available online: https://zenodo.org/records/15857137 (accessed on 30 July 2025).
21. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19), Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [CrossRef]
22. Kunjan, S.; Grummett, T.S.; Pope, K.J.; Powers, D.M.W.; Fitzgibbon, S.P.; Bastiampillai, T.; Battersby, M.; Lewis, T.W. The Necessity of Leave One Subject Out (LOSO) Cross Validation for EEG Disease Diagnosis. In Proceedings of the International Conference on Brain Informatics, Virtual Event, 17–19 September 2021; pp. 558–567. [CrossRef]
23. Giavarina, D. Understanding Bland Altman analysis. *Biochem. Med.* **2015**, *25*, 141–151. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Upper Limb Joint Angle Estimation Using a Reduced Number of IMU Sensors and Recurrent Neural Networks

Kevin Niño-Tejada [1,*], Laura Saldaña-Aristizábal [1], Jhonathan L. Rivas-Caicedo [1] and Juan F. Patarroyo-Montenegro [2,*]

[1] Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA; laura.saldana@upr.edu (L.S.-A.); jhonathanleonardo.rivas@upr.edu (J.L.R.-C.)
[2] Department of Computer Science and Engineering, University of Puerto Rico, Mayaguez, PR 00680, USA
* Correspondence: kevin.nino@upr.edu (K.N.-T.); juan.patarroyo@upr.edu (J.F.P.-M.)

Abstract

Accurate estimation of upper-limb joint angles is essential in biomechanics, rehabilitation, and wearable robotics. While inertial measurement units (IMUs) offer portability and flexibility, systems requiring multiple inertial sensors can be intrusive and complex to deploy. In contrast, optical motion capture (MoCap) systems provide precise tracking but are constrained to controlled laboratory environments. This study presents a deep learning-based approach for estimating shoulder and elbow joint angles using only three IMU sensors positioned on the chest and both wrists, validated against reference angles obtained from a MoCap system. The input data includes Euler angles, accelerometer, and gyroscope data, synchronized and segmented into sliding windows. Two recurrent neural network architectures, Convolutional Neural Network with Long-short Term Memory (CNN-LSTM) and Bidirectional LSTM (BLSTM), were trained and evaluated using identical conditions. The CNN component enabled the LSTM to extract spatial features that enhance sequential pattern learning, improving angle reconstruction. Both models achieved accurate estimation performance: CNN-LSTM yielded lower Mean Absolute Error (MAE) in smooth trajectories, while BLSTM provided smoother predictions but underestimated some peak movements, especially in the primary axes of rotation. These findings support the development of scalable, deep learning-based wearable systems and contribute to future applications in clinical assessment, sports performance analysis, and human motion research.

Keywords: deep learning; human motion analysis; inertial measurement units; joint angle estimation

1. Introduction

Human motion analysis is central to various fields, including physical rehabilitation, human–robot interaction, biomechanical research, and health monitoring [1–4]. The ability to quantify movement patterns and joint kinematics provides essential information for designing adaptive systems, assessing recovery, and developing personalized interventions [5]. In particular, the estimation of upper limb joint angles is a key metric in evaluating motor function in clinical and real-world environments [6]. These measurements support the development of robotic assistance, prosthetic control, and rehabilitation tools that rely on biomechanical feedback to function effectively. At the same time, wearable technologies have enabled real-time tracking of physiological and kinematic variables in naturalistic

settings. Recent developments include smart textiles for respiration monitoring [7], open-source smartwatches for physical activity and heart rate tracking [8], and medical health patches that provide continuous cardiorespiratory data using dry electrodes [9], illustrating the growing potential of wearable systems in integrated health assessment.

Traditionally, optical motion capture systems (e.g., Vicon®, OptiTrack®) have been used as the gold standard for obtaining ground-truth joint angles due to their sub-millimeter spatial accuracy [10]. However, these systems require fixed infrastructures, extensive camera calibration, and controlled laboratory conditions, making them impractical for deployment in everyday environments. Their spatial requirements and sensitivity to marker occlusion further limit accessibility outside research facilities [11].

In response to these limitations, IMUs have gained popularity as wearable alternatives for motion capture [12]. These low-power sensors infer segment orientation and joint angles across diverse conditions. Despite their advantages in portability and ease of integration, conventional IMU-based approaches often rely on dense sensor configurations (typically 7–10 units), which can present several challenges. These include:

- Sensor-to-segment misalignment, which can introduce substantial errors in joint angle estimation. Fan et al. reported errors exceeding 24° in the knee when misalignment exceeded 15°, highlighting the sensitivity of such systems to mounting orientation errors [13].
- Inter-sensor drift during prolonged recordings, especially in dynamic movements such as running, where orientation accuracy deteriorates over time without correction mechanisms [14].
- Although IMUs are generally rated as comfortable by users, especially when embedded in clothing or mounted non-invasively, increasing the number of sensors can complicate setup and maintenance, affecting overall usability in natural environments.

These limitations motivate the development of minimal-sensor configurations, where a reduced number of strategically placed IMUs can still deliver high-accuracy joint angle estimates through robust modeling techniques [15].

1.1. Deep Learning and Reduced Sensor Approaches

The integration of deep learning into motion estimation enables direct mapping from raw IMU data to joint kinematics. Recurrent neural networks (RNNs), particularly LSTM and CNN-LSTM hybrids, capture temporal dependencies critical for dynamic joint angle estimation. These models compensate for sparse inputs by learning latent biomechanical relationships [16,17].

Recent studies confirm that minimal IMU configurations can achieve accuracy comparable to optical systems when combined with optimized deep learning architectures. For example:

- Kim et al. introduced the Activity-in-the-loop Kinematics Estimator (AIL-KE), an end-to-end CNN-LSTM model using only two IMUs (wrist and chest). By integrating activity classification into the pipeline, they achieved shoulder joint angle errors under 6.5°—a 17 % improvement over a baseline without behavioral context [18]. However, this approach is primarily aimed at analyzing activities with periodic behaviors.
- Alemayoh et al. showed that a single IMU placed on the shank or thigh can estimate lower-limb joint angles with a mean absolute error of 3.65° using a BLSTM, demonstrating feasibility for sagittal-plane kinematics during walking [15]. No upper limbs were analyzed in this study.
- Airaksinen et al. systematically assessed trade-offs in IMU configurations and concluded that the minimal configuration with acceptable classifier performance includes at least a combination of one upper and one lower extremity sensor [19].

1.2. Problem Statement: Toward Practical and Scalable Joint Angle Estimation

Despite the growing interest in IMU-based motion tracking, achieving accurate upper limb joint angle estimation under real-world constraints remains an open challenge. Optical motion capture systems offer exceptional precision but are limited by their dependence on controlled environments. In contrast, wearable IMU solutions enable free-living monitoring but often rely on dense sensor networks that compromise usability [20].

The need for scalable and user-friendly solutions calls for approaches that maintain high accuracy while minimizing hardware complexity. Reducing the number of IMUs requires robust neural learning frameworks capable of extracting informative representations from partial data streams. This trade-off between model complexity, number of sensors, and system deployability remains a challenge. Reducing sensor count reduces observability, but intelligent modeling strategies can compensate for this loss. Evaluating different neural architectures under constrained input conditions is essential to identify scalable and efficient solutions [21].

1.3. Proposed Framework

To address these challenges, this work proposes a deep learning framework that estimates upper limb joint angles using only three IMUs located at the chest and wrists. This configuration ensures biomechanical relevance while minimizing invasiveness. We assessed the effectiveness of two different model architectures: a CNN-LSTM, which uses convolutional layers to extract spatial features followed by temporal modeling, and a Bidirectional LSTM (BLSTM), which processes sequences in both forward and backward directions to enhance context awareness. An overview of the proposed sensor placement and model pipeline is illustrated in Figure 1.

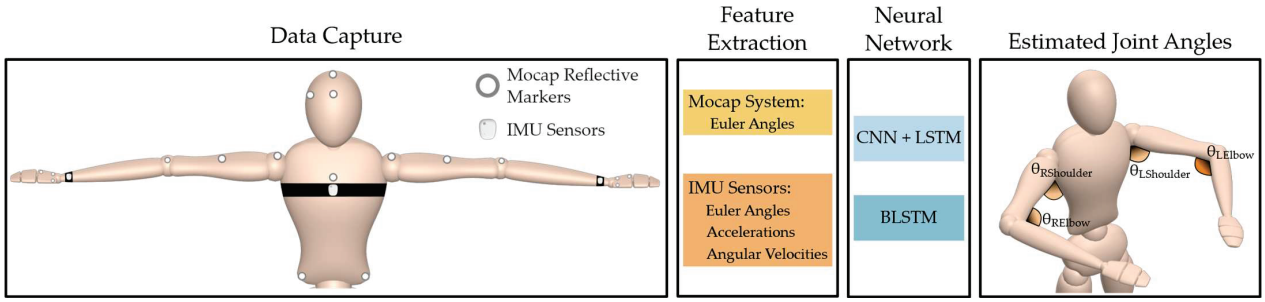


Figure 1. Proposed framework for upper limb joint angle estimation using a reduced number of IMU sensors. Both CNN + LSTM and BLSTM were compared using the same dataset.

The models were trained using IMU data synchronized with ground truth MoCap measurements, allowing the system to learn joint-specific motion patterns. By comparing both architectures, we evaluated the trade-offs between inference, accuracy, and temporal smoothness.

1.4. Contributions and Paper Organization

This work contributes to the ongoing effort to simplify joint angle estimation systems for practical deployment in health and rehabilitation contexts. We demonstrate that with a minimal sensor configuration and proper model selection, it is possible to achieve high-accuracy estimation of upper limb kinematics.

The rest of the paper is structured as follows: Section 2 describes the dataset acquisition, sensor placement, and preprocessing steps. Section 3 introduces the model architectures and training procedure. Section 4 presents quantitative results comparing CNN-LSTM and BLSTM under specific motion conditions. Section 5 discusses the implications of these

findings, highlights current limitations, and outlines future research directions toward scalable, real-world biomechanical monitoring systems.

2. Materials and Methods

2.1. Instrumentation

This study employed a dual-system configuration comprising an IMU-based wearable platform and an optical MoCap system to obtain synchronized kinematic datasets. The IMU system utilized MetaMotionS devices developed by Mbientlab Inc. (San Jose, CA, USA), capable of capturing raw linear acceleration, angular velocity, and orientation in quaternion or Euler-angle format. These sensors operate over Bluetooth Low Energy (BLE) and provide onboard fusion capabilities for orientation estimation.

For ground-truth kinematics, an OptiTrack® motion capture system (OptiTrack, Corvallis, OR, USA) was used in conjunction with Motive software (version 3.3.0.1), which handled the digitization and recording of all OptiTrack® data. This system computes segmental joint angles based on 3D marker trajectories and predefined biomechanical models. All data streams were timestamped and synchronized to ensure alignment between IMU and MoCap recordings. Although MoCap systems are considered the gold standard, they are not free from limitations such as marker displacement and soft tissue artifacts. To address these, a previous study conducted by the authors evaluated the feasibility of emulating an IMU sensor using a MoCap system by performing specific controlled movements across and along each coordinate axis, as well as combined motions [22]. The results demonstrated high consistency between the motion capture output and expected IMU behavior, supporting the reliability of MoCap-based ground truth for this application.

2.2. Data Acquisition

A single 10 min recording session was performed by a single subject, involving continuous upper limb movements executed at a natural speed and full range of motion. Prior authorization from the Institutional Review Board (IRB) was obtained to perform this study. The shoulder joint was evaluated across three degrees of freedom: abduction/adduction, flexion/extension, and internal/external rotation. The elbow joint was evaluated only in flexion/extension. No segmentation or repetition scheme was enforced, ensuring a representative sequence of natural motion.

IMU data were collected at 50 Hz and included tri-axial acceleration, angular velocity, and quaternion orientation. This sampling rate was selected based on prior studies indicating that 50 Hz is sufficient to capture the natural frequency of human motion without aliasing, even during fast transitions [23,24]. All quaternion data were converted to Euler angles using a ZYX intrinsic rotation convention. This conversion was necessary due to the limitations of directly using quaternions in neural network training.

Euler angles offer scalar orientation values aligned with anatomical axes, making them more interpretable for both biomechanical analysis and neural network learning. Unlike quaternions, which are abstract and non-intuitive, Euler angles correspond directly to clinically meaningful joint motions such as flexion or rotation. This alignment facilitates better feature learning and promotes faster convergence during model training, as the network can associate specific signal patterns with anatomically relevant movements. The dataset is available in the following reference [25].

2.3. Experimental Design

All data collection was conducted by a single subject under consistent environmental conditions. More subjects will be included for future studies to account for other body sizes, limbs, and dynamics. The subject wore three IMUs placed on the chest and both

wrists, along with reflective markers positioned according to the standardized upper-limb marker set required by the Motive skeletal solver (See Figure 1). The OptiTrack® system consisted of 7 infrared cameras surrounding the capture volume to provide high-fidelity 3D tracking. Figure 2 illustrates the complete setup used during data collection.



Figure 2. Overview of the real-world setup used for experiments.

At the beginning of each recording session, the subject adopted a T-pose posture, with both arms extended horizontally and palms facing downward. This standardized position served as the initial reference for all subsequent movements and was used to align the coordinate systems of the IMU and MoCap data. Starting from a well-defined T-pose ensured consistent orientation across both systems and facilitated the synchronization of reference frames during post-processing.

Following this posture, the subject performed a series of continuous upper limb motions at a natural speed and amplitude. These movements included:

- A shoulder rotation around the *Z*-axis (Yaw), shown in Figure 3.

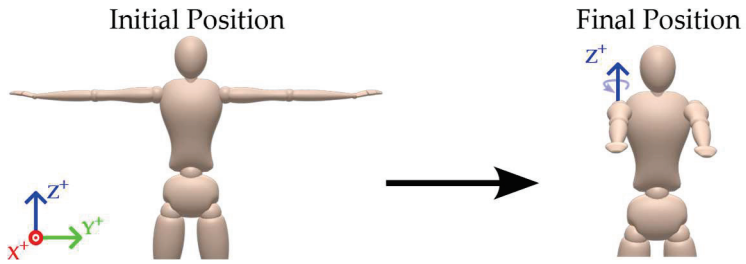


Figure 3. Illustration of the shoulder yaw (*Z*-axis) movement performed during validation.

- A shoulder flexion/extension around the *X*-axis (Roll), shown in Figure 4.

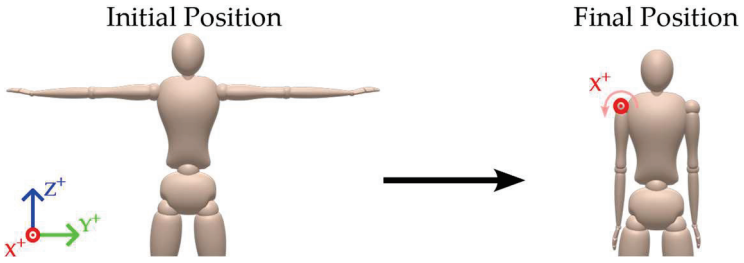


Figure 4. Illustration of the shoulder roll (*X*-axis) movement performed during validation.

- An elbow flexion/extension primarily around the *Z*-axis (Yaw), shown in Figure 5.

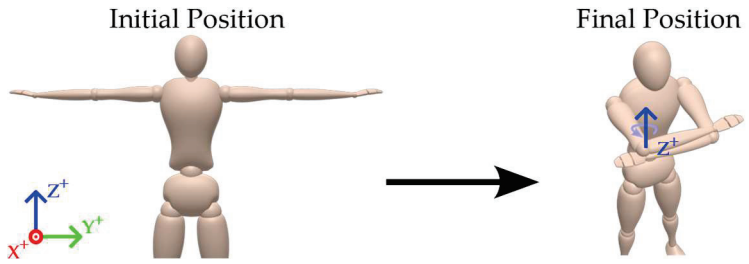


Figure 5. Illustration of the elbow yaw (*Z*-axis) movement performed during validation.

The sequence was designed to reflect a broad range of typical daily activities, ensuring sufficient variability across all rotational degrees of freedom to support robust model training.

Ground-truth joint angles were obtained from the OptiTrack® motion capture system using Motive's embedded biomechanical solver. This software computed shoulder and elbow joint angles based on the 3D trajectories of the reflective markers. The resulting time series of joint angles served as the reference targets for training and evaluating the proposed learning models.

2.4. Data Preprocessing

The raw data acquired from the wearable sensors consisted of acceleration, angular velocity, and quaternion signals from each IMU. Static angles were initially captured in quaternion format because the sensor fusion algorithm implemented in the Mbientlab MetaMotionRL devices (version r0.5) with firmware-enabled 9-axis sensor fusion, yields more accurate orientation estimates when the data is exported directly in this format. In contrast, exporting orientation directly as Euler angles from the device introduced noticeable inaccuracies. Therefore, the data was first collected in quaternion form and subsequently converted into Euler angles for model training. Preliminary experiments using quaternions as training targets yielded poor results, since the networks struggled to preserve the unit norm required for valid quaternions, and their abstract 4D representation limited the model's ability to learn meaningful geometric relationships. Euler angles provided a more interpretable and effective representation for the estimation task.

To prepare this data for use in neural network models, a structured and multi-stage preprocessing pipeline was developed, encompassing data synchronization, trimming, subsampling, segmentation into temporal windows, and normalization. This pipeline is illustrated in Figure 6.

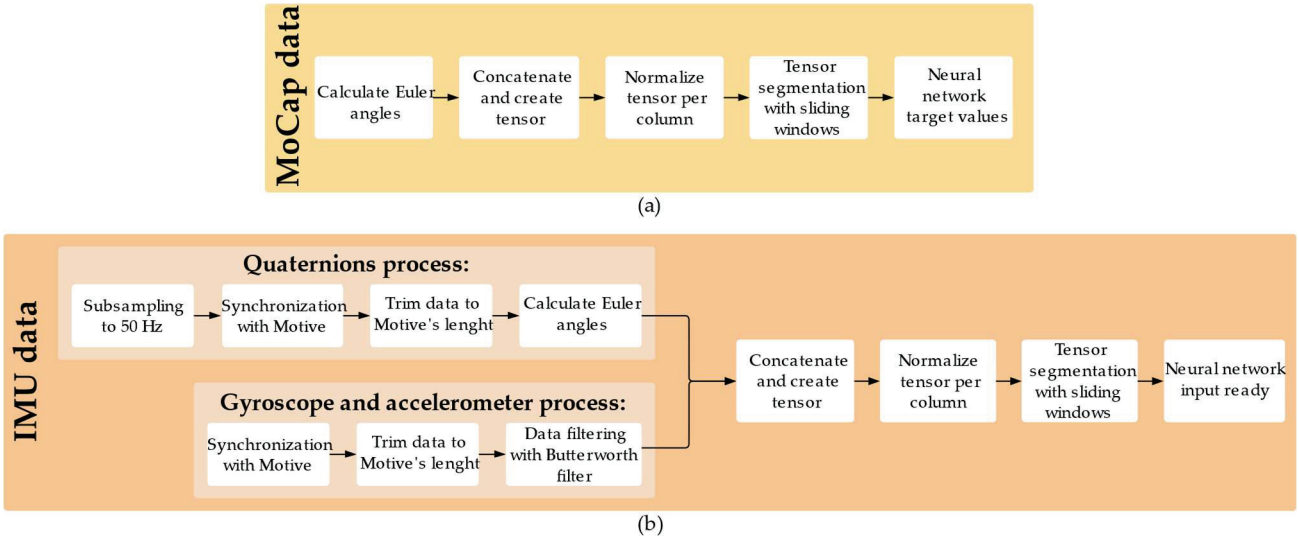


Figure 6. Data preprocessing pipeline for the proposed framework. (**a**) Describes the processing of MoCap data used to generate the target joint angles for the neural networks. (**b**) Describes the processing of IMU signals.

2.4.1. Type of Measured Data

Each IMU sensor recorded triaxial accelerometer data (g), gyroscope data (°/s), and quaternion orientation data (W, X, Y, Z). A separate CSV file was generated for each signal type and each sensor.

2.4.2. Synchronization and Trimming

Data acquisition was initiated first with the three IMU sensors. After a 5 s delay to ensure all sensors were actively recording, the MoCap system was started. This approach ensured that the MoCap system provided a consistent and precise timestamp reference to align all IMU data streams. During preprocessing, the IMU files were temporally synchronized by aligning their internal timestamps to the MoCap start time. Once synchronization was achieved, all data streams were trimmed to match the duration recorded by the MoCap system. Since the OptiTrack® Motive software allows configuration of a fixed trial duration, this ensured all signals were aligned and had identical length. Notably, IMU recorded a few extra seconds at the end due to delayed manual stopping; trimming ensured these were discarded, resulting in consistent frame alignment across all devices.

2.4.3. Subsampling

Mbientlab sensors record quaternions at a fixed sampling rate of 100 Hz, while accelerometer and gyroscope data were configured to operate at 50 Hz. Additionally, the OptiTrack® motion capture system provided ground-truth joint angles sampled at 50 Hz. To ensure temporal consistency across all input and target signals, quaternion data were subsampled to 50 Hz by retaining every other sample.

2.4.4. Filtering

A low-pass Butterworth filter (fifth-order, cutoff frequency = 5 Hz) was applied to the accelerometer and gyroscope data to attenuate high-frequency noise. This cutoff frequency was selected to preserve the dominant frequencies associated with voluntary human limb motion, which typically fall below 6 Hz [26]. In biomechanics, low-pass filtering is commonly applied to both kinematic and kinetic data to improve signal quality and reduce noise. Common cutoff frequencies range from 3 to 10 Hz [27], with 6 Hz often used as a representative value, as seen in tools like OpenSim [27,28]. The filter was

implemented using the filtfilt function to achieve zero-phase distortion. Quaternion data were not filtered.

2.4.5. Tensor Preparation and Normalization

The recordings were loaded and concatenated into a unified dataset. The input features included yaw, pitch, and roll (transformed Euler angles from quaternions), angular velocities (ωx, ωy, ωz), and linear accelerations (ax, ay, az) from each IMU sensor, resulting 27 columns across the chest, right wrist, and left wrist sensors. The output (target) variables were 10 joint angles derived from the MoCap system: 3 degrees of freedom (DOF) at each shoulder and 2 DOF at each elbow.

Min–max normalization was applied independently to the input and output variables using the MinMaxScaler from scikit-learn. The scaler was fitted on the training data and reused to normalize the validation data and later to inverse-transform the model outputs during evaluation.

2.4.6. Sliding Window Design

The normalized data were segmented into overlapping temporal windows of 50 frames (corresponding to 1 s of data at 50 Hz) using a stride of 5 frames as shown in Figure 7. This configuration results in a 90% overlap between consecutive windows, ensuring high temporal continuity across samples. The high overlap also helps preserve dynamic changes between windows, particularly important in short sequences where rapid joint movements occur. This decision was especially beneficial given the reduced number of sensors and the system's low observability, as it increased the number of training samples and improved the model's ability to capture sequential dependencies and postural transitions. While lower overlaps (e.g., 50% and 25%) were tested and yielded acceptable and poor results, respectively, 90% consistently led to better performance. Overall, this setting promoted generalization and mitigated overfitting by exposing the model to more nuanced motion patterns.

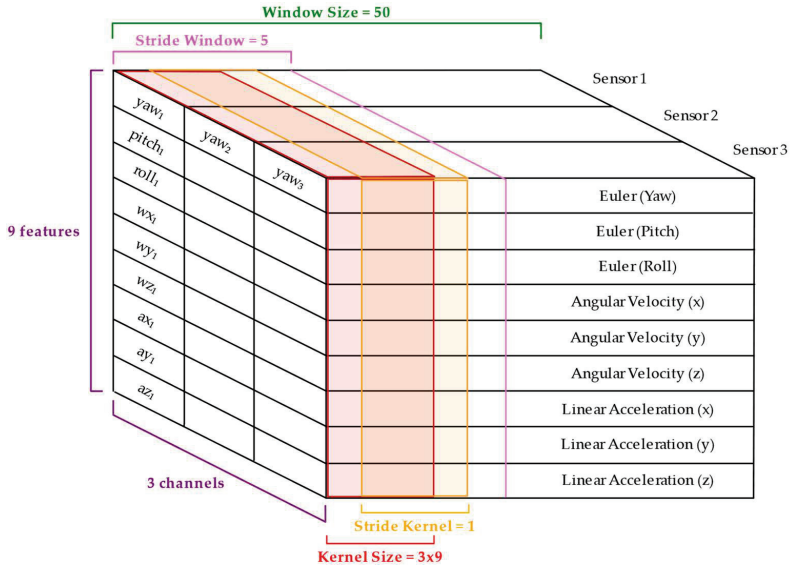


Figure 7. Input tensor representation used in the CNN-LSTM architecture. The input consists of a window of 50 and 9 features per sensor, including Euler angles, angular velocities, and linear accelerations.

Following the windowing step, the output tensors were trimmed to compensate for the shift introduced by the convolutional operations within the CNN-LSTM architecture.

This alignment guaranteed consistency between the sequence length of the inputs and the targets. A detailed explanation of this architectural adjustment is presented in Section 2.5.

2.4.7. Data Split

After preprocessing and tensor construction, the data was divided into training and validation sets. Five of the six available recordings were used for training, while the other one was exclusively reserved for validation. This take-independent split ensured that the validation set contained motion patterns not directly seen during training, allowing for a more realistic assessment of the model's generalization ability. The resulting split produced approximately 30,000 training frames and 6000 validation frames after windowing.

With the dataset structured for training and validation, the next step involved designing and evaluating neural network architectures capable of learning the complex relationship between the multivariate inertial inputs and the corresponding joint angles. Two deep learning models were assessed for this task: a CNN-LSTM network and a BLSTM network.

2.5. Comparative Analysis of Architectures

The CNN-LSTM and BLSTM architectures were assessed to estimate upper limb joint angles from inertial data; however, they differ in structure, computational complexity, and learning dynamics, which impact their respective advantages and limitations.

The CNN-LSTM model leverages convolutional layers to extract local spatiotemporal features from segmented IMU signals before passing them to the LSTM layers. This hierarchical structure enables effective encoding of short-term dynamics and reduces the dimensionality of the input prior to sequential modeling. Its use of 2D convolution allows for localized feature learning across both sensor modalities and time, promoting generalization while keeping the model relatively lightweight. Moreover, the presence of a CNN front-end introduces a degree of robustness to noise and spatiality, helping the model focus on motion-relevant features [29]. However, the choice of kernel size and stride introduces a shape mismatch between input and output sequences, requiring architectural adjustments and additional processing steps such as target cropping.

In contrast, the BLSTM model eliminates the need for convolutional preprocessing by directly operating on a flattened representation of the IMU signals. Its bidirectional nature allows it to access both past and future motion trends within the input window, enabling potentially more accurate estimation in temporally ambiguous situations [30]. This ability to integrate broader temporal context may improve performance in scenarios with delayed responses or nonlinear movement transitions. However, the absence of a CNN component also means that the model may be more sensitive to noise or irrelevant features, as it lacks an explicit mechanism for spatial feature extraction.

Overall, while the CNN-LSTM model offers a more compact and interpretable architecture that balances performance with computational efficiency, the BLSTM provides richer temporal modeling capabilities that may prove advantageous in tasks requiring higher sensitivity to motion context.

2.6. Neural Network Architectures

The selection of architectural parameters, such as the number of layers, kernel size, and hidden units, was selected based on related works and optimized through manual configuration adjustments validated experimentally. Several configurations were tested iteratively, adjusting hyperparameters such as filter size, dropout rate, and the number of recurrent layers until stable convergence and low validation error were achieved. These decisions were also informed by related works in joint angle estimation using IMUs. The

following subsections describe the final architectures and training strategies adopted for each model.

2.6.1. Convolutional-Long Short-Term Memory (CNN-LSTM) Architecture

The model input consisted of 50-frame temporal windows. For each sensor, nine features were included: three Euler angles, three angular velocities, and three linear accelerations, resulting in a total of 27 features per frame. Each input sample was thus represented as a tensor of shape (50, 27), which was expanded to (1, 50, 27) to comply with the input requirements of 2D convolutional layers. The overall structure of the CNN-LSTM architecture is shown in Figure 8.

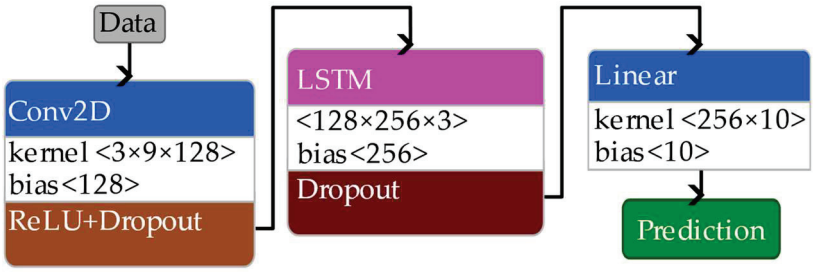


Figure 8. Architecture of CNN-LSTM regression model.

The first layer of the network was a 2D convolutional layer with 128 filters and a kernel size of (3, 10), applied without padding. This kernel spanned three timesteps and ten feature dimensions, enabling the model to extract spatiotemporal correlations across the time and all sensors. Due to the convolution operation, the time dimension was reduced from 50 to 48 frames. To maintain consistency between input and target sequences, the ground truth joint angle series was trimmed by two frames during preprocessing, as mentioned in the previous section.

Following the convolutional layer, a dropout layer with a rate of 0.3 was used to mitigate overfitting by randomly deactivating neurons during training. The output of the convolutional stage was reshaped from (batch, 64, 48, 1) to (batch, 48, 64), making it suitable for input to a recurrent layer.

To capture the temporal dynamics of the motion, the network employed a single LSTM layer with 256 hidden neurons. This layer processed the sequential features extracted by CNN and learned temporal dependencies across the 48-frame sequence. The output of the LSTM at each timestep was passed to a fully connected layer that mapped the 256-dimensional hidden state to 10 outputs per frame, representing the joint angles of interest. The target labels for the neural networks correspond to the joint angles at the final frame of each shifting window. This strategy is consistent with established practices in the literature, where the output label is aligned with the last timestep of the input sequence, ensuring temporal consistency between the observed inertial dynamics and the predicted joint angles [15].

The target angles correspond to the rotational components (Yaw, Pitch, and Roll) of the shoulder and (Yaw and Pitch) of the elbow, for both left and right arms. This results in a total of 10 predicted joint angles per window, representing the angular orientation of each joint in the anatomical frame with respect to the initial T-pose posture.

The network was trained for 150 epochs using the Adam optimizer with a learning rate of 1×10^{-5}. A batch size of 50 was used, matching the temporal length of each input window. The loss function employed was the Mean Absolute Error (MAE), selected for its interpretability in degrees and its robustness to outliers compared to squared-error-based metrics. Its exclusive use in this study ensured a simple and consistent evaluation

framework across models and joints, prioritizing clarity and interpretability over the inclusion of multiple statistical indicators.

To prevent overfitting, a dropout rate of 30% was applied between layers, and early stopping was implemented during training based on the validation loss. This training strategy enabled the model to gradually minimize the angular discrepancy between predicted and ground-truth joint angles across all degrees of freedom.

2.6.2. Bidirectional Long Short-Term Memory (BLSTM)

To assess the impact of temporal context on angle estimation, an alternative architecture was implemented using a BLSTM network. Unlike the LSTM, which only captures dependencies in the forward temporal direction, the BLSTM processes the input sequence in both forward and backward directions simultaneously. This dual traversal allows the network to access past and future context at each timestep, which can be particularly beneficial when the prediction depends on motion trends occurring before and after a given frame.

In contrast to the CNN-LSTM model, the BLSTM architecture does not include convolutional layers; thus, the input tensor is not four-dimensional. Instead, the IMU data are directly reshaped into a three-dimensional tensor of shape (batch, window size, features), where the features dimension results from concatenating all sensor channels (i.e., acceleration, angular velocity, and Euler angles) across all devices, yielding an input tensor of shape (batch, 50, 27). This flattened representation preserves the temporal structure of the signal while providing the network with simultaneous access to all sensor signals. The overall structure of the CNN-LSTM architecture is shown in Figure 9.

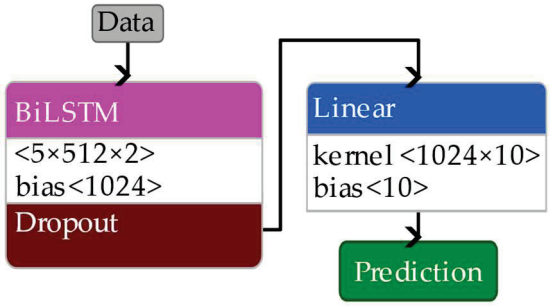


Figure 9. Architecture of the BLSTM regression model.

The BLSTM was composed of five stacked bidirectional LSTM layers, each with 512 hidden units per direction. Due to the bidirectional nature of architecture, the resulting hidden state dimension was 1024 (512 forward + 512 backward). The output of the final timestep was passed through a dropout layer ($p = 0.3$) and subsequently to a fully connected layer that mapped the 1024-dimensional hidden representation to the 10 target joint angles.

This model was trained under the same conditions as the CNN-LSTM architecture: 150 epochs, a batch size of 50, and the Adam optimizer with a learning rate of 1×10^{-5}. The Mean Absolute Error (MAE) was also used as the loss function to maintain consistency in evaluation. As with the CNN-LSTM model, a 30% dropout rate was introduced between layers, and early stopping was employed based on the validation loss to mitigate overfitting. This consistent training configuration facilitated a fair comparison between both architectures.

While the BLSTM adds complexity and computational cost compared to its unidirectional counterpart, its ability to incorporate future motion context into each prediction may enhance generalization in scenarios with temporally ambiguous or noisy input sequences.

This trade-off between temporal context and model efficiency is further explored in the discussion section through a comparative analysis of results.

3. Results

To evaluate the performance of both neural network architectures, the validation was carried out using a sequence not seen during training. As described in Section 2.3, this recording included three specific upper limb movements:

- A shoulder rotation around the *Z*-axis (Yaw).
- A shoulder flexion/extension around the *X*-axis (Roll).
- An elbow flexion/extension primarily in the *Z*-axis (Yaw).

3.1. CNN-LSTM

To evaluate the CNN-LSTM model, complete predictions for each of the three selected movements are shown alongside ground truth angles to illustrate the general behavior of the network.

Figure 10 shows the model's predictions for the three movements mentioned previously for the CNN-LSTM. The presented CNN-LSTM model demonstrates robust performance in predicting joint angles using only wrist and chest sensor data, achieving consistent prediction accuracy across multiple degrees of freedom. For both upper limbs, the model exhibits high precision in estimating planar movements, with shoulder Pitch errors as low as 3.67° (left) and 3.77° (right), and Roll errors measuring 5.08° (left) and 4.88° (right). These results indicate adequate capability in capturing fundamental movement patterns with satisfactory accuracy. Rotational (Yaw) predictions show slightly higher but still adequate performance, with shoulder errors of 8.58° (left) and 9.12° (right), and elbow errors of 9.90° (left) and 9.98° (right), this is considered satisfactory given the complexity of rotational kinematics and the minimal sensor configuration. The consistent performance across both limbs demonstrates the model's generalization capabilities. The close correspondence between predictions and ground truth throughout most movement phases highlights the effectiveness of the CNN-LSTM architecture in learning meaningful spatiotemporal relationships from limited distal sensor data. It is important to note the model's ability to accurately reconstruct proximal joint angles without direct measurements from shoulders or elbows, showcasing advanced pattern recognition capabilities.

3.2. BLSTM

The BLSTM model was evaluated using the same validation sequence as the CNN-LSTM, allowing direct comparison under identical conditions.

Figure 11 shows the model's predictions for the three movements mentioned previously for the BLSTM. The BLSTM network demonstrates effective performance in predicting joint angles from limited sensor data, with particularly strong results for certain movement types. For the right shoulder, the model achieves acceptable accuracy in Pitch estimation (MAE: 5.09°) and maintains good performance for Roll movements (MAE: 5.98°). The elbow Pitch predictions also show competitive results with an MAE of 6.82°. While the Yaw predictions for both shoulder (MAE: 17.07°) and elbow (MAE: 19.66°) exhibit higher error values, these results remain significant given the inherent complexity of rotational kinematics and the minimal sensor configuration. The consistent performance across different joint angles highlights the model's ability to capture fundamental movement patterns, with particularly accurate tracking of planar motions. The prediction curves closely follow the ground truth, highlighting the BLSTM's ability to capture temporal dependencies. Notably, it reconstructs proximal joint angles using only wrist and chest sensors, demonstrating strong pattern recognition despite the limited sensor setup.

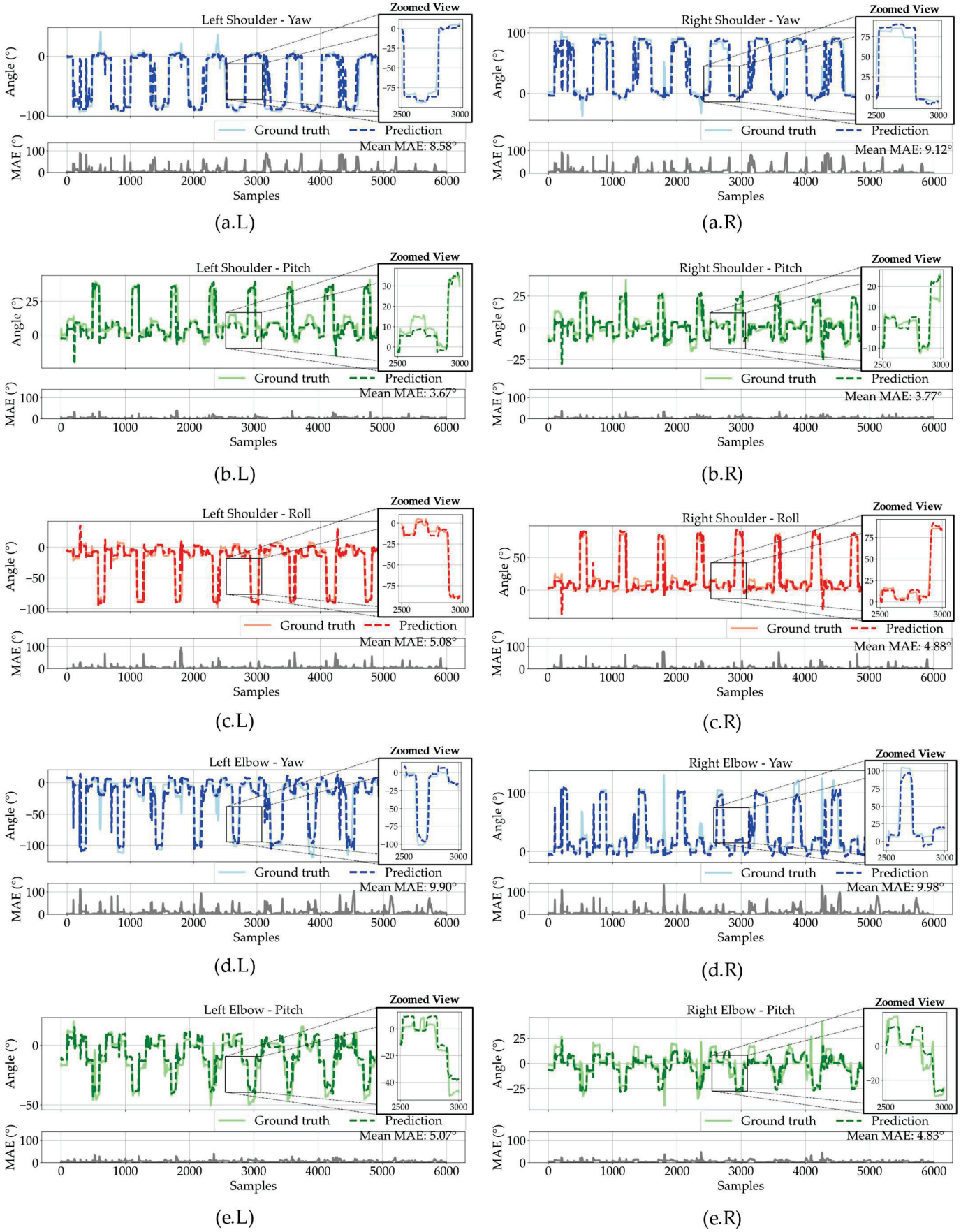


Figure 10. CNN-LSTM full prediction for the left (L) and right (R) arm. (**a**.R/L) Shoulder Yaw, (**b**.R/L) Shoulder Roll, (**c**.R/L) Shoulder Pitch, (**d**.R/L) Elbow Yaw, (**e**.R/L) Elbow Pitch. Each sub-figure shows a zoomed view of the side of the transition and the MAE at the corresponding bottom.

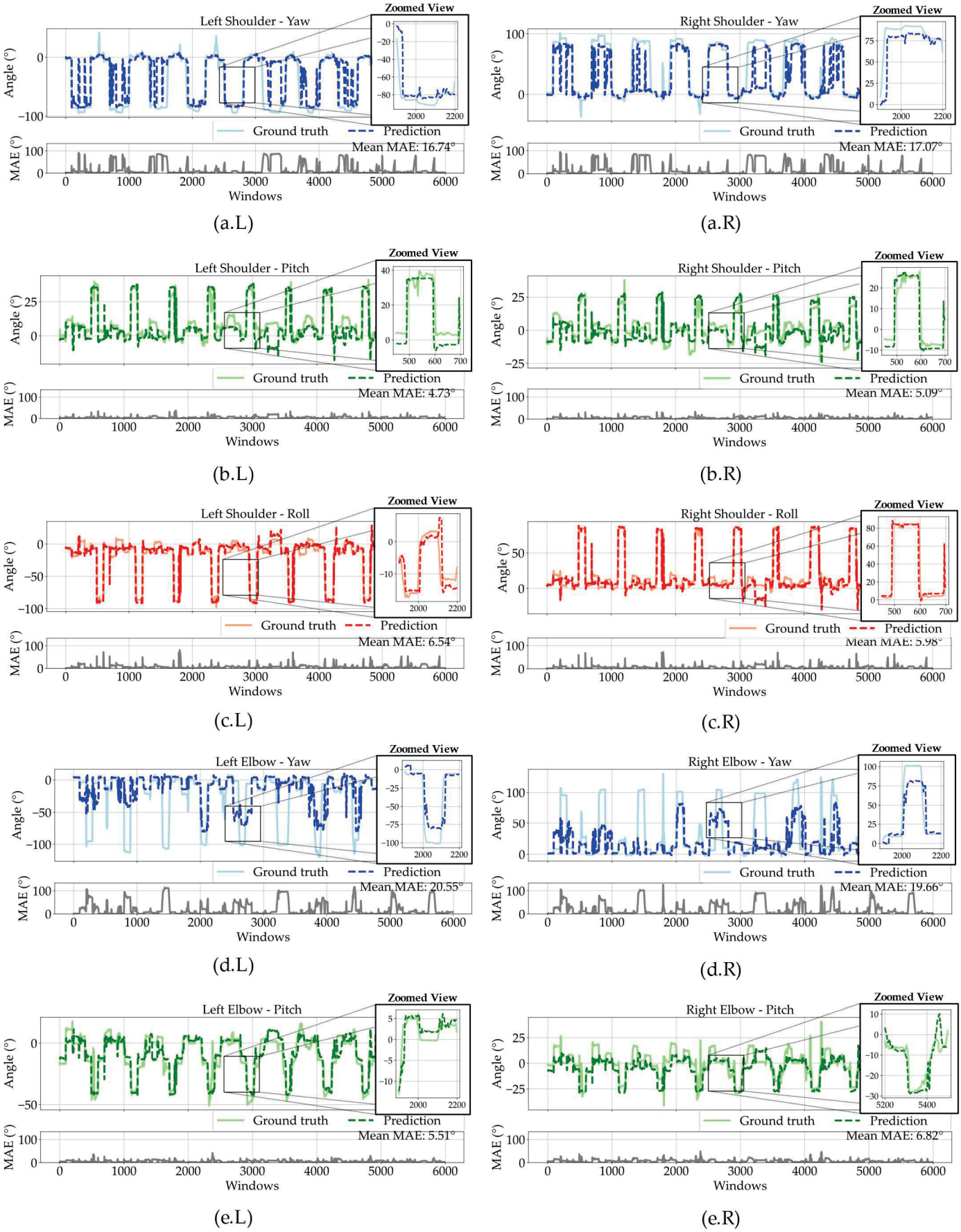


Figure 11. BLSTM full prediction for the left (L) and right (R) arm. (**a**.R/L) Shoulder Yaw, (**b**.R/L) Shoulder Roll, (**c**.R/L) Shoulder Pitch, (**d**.R/L) Elbow Yaw, (**e**.R/L) Elbow Pitch. Each sub-figure shows a zoomed view of the side of the transition and the MAE at the corresponding bottom.

The CNN-LSTM model demonstrated strong performance in planar motion prediction, achieving MAEs of 3.67–5.08° for shoulder Pitch/Roll and 4.83–5.07° for elbow Pitch, while Yaw predictions yielded higher errors (8.58–9.98°). The BLSTM network showed comparable accuracy in planar movements, with shoulder Pitch/Roll MAEs of 5.09–5.98° and elbow Pitch at 6.82°, but exhibited increased Yaw prediction errors (17.07–19.66°). Both models successfully reconstructed joint angles using only wrist and chest data, with the CNN-LSTM showing slightly better overall precision, particularly in rotational (Yaw) estimations. The BLSTM maintained robust temporal tracking but displayed marginally higher deviations in complex rotations, suggesting differences in handling kinematic dependencies between architectures.

3.3. Architectures Computational Performance

To evaluate and compare the computational performance of the tested architectures, key metrics such as the number of parameters, model size, average inference time per batch and per sample, and GPU memory usage were measured. Table 1 presents a side-by-side comparison between the CNN-LSTM and BLSTM models.

Table 1. Computational performance comparison between CNN-LSTM and BLSTM.

Metric	CNN-LSTM	BLSTM
Total Parameters	1,461,002	27,424,778
Model Size (MB)	5.58	104.62
Avg. Inference time per batch (ms)	15.894	43.456
Avg. Inference time per sample (ms)	0.318	0.869
Max GPU memory used (MB)	166.21	1122.33

Both models were evaluated under the same conditions using a batch size of 50 and a window stride of 5 (90% overlap). The results show that the CNN-LSTM model is more computationally efficient than the BLSTM across all measured aspects. It required fewer parameters, occupied less storage, had faster inference times, and consumed less memory during execution.

4. Discussion and Future Work

The results from both the CNN-LSTM and BLSTM architectures demonstrate promising capabilities for joint angle prediction using minimal sensor configurations, with each showing distinct strengths. The CNN-LSTM achieved superior performance in planar movements (Pitch/Roll MAEs: 3.67–5.08°) and notably lower Yaw errors (8.58–9.98°) compared to the BLSTM (Pitch/Roll MAEs: 5.09–5.98°; Yaw: 17.07–19.66°). However, these higher Yaw errors can be attributed to the nature of sensor placement and the reduced observability of axial rotations in this configuration, which limits the detectable range and angular variation in the Yaw axis. Additionally, abrupt motion transitions involving rapid changes in velocity or acceleration were observed to be more error-prone, as such dynamic segments are more difficult for the models to capture, especially under reduced sensor input. Although no explicit segmentation by motion phase was performed, qualitative inspection of the predicted trajectories supports this observation.

On the other hand, the CNN-LSTM model demonstrated significantly better computational efficiency compared to the BLSTM. It required fewer parameters, reduced the model size, and lowered inference time and memory usage by a considerable margin. These differences reflect the higher complexity and resource demands of the BLSTM architecture. While BLSTM may offer benefits in modeling temporal dependencies due to its bidirectional structure, the CNN-LSTM achieved a better balance between computational cost and performance.

This suggests that the CNN-LSTM's hybrid spatial-temporal feature extraction may better handle rotational kinematics. Both models effectively reconstructed proximal joint angles from distal sensors, validating their potential for wearable applications.

Future work will be focused on integrating lower limbs estimations and optimizing model architectures to combine the strengths of CNN-LSTM and BLSTM, potentially through hybrid designs or attention mechanisms, to further reduce errors in rotational axes. Architectural refinements could integrate CNN-LSTM's spatial feature extraction with BLSTM's bidirectional temporal modeling to further reduce Yaw errors. Targeted data augmentation, emphasizing rotational movements, could enhance generalization for both models. Additionally, addressing the current limitation of using a single subject, future validation will involve multi-subject datasets, including publicly available human activity recognition (HAR) datasets [31] that capture a broader range of daily movements and inter-subject variability. Although upper limb motion generally exhibits lower variability across individuals compared to lower limb movements, techniques such as dropout and early stopping were applied during training to mitigate overfitting. To further enhance model robustness, transfer learning strategies could also be explored to adapt a pretrained model to new subjects, reducing the need for extensive retraining while mitigating subject-specific biases.

Hybrid approaches incorporating attention mechanisms and biomechanical constraints may improve rotational tracking. Attention could allow the model to assign greater weight to rotational components where errors are typically higher, enhancing its ability to focus on the most informative features. In parallel, introducing biomechanical constraints based on typical human joint ranges could guide the network toward more anatomically plausible predictions, acting as a form of regularization. Together, these strategies would support more accurate and consistent angle estimation, even under sensor limitations and diverse motion patterns. Sensor fusion techniques, such as sparse upper-body sensor integration, could be explored to disambiguate complex rotations while preserving wearability. The models' complementary strengths also suggest potential for ensemble methods or adaptive switching between architectures based on motion complexity. In potential real-time applications, the CNN-LSTM architecture is more suitable due to its causal structure and lower computational demands. In contrast, BLSTM relies on future context, which limits its applicability in online scenarios; however, a short temporal buffer could be introduced to access limited future frames, enabling near real-time performance with minimal latency. Finally, validation should expand to diverse populations and real-time implementations to solidify its clinical and industrial applicability. These advancements would position such models as leading solutions for accurate, minimal-sensor motion capture across rehabilitation, sports science, and human–computer interaction.

5. Conclusions

This study demonstrates that deep learning architectures can effectively estimate upper-limb joint angles using only three IMUs placed on the chest and wrists, offering a practical alternative to multi-sensor or optical motion capture systems. The CNN-LSTM and BLSTM models both achieved competitive accuracy, with the CNN-LSTM excelling in precision for planar movements and complex rotations (Yaw), while the BLSTM provided smoother temporal predictions but exhibited higher errors in peak rotational estimations. These results validate the feasibility of minimal-sensor configurations for joint angle reconstruction, highlighting the CNN-LSTM's superior spatial-temporal feature extraction and the BLSTM's robust sequential modeling. The findings underscore the potential of such approaches to bridge the gap between laboratory-grade accuracy and real-world wearability, enabling applications in rehabilitation, sports science, and human–robot interaction.

Author Contributions: Conceptualization, K.N.-T. and J.F.P.-M.; methodology, K.N.-T., L.S.-A., J.L.R.-C. and J.F.P.-M.; software, K.N.-T.; validation, K.N.-T. and J.F.P.-M.; formal analysis, K.N.-T.; investigation, K.N.-T.; resources, J.F.P.-M.; data curation, K.N.-T.; writing—original draft preparation, K.N.-T.; writing—review and editing, K.N.-T., L.S.-A., J.L.R.-C. and J.F.P.-M.; visualization, K.N.-T.; supervision, J.F.P.-M.; project administration, J.F.P.-M.; funding acquisition, J.F.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received financial support from the NSF EPSCoR Center for the Advancement of Wearable Technologies (CAWT) under Grant No. OIA-1849243.

Institutional Review Board Statement: The study was conducted under the Declaration of Helsinki and approved by the University of Puerto Rico's Institutional Review Board (IRB) under protocol number 2024120022 (approved on 27 January 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset used in this study is openly available in Zenodo at https://doi.org/10.5281/zenodo.15778950.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IMU	Inertial Measurement Units
MoCap	Motion Capture
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
BLSTM	Bidirectional Long Short-Term Memory

References

1. Aggarwal, J.K.; Cai, Q. Human Motion Analysis: A Review. *Comput. Vis. Image Underst.* **1999**, *73*, 428–440. [CrossRef]
2. Choffin, Z.; Jeong, N.; Callihan, M.; Sazonov, E.; Jeong, S. Lower Body Joint Angle Prediction Using Machine Learning and Applied Biomechanical Inverse Dynamics. *Sensors* **2022**, *23*, 228. [CrossRef] [PubMed]
3. Coker, J.; Chen, H.; Schall, M.C.; Gallagher, S.; Zabala, M. EMG and Joint Angle-Based Machine Learning to Predict Future Joint Angles at the Knee. *Sensors* **2021**, *21*, 3622. [CrossRef]
4. Senanayake, D.; Halgamuge, S.; Ackland, D.C. Real-time conversion of inertial measurement unit data to ankle joint angles using deep neural networks. *J. Biomech.* **2021**, *125*, 110552. [CrossRef] [PubMed]
5. Roggio, F.; Ravalli, S.; Maugeri, G.; Bianco, A.; Palma, A.; Di Rosa, M.; Musumeci, G. Technological advancements in the analysis of human motion and posture management through digital devices. *World J. Orthop.* **2021**, *12*, 467–484. [CrossRef]
6. Höglund, G.; Grip, H.; Öhberg, F. The importance of inertial measurement unit placement in assessing upper limb motion. *Med. Eng. Phys.* **2021**, *92*, 1–9. [CrossRef]
7. Cay, G.; Solanki, D.; Al Rumon, M.A.; Ravichandran, V.; Fapohunda, K.O.; Mankodiya, K. SolunumWear: A smart textile system for dynamic respiration monitoring across various postures. *iScience* **2024**, *27*, 110223. [CrossRef] [PubMed]
8. Ravanelli, N.; Lefebvre, K.; Brough, A.; Paquette, S.; Lin, W. Validation of an Open-Source Smartwatch for Continuous Monitoring of Physical Activity and Heart Rate in Adults. *Sensors* **2025**, *25*, 2926. [CrossRef]
9. Wei, J.C.J.; van den Broek, T.J.; van Baardewijk, J.U.; van Stokkum, R.; Kamstra, R.J.M.; Rikken, L.; Gijsbertse, K.; Uzunbajakava, N.E.; van den Brink, W.J. Validation and user experience of a dry electrode based Health Patch for heart rate and respiration rate monitoring. *Sci. Rep.* **2024**, *14*, 23098. [CrossRef]
10. Aurand, A.M.; Dufour, J.S.; Marras, W.S. Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume. *J. Biomech.* **2017**, *58*, 237–240. [CrossRef]
11. Conconi, M.; Pompili, A.; Sancisi, N.; Parenti-Castelli, V. Quantification of the errors associated with marker occlusion in stereophotogrammetric systems and implications on gait analysis. *J. Biomech.* **2021**, *114*, 110162. [CrossRef] [PubMed]
12. Zhao, J. A Review of Wearable IMU (Inertial-Measurement-Unit)-based Pose Estimation and Drift Reduction Technologies. *J. Phys. Conf. Ser.* **2018**, *1087*, 042003. [CrossRef]
13. Fan, B.; Li, Q.; Tan, T.; Kang, P.; Shull, P.B. Effects of IMU Sensor-to-Segment Misalignment and Orientation Error on 3-D Knee Joint Angle Estimation. *IEEE Sens. J.* **2022**, *22*, 2543–2552. [CrossRef]

14. Falbriard, M.; Meyer, F.; Mariani, B.; Millet, G.P.; Aminian, K. Drift-Free Foot Orientation Estimation in Running Using Wearable IMU. *Front. Bioeng. Biotechnol.* **2022**, *8*, 65. [CrossRef] [PubMed]
15. Alemayoh, T.T.; Lee, J.H.; Okamoto, S. Leg-Joint Angle Estimation from a Single Inertial Sensor Attached to Various Lower-Body Links during Walking Motion. *Appl. Sci.* **2023**, *13*, 4794. [CrossRef]
16. Hammerla, N.Y.; Halloran, S.; Ploetz, T. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. *arXiv* **2016**. [CrossRef]
17. Ordóñez, F.J.; Roggen, D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef]
18. Kim, D.; Jin, Y.; Cho, H.; Jones, T.; Zhou, Y.M.; Fadaie, A.; Popov, D.; Swaminathan, K.; Walsh, C.J. Learning-based 3D human kinematics estimation using behavioral constraints from activity classification. *Nat. Commun.* **2025**, *16*, 3454. [CrossRef]
19. Airaksinen, M.; Räsänen, O.; Vanhatalo, S. Trade-Offs Between Simplifying Inertial Measurement Unit-Based Movement Recordings and the Attainability of Different Levels of Analyses: Systematic Assessment of Method Variations. *JMIR Mhealth Uhealth* **2025**, *13*, e58078. [CrossRef]
20. Rast, F.M.; Labruyère, R. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments. *J. Neuroeng. Rehabil.* **2020**, *17*, 148. [CrossRef]
21. Mundt, M.; Johnson, W.R.; Potthast, W.; Markert, B.; Mian, A.; Alderson, J. A comparison of three neural network approaches for estimating joint angles and moments from inertial measurement units. *Sensors* **2021**, *21*, 4535. [CrossRef] [PubMed]
22. Niño-Tejada, K.; Saldaña-Aristizábal, L.; Rivas-Caicedo, J.L.; Patarroyo-Montenegro, J.F. IMU Sensors Emulation using Motion Capture Systems. In Proceedings of the International Symposium on Intelligent Computing and Networking 2025, San Juan, Puerto Rico, 17–19 March 2025; Springer: Berlin/Heidelberg, Germany, 2025.
23. Favata, A.; Gallart-Agut, R.; Pàmies-Vilà, R.; Torras, C.; Font-Llagunes, J.M. IMU-Based Systems for Upper-Limb Kinematic Analysis in Clinical Applications: A Systematic Review. *IEEE Sens. J.* **2024**, *24*, 28576–28594. [CrossRef]
24. Pezenka, L.; Wirth, K. Reliability of a Low-Cost Inertial Measurement Unit (IMU) to Measure Punch and Kick Velocity. *Sensors* **2025**, *25*, 307. [CrossRef]
25. Niño-Tejada, K.; Saldaña-Aristizabal, L.; Rivas-Caicedo, J.L.; Patarroyo-Montenegro, J.F. MoCap and IMU Dataset for Upper Limb Joint Angle Estimation. *Zenodo* **2025**. [CrossRef]
26. Winter, D.A. *Biomechanics and Motor Control of Human Movement*; Wiley: Hoboken, NJ, USA, 2009; Available online: https://books.google.com.pr/books?id=_bFHL08IWfwC (accessed on 10 June 2025).
27. Crenna, F.; Rossi, G.B.; Berardengo, M. Filtering biomechanical signals in movement analysis. *Sensors* **2021**, *21*, 4580. [CrossRef]
28. Yeo, S.S.; Park, G.Y. Accuracy verification of spatio-temporal and kinematic parameters for gait using inertial measurement unit system. *Sensors* **2020**, *20*, 1343. [CrossRef] [PubMed]
29. Kasnesis, P.; Patrikakis, C.Z.; Venieris, I.S. PerceptionNet: A Deep Convolutional Neural Network for Late Sensor Fusion. In *Intelligent Systems and Applications*; Springer: Berlin/Heidelberg, Germany, 2018.
30. Feigl, T.; Kram, S.; Woller, P.; Siddiqui, R.H.; Philippsen, M.; Mutschler, C. A bidirectional LSTM for estimating dynamic human velocities from a single IMU. In Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation IPIN, Pisa, Italy, 30 September 2019. [CrossRef]
31. Rivas-Caicedo, J.L.; Saldaña-Aristizabal, L.; Niño-Tejada, K.; Patarroyo-Montenegro, J.F. Wearable IMU Sensor Dataset for Human Activity Recognition (HAR). *Zenodo* **2025**. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Study on Comb-Drive MEMS Acceleration Sensor Used for Medical Purposes: Monitoring of Balance Disorders

Michał Szermer * and Jacek Nazdrowicz

Department of Microelectronics and Computer Science, Faculty of Electrical, Electronic, Computer and Control Engineering, Lodz University of Technology, 93-005 Lodz, Poland; jacek.nazdrowicz@p.lodz.pl
* Correspondence: michal.szermer@p.lodz.pl

Abstract

This article presents a comprehensive modeling and simulation framework for a capacitive MEMS accelerometer integrated with a sigma-delta analog-to-digital converter (ADC), with a focus on applications in wearable health and motion monitoring devices. The accelerometer used in the system is connected to a smartphone equipped with dedicated software and will be used to assess the risk of falling, which is crucial for patients with balance disorders. The authors designed the accelerometer with special attention paid to the specification required in a system, where the acceleration is ±2 g and the frequency is 100 Hz. They investigated the sensor's behavior in the DC, AC, and time domains, capturing both the mechanical response of the proof mass and the resulting changes in output capacitance due to external acceleration. A key component of the simulation is the implementation of a second-order sigma-delta modulator designed to digitize the small capacitance variations generated by the sensor. The Simulink model includes the complete signal path from analog input to quantization, filtering, decimation, and digital-to-analog reconstruction. By combining MEMS+ modeling with MATLAB-based system-level simulations, the workflow offers a fast and flexible alternative to traditional finite element methods and facilitates early-stage design optimization for MEMS sensor systems intended for real-world deployment.

Keywords: MEMS technology; accelerometer; inertial sensor

1. Introduction

Smart electronic systems are currently one of the most growing branches of electronics. They are used in smartphones, tablets, drones, etc. They control everyday things such as lighting, window blinds, garage doors, etc., creating the Internet of Things (IoT) [1]. We can find them also in specialized equipment such as medical devices that make life easier for patients with various diseases at home. We can also find them in healthcare applications used by medical personnel in health centers. They use rapidly developing microelectromechanical systems (MEMSs), especially in activity sensors, biosensors, chemical sensors, microfluidic diagnostics, drug delivery devices, gene analyzers, etc. [2,3]. Thanks to them, medical diagnostic is faster and easier compared to previously used devices. The small medical devices used by patients are usually connected to smartphones that collect data and send it to health centers, where doctors can monitor the patient's health in real time and provide appropriate treatment immediately.

We may find many interesting usages of the MEMS sensors in wearable applications. In [4] the authors presented a wearable gait recognition system used in clinical medicine

and other fields. More advanced sensors can be used for monitoring blood pressure and heart rate [5]. The authors describe here the strain sensor used for continuous monitoring of blood pressure. In turn, work [6] is related to the wearable system for monitoring human motion and posture, where the authors presented a review through the mentioned systems. Another author presented in their work [7] a wearable human attitude tracking and visualization system based on MEMS sensors. Inertial sensors used in medical applications related to movement monitoring can also be found in [8,9].

In this article, the authors describe the operation of a MEMS sensor that can be used in a small device called Medipost (Figure 1), which is used for patient movement monitoring [10,11]. The device is a small autonomous system mounted on the patient's belt (fifth lumbar vertebra (L5) of the patient back). It collects data during its operation and sends it to a smartphone with a dedicated application installed. The main aim of Medipost is to monitor patients with balance disorders. For this purpose, we can distinguish several methods. From a medical point of view, the most important is posturography, which can be performed at medical centers only. The second most important thing is the patient's fall risk assessment, which can be conducted at home. This led to the design of the Medipost system, in which the main part consists of inertial sensors. The MEMS accelerometer used in the system should have an acceleration range of ±2 g and a sampling rate of no more than 100 Hz. It communicates with a smartphone equipped with a dedicated application via Wi-Fi or Bluetooth. Another important consideration is the power supply, which should be as small as possible, since the system is powered by a single rechargeable battery, enabling it to work for at least 2 h) [12].

While development of the device is still in progress, this article presents an analysis focused on one of the system's key components—the MEMS accelerometer. For this purpose we use a model obtained from MEMS+ and simulate it in MATLAB v. R2024a Simulink. This analysis provides many information regarding the sensor's operation.

In the following sections, the authors describe the idea of the MEMS accelerometer responsible for measuring acceleration. The simulation results of the MEMS and its interaction with the analog-to-digital converter (ADC) are discussed further in the text.

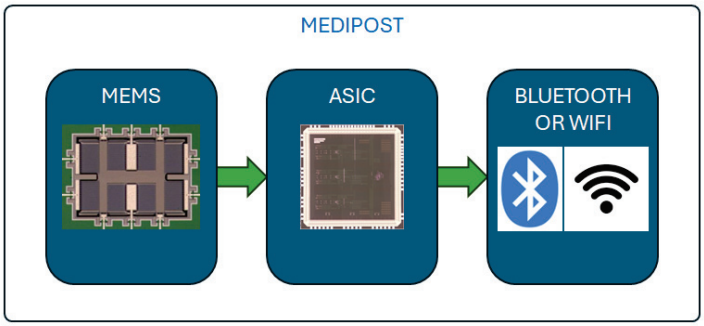


Figure 1. Block diagram of the Medipost system.

2. Materials and Methods

2.1. Accelerometer Sensor—Principle of Operation

In this section, the authors describe an acceleration sensor designed using MEMS technology. A three-axis sensor is designed by placed it on a single substrate. During the analysis, the authors focused on the performance of the single-axis sensor. The sensors operating in the X and Y planes have the same design. They are placed on a plane of silicon at 90 degrees to each other to measure the corresponding acceleration. Only the Z-axis sensor has a different design, as it operates out of plane. Some interesting information about this sensor can be found in [13].

Returning to X-axis sensors, their operation is based on differential mode. This means that the sensors contain two capacitors that operate in different ways. In the equilibrium state, both capacitances (top and bottom) have the same value. When acceleration in the X-axis occurs, the value of one capacitance increases and the other decreases (Figure 2). Therefore, the sensitivity of the sensor is better in comparison to the single capacitance.

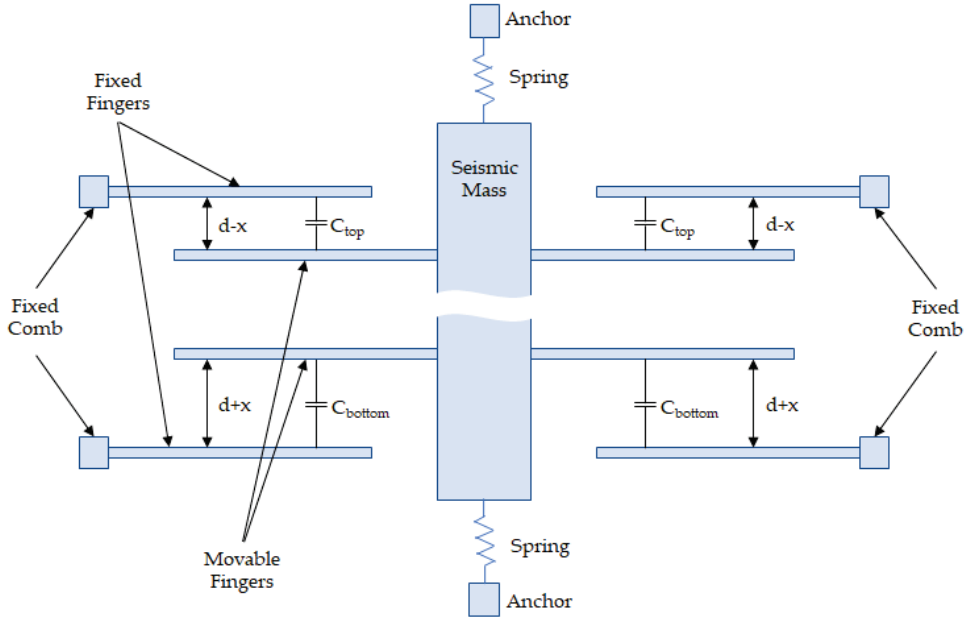


Figure 2. The principle of operation of the capacitive accelerometer acting in the X- and Y-axes (not to scale) [13].

The sensor works in an open loop, where the response is proportional to the capacitance change (the movement of the seismic mass is proportional to the inertial force). This results in lower power consumption compared to closed-loop systems, where seismic mass must be kept in equilibrium.

2.2. Technology Used for Manufacturing the Sensor

To manufacture a MEMS sensor, it is necessary to use dedicated micromachining technology. In general, the two most commonly used technologies are surface micromachining and bulk micromachining [14–16]. All of them have some advantages and disadvantages. In surface micromachining, it is possible to produce sensors and other structures by depositing thin layers on the top surface of a wafer. Among the others, the sacrificial layers were deposited, which are eventually removed to obtain the released structures. This produces some inertial sensors or other structures with moving parts, such as micromotors [17,18]. The second technology is based on the removal of some of the silicon from the silicon substrate. Here, we may produce sensors that are suspended above the cavity. For example, we are able to produce pressure sensors, accelerometers, and other structures [19]. On the other hand, there also exists LIGA (ger. Lithographie, Galvanoformung, Abformung) technology, in which we can obtain small parts of microsystems using a molding process. We prepare a stamp mold using, i.e., polymethylmethacrylate (PMMA). After exposure it to X-rays, we remove the superfluous part of the material, and then deposit a layer of metal through an electroplating process. Finally, we obtain small fragments of the mechanical parts of microsystems such as tooth wheels, etc. [20].

In our case, we used X-FAB MEMS technology with a characteristic dimension of 1.0 µm (XMB10 IC process) [21]. This technology allows us to design and manufacture inertial sensors with seismic mass and combs. We designed a triaxial accelerometer with

movable seismic mass and fixed combs. However, manufacturing the X- and Y-axis accelerometer was not a very complicated process, compared to the Z-axis accelerometer which was a really huge challenge. We had to use two types of combs (normal and reduced-height) in this sensor to correctly detect the change in acceleration.

We chose to combine this technology with the application-specific integrated circuit (ASIC), which was also manufactured using X-FAB technology (180 nm IC process), Erfurt, Germany. We used two different technologies (one for the MEMS and one for the ASIC) that were provided by the same vendor so that they could be easily connected in a single package. More interesting information about this structure can be found at [13].

2.3. Sensor Model and Design

The design of the accelerometer was conducted in the Coventor MEMS+ v. 6.2 software [22]. It is a computer-aided design/electronic design automation (CAD/EDA) software that allows users to design a prototype using a library of sensor components. It is a very useful tool that allows us to quickly design a prototype sensor. It contains a library of components such as combs, beams, anchors, frames, etc. Using these components, it is possible to design a parametrized structure that can be simulated to find the optimal structure. All components were prepared using the technology mentioned above and were used in the model. Once designed, it was possible to simulate the sensor in the same software as well in DC and AC domains to find the optimal structure. Unfortunately, it is not possible to simulate the model in the time domain in MEMS+. We could simulate it in MATLAB Simulink [23], where we imported the sensor model from MEMS+. In order to speed up the simulation, the reduced-order model (ROM) was imported. This was a huge advantage as it directly influenced the simulation time. The design automation flow used in the project is shown in Figure 3. The model of the sensor designed in MEMS+ and the photo of it are presented in Figure 4.

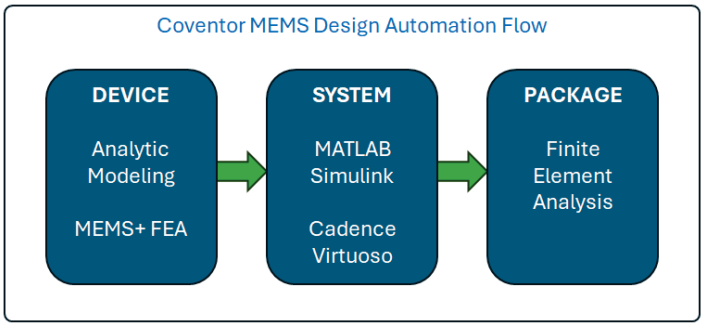
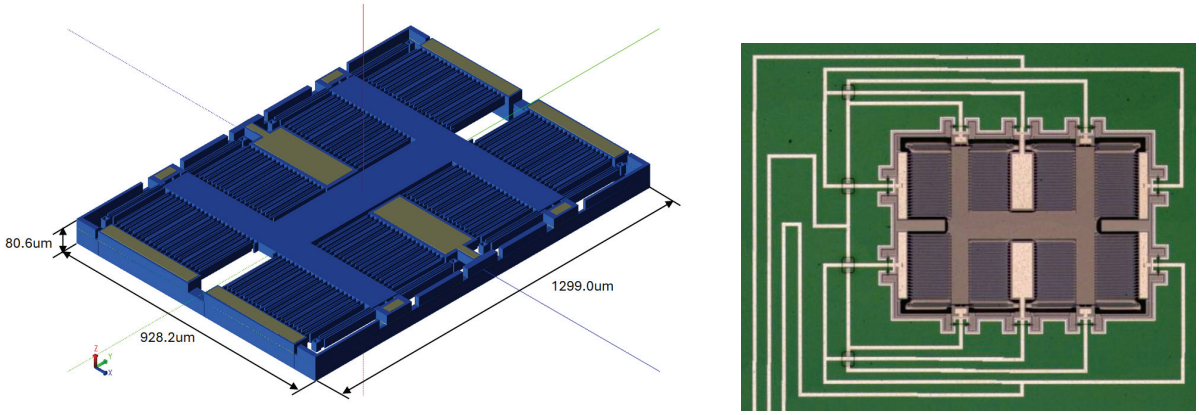


Figure 3. Design flow of the MEMS and ASIC.



(**a**) (**b**)

Figure 4. MEMS accelerometer: (**a**) model in MEMS+ Innovator. (**b**) Photo under a microscope [13].

3. Results

The first simulation models were created in the MEMS+ software [22]. In this software we used a dedicated process design kit (PDK) provided by the technology's vendor. It is a MEMS 1.0 µm micromachining technology.

It was then possible to create a model that can be used in the CADENCE IC Design environment [24] to prepare the structure layout. This model was created using the Verilog-A language and could be simulated in an analog environment [24]. After successful simulation results, a layout mask was generated. The technology uses a small number of steps to produce the sensor. It is worth emphasizing that such simulations and layouts were performed, and finally the triaxial accelerometer was fabricated at the X-FAB silicon foundry [13].

Another method of designing and simulating the sensor was to develop a model that could be used in MATLAB Simulink [23]. This made it possible to simulate the sensor's response to input signals in the time domain. Due to the complexity of the model, a ROM was used in MATLAB Simulink [25]. This is necessary to reduce the simulation time, as the simulations of the full-complexity model can take many hours or more. The reduced-order model was tuned to the first resonant frequency, which reduced the number of internal components required to reflect the behavior of the sensor in finite time. The results of the simulation of such a model are presented and discussed in this article.

The accelerometer model imported into the MATLAB Simulink environment (Figure 5) has three acceleration inputs for each axis. It also has three electrical inputs to which we applied electrical signals. For VEcapTop and VEcapBottom, we applied the square wave, but with the phase reversed. For VEcapMass we applied 2.5 volts. This was necessary to obtain optimal input conditions for the operation of the accelerometer. At the outputs of the model we have two capacitances (Ctop and Cbottom). Their variation is proportional to the input signal.

We present a complete analysis in the DC, AC and time domains further in this section. From the DC analysis we can obtain information about the change in capacitance as a function of acceleration. This dependence is linear as presented in Figure 6. As we are using a variant of two capacitances of the sensor's construction (differential mode), we can see that for the acceleration in the range between ±2 g, the top capacitance value decreases whereas the bottom capacitance value increases.

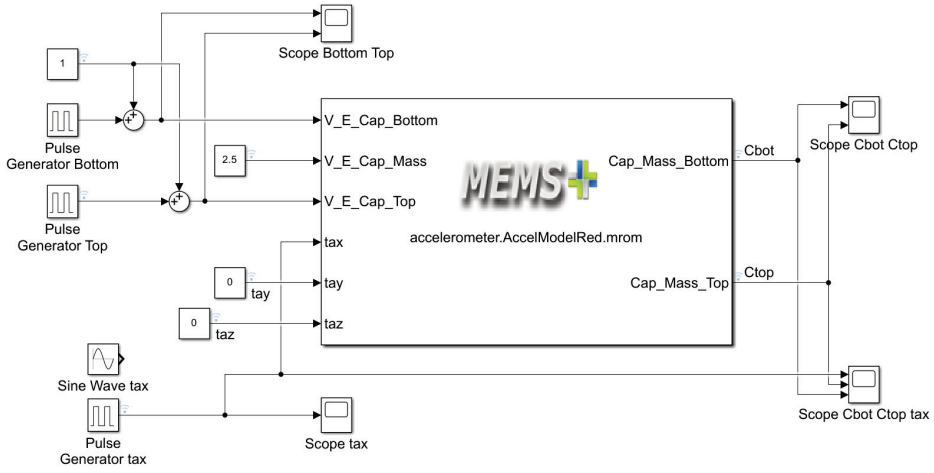
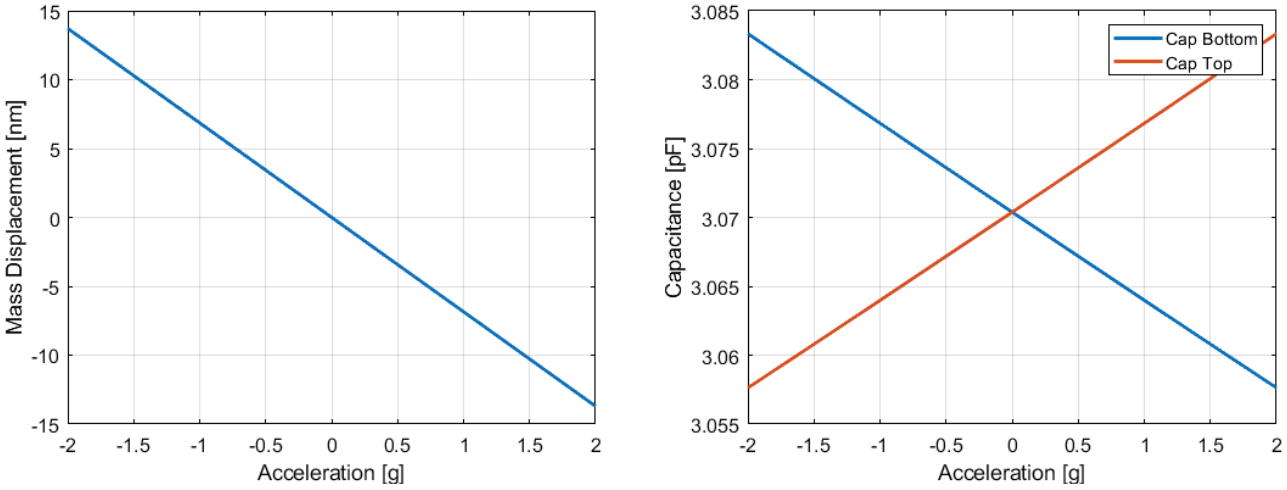


Figure 5. Accelerometer model in MATLAB Simulink.

The next simulation was the modal analysis of the accelerometer. This was necessary to find the resonant frequency of the seismic mass of the sensor. The results of the simulations can be found in Table 1. From these results we may conclude that the correct operating

range for the sensor is below the first resonant frequency, which is equal to 6.087 kHz. The same information can be obtained from the AC analysis. The results of it are shown in Figure 7. We can see the resonance for the 6.115 kHz frequency, which is very close to the modal analysis performed in MEMS+. This confirmed the limit of the operating range of the sensor to one kHz, which is below 6 kHz.



(**a**) (**b**)

Figure 6. Simulation results for ±2 g acceleration in MEMS+ for the X-axis accelerometer: (**a**) seismic mass displacement. (**b**) Capacitance change (Cbottom—blue line and Ctop—red line).

Table 1. The eigenfrequencies of the X-axis accelerometer model.

No of Mode	Mode Frequency [kHz]	Q Factor
1	6.087	10.039
2	52.698	29.803
3	54.052	28.328

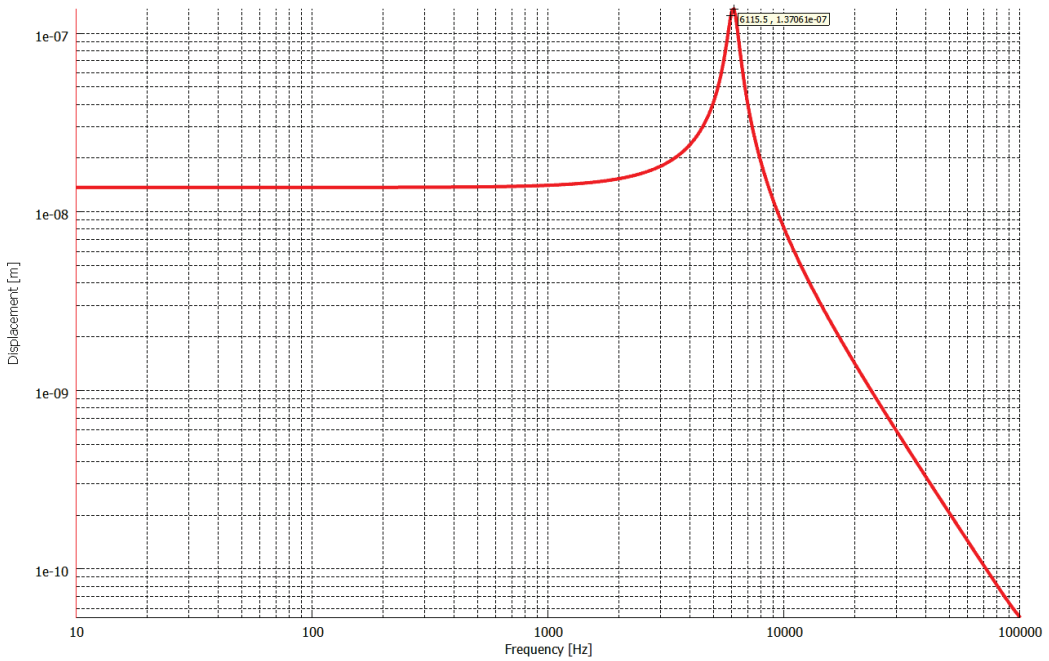


Figure 7. Frequency response of the X-axis accelerometer in MEMS+.

The aforementioned simulations can also be performed in MEMS+; however, the transient simulation must be conducted in MATLAB Simulink. Further in this section we present the simulations for some different input signals applied to the inputs of the model.

At equilibrium, where all accelerations are zero, the two capacitances are also equal, and their value is 3.07 pF. If we force the acceleration in the x-axis direction from 0 to 1 g, we see that the capacitance changes are equal to 3.08 pF and 3.06 pF (Figure 8). This is only a change of 10 fF for each capacitance. Importantly, we observe an oscillation in the sensor response of almost 2 ms. After this time, outputs of the sensor settle into equilibrium. These oscillations are related to the lower pressure inside the sensor, which is 800 Pa, meaning that it is very sensitive. In a real sensor, we have small differences in the top and bottom capacitance in the equilibrium state, which have to be trimmed by the ASIC. Only in this way can the achieved optimal state of the sensor for 0 g acceleration be achieved.

The next simulation results show the response of the sensor to a sinusoidal input signal representing external acceleration. The amplitude of the signal was equal to ±2 g, and the frequency is equal to 100 Hz. As can be seen in Figure 9, the response of the sensor is very good. The output capacitance changes along with a sine wave with an amplitude of 40 fFp-p. A total of 2.5 V has been applied to the seismic mass pin to polarize this part of the structure and eliminate the possibility of oscillations in the sensor response.

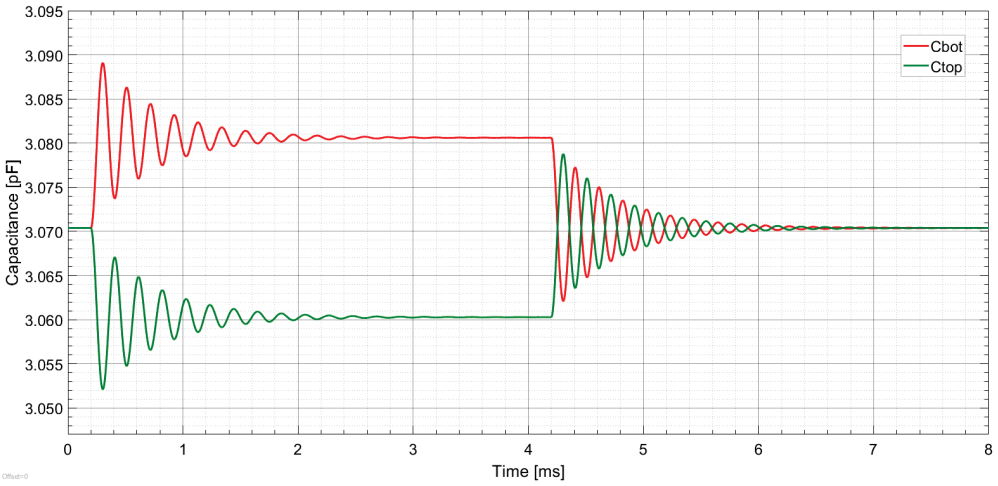


Figure 8. Simulation results of the X-axis accelerometer for input acceleration equal to a pulse of 1 g in MATLAB Simulink. Capacitance change (Cbottom—red line and Ctop—green line).

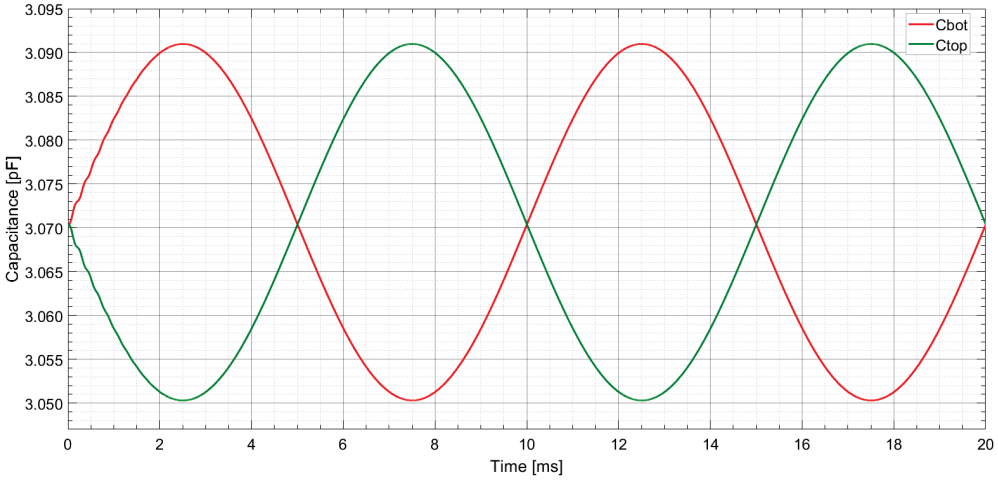


Figure 9. Simulation results of the X-axis accelerometer for sinusoidal acceleration equal to an amplitude of 2 g and a frequency of 100 Hz in MATLAB Simulink. Capacitance change (Cbottom—red line and Ctop—green line).

To test the behavior of the accelerometer in the system, we added an electronic circuit whose main task is to convert the input signal to digital. We created two models in the Simulink system. The first one is a delta modulator, and the second one is the demodulator (Figures 10 and 11).

Figure 12 shows the simulation results of the MEMS accelerometer output and the reconstructed signal from the delta modulator. The input to the system is a sinusoidal acceleration signal with an amplitude of 2 g and a frequency of 80 Hz. The vertical axis represents capacitance in femtofarads and the horizontal axis represents time in milliseconds. The red curve labeled MEMS represents the actual output of the simulated MEMS accelerometer which responds to the applied acceleration by producing a changing capacitance signal. The green curve labeled DAC shows the output of the digital-to-analog conversion process after the signal has been quantized, processed, and reconstructed. At the beginning of the plot the DAC signal slightly lags and differs in amplitude compared to the MEMS output. This initial deviation is likely due to system startup effects or delay introduced by the delta converter and filtering stages. As time progresses the DAC output more closely follows the MEMS signal in both amplitude and phase. The overall trend shows that the system is capable of accurately reconstructing the original analog signal after processing through the delta modulator and demodulator stages with only minor transient mismatches. This indicates a high level of fidelity in the signal conversion process and confirms the effectiveness of the implemented simulation model in capturing the dynamic behavior of the sensor.

Figure 13 shows the top-level Simulink model of an accelerometer system with a sigma-delta analog-to-digital converter. At the center is a MEMS accelerometer model labeled accelerometer.AccelModelRed.mrom which simulates the behavior of a capacitive sensor using pre-defined reduced-order models. The model receives excitation signals in the form of voltages applied to the bottom, top, and central sensing electrodes. These voltages come from a combination of constant values and pulse generators, simulating dynamic conditions or electrical excitation. A sine wave input labeled tax is used to simulate acceleration along the X-axis, which serves as the primary mechanical stimulus for the sensing system.

The accelerometer outputs changing capacitances Ctop and Cbot, corresponding to the top and bottom sensing electrodes. These capacitances change depending on the displacement of the proof mass due to the acceleration applied. These two signals are visualized separately and then subtracted to create a differential capacitance signal that is more sensitive to motion and less affected by noise or common-mode disturbances. This differential capacitance signal is then fed into a subsystem representing the sigma-delta ADC, which is shown in detail in Figure 14. Inside this ADC block the analog signal is processed through modulation, quantization, filtering, and decimation stages to produce a digital output. The model also includes a path that delays the analog signal to align it with the digital output, allowing accurate calculation of the error between the original and reconstructed signals. Several scope blocks are used throughout the model to monitor important signals such as the input acceleration, individual capacitance values, the differential signal, and both the analog and digital outputs of the ADC. The complete system demonstrates a realistic simulation of a MEMS accelerometer signal chain from physical excitation through capacitive sensing to digital conversion and error analysis.

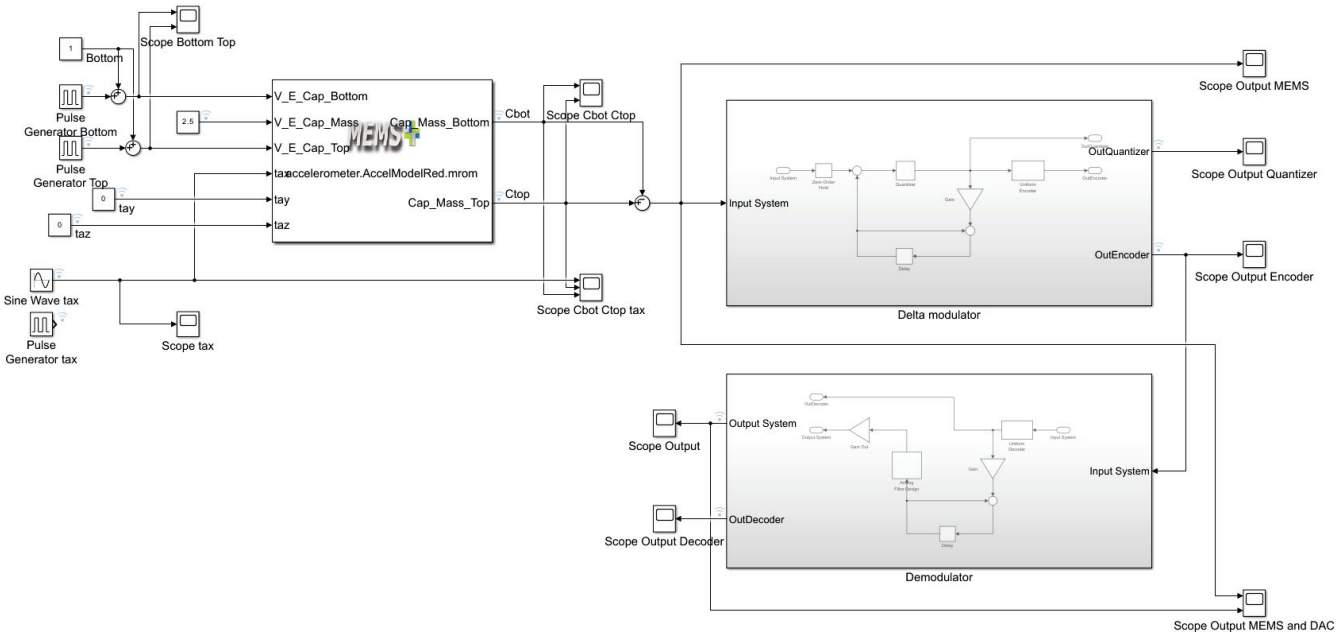


Figure 10. Accelerometer model with a delta modulator and demodulator in MATLAB Simulink.

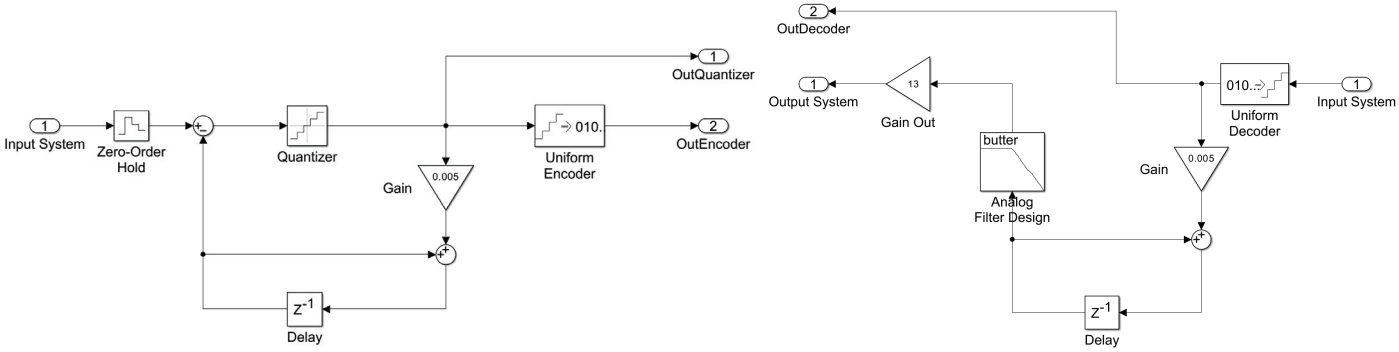


Figure 11. Accelerometer with the processing part model in MATLAB Simulink: (**a**) delta modulator; (**b**) delta demodulator.

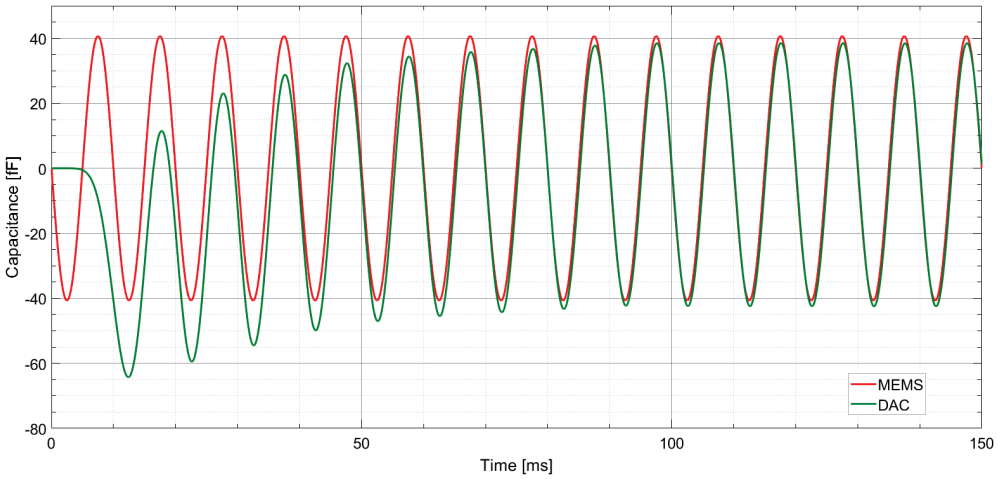


Figure 12. Simulation results of the input acceleration equal to an amplitude of 2 g and a frequency of 80 Hz in MATLAB Simulink (MEMS accelerometer output (red line) and and digital-to-analog converter output (green line)).

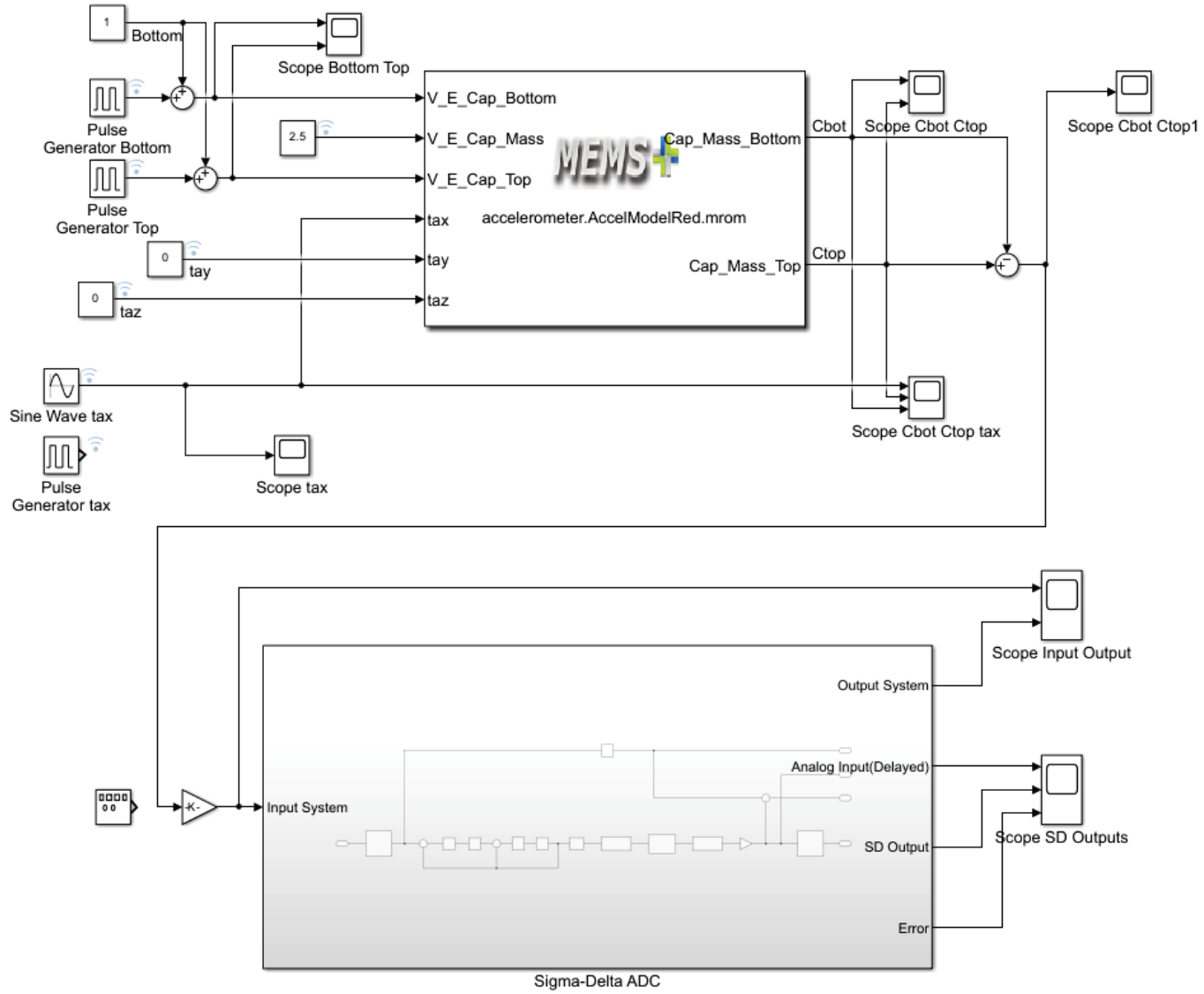


Figure 13. Accelerometer model with the sigma-delta ADC in MATLAB Simulink.

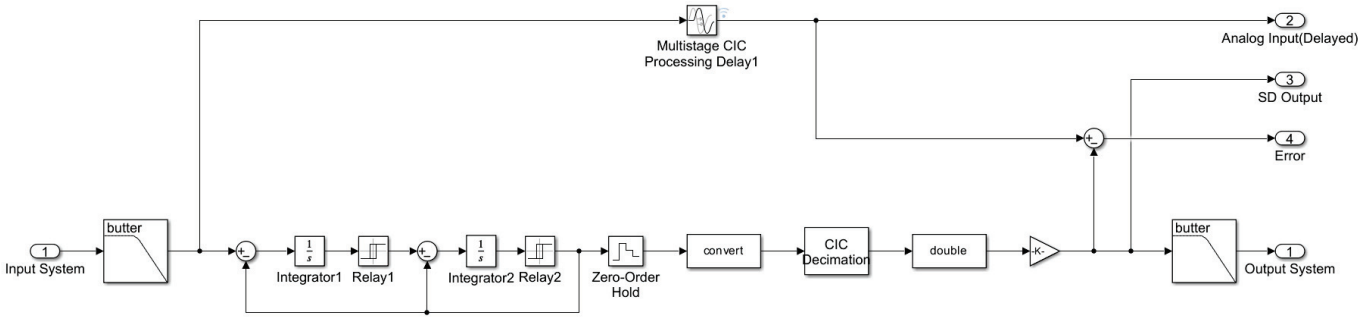


Figure 14. Sigma-delta analog-to-digital converter model in MATLAB Simulink.

In Figure 15, we can see the input and output signals obtained from the sigma-delta ADC. The top plot shows the input capacitance signal over time. It is a smooth sinusoidal waveform ranging from -50 nF to $+50$ nF with a period of about 10 ms, representing an ideal analog signal. The middle plot shows the output from the sigma-delta converter, which attempts to reproduce the input signal in a quantized digital form. Although it follows the general shape of the input, the signal is visibly stepped or quantized, indicating

limited resolution during digital conversion. The bottom plot represents the error, which is the difference between the input and output signals. The error signal fluctuates in a periodic and oscillatory manner, reaching peaks of about ±3.5 nF. The largest deviations appear during the steep slopes of the sine wave where the rate of change is highest. This behavior is typical of quantization noise and imperfections in the digital conversion process. In summary, the system reproduces the general behavior of the input signal but introduces quantization artifacts, resulting in small but noticeable errors, particularly during fast transitions. Despite this its performance appears acceptable and within expected limits for a sigma-delta-based system.

Figure 16 shows the input and output signals of the accelerometer connected to the sigma-delta ADC. The acceleration for this case was equal to 2 g in amplitude and a frequency equal to 80 Hz. The output signal is the same as the input but with a small delay. It is related to the usage of the input low-pass filter with the cut-off frequency equal to 400 Hz. It is possible to tune the output signal by using the last filter in the system. The usage of the second-order sigma-delta ADC allowed us to reduce the noise in the output signal because it is shifted to higher frequencies. This is the main advantage of using a sigma-delta ADC.

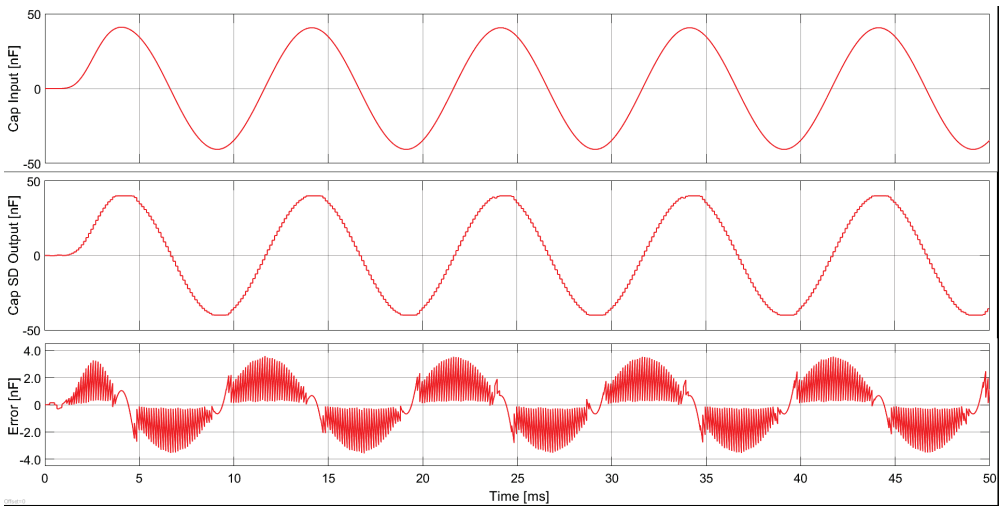


Figure 15. Simulation results of the capacitance change as an input signal (**top**), quantized output capacitance change (**middle**), and signal error (**bottom**).

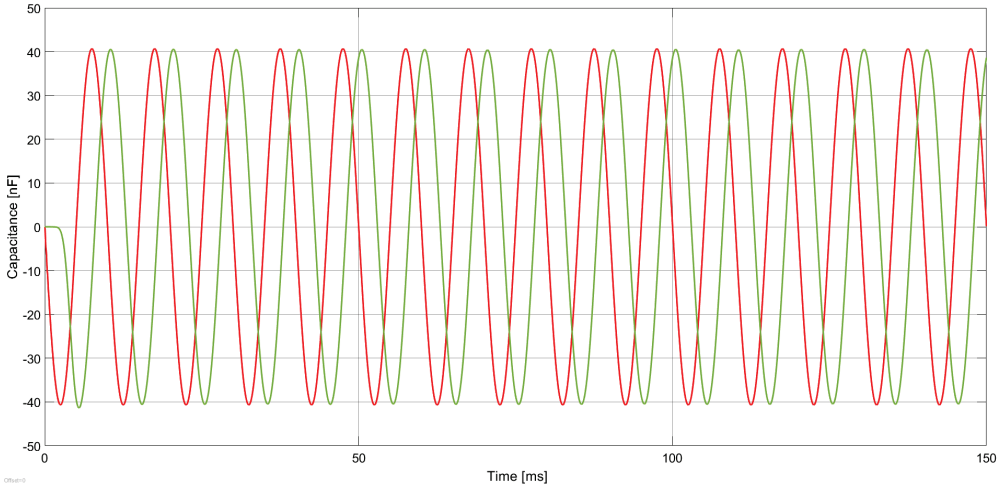


Figure 16. Simulation results of the input acceleration equal to an amplitude of 2 g and a frequency of 80 Hz in MATLAB Simulink (MEMS accelerometer output (red line) and digital-to-analog converter output (green line)).

Based on the analysis, we determined the main parameters of our ADC. For the second-order sigma-delta converter, we obtained a signal-to-noise ratio (SNR) of 54.73 dB. We calculated an effective number of bits (ENOB) of 8.61. Then, the total harmonic distortion (THD) is equal to −60.04 dB. We chose this type of converter because of the compromise between its complexity and the obtained parameters. This was influenced, as mentioned earlier, by the goal of obtaining low power consumption of the circuit and a small size.

As a preliminary measurement of the sensor, a static analysis was performed. A MEMS accelerometer with an external readout circuit on a moving lever was mounted. This simple model reflects the patient. Then, static gravitational acceleration was measured. Measurements were taken at four angles (0, 30, 60, and 90 degrees). The results were converted to acceleration and are presented in Figure 17. They show that the MEMS accelerometer is very sensitive to external interference and needs to be connected with a dedicated ASIC in a single package. The current work focuses on dynamic measurements, which will be presented in future articles.

Finally, we compared our sensor with other accelerometers available on the market (Table 2). We can see that our structure has very good linearity. This is satisfactory when we talk about monitoring patients' movement, where a small range of acceleration is required (about ±2.5 g). Note, that its size is comparable to other accelerometers. Its construction, based on the separate design of each axis sensor, reduces the sensitivity to acceleration in other directions. This is crucial when we consider the application our accelerometer to medical purposes.

Table 2. Parameters of three-axis multiple-seismic-mass accelerometers.

Ref.	Year	Device Size [mm × mm]	Range X, Y, Z [±g]	Sensitivity X, Y, Z	Nonlinearity X, Y, Z
[26]	1999	4 × 4	1.9	0.4 fF/bit	–
[27]	1999	5 × 5	–	25 fF/g 25 fF/g 100 fF/g	–
[28]	2005	7 × 9	1	6.8 pF/g 6.8 pF/g 2.9 pF/g	–
[29]	2012	4 × 4	2, 4, 8, 16	16,384 LSB/g 8192 LSB/g 4096 LSB/g 2048 LSB/g	–
[30]	2013	1.57 × 1.73	0.01 ÷ 2	105 mV/g 127 mV/g 58 mV/g	1% 0.5% 2.4%
[31]	2015	12 × 7	10 10 +12, –7	–	0.34% 0.28% 0.41%
[32]	2022	3 × 5	2, 4, 8, 16	256 LSB/g 128 LSB/g 64 LSB/g 32 LSB/g	–
This work	2018	3.95 × 6.55	2.85 2.94 4.03	13.15 fF/g, 701 mV/g 0.85 fF/g, 680 mV/g 0.31 fF/g, 496 mV/g	0.40% (R^2 = 0.9996) for ±1 g range 0.83% (R^2 = 0.9917) for ±1 g range –

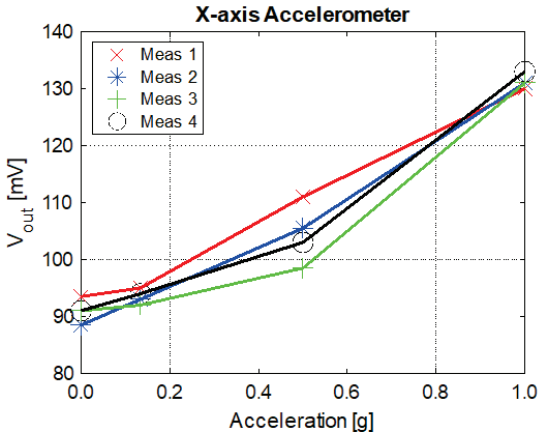


Figure 17. The measured output voltage of the external readout circuits with the MEMS accelerometer as a function of acceleration. The measured device was positioned at four angles (0, 30, 60, and 90 degrees). The same measurements were repeated four times [13].

4. Discussion

This article presents the analysis of the MEMS acceleration sensor, which constitutes the continuation of the analysis written in the article [13], which focused mainly on the static analysis of the sensor. We have analyzed the behavior of the sensor in the DC, AC, and time domains.

In the DC and AC domains, we obtained the static sensor behavior and its frequency response. It should be noted that the sensor used in the Medipost device will operate in the range of ± 2 g. Therefore, the analysis presents the behavior of the sensor in this range of acceleration. The DC response of the sensor is linear in the mentioned range. The first eigenfrequency in the AC domain is also observed near 6.1 kHz. This is sufficient to meet our requirements.

The time-domain analysis showed that the sensor was very sensitive to acceleration. The low pressure inside the sensor caused the seismic mass to oscillate with a time constant of 3 milliseconds. The size of the sensor gives a small total capacitance of 3.07 pF with a sensitivity of about 10 fF/g according to MATLAB Simulink simulations. This requires a dedicated readout circuit to be connected to the sensor.

We tested the behavior of the sensor with the delta modulator and the sigma-delta ADC. In both cases the system works correctly. However, it is not so simple to setup proper simulation solver conditions for such a complex system. After many analyses we find the optimal simulation condition, which is the *ode*14*x* Simulink solver with a fixed number of steps.

The performed analysis showed that the acceleration sensor operates correctly for the small frequencies of the input signal. For the input signal we assumed that the signal is equal to ± 2 g with a frequency of about 100 Hz. Such parameters are interesting for the Medipost device, which is used for patient movement monitoring. Such a device is connected to a smartphone with a data analysis system installed. In the future we will present further simulations and measurements results related to that sensor.

To summarize, we use a model generated in the MEMS+ software and analyze it in MATLAB Simulink. One of the important advantage of this approach is to design a scalable model with predefined components in the MEMS+ software. The design is faster in comparison to the standard approach with the usage of the finite element model (FEM) in ANSYS 15.0 [33] or COMSOL 5.2a Multiphysics [34] software. Another advantage is that in MEMS+, the PDK is presented with the design rules checked. These rules have to be met in the layout of the sensor, which is then sent to the silicon foundry for manufacturing. The link with another software such MATLAB or CADENCE 6.1.7 is also an advantage of

the MEMS+ software. Despite many advantages, there are also some disadvantages. One of them is the lack of the time-domain analysis of the structure. We need to use external software to perform such an analysis. However, the complex analysis of the structure of the sensor has to be performed with, e.g., MATLAB Simulink, so that we can obtain in a short time the full analysis of our design.

5. Conclusions

This work presents an extended dynamic analysis of a MEMS capacitive accelerometer, complementing the static evaluation discussed in earlier studies. While the previous work focused primarily on static characteristics and the frequency response, this study explores the time-domain behavior, digital signal processing performance, and system integration with a sigma-delta analog-to-digital converter (ADC). The accelerometer model was developed using the MEMS+ software, which enabled rapid system-level modeling with built-in design rule checking (DRC) and parametric component libraries. The model was then imported into MATLAB Simulink, where full-system simulations—including mechanical excitation, capacitive sensing, and analog-to-digital conversion—were performed. DC and AC analyses confirmed that the sensor operates linearly within the ± 2 g acceleration range, which matches the expected operational range for Medipost wearable devices. The first resonance frequency was observed at approximately 6.1 kHz, indicating sufficient bandwidth for low-frequency physiological signal acquisition, such as human motion, gait, or posture. Time-domain simulations show that the sensor is highly responsive to applied acceleration, with a time constant of about 3 ms due to the low-pressure damping environment. The resulting signal exhibits relatively low capacitance variation (on the order of 10 fF/g), which requires highly sensitive and noise-resilient readout circuitry. To digitize the signal, a second-order sigma-delta ADC was designed in Simulink, as shown in Figure 14. The ADC was tested using synthetic acceleration signals (± 2 g at 80 Hz and 100 Hz), which are representative of the target application domain. The modulator and decimator chain within the ADC effectively captured the analog input, and its digital output was validated against the original signal.

The use of MEMS+ and MATLAB Simulink enabled efficient, multi-domain modeling of a capacitive MEMS accelerometer with integrated signal processing. The simulations confirm that the sensor exhibits linear and stable behavior within its operational range and that the designed sigma-delta ADC can accurately digitize and reconstruct dynamic signals with high resolution and low noise. This modeling workflow offers several advantages over conventional FEM approaches, such as faster prototyping, a DRC-compliant design, and streamlined simulation integration with software environments like MATLAB. While MEMS+ lacks direct time-domain simulation capabilities, coupling it with Simulink allows full behavioral analysis, including signal fidelity and ADC performance. The analysis confirms that the proposed sensor–ADC system is well-suited for wearable applications involving human motion monitoring. Future work will involve validation against physical measurements and further optimization of the readout circuitry for ultra-low-power embedded systems.

Author Contributions: Conceptualization, M.S. and J.N.; methodology, M.S.; validation, M.S. and J.N.; formal analysis, M.S.; investigation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, M.S. and J.N.; visualization, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: The results presented in the paper are supported by the project STRATEGMED2/266299/19NCBR/2016, which is funded by The National Centre for Research and Development in Poland.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to express their deep gratitude to the late Andrzej Napieralski from Lodz University of Technology for their support and deep commitment during the implementation of the STRATEGMED project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sundmaeker, H.; Guillemin, P.; Friess, P.; Woelfflé, S. Vision and challenges for realising the Internet of Things. *Clust. Eur. Res. Proj. Internet Things Eur. Commision* **2010**, *3*, 34–36.
2. Napieralski, A.; Napieralska, M.; Szermer, M.; Maj, C. The evolution of MEMS and modelling methodologies. *COMPEL-Int. J. Comput. Math. Electr. Electron. Eng.* **2012**, *31*, 1458–1469.
3. Bryzek, J. MEMS Tornado to Challenge Next Generation Packaging. *Addit. Conf. (Device Packag. HiTEC HiTEN CICMT)* **2012**, *2012*, 000417–000456. [CrossRef]
4. Li, W.; Lu, W.; Sha, X.; Xing, H.; Lou, J.; Sun, H.; Zhao, Y. Wearable Gait Recognition Systems Based on MEMS Pressure and Inertial Sensors: A Review. *IEEE Sens. J.* **2022**, *22*, 1092–1104. [CrossRef]
5. Jiang, X.; Shokouhmand, A.; Ebadi, N.; Ayazi, F. Fingertip Plethysmography Using A Sensitive Wafer-Level-Packaged Capacitive MEMS Strain Sensor. In Proceedings of the 2024 IEEE 37th International Conference on Micro Electro Mechanical Systems (MEMS), Austin, TX, USA, 21–25 January 2024; pp. 348–351. [CrossRef]
6. Huang, X.; Xue, Y.; Ren, S.; Wang, F. Sensor-Based Wearable Systems for Monitoring Human Motion and Posture: A Review. *Sensors* **2023**, *23*, 9047. [CrossRef] [PubMed]
7. Xu, T.; Wang, J.; Yu, J.; Fu, Y.; Bai, R. Wearable Human Attitude Visualization Tracking System Based on MEMS Inertial Sensors. In Proceedings of the 2025 6th International Conference on Electrical, Electronic Information and Communication Engineering (EEICE), Shenzhen, China, 18–20 April 2025; pp. 1467–1472. [CrossRef]
8. Alessandrini, M.; Micarelli, A.; Viziano, A.; Pavone, I.; Costantini, G.; Casali, D.; Paolizzo, F.; Saggio, G. Body-worn triaxial accelerometer coherence and reliability related to static posturography in unilateral vestibular failure. *J. Vestib. Res.* **2017**, *37*, 231–236. [CrossRef] [PubMed]
9. Kim, S.C.; Kim, M.J.; Kim, N.; Hwang, J.H.; Han, G.C. Ambulatory balance monitoring using a wireless attachable three-axis accelerometer. *J. Vestib. Res. Equilib. Orientat.* **2013**, *23*, 217–225 . [CrossRef] [PubMed]
10. Tylman, W.; Kotas, R.; Kamiński, M.; Marciniak, P.; Woźniak, S.; Napieralski, J.; Sakowicz, B.; Janc, M.; Józefowicz-Korczyńska, M.; Zamysłowska-Szmytke, E. Fully Automatic Fall Risk Assessment Based on a Fast Mobility Test. *Sensors* **2021**, *21*, 1338. [CrossRef] [PubMed]
11. Jankowski, M.; Szermer, M.; Amrozik, P. Concept of an Electronic Device Dedicated to Imbalance Disorders Monitoring. In Proceedings of the 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Polyana, Ukraine, 26 February–2 March 2019; pp. 23–26. [CrossRef]
12. Kotas, R.; Janc, M.; Kamiński, M.; Marciniak, P.; Zamysłowska-Szmytke, E.; Tylman, W. Evaluation of Agreement Between Static Posturography Methods Employing Tensometers and Inertial Sensors. *IEEE Access* **2019**, *7*, 164120–164126. [CrossRef]
13. Szermer, M.; Zając, P.; Amrozik, P.; Maj, C.; Jankowski, M.; Jabłoński, G.; Kiełbik, R.; Nazdrowicz, J.; Napieralska, M.; Sakowicz, B. A Capacitive 3-Axis MEMS Accelerometer for Medipost: A Portable System Dedicated to Monitoring Imbalance Disorders. *Sensors* **2021**, *21*, 3564. [CrossRef] [PubMed]
14. Maluf, N.; Williams, K. *Introduction to Miroelectromechanical Systems Engineering*, 2nd ed.; Artech: Singapore, 2004.
15. Zhu, X. Post-CMOS Micromachining of Surface and Bulk Structures. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2002.
16. French, P.J.; Sarro, P.M. Surface versus bulk micromachining: The contest for suitable applications. *J. Micromech. Microeng.* **1998**, *8*, 45. [CrossRef]
17. Leinenbach, C.; Kattelus, H.; Knechtel, R. Chapter 19—Surface micromachining. In *Handbook of Silicon Based MEMS Materials and Technologies*, 3rd ed.; Tilli, M., Paulasto-Krockel, M., Petzold, M., Theuss, H., Motooka, T., Lindroos, V., Eds.; Micro and Nano Technologies; Elsevier: Amsterdam, The Netherlands, 2010; pp. 503–517. [CrossRef]
18. Esteve, D.; Alderman, J.; Cane, C.; Ciurtois, B.; Glesner, M.; Napieralski, A.; Rencz, M.; Samitier, J.; Troccaz, J.; Cinquin, P.; Dillman, J. *Basic Research for Microsystems Integration*; Cepadues-Editions: Toulouse, France, 1997.
19. Rao, A.V.N.; Pal, P.; Pandey, A.K.; Menon, P.K.; Tanaka, H.; Sato, K. High Speed Silicon Wet Bulk Micromachining of Si111 in KOH Based Solution. In Proceedings of the 2020 Symposium on Design, Test, Integration and Packaging of MEMS and MOEMS (DTIP), Lyon, France, 15–26 June 2020; pp. 1–5. [CrossRef]

20. Menz, W. Liga And Related Technologies For Industrial Application. In Proceedings of the International Solid-State Sensors and Actuators Conference—TRANSDUCERS'95, Stockholm, Sweden, 25–29 June 1995; Volume 1, pp. 552–555. [CrossRef]
21. X-FAB Prototyping Support. 2025. Available online: https://www.xfab.com/manufacturing/prototyping (accessed on 18 March 2025).
22. A Lam Research, C. Introduction to MEMS+. 2025. Available online: https://www.coventor.com/courses/introduction-to-mems/ (accessed on 18 March 2025).
23. MathWorks. MATLAB Simulink. 2025. Available online: https://www.mathworks.com/products/simulink.html (accessed on 18 March 2025).
24. CADENCE IC Design Suite. 2025. Available online: https://www.cadence.com/en_US/home/tools/custom-ic-analog-rf-design.html (accessed on 18 March 2025).
25. MathWorks. MATLAB Simulink—Reduced Order Modeling. 2025. Available online: https://www.mathworks.com/discovery/reduced-order-modeling.html (accessed on 18 March 2025).
26. Lemkin, M.; Boser, B. A three-axis micromachined accelerometer with a CMOS position-sense interface and digital offset-trim electronics. *IEEE J. Solid-State Circuits* **1999**, *34*, 456–468. [CrossRef]
27. Matsumoto, Y.; Nishimura, M.; Matsuura, M.; Ishida, M. Three-axis SOI capacitive accelerometer with PLL C-V converter. *Sens. Actuators A Phys.* **1999**, *75*, 77–85. [CrossRef]
28. Chae, J.; Kulah, H.; Najafi, K. A monolithic three-axis micro-g micromachined silicon capacitive accelerometer. *J. Microelectromech. Syst.* **2005**, *14*, 235–242. [CrossRef]
29. InvenSense. MPU-6000 and MPU-6050 Product Specification Revision 3.3. 2012. Available online: https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf (accessed on 13 May 2025).
30. Liu, Y.; Tsai, M.; Li, S.; Fang, W. A Fully-Differential, Multiplex-Sensing Interface Circuit Monolithically Integrated with Tri-Axis Pure Oxide Capacitive CMOS-MEMS Accelerometers. In Proceedings of the 17th International Conference on Solid-State Sensors, Actuators and Microsystems, Barcelona, Spain, 16–20 June 2013.
31. Tez, S.; Aykutlu, U.; Torunbalci, M.; Akin, T. A Bulk-Micromachined Three-Axis Capacitive MEMS Accelerometer on a Single Die. *J. Microelectromech. Syst.* **2015**, *24*, 1264–1274. [CrossRef]
32. Devices, A. ADXL345 Data Sheet. 2022. Available online: https://www.analog.com/media/en/technical-documentation/data-sheets/adxl345.pdf (accessed on 13 May 2025).
33. ANSYS, a CAD Software Suite. 2025. Available online: https://www.ansys.com/ (accessed on 9 June 2025).
34. COMSOL Multiphysics. 2025. Available online: https://www.comsol.com/ (accessed on 9 June 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

Wearable Sensor-Based Analysis of Human Biomechanics in Manual and Cobot-Assisted Agricultural Transplanting

Yuetong Wu [1], Xiangrui Wang [2] and Boyi Hu [1,*]

[1] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA; wuy2@ufl.edu
[2] Department of Mechanical, Materials and Aerospace Engineering, West Virginia University, Morgantown, WV 26506, USA; xw00016@mix.wvu.edu
* Correspondence: boyihu@ise.ufl.edu

Abstract: Work-related musculoskeletal disorders (WMSDs) are common in the agricultural industry due to repetitive tasks, like plant transplanting, which involve sustained bending, squatting, and awkward postures. This study uses wearable sensors to evaluate human biomechanics during simulated transplanting and assesses the potential of collaborative robot (cobot) assistance to reduce physical strain. Sixteen participants performed transplanting tasks under manual and cobot-assisted conditions. Kinematic and electromyographic (EMG) data were collected using Xsens motion capture and Trigno EMG systems. Cobot assistance significantly reduced the segment velocity and acceleration in key spinal regions (L5/S1, L1/T12, T1/C7), indicating lower dynamic spinal loading. It also altered muscle activation, decreasing biceps brachii use while increasing activation in stabilizing muscles such as the flexor carpi radialis, brachioradialis, and upper trapezius. Task duration decreased by 59.46%, suggesting improved efficiency. These findings highlight cobots' potential to enhance ergonomic outcomes by encouraging controlled movements and reducing postural stress. However, the shift in muscle activation underscores the need for task-specific cobot tuning. This research supports the use of integrated IMU and EMG systems to inform cobot design and enable real-time biomechanical monitoring in labor-intensive settings.

Keywords: collaborative robotics; human–robot interaction (HRI); wearable sensors; agricultural ergonomics; biomechanical analysis; musculoskeletal strain reduction

1. Introduction

Work-related musculoskeletal disorders (WMSDs) have become a prevalent issue in the agricultural industry, particularly due to the physically demanding nature of tasks that require repetitive bending and prolonged awkward postures. These biomechanical stressors place workers at a heightened risk of developing low back pain (LBP), one of the most reported WMSDs in this sector [1]. The physical burden associated with these tasks is exacerbated by prolonged exposure to high-risk postures, repetitive force exertions, and external loads, all of which contribute to musculoskeletal fatigue and injury over time. Kang et al. [2] investigated WMSD prevalence and risk factors among Korean farmers, finding that over 26% reported back-related WMSDs. Similarly, Shivakumar et al. [3] found that the 12-month pooled prevalence of LBP among agricultural workers was 61.96% in Africa, 54.16% in Asia, and 28.52% in South/Central America. These findings underscore the global nature of this issue and the need for targeted ergonomic interventions to address the unique demands of agricultural labor, a sector that remains underserved

by conventional ergonomic solutions [4–6]. Addressing these challenges is particularly critical as agricultural workers often experience longer-than-average work hours with little opportunity for physical recovery, further compounding their risk of injury.

In recent years, the human–robot interaction (HRI) has gained increasing attention as a viable solution to help alleviate physical strain and cognitive workload in labor-intensive industries [7,8]. Collaborative robots, or cobots, are designed to work alongside humans, augmenting their capabilities and reducing high-risk postures. Seyyedhasani et al. [9] developed a simulation model to optimize collaboration between human workers and robots for high-volume harvesting, demonstrating that cobots can increase efficiency and reduce physical demands by minimizing non-productive walking times. In terms of agricultural settings, Adamides et al. [10] presented a semi-autonomous agricultural robotic sprayer that is designed to work in collaboration with a human teleoperator. Similarly, Yerebakan et al. [11] investigated the biomechanical and cognitive workload effects of HRIs in agricultural pollination tasks, highlighting the potential of HRIs to mitigate physical and mental fatigue. While these studies establish a foundation for integrating robotic systems in the agricultural sector, further research is needed to explore their impact on the ergonomic challenges faced in specialized, small-scale farming operations.

While fully automated plant-transplanting systems have become increasingly prevalent in large-scale industrial farming, their applicability remains limited in several key scenarios. Small-scale farming operations, especially those with limited financial resources, often lack the infrastructure and capital required to implement and maintain fully automated systems [12]. In these contexts, human labor remains indispensable, yet the physical demands of manual tasks exacerbate the risk of WMSDs [5]. Cobots provide a middle-ground solution, combining the adaptability of human labor with the mechanical assistance of automation. While not a replacement for fully manual or fully automated approaches, they offer a flexible alternative that can reduce physical demands in certain agricultural settings with lower investment requirements than large-scale automation [13,14]. To rigorously evaluate such alternatives, especially in the early stages of cobot integration, robust sensor-based approaches are essential. Advances in wearable sensor systems, particularly inertial measurement units (IMUs) and surface electromyography (sEMG), now allow for detailed and quantitative analysis of human movement and muscle activation during task execution. These technologies provide precision, portability, and real-time feedback necessary to investigate ergonomic performance in dynamic work environments, including the agriculture sector.

Plant transplanting presents unique ergonomic challenges due to its repetitive nature and the need for workers to maintain sustained postural deviations. These challenges place considerable biomechanical strain on the lower back, particularly in manual transplanting tasks that require prolonged kneeling, squatting, or trunk flexion. Such conditions contribute to the high prevalence of WMSDs among agricultural workers, especially in small-scale and traditional farming settings [15,16]. While recent studies have demonstrated the potential of the human–robot interaction (HRI) in reducing physical burden across industrial applications [17,18], its application to fine-grained, labor-intensive agricultural tasks like plant transplanting remains largely unexplored.

To address this gap, our research group is actively developing a generalizable, mobile agricultural cobot platform, which has shown early success in controlled-environment pollination tasks [19–21]. However, extending this system to support plant transplanting requires a nuanced understanding of the biomechanical demands and physical strategies currently used by human workers. Unlike in prior studies, our approach does not begin with robotic intervention but rather seeks to characterize plant-transplanting-specific motion and posture data that can inform future cobot design. Specifically, through this study,

we aim to (1) collect foundational human-centered data on segment velocities, joint accelerations, and muscle activation patterns under normal manual transplanting conditions; and (2) evaluate whether cobot integration can improve human postures and movement profiles that reduce physical load while maintaining task efficiency. The findings will not only guide controller development and adaptive behavior strategies for plant-transplanting cobots but will also contribute empirical insights to the broader HRI-in-agriculture literature. By bridging human biomechanics with collaborative robotics, we lay the groundwork for safer, more efficient cobot-supported farming systems in the future.

2. Materials and Methods

2.1. Apparatus

The experimental setup (Figure 1) was designed to simulate a plant-transplanting task in a controlled laboratory environment. The apparatus included four pots positioned on the ground at angles of 0°, 30°, 60°, and 90° relative to the participant's body to mimic potential real-world plant-transplanting conditions and postures. Small artificial succulent plants were used as plant substitutes, which participants placed into the pots during the task. Participants were instructed to stand with their feet shoulder-width apart, ensuring a stable base throughout the experiment. Each pot was placed 45 cm away from the participant's feet, ensuring a consistent reach distance across all angle variations. Kinematics data were collected using the Xsens Awinda system (Xsens, Enschede, The Netherlands), which consisted of 17 inertial measurement unit (IMU) sensors placed on the participant's body according to the manufacturer's guidelines. The system recorded segment velocities and accelerations at a sampling rate of 60 Hz. Muscle activation was measured using Trigno EMG sensors (Delsys Inc., Boston, MA, USA), which were placed on the flexor carpi radialis (FCR), brachioradialis (BR), biceps brachii (BB), and upper trapezius (UT) muscles.

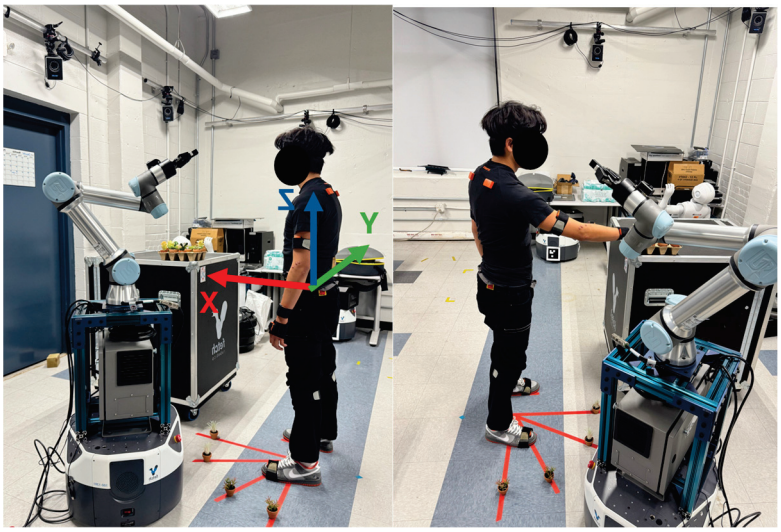


Figure 1. Lab-Based-Simulation Experimental Setup.

This study employed a UR5 robot (Universal Robots, Odense, Denmark) and a Fetch mobile base (Fetch Robotics, San Jose, CA, USA) as the cobot assistant to support plant-transplanting tasks (Figure 1). The UR5-based setup was used as the proxy of the Stickbug robot [21], with comparable dimensions and major design features, and was operated as the Wizard of Oz method. In this study, the cobot was operated without autonomous motion planning. While the robot did not execute real-time path planning, its pose and positioning were selected to reflect plausible ergonomic placements. The biomechanical

data collected here will inform future control strategies that can adapt robot trajectories to minimize dynamic spinal loading and maximize human comfort.

2.2. Participants

A total of 16 participants (8 males and 8 females) were recruited for this study from the local university and community. All participants reported good health and had no history of musculoskeletal disorders. The inclusion criteria required no dysfunction or recent injuries that could affect their physical capabilities and task completion. The study received approval from the University of Florida Institutional Review Board (IRB), and all participants provided informed consent prior to the experiment. Detailed characteristics of the subjects are summarized in Table 1.

Table 1. Participant's Demographic Information (N = 16).

Variables	
Age (yrs)	23.1 ± 2.43
Height (cm)	173.8 ± 10.2

2.3. Experimental Design

There was a single independent variable in the current study, which was the HRI mode. HRI modes included manual transplanting, where participants performed tasks independently. In the robot-assisted transplanting mode, a collaborative robot passively held the plants at a consistent, ergonomically favorable height. This condition was not intended to test a finalized robotic control strategy, but rather to allow for the collection of human biomechanical data across different postures. These data serve both to characterize current biomechanical demands in manual transplanting and to inform the development of future cobot motion planning algorithms that optimize ergonomic outcomes based on bending, twisting, and physical strain during transplanting tasks. Details of the mechanical platform and planned control framework will be presented in a separate manuscript. Dependent variables were grouped into two categories: (1) segment velocities and accelerations for L5/S1, T1/C7, and L1/T12, and the pelvis in the x-, y-, and z-axes. The x-axis represents the anteroposterior direction, with positive values indicating forward movement and negative values indicating backward movement. The y-axis corresponds to the mediolateral direction, where positive values represent rightward movement (having the participants facing forward) and negative values represent leftward movement. The z-axis aligns with the vertical direction, with positive values indicating upward movement and negative downward values; (2) muscle activation, measured via EMG from the flexor carpi radialis, brachioradialis, biceps brachii, and upper trapezius. These metrics were selected to assess the biomechanical and physiological impacts on participants. The inclusion of four pot angles (0°, 30°, 60°, and 90°) was intentional to simulate the spatial variability encountered in real-world transplanting environments. Although the angles used in this study do not capture the full irregularity of real-world terrain and plant spacing, they offer a controlled way to simulate variations in the reaching, twisting, and bending required in field transplanting. The chosen configurations reflect common planting layouts and introduce lateral asymmetry that approximates realistic planting conditions while maintaining repeatability. Limiting the study to a single pot angle would have simplified the protocol but failed to represent the asymmetrical and dynamic motions typical of actual agricultural work. Training a controller using biomechanical data derived only from symmetric, forward-reaching postures would hinder its ability to adapt to diverse field configurations. By incorporating multiple angles, this study ensures that both the characterization of human movement and the future cobot behaviors are grounded in realistic and varied transplanting scenarios.

2.4. Protocol

Upon arrival, participants were briefed on the study's objectives and procedures before providing informed consent. Seventeen inertial measurement unit (IMU) sensors were then positioned on each participant following the manufacturer's guidelines, as shown in Figure 2. Anthropometric data were collected, and participants underwent a calibration process using the Xsens MVN Analyze/Animate software (Version 2021.0.1). Calibration consisted of a static pose held for five seconds, followed by a brief walking sequence to ensure precise motion capture alignment. Next, Trigno EMG sensors were attached to the participants, as shown in Figure 3, to measure muscle activity. The maximum voluntary contraction (MVC) of each muscle was assessed following established guidelines [22] to normalize muscle activation data across individuals.



Figure 2. Xsens Sensor Placements.

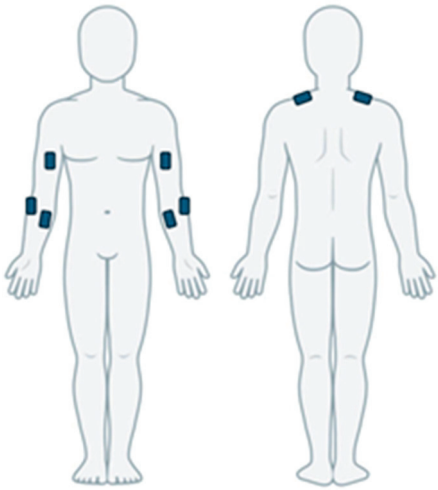


Figure 3. EMG Sensor Placements.

In the manual plant-transplanting task, participants were required to stoop to place a small plant into each of the four pots positioned on the ground at angles of 0°, 30°, 60°, and 90° relative to their body. After placing a plant in each pot, participants returned to a neutral standing posture before stooping again to reach the next pot. This required participants to repeatedly stoop, return to neutral, and stoop again until all four pots were planted, simulating the repetitive bending motions commonly involved in manual plant-transplanting tasks. In contrast, in the robot-assisted mode, participants were asked to stoop once to plant all four pots without returning to a neutral posture. The robot held the

plant container in an ergonomically friendly position, allowing participants to reach each pot sequentially without the need for repetitive stooping and standing, thereby reducing the physical demands of the task.

Each plant-transplanting task was repeated twice to ensure data reliability, resulting in 4 trials per participant (2 HRI modes × 2 repetitions). Each trial lasted for around 20 s, during which motion and muscle data were recorded. The total duration of the experiment lasted around 1 h.

2.5. Data Processing and Statistical Analysis

Kinematic data were collected using the Xsens Awinda system and processed through its built-in software. Participant movements were recorded at a frame rate of 60 Hz, and the processed data were exported for further analysis. The final kinematic values for each condition were obtained by computing the mean across the two repeated trials. The normality of the datasets was assessed prior to statistical testing. EMG signals were pre-processed using MATLAB R2018b. The DC offset was removed, and the signals were filtered using a 60 Hz notch filter and a fourth-order Butterworth band-pass filter (15–450 Hz) to eliminate environmental noise and retain the relevant frequency content. Full-wave rectification was applied to the signals, followed by smoothing with a fourth-order Butterworth low-pass filter at a cutoff frequency of 6 Hz to extract the linear envelope. The signals were then normalized to the maximum value of the maximum voluntary contraction (MVC) for each participant. Out of the 16 participants, EMG data were recorded and analyzed for 10 participants due to an unexpected hardware issue. A one-way ANOVA was then conducted to compare the mean values between conditions, yielding F-statistics and p-values to determine the significance of differences. A significance threshold of $p < 0.05$ was chosen to determine statistical significance.

3. Results

3.1. Segment Velocity and Acceleration

Cobot assistance significantly reduced segment velocities across key spinal regions, with notable differences observed in the L5/S1, L1/T12, and T1/C7 joints (Table 2, Figure 4). At the L5/S1 joint, velocity in the y-axis decreased significantly from 10.1 to 5.5 cm/s (F = 8.207, p = 0.006), with no significant differences in other axes. L1/T12 exhibited significant reductions in the y-axis velocity from 12.5 to 6.9 cm/s (F = 7.561, p = 0.01), and from 11.9 to 9.7 (F = 5.973, p = 0.017) cm/s in the z-axis, while no significant changes were noted in the x-axis. At T1/C7, the velocity in the y-axis decreased significantly from 16.1 to 9.4 cm/s (F = 6.987, p = 0.011), and the z-axis velocity showed a significant reduction from 18.2 to 14.3 cm/s (F = 14.150, $p < 0.001$). At the pelvis, the mean velocity in all three axes decreased but the values showed no significant differences.

Table 2. Segment Velocity of Joints.

Segment Velocity	**Joints**		
		Pelvis	
	x	y	z
Manual Mean ± SD (cm/s)	4.3 ± 0.4	3.1 ± 0.4	1.5 ± 0.2
Robot Mean ± SD (cm/s)	3.0 ± 4.0	2.5 ± 3.0	1.1 ± 1.1
F-Statistic	2.949	1.263	1.36
p-value	0.091	0.265	0.248
Power	0.831	0.508	0.5

Table 2. *Cont.*

Segment Velocity	**Joints**		
		L5S1	
	x	y	z
Manual Mean ± SD (cm/s)	7.0 ± 1.4	**10.1 ± 1.6**	8.1 ± 1.4
Robot Mean ± SD (cm/s)	5.8 ± 5.2	**5.5 ± 5.4**	6.2 ± 5.3
F-Statistic	1.48	**8.207**	2.956
p-value	0.228	**0.006 ***	0.091
Power	0.607	**0.989**	0.841
		L1T12	
	x	y	z
Manual Mean ± SD (cm/s)	7.6 ± 1.8	**12.5 ± 2**	**11.9 ± 1.3**
Robot Mean ± SD (cm/s)	7.2 ± 6.5	**6.9 ± 7.0**	**9.7 ± 9.8**
F-Statistic	0.11	**7.561**	**5.973**
p-value	0.742	**0.01 ***	**0.017 ***
Power	0.091	**0.978**	**0.992**
		T1C7	
	x	y	z
Manual Mean ± SD (cm/s)	11.2 ± 2.4	**16.1 ± 2.7**	**18.2 ± 1.4**
Robot Mean ± SD (cm/s)	10.5 ± 10.7	**9.4 ± 10.3**	**14.3 ± 15.7**
F-Statistic	0.16	**6.987**	**14.150**
p-value	0.691	**0.011 ***	**<0.001 ***
Power	0.117	**0.975**	**1**

The bolded values mean that those are significant ($p < 0.05$). An * is placed next to values that are statistically significant ($p < 0.05$).

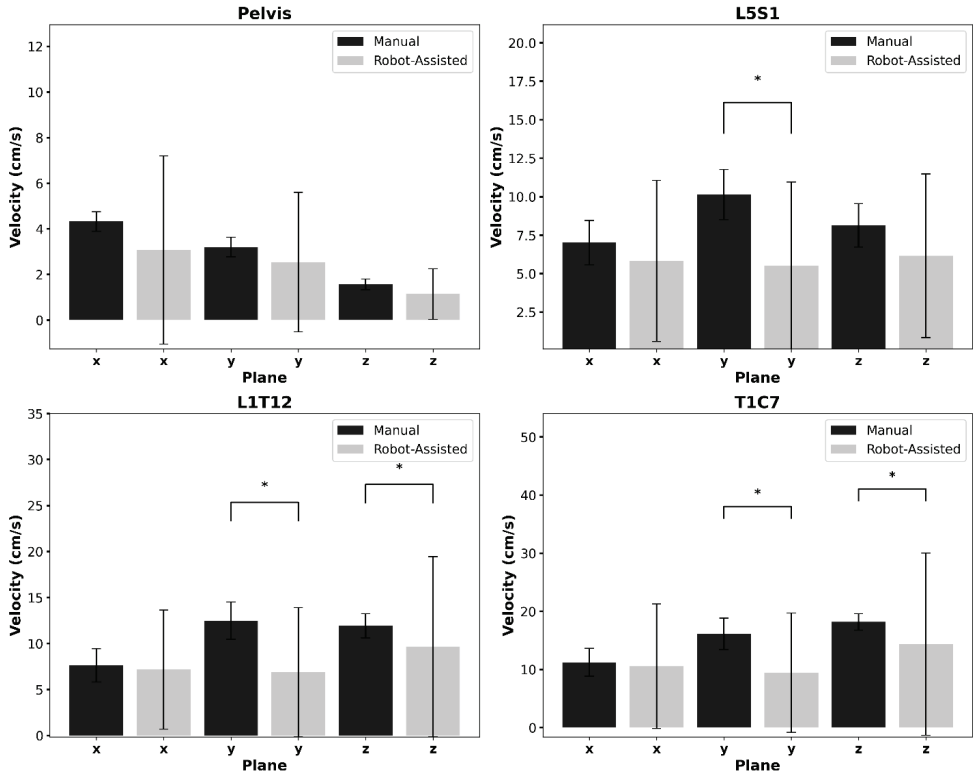


Figure 4. Graph of Segment Velocity for Different Joints. An * means a *p*-value of <0.05.

The effects of cobot assistance on segment acceleration were similar across spinal regions (Table 3, Figure 5). At the L5/S1 joint, the *y*-axis showed a significant reduction from 15.6 to 11.8 cm/s^2 (F = 6.872, $p = 0.005$), with no significant changes in the *x*- and

z-axes. At L1/T12, acceleration decreased significantly in the *y*-axis from 26.6 to 20.3 cm/s^2 (F = 7.143, p = 0.001) and in the *z*-axis from 32.5 to 24.4 cm/s^2 (F = 4.752, p = 0.027), while the *x*-axis remained unchanged. At T1/C7, significant reductions in acceleration were observed in the *y*-axis from 32.7 to 24.0 cm/s^2 (F = 7.068, p = 0.005) and in the *z*-axis from 44.5 to 32.3 cm/s^2 (F = 12.670, p = 0.001), with no significant changes in the *x*-axis. At the pelvis, the mean acceleration in all three axes showed no significant differences.

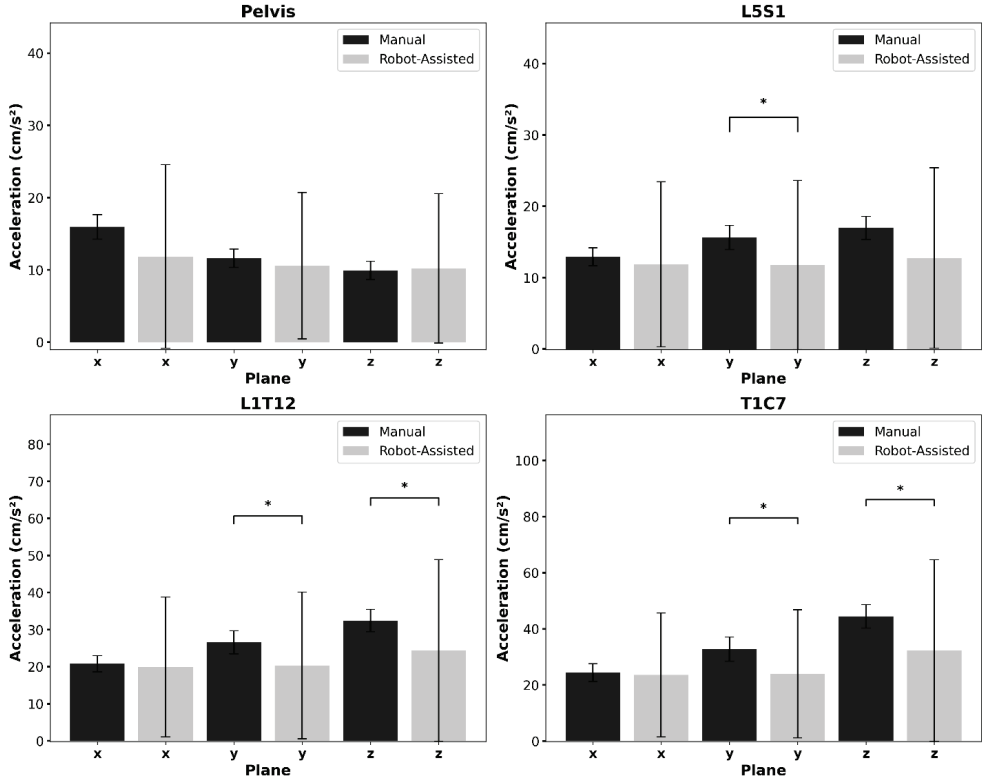


Figure 5. Graph of Segment Acceleration for Different Joints. An * means a p-value of <0.05.

Table 3. Segment Acceleration of Joints.

Segment Acceleration		**Joints**	
		Pelvis	
	x	y	z
Manual Mean ± SD (cm/s^2)	15.96 ± 1.7	11.61 ± 1.27	9.95 ± 1.26
Robot Mean ± SD (cm/s^2)	11.85 ± 12.71	10.59 ± 10.13	10.22 ± 10.34
F-Statistic	1.984	0.176	0.015
p-value	0.164	0.676	0.903
Power	0.685	0.122	0.056
		L5S1	
	x	y	z
Manual Mean ± SD (cm/s^2)	12.9 ± 1.3	**15.6 ± 1.7**	17 ± 1.6
Robot Mean ± SD (cm/s^2)	11.9 ± 11.6	**11.8 ± 11.9**	12.8 ± 12.6
F-Statistic	2.946	**6.872**	1.805
p-value	0.217	**0.005 ***	0.109
Power	0.61	**0.994**	0.811
		L1T12	
	x	Y	z
Manual Mean ± SD (cm/s^2)	20.8 ± 2.2	**26.6 ± 3.1**	**32.5 ± 3**
Robot Mean ± SD (cm/s^2)	19.9 ± 18.8	**20.3 ± 19.8**	**24.4 ± 24.5**
F-Statistic	0.1	**7.143**	**4.752**
p-value	0.712	**0.001 ***	**0.027 ***
Power	0.183	**1**	**0.894**

Table 3. *Cont.*

Segment Acceleration	**Joints**		
		T1C7	
	x	y	z
Manual Mean ± SD (cm/s^2)	24.4 ± 3.1	**32.7 ± 4.3**	**44.5 ± 4.2**
Robot Mean ± SD (cm/s^2)	23.6 ± 21.1	**24 ± 22.9**	**32.3 ± 32.4**
F-Statistic	0.1	**7.068**	**12.67**
p-value	0.701	**0.005 ***	**<0.001 ***
Power	0.154	**0.96**	**0.977**

The bolded values mean that those are significant ($p < 0.05$). An * is placed next to values that are statistically significant ($p < 0.05$).

3.2. Muscle Activation Patterns

Cobot assistance led to a significant reduction in the activation of the BB muscle during stooping, decreasing from 12.09 µV (±6.9) to 10.72 µV (±6.5) (F = 6.29, $p = 0.018$). Conversely, activation significantly increased with cobot assistance for the FCR muscle (8.71 to 12.97 µV, F = 7.15, $p = 0.012$). Similarly, for the BR muscle, activation increased significantly (9.93 to 13.0 µV, F = 7.4, $p = 0.01$). For the UT muscle, significant increases were observed from 46.5 to 59.23 µV (F = 11.233, $p = 0.002$) (Table 4, Figure 6).

Table 4. iEMG Values for Different Muscles.

Muscle					
	Manual	Robot-Assisted	F-Statistic	*p*-value	Power
FCR	8.71 (6.7)	12.97 (13.8)	7.15	0.012 *	0.086
BR	9.93 (6.9)	13.0 (9.8)	7.40	0.01 *	0.088
BB	12.09 (6.9)	10.72 (6.5)	6.29	0.018 *	0.079
UT	46.5 (35.6)	59.23 (45.7)	11.233	0.002 *	0.122

An * is placed next to values that are statistically significant ($p < 0.05$).

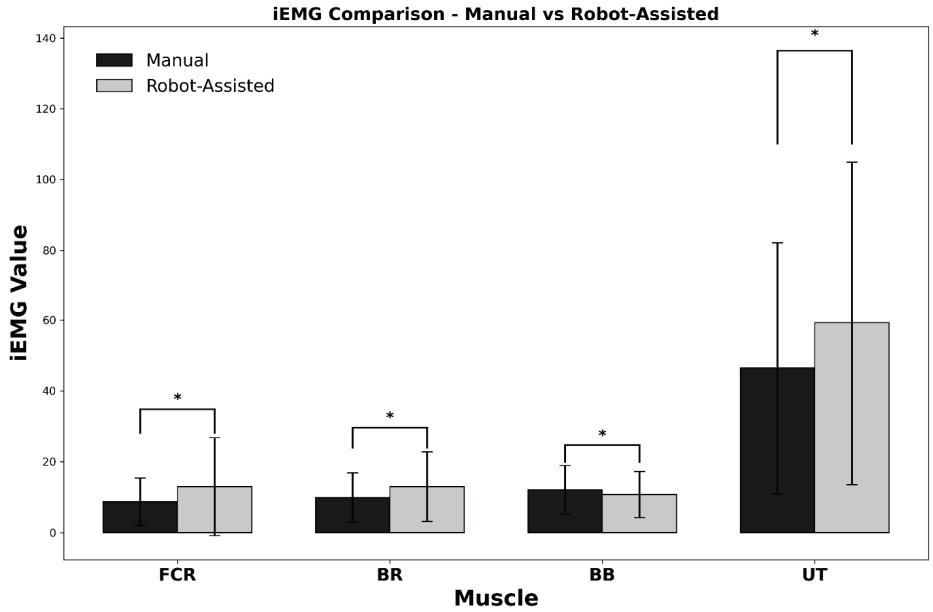


Figure 6. Graph of iEMG Values for Different Muscles. An * means a *p*-value of <0.05.

3.3. Task Performance Parameters

Task performance analysis revealed that the cobot-assisted condition reduced the overall task duration by 59.85% decreasing from 74.32 s (±14.09) to 29.84 s (±4.68). In the manual condition, 54.49% (±13.21%) of movements were classified as low risk (<20°),

9.28% (±3.32%) as moderate risk (20–45°), and 36.23% (±10.71%) as high risk (>45°). In the cobot-assisted condition, 16.61% (±5.49%) of movements were classified as low risk, 4.42% (±0.84%) as moderate risk, and 78.97% (±6.08%) as high risk. Based on prior research, we adopted 20° and 45° as critical cutoffs for evaluating trunk and joint posture risks in physically demanding tasks [23,24]. This analysis specifically pertains to vertical pelvis flexion/extension, which is the primary focus of this study as it represents stooping postures relevant to trunk flexion–extension.

Cobot assistance led to a significant reduction in the time spent in low-risk postures, decreasing from 39.44 ± 8.70 s in the manual condition to 4.79 ± 1.04 s (F = 140.61, $p < 0.001$). Similarly, the time spent in moderate-risk postures significantly decreased from 7.00 ± 3.25 s to 1.30 ± 0.20 s with cobot assistance (F = 27.62, $p < 0.001$). In contrast, the time spent in high-risk postures slightly decreased from 27.88 ± 12.37 s in the manual condition to 23.76 ± 5.05 s; however, this difference was not statistically significant (F = 0.86, $p = 0.367$). Detailed results are summarized in Table 5.

Table 5. Stooping Time Spent in Various Categories during Manual and HRI modes.

Body Posture	Manual		Robot-Assisted		Comparison
	% Time ± SD	Time (s) ± SD	% Time	Time (s) ± SD	
Pelvis Neutral	54.49% ± 13.21%	39.44 ± 8.70	16.61% ± 5.49%	4.79 ± 1.04	F-Statistic—140.61 *p*-value—<0.001 * Power—1
Pelvis Moderate Flexion (>20°)	9.28% ± 3.32%	7.00 ± 3.25	4.42% ± 0.84%	1.30 ± 0.20	F-Statistic—27.62 *p*-value—<0.001 * Power—1
Pelvis Severe Flexion (>45°)	36.23% ± 10.71%	27.88 ± 12.37	78.97% ± 6.08%	23.76 ± 5.05	F-Statistic—0.86 *p*-value—0.3668 Power—0.2299

An * is placed next to comparisons that are statistically significant ($p < 0.05$).

4. Discussion

The integration of cobots into agricultural tasks has the potential to mitigate some of the ergonomic risks associated with repetitive bending, side bending, or twisting, though their effectiveness depends on task complexity and user adaptability. Our prior research has demonstrated the potential of cobots to reduce awkward postures and excessive joint angles during pollination operations, thereby enhancing ergonomics and improving worker safety and comfort in HRI systems [11]. Building on this foundation, the present study examines the biomechanical effects of cobot-assisted plant transplanting, with a specific focus on joint kinematics as well as EMG activity. By reframing the analysis from merely evaluating cobot utility to exploring movement strategies in repetitive transplanting tasks, the findings contribute baseline data for developing ergonomically adaptive cobots. This study uniquely demonstrates how wearable IMU and EMG sensors can capture high-resolution human biomechanics in agricultural tasks, providing a replicable approach for evaluating cobot-assisted interventions beyond traditional lab setups.

For segment velocities across the pelvis, and L5/S1, L1/T12, and T1/C7 joints, all showed decreases across the x, y, and z planes (Table 2), with a portion being statistically significant. This reduction is particularly relevant from a biomechanical perspective, as lower segment velocities can potentially minimize dynamic spinal loading, thereby reducing excessive shear and compression forces that contribute to musculoskeletal strain and injury. Davis and Marras [25] conducted a comprehensive review of the effects of trunk motion on musculoskeletal loading and injury risk, demonstrating that increased movement velocity leads to greater muscle coactivation, higher spinal loads, and reduced trunk strength—factors known to elevate the risk of LBP. Their findings identified three-

dimensional trunk velocity as a significant risk factor, as higher velocities were directly associated with increased spinal compression and shear forces. Similarly, Marras et al. [26] reported that elevated trunk velocities amplify trunk muscle coactivation and lateral shear forces, further exacerbating musculoskeletal strain. More recent studies have reinforced these findings. Allread and Vossenas [27] examined trunk motions and LBP risk among hotel-room cleaners, demonstrating that an increased speed of work was associated with higher spinal loading and an increased likelihood of developing LBDs. These findings reinforce the significance of the present study, where the observed reductions in segment velocity across spinal regions suggest a potential ergonomic advantage of cobot assistance. By lowering segment velocity, cobot-assisted interventions may contribute to reduced spinal loading and mitigate cumulative strain, particularly in highly repetitive agricultural tasks.

Compared to segment velocity, segment acceleration also exhibited variation across most planes and joints following the introduction of the cobot (Table 3). This suggests that cobot assistance may help regulate movement, reducing the likelihood of abrupt changes in acceleration that could contribute to increased mechanical loading on the spine. Tsang et al. [28] investigated the effects of different bending speeds on lumbo-pelvic kinematics and found that individuals with LBP exhibited compromised control over segment acceleration, particularly at slower speeds, leading to less efficient movement coordination. Their study demonstrated that uncontrolled acceleration changes increased mechanical demand on the lumbar spine, reinforcing the importance of regulating segment accelerations to reduce spinal stress. Similarly, in the present study, the reduction in segment acceleration suggests that cobot assistance maintained controlled and stable movement patterns, potentially mitigating sudden mechanical stress on the spine and reducing the risk of cumulative strain during repetitive tasks. The findings align with prior research emphasizing that higher movement accelerations contribute to increased spinal loading, leading to potential long-term musculoskeletal strain. By moderating movement acceleration, cobot assistance may provide an ergonomic advantage in reducing the risk of low back disorders (LBDs), particularly in physically demanding tasks that require repetitive trunk flexion.

While reductions in segment velocity and acceleration indicate that cobot assistance may reduce dynamic spinal loading, EMG analysis revealed increased activation in the flexor carpi radialis (FCR), brachioradialis (BR), and upper trapezius (UT) muscles, suggesting a redistribution rather than a universal reduction in muscular demands (Table 4). In contrast, activation of the biceps brachii (BB) decreased significantly from 12.09 ($\pm$6.9) to 10.72 ($\pm$6.5) ($F = 6.29$, $p = 0.018$), suggesting that cobot support may reduce the reliance on primary movers while shifting muscular demand to the stabilizing and fine-motor-control muscles. A similar redistribution of muscle activity has been reported in exoskeleton-assisted tasks, where external support alleviates the load on primary muscles but leads to increased activation in the stabilizing muscles due to postural adjustments and task-specific demands [29]. It was demonstrated that exoskeletons reduced trunk muscle activation in prolonged surgical postures but did not alleviate shoulder and neck muscle strain, highlighting the complex neuromuscular adaptations that occur with assistive devices. This shift may explain why FCR, BR, and UT activation increased despite the reduction in segment velocity. Notably, the flexion–relaxation phenomenon (FRP), which typically allows for muscle relaxation during certain movement phases, has been shown to be disrupted by assistive devices, leading to sustained muscle activation [30]. This redistribution pattern suggests that while the cobot reduced gross movement demands (e.g., for biceps brachii), it simultaneously increased the need for fine motor and stabilizer muscle activity to manage reach coordination and postural control during sequential pot placements. While this redistribution may alleviate overuse in primary movers, prolonged activation of stabilizers

like the UT and BR may lead to localized fatigue. Chronic overuse of these muscles could contribute to discomfort or even overuse injury if not managed with appropriate rest cycles or cobot control adjustments. Longitudinal studies are needed to determine the cumulative effects of these activation patterns.

4.1. Impact of Cobot Assistance on Task Dynamics and Efficiency

The robot-assisted approach fundamentally alters task dynamics by allowing participants to depend on the robot for assistance, leading to greater variability in segment velocity and accelerations as they adapt to the robot's presence. This variability indicates less-consistent movement patterns, as the robot reduces some of the biomechanical demands. In contrast, manual tasks enforce more uniform motion patterns due to their repetitive nature and the direct control exerted by participants. Research has shown that adaptive robotic systems not only modify physical behaviors but also enhance task efficiency by decreasing task completion time and improving team fluency [31]. In our study, the cobot reduced the total task time by an average of 59.46%, a reduction that significantly outweighed the observed increases in iEMG values. Similarly, Liu et al. [32] found that while cobots enhance efficiency, they may redistribute physical demands rather than uniformly reduce them. These findings highlight the potential of cobot assistance to not only balance physical demands but also improve overall task performance in repetitive, collaborative environments [33,34].

Although the cobot-assisted condition resulted in a higher percentage of movements classified as high-risk postures, this increase should be interpreted in the context of overall task efficiency and exposure duration. While participants spent a larger proportion of time in high-risk postures during the cobot-assisted tasks (78.97% $\pm$ 6.08%) compared to the manual tasks (36.23% $\pm$ 10.71%), the shorter task duration plays a crucial role in mitigating prolonged exposure to these postures. When examining the absolute time spent in each risk category, participants in the manual condition spent an average of 39.44 $\pm$ 8.70 s in low-risk postures, 7.00 $\pm$ 3.25 s in moderate-risk postures, and 27.88 $\pm$ 12.37 s in high-risk postures. In contrast, the cobot-assisted condition significantly reduced time spent in low- and moderate-risk categories to 4.79 $\pm$ 1.04 s ($F = 140.61$, $p < 0.001$) and 1.30 $\pm$ 0.20 s ($F = 27.62$, $p < 0.001$), respectively. Although the percentage of high-risk postures increased, the absolute time spent in high-risk postures still decreased to 23.76 $\pm$ 5.05 s in the cobot-assisted condition, albeit not significantly ($F = 0.86$, $p = 0.367$). Our results demonstrate that, while cobot integration alters posture dynamics, it significantly reduces the overall task duration and moderates segmental motion. These findings support the second objective of this study, showing that cobot assistance can improve movement efficiency and potentially reduce long-term physical strain. Taken together with the detailed segmental and muscular analyses, this study offers foundational biomechanical data on human movement strategies in plant transplanting, fulfilling our first objective and paving the way for informed cobot control design and future HRI optimization.

4.2. Limitations

This study provides evidence supporting the potential benefits of cobot assistance in reducing biomechanical strain during plant-transplanting tasks. However, it is important to acknowledge that manual labor remains valuable in certain agricultural contexts. The controlled laboratory environment, while ideal for precise data collection, does not fully capture the variability of real-world farming conditions. Additionally, participants were not professional agricultural workers, which may limit the generalizability of our findings. Experienced workers may demonstrate optimized movement strategies developed over time. Although recruitment focused on healthy individuals with no musculoskeletal im-

pairments, the lack of farming experience may underrepresent biomechanical adaptations acquired through real-world exposure. Future studies should include professional agricultural laborers to better understand how expertise influences biomechanics and to validate cobot assistance under more ecologically valid conditions. These real-world complexities may influence the effectiveness of cobot assistance and underscore situations where manual tasks remain practical or even preferable, particularly in small-scale or highly variable farming operations.

While the present study focused on controlled simulations to replicate real-world environments, it is important to recognize the inherent limitations of this approach. Simulated tasks, although valuable for isolating specific biomechanical and neuromuscular parameters, cannot fully capture the dynamic and unpredictable nature of agricultural work. However, the biomechanical components examined here remain relevant to field applications. For instance, Lüling et al. [35] demonstrated that a robotic planting unit with AI-based adjustment still required human oversight, underscoring that manual movements remain integral in real-world settings. This supports the relevance of our simulation for informing ergonomic design. Moreover, the cobot used in this study was operated under a Wizard-of-Oz protocol to simulate assistance in a controlled manner. While effective for initial feasibility evaluation, this method does not reflect adaptive or autonomous cobot behavior and may introduce experimenter bias. Future studies should integrate real-time human-aware control strategies to better reflect the intended system performance.

To further address the ecological validity gap, future studies should include parallel data collection under both simulated and real agricultural conditions using the same wearable sensor configuration. Comparative analyses of biomechanical variability could quantify how uneven terrain and spatial irregularity influence human motion during transplanting tasks. Additionally, sensor performance metrics such as EMG signal noise, IMU drift, and data dropout rates could serve as indirect indicators of field-related disruptions. These comparisons would provide a quantitative framework for calibrating lab-based results and guiding cobot development that is robust to environmental variability. While this study does not directly quantify the difference between laboratory and real-world biomechanical responses, its simulated design was intentionally chosen to isolate key neuromuscular and postural features without the interference of environmental confounds. This level of control allows for the high-resolution analysis of cobot–human interaction dynamics and strain-reduction mechanisms that may otherwise be masked in complex field settings. Importantly, while this study was conducted in a controlled laboratory setting, the biomechanical patterns observed still reflect meaningful human responses to the physical demands of transplanting tasks. Prior research has shown that controlled environments are necessary to isolate key movement features and reduce confounding variables, allowing for the high-resolution assessment of postural strategies, muscle activation patterns, and joint dynamics. As is supported by Moore and Clifton [36], laboratory- and field-based studies should be seen as complementary: findings from simulated conditions provide essential mechanistic insight that informs and enhances the translational value of future field-based validations. Therefore, the current findings offer foundational insights into physical strain and human–robot interactions that can inform future field-validated cobot designs.

Building on these considerations, future research should also explore adjustable robotic configurations tailored to user-specific ergonomic needs, enhancing the adaptability of cobots across diverse agricultural settings. Although both IMU and EMG systems used have been validated in prior studies, variability due to soft tissue movement, sensor placement, and electrode impedance may affect data consistency. Future studies will incorporate reliability testing to quantify these potential sources of error and consider multi-sensor fusion to improve robustness. In addition to traditional motion and muscle activity tracking,

emerging wearable sensor technologies such as photonic smart wristbands, which enable real-time cardiorespiratory monitoring and biometric identification, could be integrated to provide a more comprehensive understanding of user physiology and task demands. These technologies may offer new insights into human–cobot interactions by capturing both physical and physiological stressors in field applications. For example, physiological indicators such as heart rate variability, respiration rate, and electrodermal activity [37] could help to identify systemic stress responses associated with cobot use, particularly under prolonged or repetitive tasks. Incorporating these metrics alongside biomechanical data would enable a holistic assessment of user strain and support the development of cobot systems that optimize both physical load distribution and physiological recovery. Additionally, due to hardware instability, EMG data from six participants were excluded from the final analysis. While this did not affect the kinematic dataset, it reduced the EMG sample size to 10 participants. Future studies should employ improved sensor calibration protocols and real-time signal monitoring to mitigate data loss. Expanding the sample size to include experienced agricultural workers could further improve the generalizability of the results, as their expertise may reveal subtle yet significant differences in how manual and cobot-assisted techniques are executed. Furthermore, integrating real-world task simulations could provide deeper insights into the practical applications of cobot technology in agriculture. Factors such as uneven terrain, soil softness, and spatial irregularities, along with fluctuating weather and varying plant heights, were not accounted for, which may limit the ecological validity of the findings. These environmental factors can alter user posture, compromise sensor stability, and affect robot navigation, highlighting the need for future field-based trials to evaluate cobot performance under realistic agricultural conditions. These efforts will also help to identify scenarios where manual labor may offer advantages in terms of flexibility, cost-effectiveness, or task adaptability.

5. Conclusions

This study provides foundational biomechanical evidence on human movement strategies during plant transplanting, fulfilling the first objective of characterizing manual task demands prior to cobot integration. Using wearable sensor systems, including inertial measurement units (IMUs) and surface electromyography (sEMG), we analyzed segment velocities, joint accelerations, and muscle activation patterns to identify key contributors to biomechanical strain in repetitive transplanting tasks. The second objective, assessing the impact of cobot assistance, was supported by findings that demonstrated significant reductions in movement velocity and acceleration across spinal segments, along with improved task efficiency. However, these benefits were accompanied by a redistribution of muscular demand, particularly in the stabilizing muscle groups. These results underscore the need to design cobot systems that are biomechanically informed, not simply optimized for speed or automation, so that they can better support human physical needs. Overall, this study contributes both empirical insights and practical sensor-derived data to guide the ergonomic design of collaborative robots, particularly in small-scale, labor-intensive agricultural environments where manual strategies still prevail. These findings can directly inform cobot design by identifying optimal arm-reach limits, segment velocity thresholds, and ergonomic delivery positions. Such sensor-informed design principles can guide motion planning algorithms that adapt in real time to user kinematics, ensuring that collaborative assistance enhances rather than hinders biomechanical efficiency. In doing so, it establishes a sensor-integrated framework for real-time human monitoring and cobot adaptation in dynamic field applications.

Author Contributions: Conceptualization, Y.W. and B.H.; methodology, Y.W.; software, Y.W.; validation, Y.W., X.W. and B.H.; formal analysis, Y.W. and X.W.; investigation, Y.W.; resources, B.H.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W. and B.H.; visualization, X.W.; supervision, B.H.; project administration, B.H.; funding acquisition, B.H. All authors have read and agreed to the published version of the manuscript.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work has been partially supported by the US Department of Agriculture (NIFA 2022-67021-36125).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Florida (IRB202400216, approved on 11 March 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EMG	Electromyography
IMU	Inertial Measurement Unit
LBP	Low Back Pain
WMSD	Work-related Musculoskeletal Disorder

References

1. Lee, H.J.; Oh, J.-H.; Yoo, J.R.; Ko, S.Y.; Kang, J.H.; Lee, S.K.; Jeong, W.; Seong, G.M.; Kang, C.H.; Song, S.W. Prevalence of Low Back Pain and Associated Risk Factors among Farmers in Jeju. *Saf. Health Work.* **2021**, *12*, 432–438. [CrossRef]
2. Kang, M.-Y.; Lee, M.-J.; Chung, H.; Shin, D.-H.; Youn, K.-W.; Im, S.-H.; Chae, H.S.; Lee, K.S. Musculoskeletal Disorders and Agricultural Risk Factors Among Korean Farmers. *J. Agromed.* **2016**, *21*, 353–363. [CrossRef]
3. Shivakumar, M.; Welsh, V.; Bajpai, R.; Helliwell, T.; Mallen, C.; Robinson, M.; Shepherd, T. Musculoskeletal disorders and pain in agricultural workers in Low- and Middle-Income Countries: A systematic review and meta-analysis. *Rheumatol. Int.* **2024**, *44*, 235–247. [CrossRef] [PubMed]
4. Marras, W.S.; Allread, W.G.; Burr, D.L.; Fathallah, F.A. Prospective validation of a low-back disorder risk model and assessment of ergonomic interventions associated with manual materials handling tasks. *Ergonomics* **2000**, *43*, 1866–1886. [CrossRef] [PubMed]
5. Fathallah, F.A. Musculoskeletal disorders in labor-intensive agriculture. *Appl. Ergon.* **2010**, *41*, 738–743. [CrossRef]
6. Fathallah, F.A.; Tang, S.C.H.; Waters, T. Development and Evaluation of Ergonomic Interventions for Bucket Handling on Farms. *Hum. Factors* **2016**, *58*, 758–776. [CrossRef] [PubMed]
7. Wang, X.; Grimaldi, N.S.; Behdad, S.; Hu, B. Evaluating the Impact of Collaborative Robots in E-Waste Disassembly Through EMG-EMG Coherence Analysis. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Phoenix, AZ, USA, 8–13 September 2024; Volume 68, pp. 677–682. [CrossRef]
8. Chen, Y.; Luo, Y.; Yerebakan, M.O.; Xia, S.; Behdad, S.; Hu, B. Human Workload and Ergonomics during Human-Robot Collaborative Electronic Waste Disassembly. In Proceedings of the 2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS), Orlando, FL, USA, 17–19 November 2022; pp. 1–6. [CrossRef]
9. Seyyedhasani, H.; Peng, C.; Jang, W.-J.; Vougioukas, S.G. Collaboration of human pickers and crop-transporting robots during harvesting–Part I: Model and simulator development. *Comput. Electron. Agric.* **2020**, *172*, 105324. [CrossRef]
10. Adamides, G.; Katsanos, C.; Constantinou, I.; Christou, G.; Xenos, M.; Hadzilacos, T.; Edan, Y. Design and development of a semi-autonomous agricultural vineyard sprayer: Human–robot interaction aspects. *J. Field Robot.* **2017**, *34*, 1407–1426. [CrossRef]
11. Yerebakan, M.O.; Gu, Y.; Gross, J.; Hu, B. Evaluation of Biomechanical and Mental Workload During Human–Robot Collaborative Pollination Task. *Hum. Factors* **2024**, *67*, 100–114. [CrossRef]
12. Lowenberg-DeBoer, J.; Huang, I.Y.; Grigoriadis, V.; Blackmore, S. Economics of robots and automation in field crop production. *Precis. Agric.* **2020**, *21*, 278–299. [CrossRef]

13. Berenstein, R.; Edan, Y. Automatic adjustable spraying device for site-specific agricultural application. *IEEE Trans. Autom. Sci. Eng.* **2017**, *15*, 641–650. [CrossRef]
14. Matheson, E.; Minto, R.; Zampieri, E.G.G.; Faccio, M.; Rosati, G. Human–robot collaboration in manufacturing applications: A review. *Robotics* **2019**, *8*, 100. [CrossRef]
15. Silverstein, B.A.; Bao, S.S.; Russell, S.; Stewart, K. Water and Coffee: A Systems Approach to Improving Coffee Harvesting Work in Nicaragua. *Hum. Factors* **2012**, *54*, 925–939. [CrossRef]
16. Brumitt, J.; Reisch, R.; Krasnoselsky, K.; Welch, A.; Rutt, R.; Garside, L.I.; McKay, C. Self-Reported Musculoskeletal Pain in Latino Vineyard Workers. *J. Agromed.* **2010**, *16*, 72–80. [CrossRef] [PubMed]
17. Wang, X.; Grimaldi, N.S.; Zheng, M.; Behdad, S.; Yang, C.; Hu, B. Enhancing ergonomics in E-waste disassembly: The impact of collaborative robotics on muscle activation and coordination. *Ergonomics* **2024**, 1–13. [CrossRef]
18. Yerebakan, M.O.; Hu, B. Wearable sensors assess the effects of human–robot collaboration in simulated pollination. *Sensors* **2024**, *24*, 577. [CrossRef]
19. Strader, J.; Xie, Z.; Kothari, M.; Kumar, V. Flower interaction subsystem for a precision pollination robot. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 4–8 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5534–5541. [CrossRef]
20. Ohi, N.; Lassak, K.; Watson, R.; Strader, J.; Du, Y.; Yang, C.; Hedrick, G.; Nguyen, J.; Harper, S.; Reynolds, D.; et al. Design of an autonomous precision pollination robot. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 7711–7718. [CrossRef]
21. Smith, T.; Rijal, M.; Tatsch, C.; Butts, R.M.; Beard, J.; Cook, R.T.; Chu, A.; Gross, J.; Gu, Y. Design of stickbug: A six-armed precision pollination robot. In Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 69–75. [CrossRef]
22. Brown, M.; Avers, D. (Eds.) *Daniels and Worthingham's Muscle Testing-E-Book: Techniques of Manual Examination and Performance Testing*, 9th ed.; Saunders: Philadephia, PA, USA, 2013.
23. Gregory, D.E.; Laughton, C.; Carman, A.; Milosavljevic, S.; Callaghan, J.P. Trunk postures and peak and cumulative low back kinetics during upright posture sheep shearing. *Ergonomics* **2009**, *52*, 1576–1583. [CrossRef]
24. Earle-Richardson, G.; Jenkins, P.; Fulmer, S.; Mason, C.; Burdick, P.; May, J. An ergonomic intervention to reduce back strain among apple harvest workers in New York State. *Appl. Ergon.* **2005**, *36*, 327–334. [CrossRef]
25. Davis, K.; Marras, W. The effects of motion on trunk biomechanics. *Clin. Biomech.* **2000**, *15*, 703–717. [CrossRef] [PubMed]
26. Marras, W.S.; Lavender, S.A.; Leurgans, S.E.; Rajulu, S.L.; Allread, W.G.; Fathallah, F.A.; Ferguson, S.A. The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders. The effects of work-place factors, trunk position, and trunk motion characteristics on risk of injury. *Spine* **1993**, *18*, 617–628. [CrossRef] [PubMed]
27. Allread, W.G.; Vossenas, P. Comparisons of Trunk Motions and Low Back Injury Risk between Alternative Hotel Room Cleaning Methods. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14907. [CrossRef] [PubMed]
28. Tsang, S.M.H.; Szeto, G.P.Y.; Li, L.M.K.; Wong, D.C.M.; Yip, M.M.P.; Lee, R.Y.W. The effects of bending speed on the lumbo-pelvic kinematics and movement pattern during forward bending in people with and without low back pain. *BMC Musculoskelet. Disord.* **2017**, *18*, 157. [CrossRef] [PubMed]
29. Tetteh, E.; Hallbeck, M.S.; Mirka, G.A. Effects of passive exoskeleton support on EMG measures of the neck, shoulder and trunk muscles while holding simulated surgical postures and performing a simulated surgical procedure. *Appl. Ergon.* **2022**, *100*, 103646. [CrossRef] [PubMed]
30. Thamsuwan, O.; Milosavljevic, S.; Srinivasan, D.; Trask, C. Potential exoskeleton uses for reducing low back muscular activity during farm tasks. *Am. J. Ind. Med.* **2020**, *63*, 1017–1028. [CrossRef] [PubMed]
31. Liu, R.; Natarajan, M.; Gombolay, M.C. Human-Robot Team Coordination with Dynamic and Latent Human Task Proficiencies: Scheduling with Learning Curves. *arXiv* **2020**, arXiv:2007.01921.
32. Liu, L.; Schoen, A.J.; Henrichs, C.; Li, J.; Mutlu, B.; Zhang, Y.; Radwin, R.G. Human Robot Collaboration for Enhancing Work Activities. *Hum. Factors* **2024**, *66*, 158–179. [CrossRef]
33. Lasota, P.A.; Shah, J.A. Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human–Robot Collaboration. *Hum. Factors* **2015**, *57*, 21–33. [CrossRef]
34. Bechar, A.; Edan, Y. Human-robot collaboration for improved target recognition of agricultural robots. *Ind. Robot. Int. J. Robot. Res. Appl.* **2003**, *30*, 432–436. [CrossRef]
35. Lüling, N.; Straub, J.; Stana, A.; Brodbeck, M.; Reiser, D.; Berner, P.; Griepentrog, H.W. Development and evaluation of a self-adaptable planting unit for an autonomous planting process of field vegetables. *Smart Agric. Technol.* **2024**, *9*, 100578. [CrossRef]

36. Moore, T.Y.; Clifton, G.T.; Patek, S.N.; Daley, M.A.; Sane, S.P. Jumping over fences: Why field- and laboratory-based biomechanical studies can and should learn from each other. *J. Exp. Biol.* **2023**, *226* (Suppl. 1), jeb245284. [CrossRef]
37. Li, K.; Cardoso, C.; Moctezuma-Ramirez, A.; Elgalad, A.; Perin, E. Heart rate variability measurement through a smart wearable device: Another breakthrough for personal health monitoring? *Int. J. Environ. Res. Public Health* **2023**, *20*, 7146. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

electronics

MDPI

Article

On-Device Motion Activity Intensity Recognition Using Smartwatch Accelerator

Seungyeon Kim and Jaehyun Yoo *

School of AI Convergence, Sungshin Women's University, 34 da-gil 2, Bomun-ro, Seongbuk-gu, Seoul 02844, Republic of Korea
* Correspondence: jhyoo@sungshin.ac.kr

Abstract

Wearable device-based Human Activity Recognition (HAR) is widely used in health management, rehabilitation, and personal safety. While contemporary HAR research effectively classifies a wide range of discrete activities, there remains a significant gap in organizing these heterogeneous motions into a structured intensity framework suitable for continuous risk assessment. Furthermore, many high-performing models rely on computationally intensive architectures that hinder real-time deployment on resource-constrained wearables. We propose an on-device method for estimating five-level activity intensity in real time using only accelerometer signals from a commercial smartwatch. To bridge the gap between simple identification and intensity modeling, 13 dynamic and emergency-like wrist motions were integrated with 11 daily activities from the PAMAP2 dataset, yielding 21 activities mapped onto an ordinal five-level intensity scale. A finetuned Multi-Layer Perceptron (MLP) classifier trained on this integrated dataset achieved 0.939 accuracy and a quadratic weighted kappa (QWK) of 0.971. The model was deployed on a Galaxy Watch 7, achieving < 1 ms inference latency and a size < 0.1 MB, confirming real-time feasibility. This approach demonstrates that organizing diverse activities into a lightweight, intensity-aware framework provides a robust foundation for safety-aware monitoring systems under real-world, on-device constraints.

Keywords: wearable devices; accelerometer; Human Activity Recognition (HAR); activity intensity recognition; risk assessment; on-device inference

1. Introduction

As societal concerns about personal safety grow, the need for real-time monitoring technology using wearable devices is increasing. Smartwatches incorporate sensors such as accelerometers, heart rate monitors, and GPS, enabling them to sensitively capture changes in user activity. They hold significant potential not only for health management and rehabilitation but also for detecting dangerous situations. Importantly, danger often manifests not as a single abrupt event but through gradual changes in activity intensity that serve as early indicators. Therefore, beyond simply classifying activities, a methodology is needed to represent human behavior along discretized ordinal intensity levels, providing a foundation for subsequent risk assessment.

Recent safety-oriented HAR research has explored various applications, including fall detection, personal protection monitoring, and workplace safety management. However, many existing approaches depend on the offline analysis of data collected in controlled laboratory environments, with limited validation on actual wearable devices. Moreover,

for users who are highly likely to be exposed to dangerous situations—such as individuals under personal protection—technology that detects risks in real time through a worn smartwatch is crucial for enabling immediate response. These limitations underscore the need for real-time HAR models capable of recognizing multi-level activity intensity.

To address this, we constructed a new smartwatch accelerometer dataset containing dynamic and emergency-like wrist motions that are largely absent from existing public HAR datasets. Designed to capture a wide range of real-world intensity variations, this dataset complements established activity corpora used to represent lower-intensity daily and static behaviors. Extending this perspective to the protection of individuals under wearable-based personal safety monitoring supports the development of situation-specific, tiered alert systems.

Motivated by these needs, this study proposes an on-device accelerometer-only framework for five-level activity intensity estimation, with the objective of determining its feasibility under the computational and memory constraints of a commercial smartwatch. In this framework, the spectrum of activity intensity is presented by five ordinal levels, where the highest level includes aggressive or emergency-like wrist movements that may be associated with potentially dangerous situations. This study supports the development of a hierarchical warning system by aiming to develop threat detection technology for women in their 20s and 30s, the primary target group of personal protection programs, as part of public safety R&D efforts.

The main contributions of this study are as follows:

(1) Development of a multi-level activity intensity recognition model: Using only accelerometer signals from a smartwatch, we mapped 21 activities to five intensity levels and designed a model capable of real-time inference with four features.

(2) Dataset construction for activity intensity recognition: We built the core of our dataset using 13 Dyn-Intensity activities designed to capture dynamic and emergency-like wrist motions, and complemented them with 11 low-intensity PAMAP2 activities. This integration produced a unified 21-class dataset spanning an activity intensity spectrum.

(3) On-device real-time verification: The proposed model was deployed on a Galaxy Watch 7, achieving $<$1 ms inference latency and a memory footprint of $<$0.1 MB, demonstrating real-time feasibility on commercial devices.

The remainder of this paper is organized as follows. First, we provide an overview of prior research on wearable-based activity recognition and intensity estimation. We then describe the datasets, feature engineering process, and model design employed in this study. Next, we outline the experimental setup and evaluation procedures used to validate the proposed model. The results are subsequently presented and analyzed, including both model performance and on-device evaluation. We then discuss the main findings and their broader implications. Finally, we conclude the paper with closing remarks and potential directions for future research.

2. Related Work

HAR using IMU-based sensing has advanced significantly across healthcare, fitness, and safety monitoring [1–4]. Early studies primarily focused on classifying fundamental activities—such as walking, running, and sitting—using accelerometer and IMU time-series data. Classic reviews, such as Bulling et al. [5], summarized the general HAR pipeline from pre-processing to classification, while Attal et al. [6] benchmarked machine learning methods for accelerometer-based activity recognition. More recently, deep learning architectures such as CNN, DeepConvLSTM, and Transformer-based models, including Conformer, have demonstrated outstanding performance on datasets containing wearable

https://doi.org/10.3390/electronics15071351

IMU sensor data for HAR [7,8]. These approaches have established a robust research foundation for reliable nominal multi-class activity recognition, where activity labels are treated as discrete and unordered categories.

Alongside advances in activity-type classification, increasing attention has been directed toward estimating activity intensity from wrist-worn accelerometers due to its relevance to health monitoring and behavioral assessment [9,10]. In physical-activity research, intensity is commonly structured using a four-level taxonomy based on Metabolic Equivalents (METs)—sedentary, light, moderate, and vigorous (SLMV) [11]. Despite the inherently ordered nature of such scales, many computational approaches formulate intensity estimation as a standard multi-class classification problem. In practice, several studies further simplify the task into the binary detection of moderate-to-vigorous physical activity (MVPA) versus non-MVPA [12,13], thereby reducing the granularity of the intensity spectrum and not explicitly incorporating ordinal relationships into the modeling objective.

To address this limitation, a limited number of studies have explored ordinal modeling strategies. Hossain et al. [14] applied ordinal random forest and partial proportional odds models to four-level activity intensity classification from accelerometer data. Similarly, Bağcı Daş and Birant [15] introduced HAROC, an ordinal classification framework for HAR based on band-power-derived activity ordering. These works represent meaningful efforts to incorporate ordered structure into activity modeling. However, most evaluations have been conducted in controlled laboratory environments using multi-sensor configurations or offline processing pipelines, and real-time deployment on single wrist-worn devices has been less extensively investigated.

In parallel, substantial progress has been made in real-time on-device HAR. Lightweight architectures such as TinierHAR [16] and smartwatch inference pipelines achieving sub-10 ms latency [17] have demonstrated the feasibility of executing nominal activity classification under strict computational constraints. Advances in edge-oriented deep learning design [18] and efficient triggering mechanisms [19] have further enhanced deployment practicality in wearable systems. Nevertheless, these developments have predominantly focused on activity-type recognition, while the preservation of ordinal intensity structure within a single accelerometer-based, real-time inference framework remains comparatively underexplored.

Taken together, nominal multi-class activity recognition and lightweight real-time HAR have achieved considerable maturity. However, the integration of ordinal activity intensity modeling with single-sensor, wrist-worn, real-time on-device deployment remains relatively limited. To address this gap, the present study proposes an accelerometer-only framework for five-level ordinal activity intensity estimation and validates its feasibility under the computational and memory constraints of a commercial smartwatch platform.

3. Methods

3.1. System Overview

Figure 1 illustrates the overall pipeline of the proposed on-device multi-level activity intensity recognition framework. The 3-axis accelerometer signals from the smartwatch are sampled at 50 Hz and processed in four stages:

Pre-processing: Raw signals (a_x, a_y, a_z) are segmented into 6 s sliding windows with 50% overlap, providing fixed intervals for analysis.

Extracting features from raw data: Within each window, two derived signals are computed: the Signal Vector Magnitude (SVM) and its temporal difference (ΔSVM). The mean and standard deviation of each are computed, resulting in four lightweight statistical features (SVM mean, SVM std, ΔSVM mean, ΔSVM std).

https://doi.org/10.3390/electronics15071351

Modeling: The extracted features were normalized according to each model's requirements and subsequently fed into two types of predictive models: non-MLP methods (Linear Regression [20], Logistic Regression [21]) and MLP methods (regression and classification variants [22]). Each model estimates the activity intensity on a five-level scale.

Post-processing: MLP model implementations employ temporal smoothing, hysteresis rules, median filtering, and jerk suppression, whereas non-MLP model implementations apply probability thresholds, margin conditions, and standardized feature gating. This pipeline is optimized for lightweight, real-time operation on commercial smartwatch platforms and is validated through on-device evaluation.

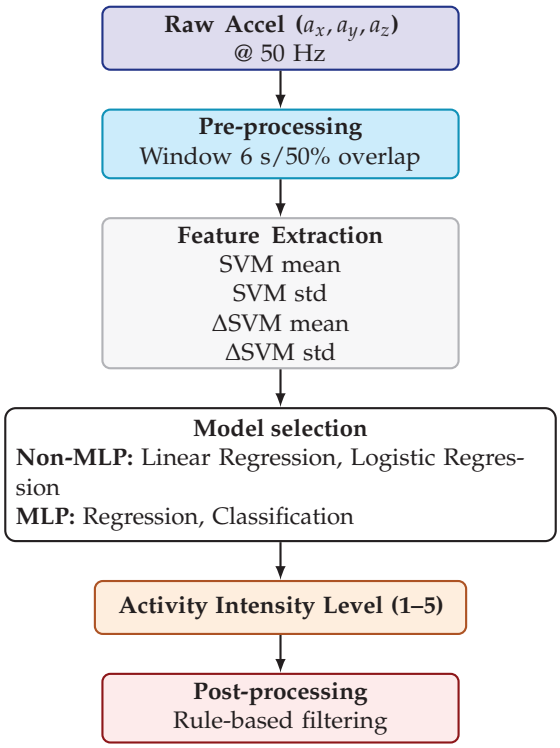


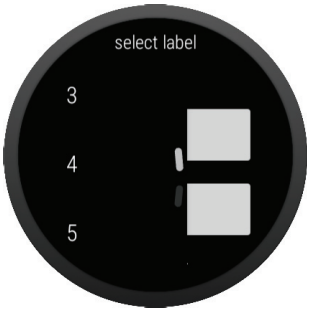
Figure 1. Compressed system pipeline for on-device multi-level activity intensity recognition.

3.2. Data Acquisition

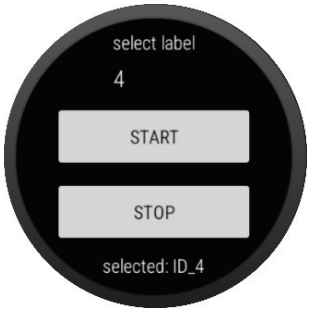
This study utilized two complementary datasets: a newly collected smartwatch accelerometer dataset, Dyn-Intensity, and a publicly available dataset, PAMAP2. Dyn-Intensity was designed to capture dynamic and emergency-like wrist motions that are generally absent from existing HAR datasets, while PAMAP2 provides representative static and low-intensity daily activities. Together, these datasets offer a broad range of movements needed to model continuous activity intensity. Data can be found at https://github.com/AILAB-IMU/Dyn-Intensity (accessed on 9 February 2026).

Dyn-Intensity Dataset. Dyn-Intensity was collected from eight female participants aged 20–23, each wearing a Samsung Galaxy Watch 7 on the dominant wrist. The dataset contains three static activities (lying, sitting, watching TV), nine dynamic activities (running, ascending stairs, descending stairs, rope jumping, slalom running, jumping jacks, board work, palm pushing, dynamic handclap) and several emergency-like movements (sudden sprints, irregular arm motions, violent whole-body gestures) to represent high-intensity behaviors. These emergency-like motions were grouped into a single emergency-like motion category and subsequently mapped to the highest activity intensity level (Level 5). Participants selected activity labels directly on the watch, and Figure 2 illustrates the recorded accelerometer log structure. A detailed list of Dyn-Intensity activities is provided in Table 1. To ensure precise temporal alignment between activity labels and sensor signals,

the data segments recorded before the actual onset of each activity (after pressing the start button) and after the completion of the activity (before pressing the stop button) were manually removed. This trimming process was performed by synchronizing the recorded accelerometer logs with simultaneously captured video recordings, allowing non-activity intervals to be accurately identified and excluded from the dataset.



(**a**)



(**b**)

	Timestamp	X	Y	Z
2	2025-07-23-15:17:32.445	−7.39569	6.251261	−2.53067
3	2025-07-23-15:17:32.456	−7.41245	6.215348	−2.478
4	2025-07-23-15:17:32.465	−7.40287	6.157887	−2.63362
5	2025-07-23-15:17:32.476	−7.41245	6.124368	−2.63123
6	2025-07-23-15:17:32.485	−7.3502	6.203377	−2.52109
7	2025-07-23-15:17:32.495	−7.49146	6.241684	−2.45885
8	2025-07-23-15:17:32.505	−7.50343	6.270414	−2.4373
9	2025-07-23-15:17:32.516	−7.52737	6.339846	−2.44687
10	2025-07-23-15:17:32.525	−7.50822	6.227319	−2.50912

(**c**)

Figure 2. On-watch labeling and logging during data acquisition. (**a**) On-watch label selection; (**b**) start/stop control; (**c**) sample CSV log (a_x, a_y, a_z).

Table 1. Descriptions of activities included in the Dyn-Intensity dataset.

Activity	Description
Lying	Lying quietly while doing nothing; small posture changes are allowed.
Sitting	Sitting in a chair in a comfortable position; posture changes are allowed.
Watching TV	Watching television in a relaxed posture (sitting or lying).
Board work	Standing in front of a whiteboard and writing or erasing text; involves moderate forward–backward arm swings.
Palm pushing	Pushing forward repeatedly with both palms to simulate resistance or self-defense motion; involves abrupt upper-body acceleration.
Dynamic handclap	Clapping hands rapidly with large arm movements; produces strong upper-body motion.
Jumping jacks	Performing full-body jumping movements, spreading and closing arms and legs simultaneously.
Running	Sprinting for short distances at maximum speed; generates intense whole-body acceleration and impact.
Slalom running	Sprinting in a zig-zag pattern between cones or markers; requires rapid direction changes at high speed.
Rope jumping	Jumping rope using basic or alternate-foot jumps.
Ascending stairs	Walking upstairs between the ground and top floors (five floors total).
Descending stairs	Walking downstairs between the ground and top floors (five floors total).
User_level5	Performing extreme resistance or escape-like motion under simulated threat; includes rapid arm swings, pushing, and evasive body movements to represent intense struggle.

PAMAP2 Dataset. To complement Dyn-Intensity with additional static and low-intensity daily activities, we selected 11 activities from the PAMAP2 dataset [23], including lying, sitting, standing, walking, cycling, watching TV, computer work, car driving, ironing, folding laundry, and vacuum cleaning. The rationale for this selection is revisited in the Activity Intensity-Level Mapping section, where feature distributions are compared between datasets. Detailed descriptions appear in Table 2. Since the PAMAP2 accelerometer signals are originally sampled at 100 Hz, an anti-aliasing low-pass FIR filter (cutoff = 22.5 Hz, below the Nyquist frequency of 25 Hz) was applied prior to decimation, and the signals were subsequently downsampled to 50 Hz to ensure consistent temporal resolution with the Dyn-Intensity dataset.

https://doi.org/10.3390/electronics15071351

Table 2. Descriptions of PAMAP2 activities used in this study.

Activity	Description
Lying	Lying quietly while doing nothing; small posture changes are allowed.
Sitting	Sitting in a chair in a comfortable position; posture changes are allowed.
Standing	Standing still or while talking, possibly with gestures.
Watching TV	Watching television in a relaxed posture (sitting or lying).
Computer work	Performing normal office-related computer tasks.
Car driving	Driving between the office and home.
Ironing	Ironing one or two shirts or T-shirts.
Folding laundry	Folding shirts, T-shirts, and bed linens.
Vacuum cleaning	Cleaning one or two office rooms, including moving chairs or light furniture.
Cycling	Cycling outdoors at a slow to moderate pace, similar to commuting or leisure biking.
Walking	Walking outdoors at a moderate to brisk pace (4–6 km/h), comfortable for the subject.

Activity Class Integration. To build a unified dataset, overlapping activities (lying, sitting, watching TV) were merged after confirming that their feature distributions were nearly identical across datasets. The final integrated dataset consists of 21 activity classes, spanning static, daily, dynamic, and emergency-level movements.

3.3. Signal Pre-Processing

The collected acceleration signals were segmented into fixed time intervals for analysis. Each interval was defined as 6 s (300 samples), with adjacent windows overlapping by 3 s (150 samples). This sliding-window strategy enables the capture of abrupt short-duration motions while maintaining stable feature extraction over longer temporal contexts. A 6 s window length was selected to balance responsiveness and statistical stability for wrist-worn accelerometer signals, as shorter windows are more sensitive to transient fluctuations, whereas longer windows increase decision latency. This choice is further supported by a systematic sensitivity analysis reported in study [24], which demonstrated progressive performance improvement with increasing window duration in wrist-based accelerometer activity recognition. The 50% overlap further reduces effective latency and ensures smoother temporal transitions between consecutive predictions without incurring significant computational overhead, making this configuration suitable for real-time on-device intensity estimation.

3.4. Feature Extraction

Each segmented window was processed to derive two motion-related signals from the tri-axial acceleration data (a_x, a_y, a_z):

- Signal Vector Magnitude (SVM):

$$\mathrm{SVM}_t = \sqrt{a_{x,t}^2 + a_{y,t}^2 + a_{z,t}^2}, \tag{1}$$

  representing the overall magnitude of motion at time t.
- Delta Signal Vector Magnitude (ΔSVM):

$$\Delta\mathrm{SVM}_t = \mathrm{SVM}_t - \mathrm{SVM}_{t-1}, \quad (t > 1) \tag{2}$$

  capturing instantaneous changes between consecutive samples.

For both signals, the mean and standard deviation were computed within each window, yielding four features:

https://doi.org/10.3390/electronics15071351

- SVM mean and SVM standard deviation: indicators of overall motion intensity and stability.
- ΔSVM mean and ΔSVM standard deviation: descriptors of rapidity and irregularity of temporal changes.

These features require only $O(L)$ computation per window, where L is the number of samples in a window, making them suitable for lightweight on-device processing.

Feature distributions of identical activities across participants in the Dyn-Intensity dataset showed consistent patterns (Figures 3–15), confirming inter-subject reliability. Based on this consistency, data corresponding to the same activity were merged into representative categories. The merged Dyn-Intensity dataset was then integrated with PAMAP2 to compare average feature characteristics and establish the foundation for intensity-level mapping. As shown, box colors represent the extracted features in all feature distribution plots: blue for SVM mean, orange for SVM standard deviation, green for ΔSVM mean, and red for ΔSVM standard deviation.

These comparisons (Figures 16–19) show that static activities exhibit nearly overlapping distributions across all four features, supporting their integration into unified categories. To differentiate Dyn-Intensity activities from PAMAP2 activities in the boxplot visualizations, Dyn-Intensity activity names include the prefix "c_", although this naming convention is used solely for figure labeling and does not affect any modeling procedure. In contrast, high-intensity activities show larger mean values and greater variability, making them clearly distinguishable from lower-intensity motions. Furthermore, change-based features such as ΔSVM standard deviation effectively capture irregular activities like slalom running and palm pushing. These findings confirm that the four selected features are robust across data sources and suitable for modeling the five discrete ordinal levels of activity intensity.

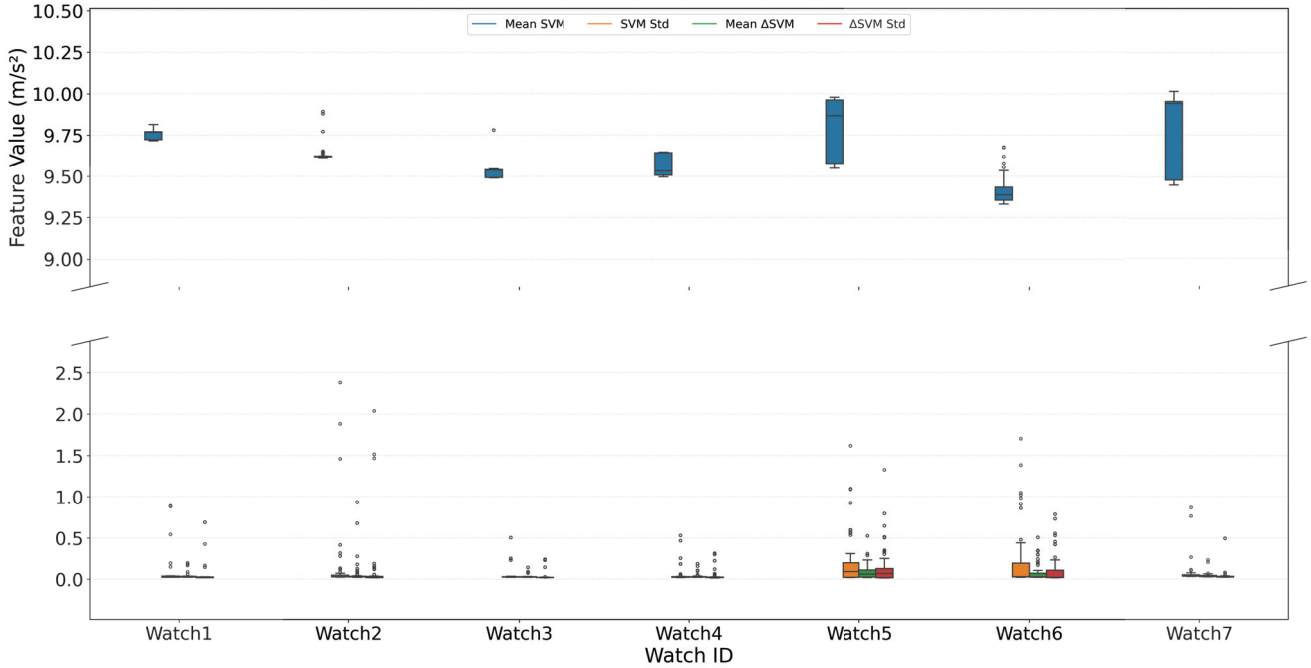


Figure 3. Feature distribution for lying.

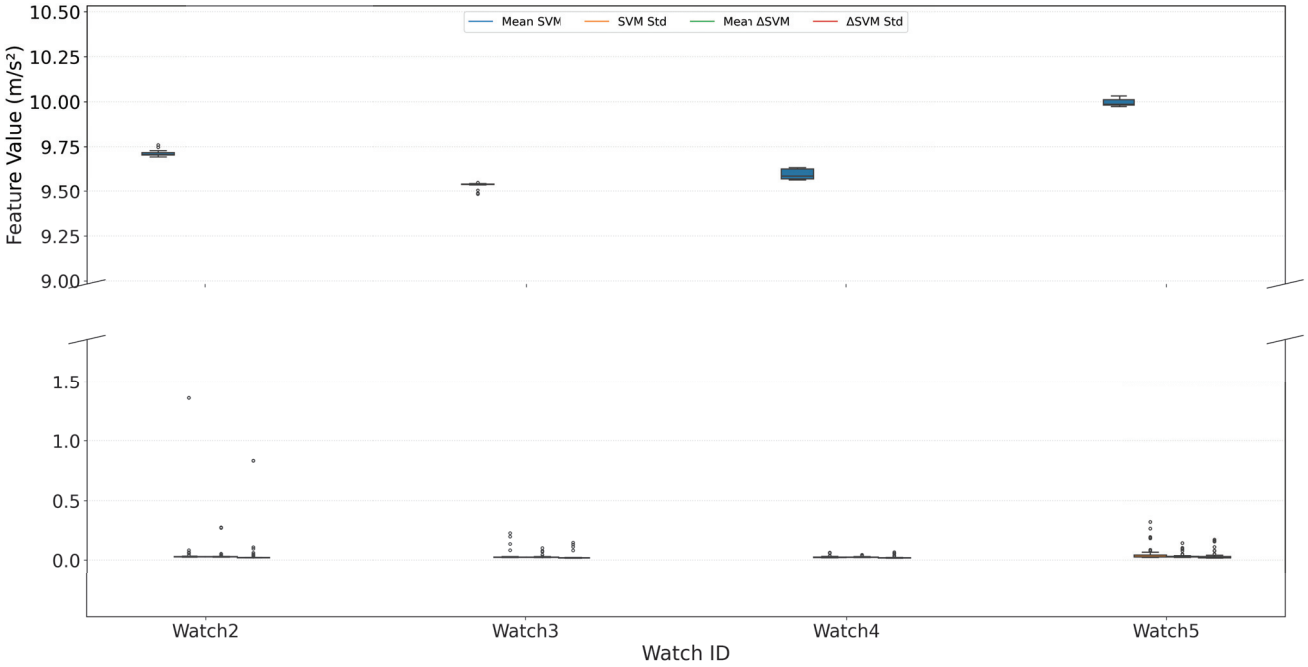


Figure 4. Feature distribution for sitting.

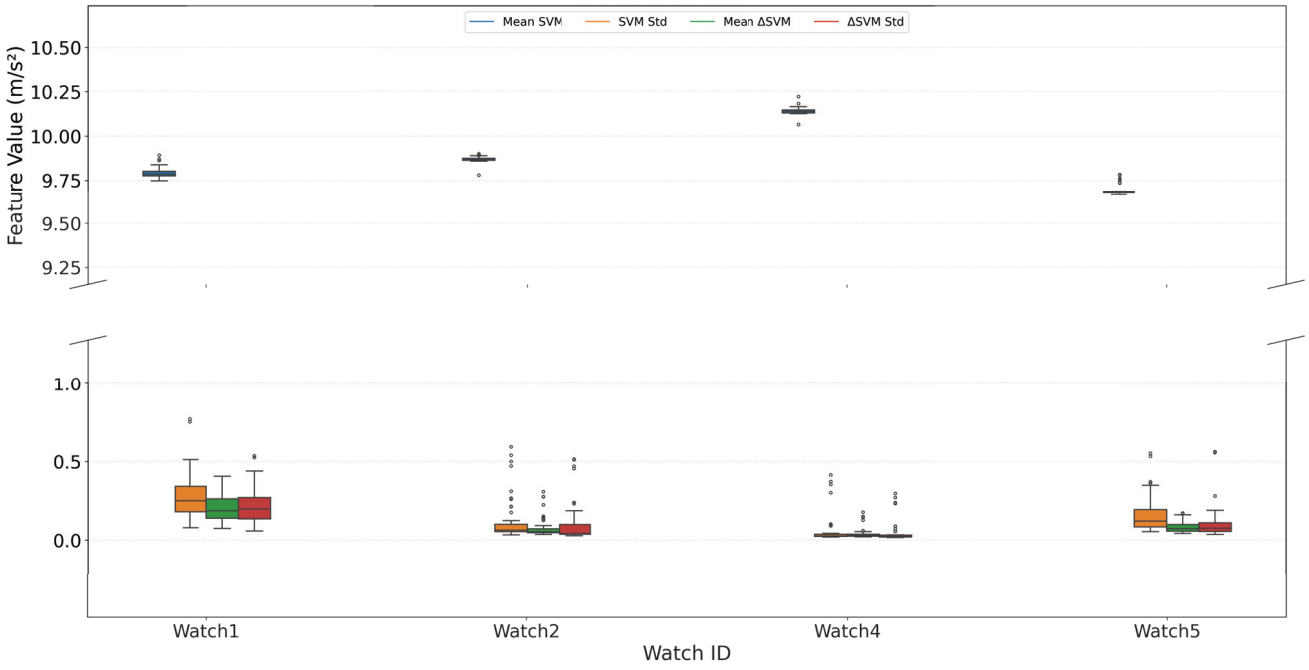


Figure 5. Feature distribution for watching TV.

https://doi.org/10.3390/electronics15071351

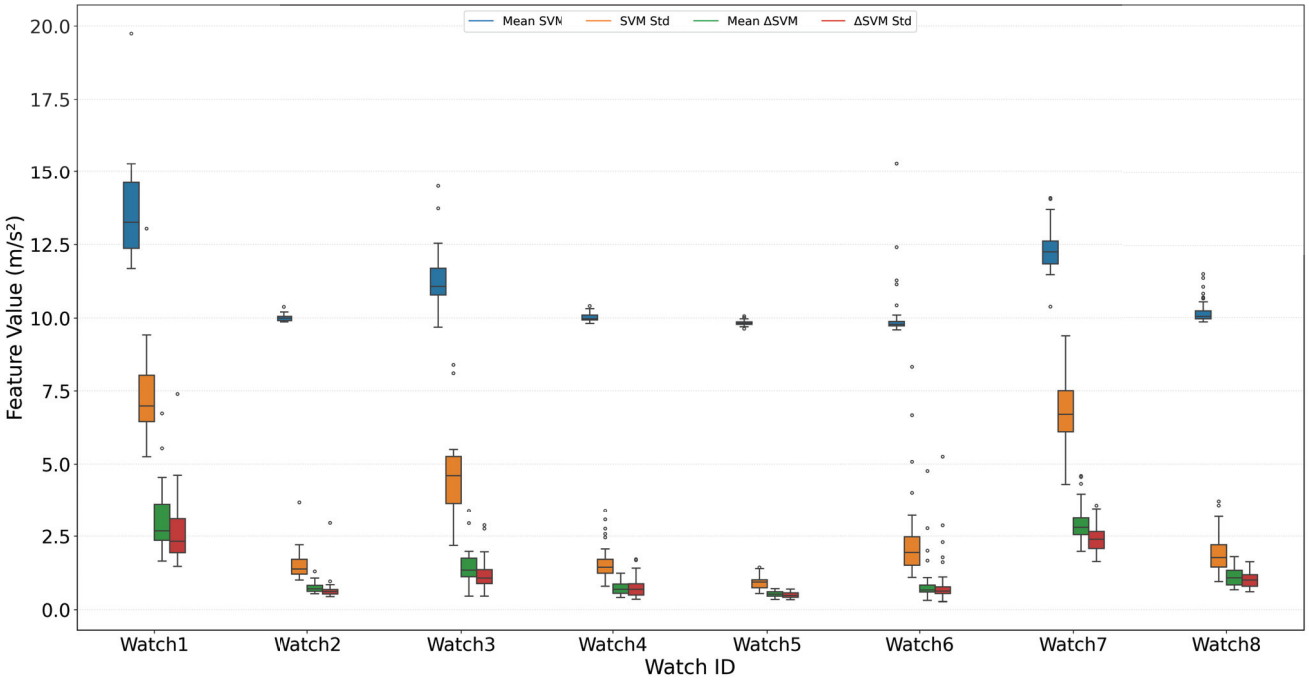


Figure 6. Feature distribution for board work.

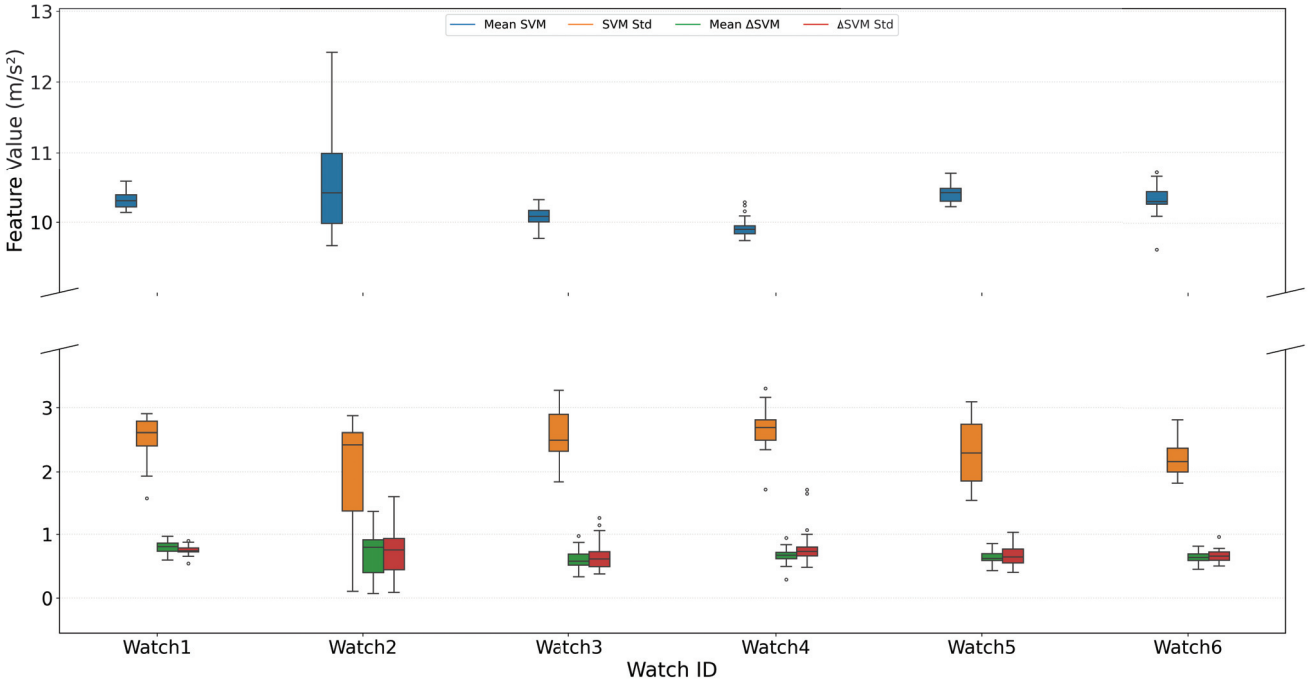


Figure 7. Feature distribution for ascending stairs.

https://doi.org/10.3390/electronics15071351

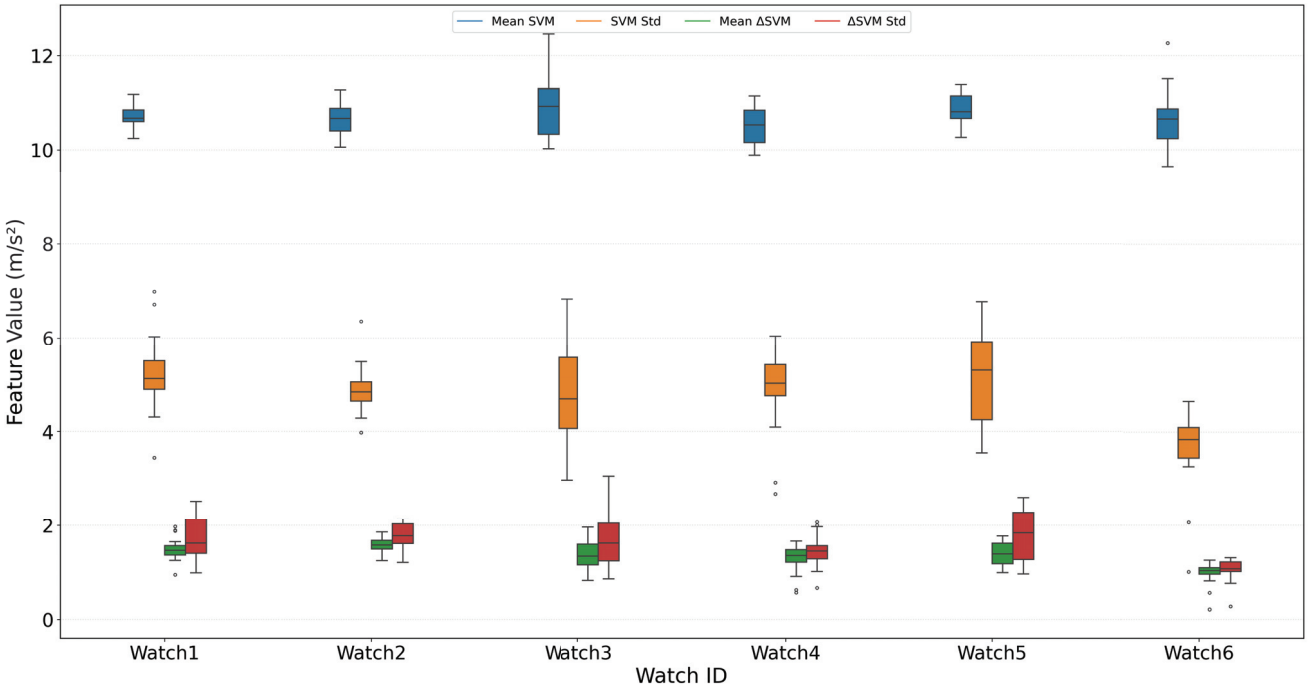


Figure 8. Feature distribution for descending stairs.

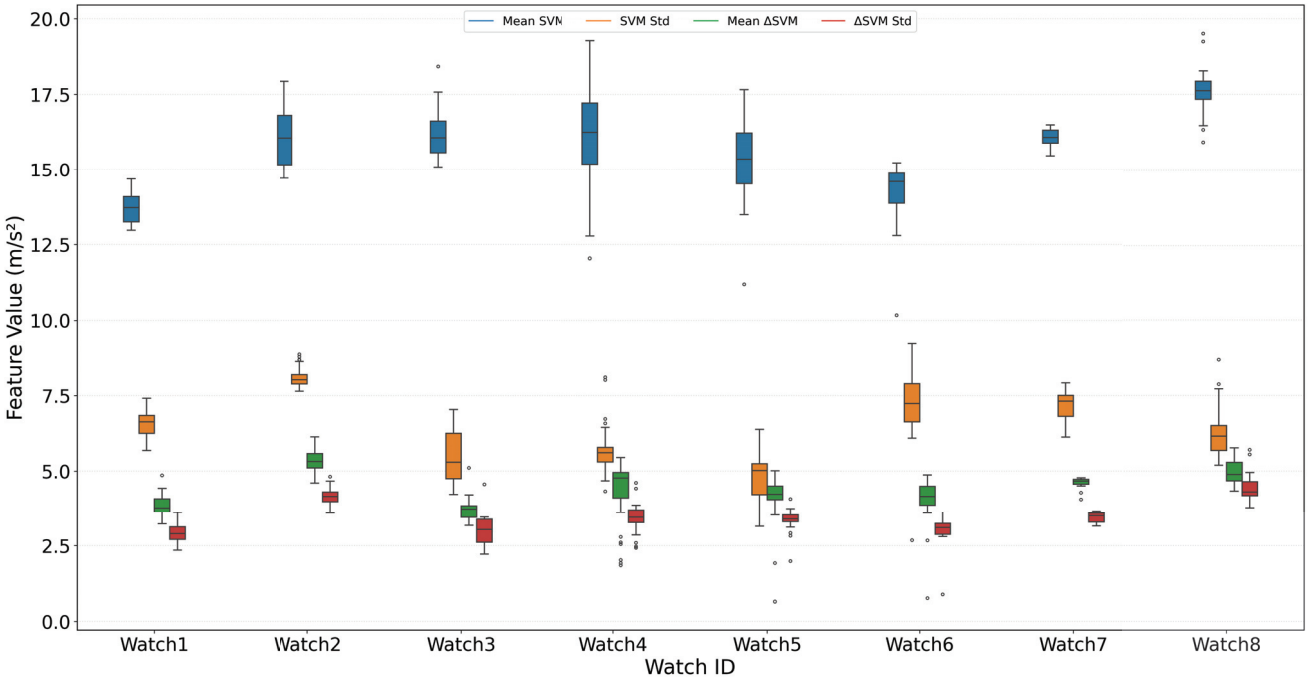


Figure 9. Feature distribution for rope jumping.

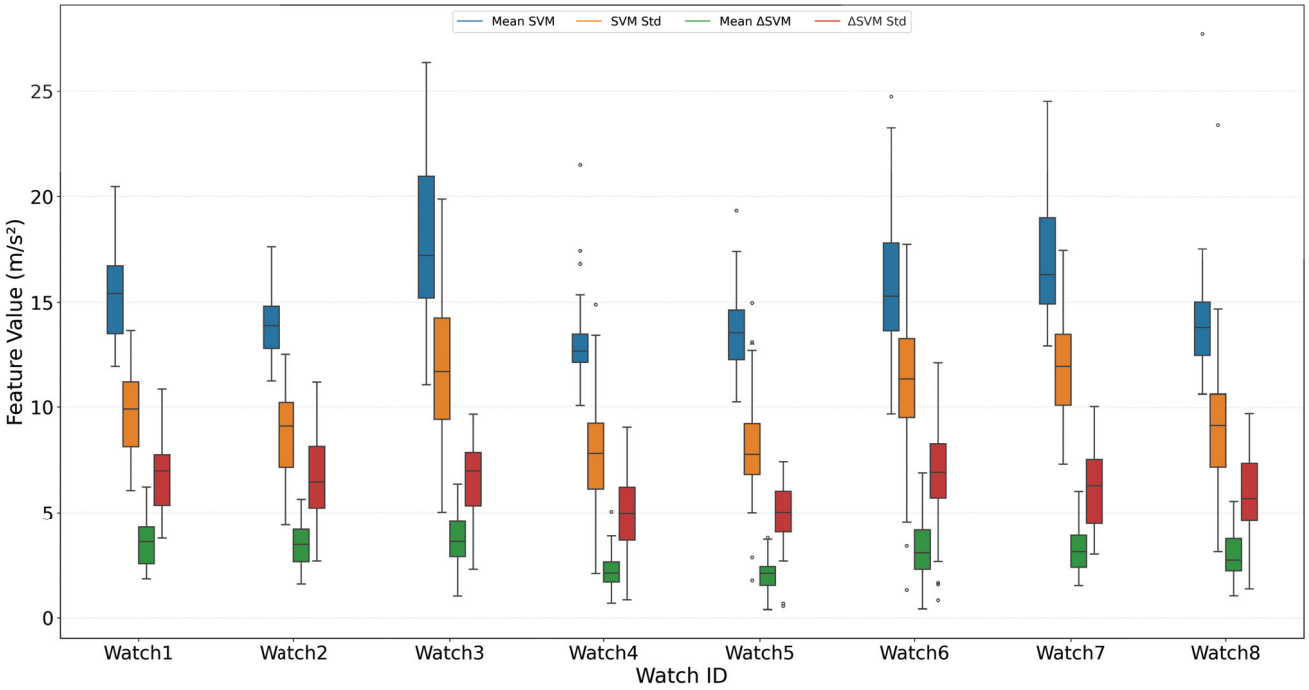


Figure 10. Feature distribution for palm pushing.

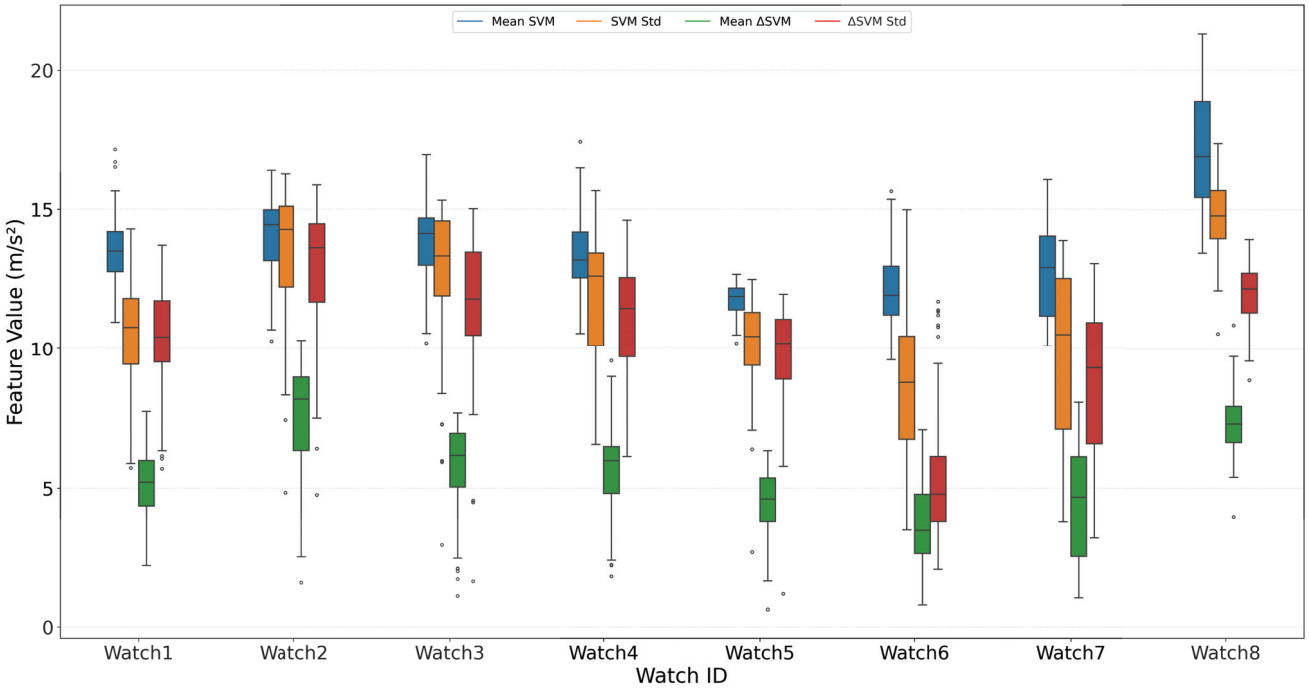


Figure 11. Feature distribution for dynamic handclap.

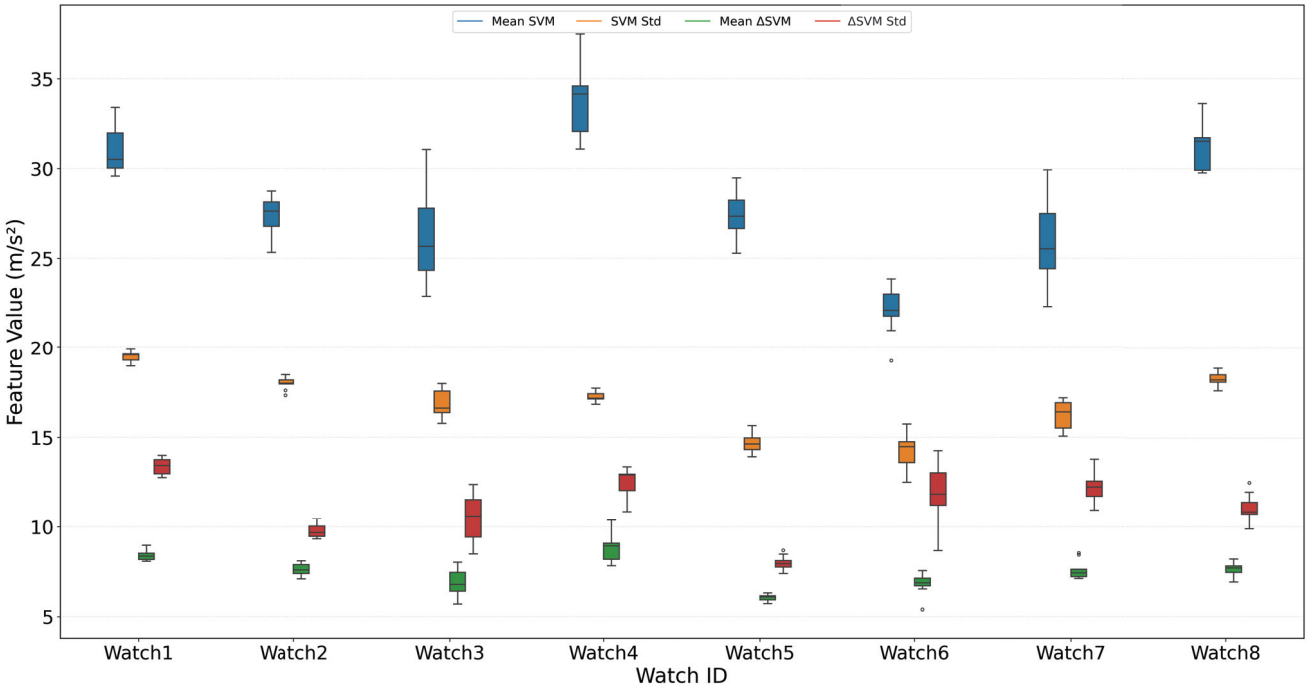


Figure 12. Feature distribution for jumping jacks.

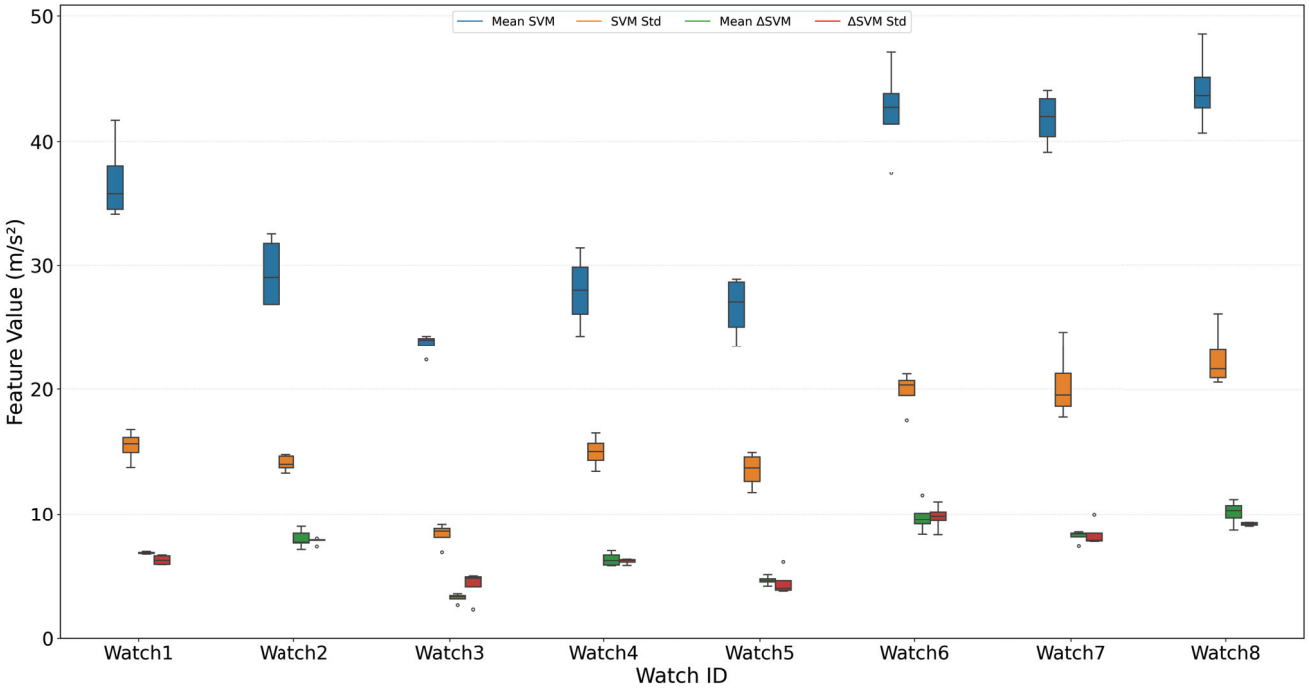


Figure 13. Feature distribution for slalom running.

https://doi.org/10.3390/electronics15071351

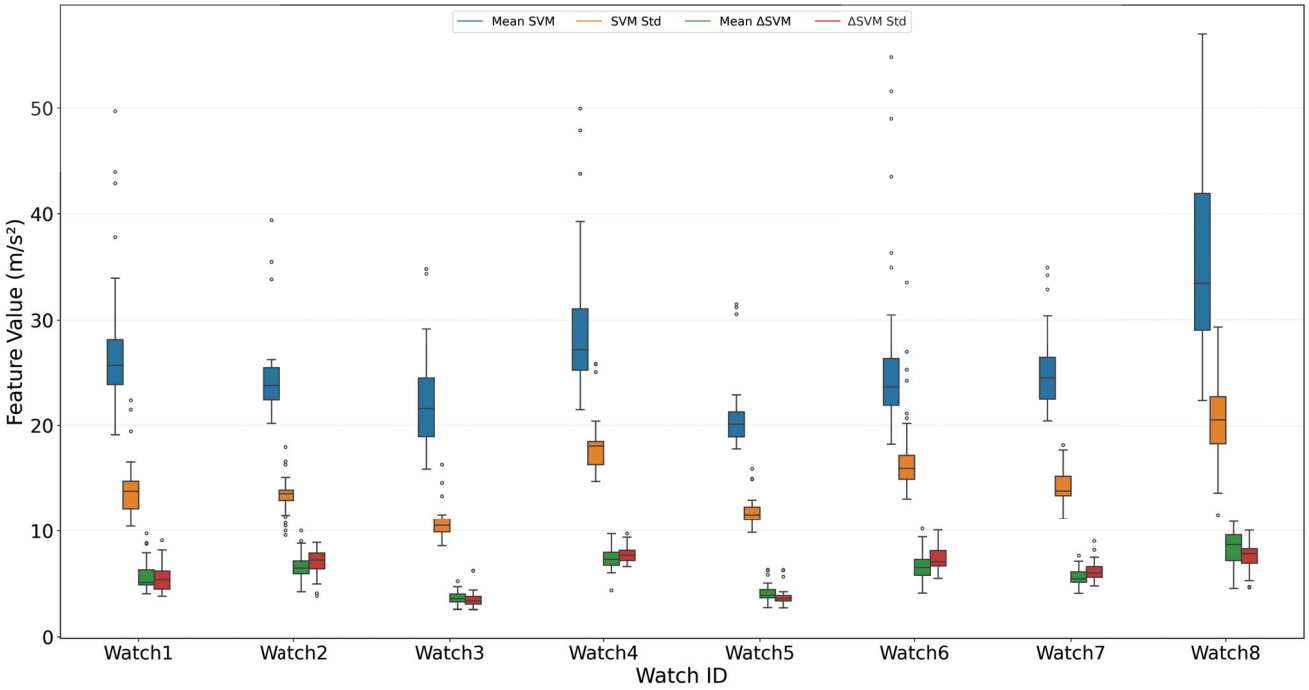


Figure 14. Feature distribution for running.

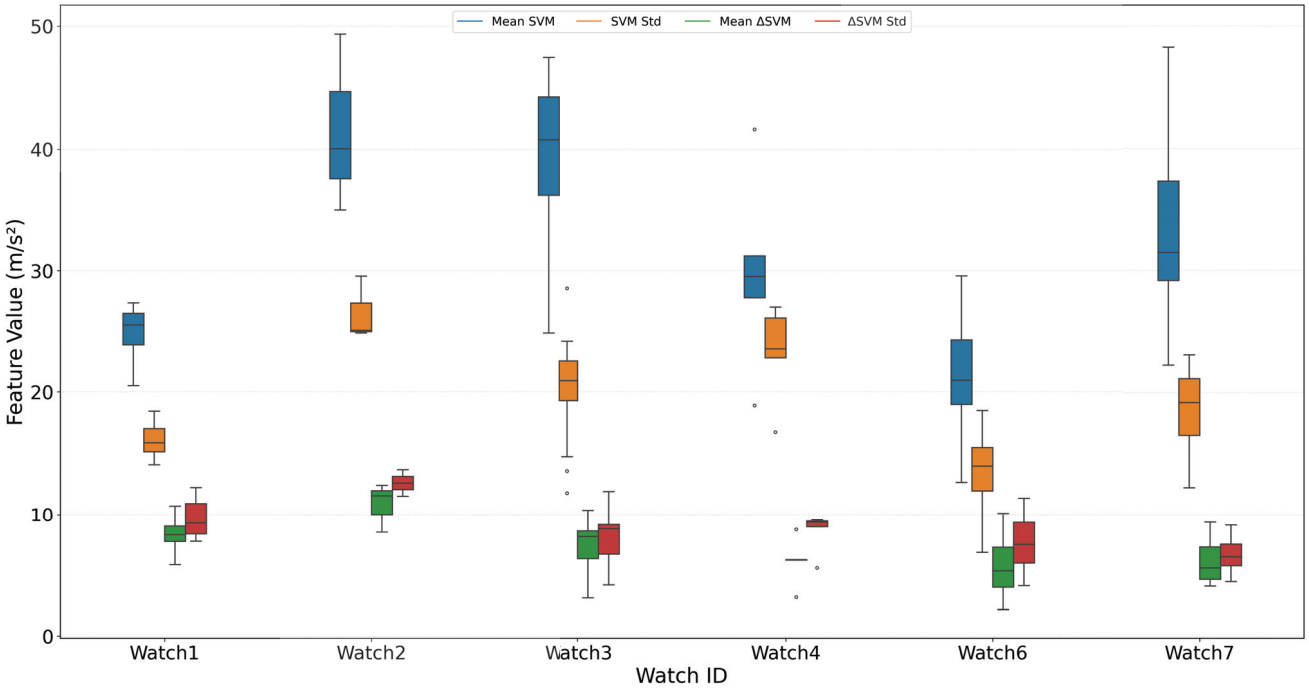


Figure 15. Feature distribution for emergency-level movements.

https://doi.org/10.3390/electronics15071351

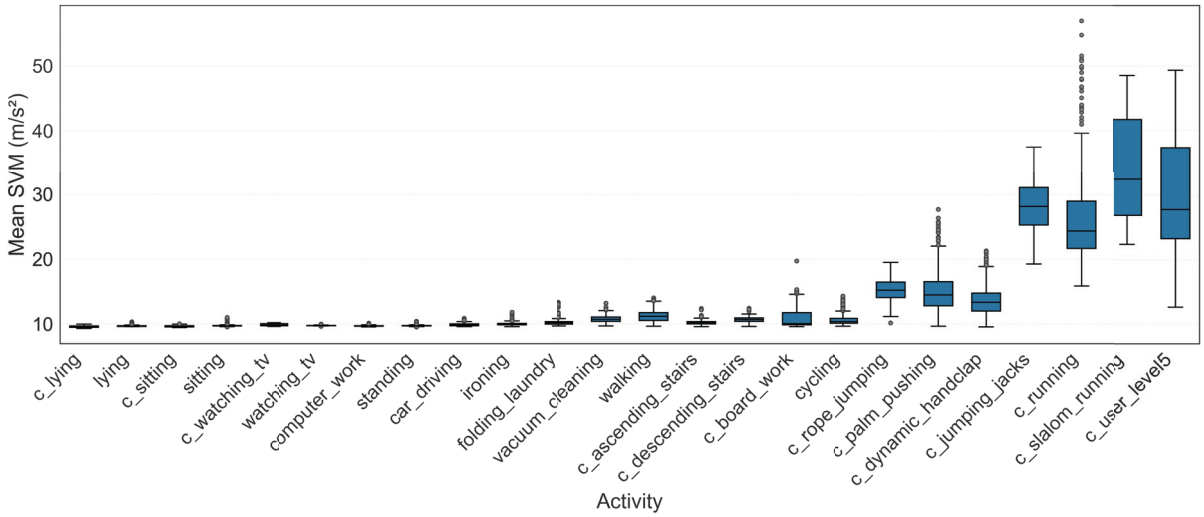


Figure 16. Mean SVM feature values extracted from the activities in the Dyn-Intensity and PAMAP2 datasets.

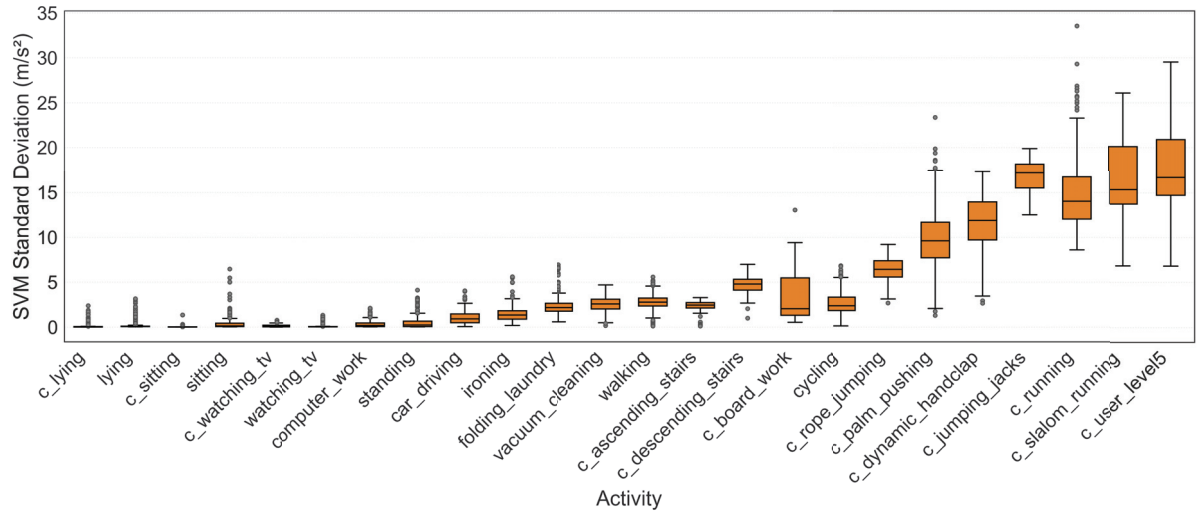


Figure 17. Standard deviation of the SVM feature observed across the activities in the Dyn-Intensity and PAMAP2 datasets.

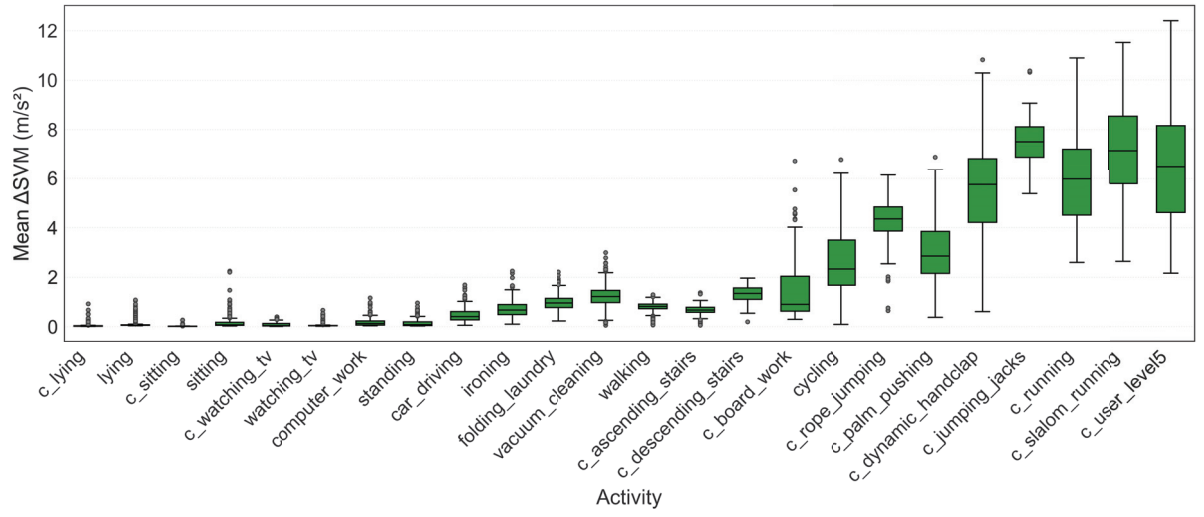


Figure 18. Mean ΔSVM feature values computed for the activities in the Dyn-Intensity and PAMAP2 datasets.

https://doi.org/10.3390/electronics15071351

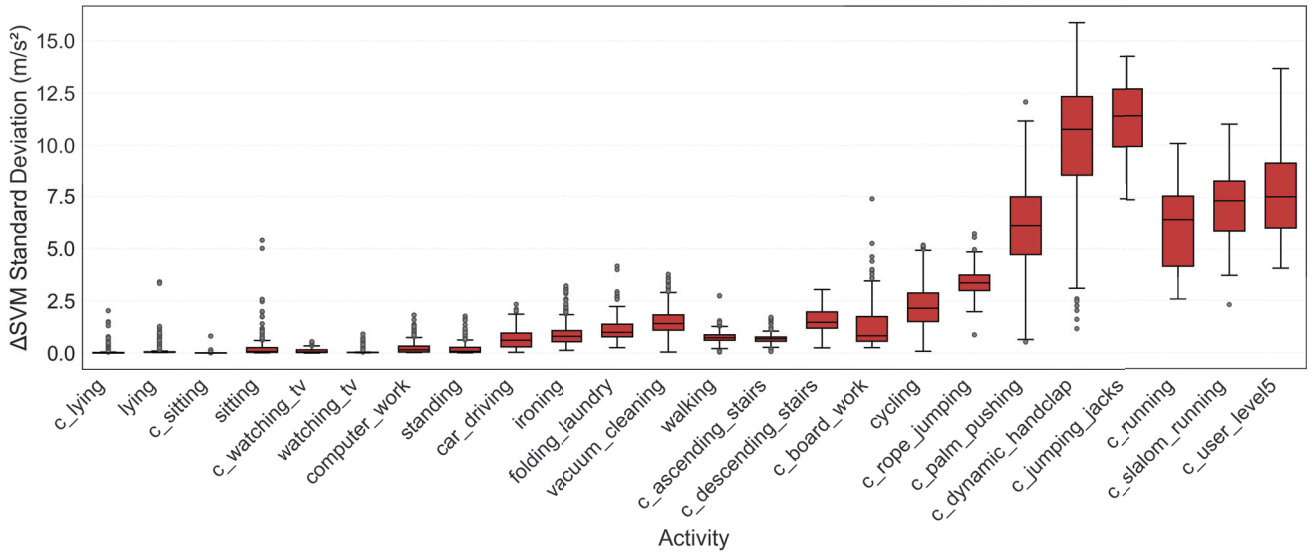


Figure 19. Standard deviation of the ΔSVM feature across the activities in the Dyn-Intensity and PAMAP2 datasets.

3.5. Activity Intensity-Level Mapping

In this study, the 21 activities were reorganized into five intensity levels according to activity intensity. This restructuring enabled activity intensity to be represented using five ordered levels, ranging from daily behaviors to vigorous and atypical emergency situations, thereby moving beyond simple activity classification.

As described in the Data Acquisition section, the integrated dataset comprised 21 activities. Among these, overlapping static activities were merged into unified representative categories, supported by feature-level consistency analysis. Figure 20 compares the feature distributions of a static activity (lying) and a dynamic activity (running) between the PAMAP2 and Dyn-Intensity datasets.

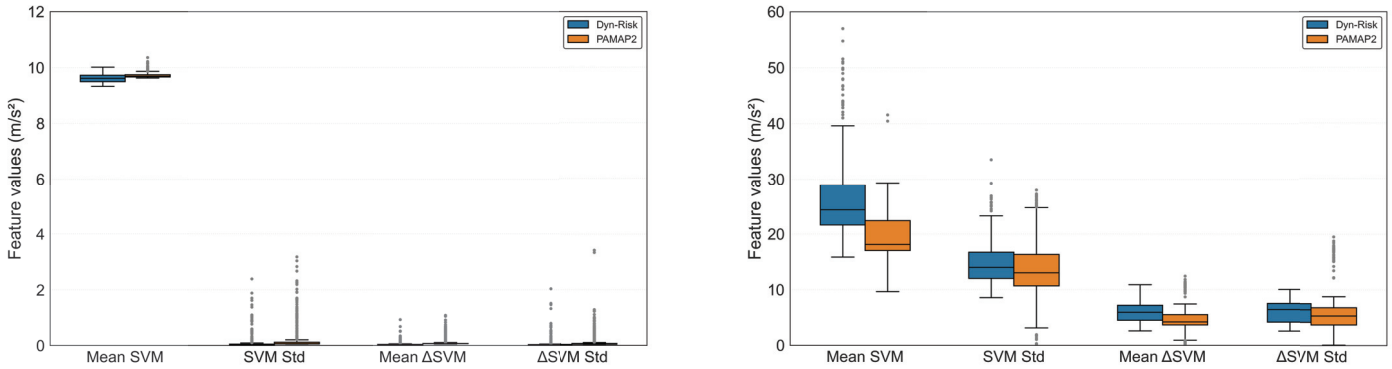


Figure 20. Comparison of feature distributions between the PAMAP2 and self-collected datasets.The static activity (lying) shows nearly identical distributions across all four features, validating its integration into a unified class. In contrast, the dynamic activity (running) exhibits noticeable differences, likely reflecting individual variability in movement intensity and sensor placement. (**a**) Static activity, lying: Similar distributions across datasets, consolidated into one category. (**b**) Dynamic activity, running: Distinct distributions observed, only self-collected data used.

For the static activity, all four features—SVM mean, SVM standard deviation, ΔSVM mean, and ΔSVM standard deviation—exhibited nearly identical central tendencies and ranges across the two datasets. This consistency indicates that sensor characteristics and

https://doi.org/10.3390/electronics15071351

signal magnitudes remained stable despite differences in devices and participants, thereby validating their integration into a single representative class. Therefore, each instance of lying, sitting, and watching TV was integrated across datasets into a single, consistent category for subsequent intensity-level mapping and model training.

In contrast, the dynamic activity (running) exhibited noticeable differences primarily in SVM-based features, both in mean values and distributional variability, whereas the ΔSVM-derived features showed relatively consistent patterns between groups. These discrepancies are likely attributable to variations in running speed, arm swing intensity, and sensor placement among participants rather than inconsistencies in data collection. In particular, the Dyn-Intensity running data were designed to emulate an emergency-like situation, where participants performed rapid, full-speed running to replicate realistic high-risk movements. Consequently, dynamic activities were retained as independent categories based solely on the Dyn-Intensity dataset to preserve the individuality of high-intensity motion patterns.

Ultimately, the 21 activities were mapped into five hierarchical intensity levels—static, low-intensity daily, moderate-intensity, high-intensity, and vigorous/emergency-like—as summarized in Table 3. It should be noted that this five-level intensity mapping was defined as a heuristic structuring approach rather than derived from externally validated physiological intensity scales, and it was informed by exploratory inspection of the feature distributions across activities, as illustrated in Figures 16–19. This five-level structure establishes a unified labeling framework that supports both classification-based (discrete levels) and regression-based (continuous-scale) modeling in subsequent analysis.

Table 3. Five-level intensity mapping of 21 integrated activities constructed from the PAMAP2 and Dyn-Intensity datasets.

Level	Description	Activity Classes
Level 1	Static (P + D)	lying, sitting, standing, watching TV, computer work, car driving
Level 2	Low-intensity daily (P + D)	walking, ironing, folding laundry, vacuum cleaning, board work, ascending stairs, descending stairs
Level 3	Moderate intensity (P + D)	cycling, rope jumping
Level 4	High intensity (D)	palm pushing, dynamic handclap
Level 5	Vigorous/emergency-like (D)	jumping jacks, running, slalom running, user_level5 (emergency-like vigorous motions)

P = PAMAP2 dataset; D = Dyn-Intensity dataset. Overlapping static activities were merged.

3.6. Activity Intensity Estimation Model

The proposed modeling framework was designed to estimate activity intensity levels in both continuous and discrete forms. To comprehensively evaluate model performance and on-device applicability, two approaches were employed: regression-based models for continuous intensity estimation and classification-based models for discrete level prediction.

3.6.1. Motivation for Comparing Non-MLP and MLP Approaches

Wearable devices such as smartwatches operate under strict resource constraints, including limited computational power, memory capacity, and battery life. Therefore, activity intensity estimation models must simultaneously achieve high accuracy, real-time processing capability, and lightweight implementation. To address these requirements, this study comparatively evaluates two fundamentally different modeling paradigms—the model-free approach and the MLP-model approach.

The non-MLP models (both Linear Regression and Logistic Regression) offer extremely low computational complexity and can be implemented using simple mathematical formulations. Conversely, the MLP models (both regression and classification) offer greater

predictive accuracy by learning nonlinear relationships within the data. However, these models typically incur higher computational costs, making direct deployment on wearable devices challenging. To overcome this limitation, the present study applies TensorFlow Lite (float16) quantization to the MLP models, enabling efficient on-device inference while maintaining accuracy.

Because the two approaches differ in model complexity, inference cost, memory requirements, and predictive performance, a systematic comparison is necessary. Moreover, since activity intensity can be formulated as either a continuous variable (regression) or a discrete set of levels (classification), each modeling paradigm was implemented in both forms. This design enables a comprehensive evaluation across predictive accuracy, computational efficiency, and real-time feasibility on wearable devices.

3.6.2. Non-MLP Approaches

(1) Linear Regression

A second-order polynomial feature transformation (interaction-only) was applied before training a linear regressor. The model outputs continuous intensity values, reflecting its formulation as an ordinal regression estimator. For evaluation under the unified five-level framework, these continuous predictions were discretized using a lightweight validation-based threshold optimization procedure. Because the output of the Linear Regression model exhibits a smooth and approximately monotonic structure, a simple grid-based search around median-derived initial thresholds was sufficient to identify a strictly increasing set of thresholds that maximized the quadratic weighted kappa (QWK) on the validation set. This optimization is applied only during evaluation and does not affect model training. The dataset was partitioned into disjoint training, validation, and test sets. The threshold optimization was performed by conducting a grid search on the validation set to find the cut-off points that maximize the QWK. The optimized thresholds obtained on the validation set were fixed prior to final test evaluation, and the test set was not involved in threshold selection at any stage.

(2) Logistic Regression

A multinomial Logistic Regression model was trained using quadratic interaction-only polynomial features and StandardScaler normalization. As a pure classification model, it directly predicts one of five discrete intensity levels without requiring any post-processing.

3.6.3. MLP Approaches

(3) MLP (Regression)

The regression MLP was composed of four hidden layers (256–128–64–16 units) with ReLU activations, followed by batch normalization and dropout for regularization. The output layer used a linear activation to generate continuous intensity values, and the model was trained using the Huber loss ($\delta = 3.0$) to enhance robustness to outliers. After training, the model was quantized to float16 precision and converted into a TFLite format to enable real-time on-device inference. To allow direct comparison with classification models, continuous outputs were discretized into five levels using a validation-driven threshold optimization procedure. Unlike the linear model, the MLP produced nonlinear and broader output distributions; therefore, a two-stage coordinate-ascent search (coarse-to-fine) was employed to efficiently locate a strictly increasing set of thresholds that maximized QWK. The optimized thresholds were fixed and applied unchanged to the test set. As with the linear model, this discretization step does not influence training and serves solely to provide a consistent ordinal mapping for evaluation.

https://doi.org/10.3390/electronics15071351

(4) MLP (Classification)

The classification MLP consisted of three hidden layers (256–128–64 units) with ReLU activations (tanh in the final hidden layer), combined with batch normalization and a dropout rate of 0.2 for regularization. The output layer used a softmax activation to produce one of five discrete intensity levels. Training was performed using categorical cross-entropy loss with class weights to address class imbalance. After training, the model was quantized to float16 precision and exported as a TFLite model for efficient real-time inference on the smartwatch.

3.6.4. Model Efficiency and On-Device Feasibility

All four models balance predictive accuracy and computational efficiency. Linear Regression and Logistic Regression are implemented as direct mathematical formulae for reliable execution on the device. MLP models achieve higher predictive performance while remaining feasible for real-time operation. A summary of the model configurations is presented in Table 4.

Table 4. Summary of the implemented models and their main configurations.

Approach	Model	Key Characteristics	Scaler	Loss/Objective
Non-MLP Approaches				
	Linear Regression	Second-order polynomial features (interaction only); threshold optimization via validation grid search.	RobustScaler	MSE/QWK maximization
	Logistic Regression	Multinomial Logistic Regression; polynomial features (interaction only).	StandardScaler	Cross-entropy
MLP Approaches				
	MLP (Regression)	Four hidden layers (256–128–64–16); ReLU activation, batch normalization, dropout (0.2–0.3).	Min–Max	Huber ($\delta = 3.0$)
	MLP (Classification)	Three hidden layers (256–128–64); ReLU activation (tanh for last layer), batch normalization, dropout (0.2); softmax output.	StandardScaler	Cross-entropy

4. Experiments

4.1. Experimental Environment

Offline training was performed on a workstation equipped with an AMD (Advanced Micro Devices, Inc., Santa Clara, CA, USA) Ryzen 9 9950X CPU and 64 GB RAM. On-device experiments were conducted on a commercial smartwatch, the Samsung Galaxy Watch 7 (SM-L300, Exynos W1000 CPU, 2 GB RAM, Wear OS 5) (Samsung Electronics Co., Ltd., Suwon-si, Republic of Korea).

All on-device experiments were conducted in an indoor sports facility. Participants wore the smartwatch on their dominant wrist and repeatedly performed activities of varying intensity (e.g., running, dynamic clapping, and palm pushing). Examples of the experimental environment and performed activities are shown in Figure 21.



(**a**) (**b**) (**c**)

Figure 21. Examples of the experimental environment and activities performed. (**a**) Running; (**b**) dynamic handclap; (**c**) palm pushing.

4.2. Training Configuration

Specifically, the model-free approaches were implemented as formula-based algorithms (direct mathematical expressions), whereas the MLP-model approaches were deployed as quantized TFLite models. Both formula-based and TFLite-based implementations were executed on the smartwatch to enable consistent performance comparison.

All models were trained using the Adam optimizer (learning rate = 0.0005), with Early Stopping (patience = 25) and ReduceLROnPlateau (patience = 8) applied to ensure stable convergence. To account for class imbalance, sample weights were used during the training of regression-formulated models, whereas class weights were assigned for classification-formulated models. The continuous outputs of regression models were discretized into five levels using validation-optimized thresholds, ensuring compatibility with the classification-based five-level prediction framework.

4.3. Evaluation Protocol

Both non-MLP and MLP approaches were implemented in two prediction formulations: regression and classification. The dataset was split in a stratified manner at a ratio of 64% for training, 16% for validation, and 20% for testing. All scalers were fitted exclusively to the training set to avoid data leakage.

For regression-based models, MSE, Mean Absolute Error (MAE), and the coefficient of determination (R^2) were used to assess continuous prediction performance. For discretized regression results, accuracy and the quadratic weighted kappa (QWK) were additionally calculated. QWK is particularly suitable for ordinal prediction because it reflects the degree of sequential agreement between predicted and true labels.

The QWK is formally defined as

$$\text{QWK} = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}}, \tag{3}$$

where i and j index the true and predicted intensity levels, respectively, and span all K ordinal classes (in this study, $K = 5$). Thus, both O_{ij} and E_{ij} form $K \times K$ matrices representing the observed and expected confusion matrices. The weight matrix is defined as $w_{ij} = \frac{(i-j)^2}{(K-1)^2}$, which assigns larger penalties as the ordinal difference between classes increases, while assigning zero penalty when $i = j$. This formulation ensures that ordinal misclassifications are penalized proportionally to their severity.

Classification-based models were evaluated using accuracy, precision, recall, F1-score, and QWK. This unified evaluation protocol ensured fair and consistent comparison between regression-based and classification-based approaches, enabling quantitative verification of prediction tendencies across model types.

5. Results

5.1. Regression Model Performance

Table 5 compares the performance of the two regression models: the non-MLP Linear Regression and the MLP regressor. As a result, the Linear Regression model achieved a QWK of 0.933, demonstrating that the model can preserve ordinal relationships between activity intensity levels. In contrast, the MLP regressor trained with Huber loss ($\delta = 3.0$) and Min–Max scaling demonstrated superior performance, achieving an MSE of 0.075, an MAE of 0.151, an R^2 of 0.95, and a QWK of 0.969. This indicates that while the non-MLP Linear Regression model can model ordinal continuity efficiently, the MLP regression model provides more robust and precise estimation across nonlinear transitions between activity intensity levels.

https://doi.org/10.3390/electronics15071351

Table 5. Performance comparison of regression-based models (non-MLP vs. MLP).

Model	MSE	MAE	R^2	QWK
Linear Regression	0.189	0.333	0.874	0.933 ± 0.004
MLP Regressor	0.075	0.151	0.95	0.969 ± 0.003

5.2. Classification Model Performance

Table 6 compares the performance of the two classification models: the non-MLP Logistic Regression and the MLP classifier. The Logistic Regression model achieved an accuracy of 0.931 and a QWK of 0.968. The MLP classifier achieved the best performance with an accuracy of 0.939, an F1-score of 0.941, and a QWK of 0.971. These results indicate that although the classification models do not explicitly model sequential dependencies between intensity levels, the MLP classifier achieved an ordinal consistency comparable to that of regression-based models. This demonstrates that the proposed MLP classification model can provide stable multi-level activity intensity recognition while maintaining high efficiency in resource-constrained on-device environments.

Table 6. Performance comparison of classification-based models (non-MLP vs. MLP).

Model	Accuracy	Precision	Recall	F1	QWK
Logistic Regression	0.931 ± 0.007	0.917	0.924	0.92	0.968 ± 0.004
MLP Classifier	0.939 ± 0.006	0.94	0.943	0.941	0.971 ± 0.003

To provide further insight into the behavior of the best-performing model, the per-class precision, recall, and F1-score of the MLP classifier are reported in Table 7.

Table 7. Per-class performance of the MLP classifier across the five activity intensity levels.

Level	Precision	Recall	F1-Score	Support
1	0.9599	0.9423	0.9511	763
2	0.9040	0.9253	0.9146	509
3	0.9270	0.9203	0.9236	138
4	0.9551	0.9659	0.9605	176
5	0.9524	0.9615	0.9569	104

5.3. Feature Contribution Analysis

To further examine the behavior of the coefficient-based models, feature contributions were analyzed for both Linear Regression and Logistic Regression. Feature importance was estimated based on the absolute magnitude of the learned model coefficients, which reflects the relative influence of each feature on the model predictions. Because polynomial feature expansion was applied during pre-processing, additional interaction terms derived from the original four features were generated. Each expanded feature was mapped back to its original feature group based on the feature name, and the absolute coefficient magnitudes were aggregated into two groups: SVM-based features and ΔSVM-based features. The contribution ratio was computed as the proportion of the summed absolute coefficients within each feature group relative to the total coefficient magnitude.

For Linear Regression, ΔSVM-based features accounted for 0.6447 of the total coefficient magnitude, compared to 0.3553 for SVM-based features. A similar pattern was observed in Logistic Regression, where ΔSVM-based features contributed 0.688, whereas SVM-based features contributed 0.312, as summarized in Table 8.

https://doi.org/10.3390/electronics15071351

Table 8. Feature contribution ratios based on the absolute magnitude of model coefficients.

Feature Group	Linear Regression	Logistic Regression
SVM-based features	0.3553	0.3120
ΔSVM-based features	0.6447	0.6880

To further validate these results, permutation importance analysis [25] was conducted on the test set using classification accuracy as the evaluation metric. This method evaluates feature importance by measuring the decrease in model performance when the values of a specific feature are randomly permuted. In the Logistic Regression model, permuting the mean of ΔSVM resulted in the largest decrease in accuracy, substantially exceeding that of representative SVM-based features such as the mean of SVM, as shown in Table 9.

Table 9. Permutation importance of representative features in the Logistic Regression model.

Feature	Accuracy Decrease
Mean of ΔSVM	0.4767
Mean of SVM	0.1312

Overall, these findings indicate that variation-oriented dynamics captured by ΔSVM play a more influential role in distinguishing activity intensity levels than magnitude-oriented SVM statistics.

5.4. Confusion Matrix and Error Analysis

Figure 22 shows the normalized confusion matrix of the MLP classifier on the test set. Unlike conventional HAR studies that mainly focus on recognizing daily activities, the evaluation dataset in this study includes customized emergency-like wrist motions collected for wearable-based personal safety monitoring. This design allows the proposed model to capture activity intensity patterns relevant to potential risk situations that are rarely represented in existing public HAR datasets.

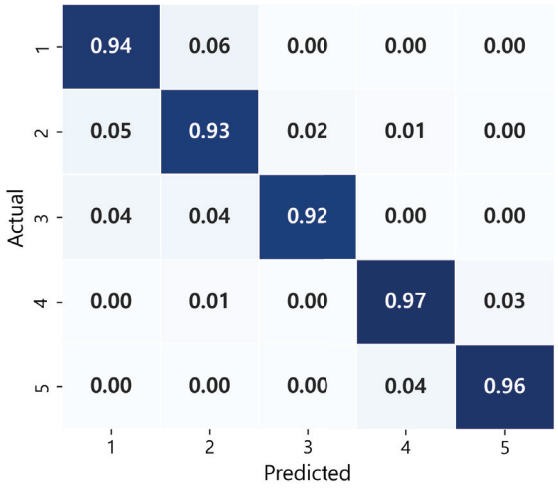


Figure 22. Confusion matrix of the MLP classifier model on the test set (activity intensity Levels 1–5). The color intensity represents the normalized value, where darker shades indicate a higher classification frequency.

The matrix exhibits strong concentration along the diagonal, confirming that most samples were correctly classified into their corresponding activity intensity levels. Misclassifications were primarily observed between adjacent levels (Level 2↔3 and Level 4↔5),

indicating that the model successfully captured the ordinal relationships between intensity levels. Furthermore, the limited confusion observed between adjacent levels can be interpreted as a reflection of natural boundary transitions that occur within the continuous spectrum of human motion. While such cases represent ordinal misclassification between neighboring levels, they are consistent with the gradual transitions that occur between similar activity intensities. Notably, confusion was minimal between Level 1 (static activities) and Level 5 (vigorous or high-intensity activities), demonstrating the model's ability to clearly distinguish extreme activity levels.

In conclusion, the proposed MLP classifier consistently distinguished multi-level activity intensities—from static to highly vigorous activities—while stably learning the sequential structure between intensity levels. For completeness, additional confusion matrices for the remaining three models—classification (non-MLP), regression (MLP), and regression (non-MLP)—are provided in Figures 23–25.

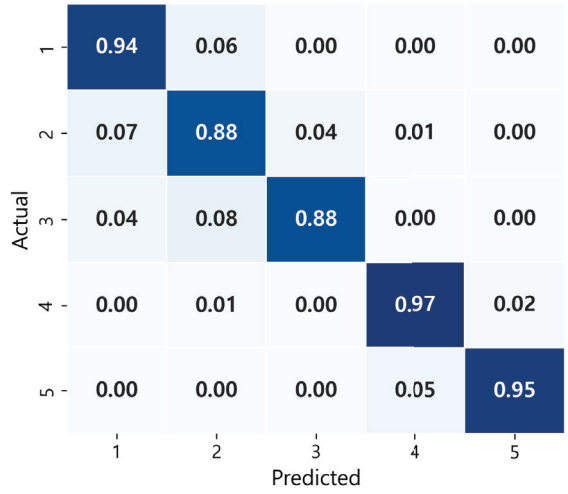


Figure 23. Confusion matrix of the Logistic Regression model on the test set (activity intensity Levels 1–5). The color intensity represents the normalized value, where darker shades indicate a higher classification frequency.

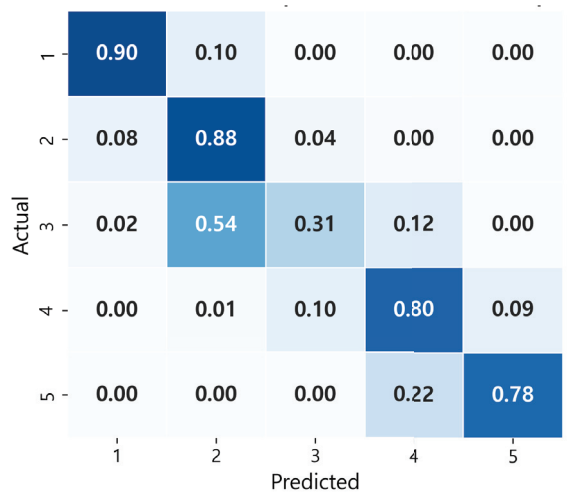


Figure 24. Confusion matrix of the Linear Regression model on the test set (activity intensity Levels 1–5). The color intensity represents the normalized value, where darker shades indicate a higher classification frequency.

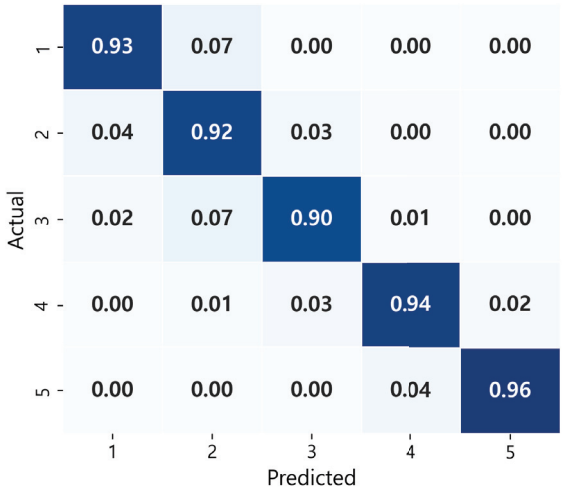


Figure 25. Confusion matrix of the MLP regression model on the test set (activity intensity Levels 1–5). The color intensity represents the normalized value, where darker shades indicate a higher classification frequency.

5.5. On-Device Evaluation

To verify the real-time applicability of the proposed activity intensity recognition model, on-device inference experiments were conducted on a Galaxy Watch 7. All experiments focused on real-time processing performance (inference latency, memory usage, and model size) as well as prediction stability, providing a comprehensive evaluation of the device-independence and real-time inference capability of the proposed approach.

5.5.1. Evaluation of MLP Classifier

The MLP-model classifier, deployed in its quantized TFLite form, was executed on the device. After 2000 iterations, the model achieved sub-millisecond inference latency (p50 = 0.038 ms) with minimal memory overhead, as shown in Table 10. It should be noted that this value reflects only the neural network computation time per window. The overall decision latency of the system is primarily determined by the fixed 6 s sliding window used for feature extraction, whereas the model inference introduces negligible additional delay. Here, p50, p90, and p99 denote the 50th, 90th, and 99th percentile latencies, respectively, characterizing both typical and worst-case inference times. Peak memory indicates the maximum runtime memory consumption during inference, while model size represents the storage footprint of the quantized TFLite model on the device.

Table 10. On-device inference performance of the MLP classifier deployed in its quantized TFLite form on the Galaxy Watch 7. The table reports inference latency (p50, p90, p99), peak memory usage, and model size under the default XNNPACK CPU delegate.

Configuration	p50 (ms)	p90 (ms)	p99 (ms)	Peak Memory (MB)	Model Size (MB)
XNNPACK (default)	0.038	0.073	0.226	50.38	0.088

The model enabled predictions directly on the watch; however, temporary fluctuations in predicted intensity levels were occasionally observed during wrist shaking or small stationary motions. To mitigate this issue, post-processing rules were applied. As a post-processing step, the prediction probabilities from recent windows were smoothed using a moving average, and a hysteresis condition was introduced so that level transitions occurred only when consistent predictions were maintained for a certain duration. In addition, a median filter and probability-based constraints were applied to suppress transient spikes caused by abrupt jerks.

https://doi.org/10.3390/electronics15071351

These post-processing rules consist only of simple arithmetic operations, enhancing prediction stability without affecting inference latency. Experimental results showed that erroneous high-intensity predictions during low-intensity activities were effectively suppressed. The model maintained stable multi-level predictions across various activity scenarios.

To further evaluate generalization performance under real-world conditions, a real-time subject-independent validation was conducted using three newly recruited participants (Subjects 09–11) who were entirely excluded from the training and hyperparameter tuning processes. The deployed quantized MLP model was executed directly on the Galaxy Watch without retraining, and predictions were generated in real time while the participants performed activities corresponding to each intensity level. The unseen participants performed representative activities for each intensity level, including lying (Level 1), walking (Level 2), dynamic handclap (Level 3), rope jumping (Level 4), and running (Level 5).

As shown in Table 11, for these unseen users, the model maintained consistent multi-level intensity recognition without observable performance degradation compared to the original test results. These findings demonstrate that the proposed MLP classifier generalizes robustly to new users under fully on-device, real-time operating conditions, thereby supporting its practical applicability in wearable deployment scenarios.

Table 11. Performance comparison between the original test set and real-time subject-independent validation (MLP classifier). Unseen participants performed representative activities across all five intensity levels.

Evaluation Setting	Accuracy	QWK
Original Test Set	0.939	0.971
Real-Time Unseen Subjects	0.952	0.988

5.5.2. Evaluation of the Non-MLP Classifier

A comparative experiment was conducted by implementing a non-MLP classifier on the same device. Due to fundamental differences in model architecture and processing pipelines, the evaluation metrics for the MLP classifier and the non-MLP classifier differ. While the non-MLP evaluation emphasizes inference latency and resource usage under optimized CPU delegation, as shown in Table 10, the non-MLP model assessment focuses on end-to-end latency including the fixed processing window, with detailed measurements presented in Table 12. This complementary evaluation approach provides a holistic understanding of on-device execution feasibility, distinguishing between model computation time and end-to-end system latency.

This model was implemented using explicit mathematical formulas as its internal logic, with all computations executed in real-time on the watch CPU. The end-to-end latency, which includes a single 6 s processing window, was primarily determined by the fixed window duration rather than computational overhead. This result confirms that the model can perform efficient on-device inference under smartwatch constraints, with Table 12 presenting the detailed end-to-end latency measurements. Here, end-to-end (E2E) latency refers to the total decision delay, including the fixed 6 s processing window and the subsequent computation time. As in the previous analysis, p50, p90, and p99 correspond to the 50th, 90th, and 99th percentile E2E latencies, respectively, highlighting both typical and worst-case decision delays on the device.

https://doi.org/10.3390/electronics15071351

Table 12. End-to-end (decision) latency of the non-MLP classification pipeline on the Galaxy Watch 7. This table reports the total latency including the fixed 6 s processing window, thus focusing on overall pipeline delay rather than resource consumption.

Scenario	E2E (p50, ms)	E2E (p90, ms)	E2E (p99, ms)
Real-time monitoring	6014.5	6020.3	6070.0

The post-processing of the non-MLP classifier was designed to apply additional probability thresholds and margin conditions between classes. This design holds predictions with low confidence and only allows level transitions when a certain level of cumulative change is reached. Furthermore, a feature gate was applied to temporarily hold results when input features deviated from the normal range, preventing unstable judgments. These rules require negligible computation and were optimized to operate effectively within the real-time inference cycle. The model maintained prediction trends similar to the MLP while requiring significantly lower computational cost.

6. Discussion

This study proposed a model capable of multi-level activity intensity recognition, ranging from low-intensity daily activities to vigorous behaviors associated with higher physical risk, using only the accelerometer sensor of a smartwatch. The proposed MLP Approaches successfully learned a five-level continuous intensity structure, showing limited confusion only between adjacent intensity intervals. Moreover, the ΔSVM-based feature sensitively captured instantaneous motion changes, effectively distinguishing atypical or vigorous movements.

Unlike previous wearable HAR studies that primarily focused on single-purpose binary classification tasks such as fall detection or exercise intensity estimation, this work primarily focused on self-collected dynamic behaviors and further incorporated static and daily activities to extend the dynamic intensity scale into a continuous spectrum. This approach enabled the recognition of gradual transitions in activity intensity rather than discrete class distinctions. Furthermore, the same algorithm was implemented in both MLP and non-MLP approaches, demonstrating real-time operation. These results demonstrate that the proposed model can support efficient on-device activity intensity recognition on the Galaxy Watch 7 platform.

Another notable contribution of this study lies in its potential for personalization. The User_level5 label, representing user-defined vigorous or hazardous situations, can be re-collected and re-trained to reflect individual movement patterns. The personalized model can be directly deployed in models, with adjustable thresholds for customized sensitivity, enabling user-driven risk detection based on activity intensity.

Nevertheless, several limitations should be acknowledged. First, the data collection was limited to female participants in their 20s, and further validation is required to generalize the findings to broader populations across different ages, genders, and wearing positions. In addition, the data were collected in controlled indoor environments; therefore, robustness against environmental noise and variations in sensor placement in real-world and industrial settings remains to be evaluated.

7. Conclusions

This study proposed an on-device model capable of real-time recognition of five activity intensity levels using only three-axis acceleration data from smartwatches. The MLP classification model, based on four statistical features (SVM mean and standard deviation, ΔSVM mean and standard deviation), achieved 0.939 accuracy and a QWK of 0.971. It achieved sub-millisecond inference latency ($<$1 ms) and a compact model size

(<0.1 MB), experimentally confirming stable real-time operation even under the limited computational resources of smartwatch environments.

The proposed framework demonstrated the ability to recognize risk levels across ordered activity intensity levels, enabling stepwise interpretation of activity intensity beyond simple activity classification. This provides a technical foundation for implementing real-time risk awareness and stepwise alert systems in wearable environments. Furthermore, by re-collecting and retraining the User_level5 data defined by individual users, the model can be extended into a personalized risk detection system tailored to each user's motion characteristics. In addition, fine-tuning post-processing thresholds allows customization of alert sensitivity to reflect the user's personal risk perception or behavioral tendencies.

Future research could consider the (i) validation of generalization across diverse populations, wear locations, and activity environments; (ii) fusion of multimodal signals such as heart rate and gyroscope data; (iii) automatic optimization of user-specific retraining and post-processing thresholds; and (iv) adaptive update techniques to address data drift during long-term use. Through these directions, the proposed system could evolve into a personalized risk detection and prediction platform that adapts to real-world conditions.

In conclusion, the proposed approach simultaneously achieved high accuracy and real-time performance using a single sensor, demonstrating its potential for future development into wearable-based personal safety management and industrial safety monitoring systems.

Author Contributions: Conceptualization, J.Y.; methodology, S.K. and J.Y.; software, S.K.; validation, S.K. and J.Y.; formal analysis, S.K.; investigation, S.K.; resources, S.K.; data curation, J.Y.; writing—original draft preparation, S.K.; writing—review and editing, S.K. and J.Y.; visualization, S.K.; supervision, J.Y.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Sungshin Women's University Research Grant of 2026.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Sungshin Women's University (SSWUIRB-2024-061).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The Dyn-Risk dataset generated in this study, along with the pre-processed PAMAP2 subsets used for model integration, is available at https://github.com/AILAB-IMU/Dyn-Intensity (accessed on 9 February 2026).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Del Rosario, M.B.; Lovell, N.H.; Redmond, S.J. Learning the Orientation of a Loosely-Fixed Wearable IMU Relative to the Body Improves the Recognition Rate of Human Postures and Activities. *Sensors* **2019**, *19*, 2845. [CrossRef] [PubMed]
2. Mao, Y.; Yan, L.; Guo, H.; Hong, Y.; Huang, X.; Yuan, Y. A Hybrid Human Activity Recognition Method Using an MLP Neural Network and Euler Angle Extraction Based on IMU Sensors. *Appl. Sci.* **2023**, *13*, 10529. [CrossRef]
3. Fridriksdottir, E.; Bonomi, A.G. Accelerometer-Based Human Activity Recognition for Patient Monitoring Using a Deep Neural Network. *Sensors* **2020**, *20*, 6424. [CrossRef]
4. Bibbo, L.; Angiulli, G.; Laganà, F.; Pratticò, D.; Cotroneo, F.; La Foresta, F.; Versaci, M. MEMS and IoT in HAR: Effective Monitoring for the Health of Older People. *Appl. Sci.* **2025**, *15*, 4306. [CrossRef]
5. Bulling, A.; Blanke, U.; Schiele, B. A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. *ACM Comput. Surv.* **2014**, *46*, 33. [CrossRef]
6. Attal, F.; Mohammed, S.; Dedabrishvili, M.; Chamroukhi, F.; Oukhellou, L.; Amirat, Y. Physical Human Activity Recognition Using Wearable Sensors. *Sensors* **2015**, *15*, 31314–31338. [CrossRef] [PubMed]
7. Seenath, S.; Dharmaraj, M. Conformer-Based Human Activity Recognition Using Inertial Measurement Units. *Sensors* **2023**, *23*, 7357. [CrossRef] [PubMed]

8. Zhou, H.; Zhang, X.; Feng, Y.; Zhang, T.; Xiong, L. Efficient Human Activity Recognition on Edge Devices Using DeepConv LSTM Architectures. *Sci. Rep.* **2025**, *15*, 13830. [CrossRef] [PubMed]
9. Gomes, E.; Bertini, L.; Campos, W.R.; Sobral, A.P.; Mocaiber, I.; Copetti, A. Machine Learning Algorithms for Activity-Intensity Recognition Using Accelerometer Data. *Sensors* **2021**, *21*, 1214. [CrossRef]
10. Chen, D.; Du, Y.; Liu, Y.; Hong, J.; Yin, X.; Zhu, Z.; Wang, J.; Zhang, J.; Chen, J.; Zhang, B.; Du, L.; et al. Development and Validation of a Smartwatch Algorithm for Differentiating Physical Activity Intensity in Health Monitoring. *Sci. Rep.* **2024**, *14*, 9530. [CrossRef] [PubMed]
11. Wei, B.; Romano, C.; Pedram, M.; Nolan, B.; Morelli, W.A.; Alshurafa, N. Developing and Comparing a New BMI-Inclusive Energy Expenditure Algorithm on Wrist-Worn Wearables. *Sci. Rep.* **2025**, *15*, 20060. [CrossRef] [PubMed]
12. Wang, X.; Yu, H.; Kold, S.; Rahbek, O.; Bai, S. Wearable Sensors for Activity Monitoring and Motion Control: A Review. *Biomim. Intell. Robot.* **2023**, *3*, 100089. [CrossRef]
13. Pernek, I.; Kurillo, G.; Stiglic, G.; Bajcsy, R. Recognizing the Intensity of Strength Training Exercises with Wearable Sensors. *J. Biomed. Inform.* **2015**, *58*, 145–155. [CrossRef]
14. Hossain, E.; Hossain, M.; Alam, M.; Haque, M.; Uddin, M.; Haque, M. Ordinal statistical models of physical activity levels from accelerometer data. *Int. J. Exerc. Sci.* **2021**, *14*, 1460–1475. [CrossRef] [PubMed]
15. Bağcı Daş, D.; Birant, D. Ordered physical human activity recognition based on ordinal classification. *Turk. J. Electr. Eng. Comput. Sci.* **2021**, *29*, 2478–2496. [CrossRef]
16. Bian, S.; Liu, M.; Rey, V.F.; Geissler, D.; Lukowicz, P. TinierHAR: Towards Ultra-Lightweight Deep Learning Models for Efficient Human Activity Recognition on Edge Devices. In Proceedings of the 2025 ACM International Symposium on Wearable Computers (ISWC), Espoo, Finland, 12–16 October 2025.
17. Yeon, T.; Xu, V.; Hoffmann, H.; Ahuja, K. WatchHAR: Real-Time On-Device Human Activity Recognition System for Smartwatches. In Proceedings of the 27th International Conference on Multimodal Interaction (ICMI), Canberra, Australia, 13–17 October 2025. [CrossRef]
18. Huang, X.; Chen, Y.; Wang, W.; Wu, H.; Song, A. Deep Ensemble Learning for Human Activity Recognition Using Wearable Sensors via Filter Activation. *J. ACM* **2022**, *1*, 1–24. [CrossRef]
19. Lattanzi, E.; Calisti, L.; Capellacci, P. Lightweight Accurate Trigger to Reduce Power Consumption in Sensor-Based Continuous Human Activity Recognition. *Pervasive Mob. Comput.* **2023**, *96*, 101848. [CrossRef]
20. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; Wiley: Hoboken, NJ, USA, 2012.
21. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; Wiley: New York, NY, USA, 2000.
22. TensorFlow. TensorFlow Lite Guide. Available online: https://www.tensorflow.org/lite/guide (accessed on 16 October 2025).
23. Reiss, A.; Stricker, D. Introducing a New Benchmarked Dataset for Activity Monitoring. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 108–109.
24. Oh, H.; Yoo, J. Real-Time Detection of At-Risk Movements Using Smartwatch IMU Sensors. *Appl. Sci.* **2025**, *15*, 1842. [CrossRef]
25. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Electronics Editorial Office
E-mail: electronics@mdpi.com
www.mdpi.com/journal/electronics



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-8092-8