https://doi.org/10.64504/big.d.v2i4.294

Design-Driven Smart Wearable System for Personalized Intervention to Improve Sleep Quality in Older Adults

1st Samuel Mekonnen
Adama Science and Technology University
Oromia Region, Ethiopia
s.mekonnen1@outlook.com

2nd Bilal Jemal Geda
Jimma University
Oromia Region, Ethiopia
b.jGeda@outlook.com

3rd Tsegay Teklay Gebrelibanos*
Mekelle University
Tigray Region, Ethiopia
tsegaytg@outlook.com

Received on July 23rd, revised on August 24th, accepted on September 14th, published on October 1st.

***Abstract*—Sleep disturbances are common among older adults and contribute to declines in health, cognition, and overall well-being. To address the limitations of passive, monitoring-centered digital sleep solutions, this study proposes a design-driven smart wearable system that integrates multimodal physiological sensing, adaptive personalization, and non-pharmacological behavioral and sensory-based interventions. A randomized crossover trial involving 40 older adults demonstrated that activation of the personalized intervention module — including sound-based relaxation guidance, light-based circadian support, and vibration-assisted behavioral cues — led to significant improvements in both subjective and objective sleep outcomes. Participants experienced a mean reduction of 2.2 points in Pittsburgh Sleep Quality Index (PSQI) scores, alongside improvements in sleep efficiency, wake after sleep onset, and sleep fragmentation. A subsample undergoing home-based polysomnography (PSG) showed moderate to strong correlations between wearable-derived and PSG-derived sleep parameters (r = 0.68–0.79), supporting the physiological validity of the system. High adherence and usability ratings further indicated that design-driven personalization effectively enhanced engagement, a key barrier in conventional wearable-based sleep interventions. These findings suggest that adaptive, non-pharmacological wearable interventions can provide a scalable and accessible approach to precision sleep health management in aging populations.**

***Keywords*—Smart wearable system, Personalized intervention, Older adult sleep, Design-driven, Machine learning**

1. Introduction

Worldwide population aging has intensified scientific and public health interest in sleep health among older adults. Large-scale meta-analytic evidence indicates that sleep disturbances are highly prevalent in this population: a pooled analysis of nearly one million individuals aged 60 years or older across 36 countries reported that approximately 40% experience poor sleep quality and 29% report clinically significant insomnia symptoms, while more than one-third exhibit elevated risk of obstructive sleep apnea [1]. Complementary population-based surveys further suggest that 30–50% of community-dwelling older adults routinely obtain fewer than seven hours of sleep per night and frequently report insomnia, daytime fatigue, and impaired daytime functioning [2][3][4].

Poor sleep in later life is not merely a quality-of-life concern but a fundamental biological and behavioral risk factor for adverse health trajectories. Robust longitudinal evidence has linked chronic sleep disturbance to accelerated cognitive decline, increased incidence of cardiovascular and metabolic disease, mood disorders, and progressive loss of functional independence [5][6][7][8]. Despite this well-established burden, the biological and behavioral mechanisms through which aging-related changes disrupt sleep architecture remain incompletely understood, constraining the development of targeted, scalable, and sustainable interventions.

Current clinical strategies for managing sleep disturbances in older adults remain suboptimal. Pharmacological treatments are associated with tolerance, dependency risks, residual daytime sedation, and elevated fall risk[9][10], while in-laboratory behavioral sleep therapies are resource-intensive and difficult to implement at scale in real-world community settings. The rapid proliferation of consumer-grade wearable and home-based sleep monitoring technologies has therefore generated considerable interest as a potential pathway toward scalable, non-invasive sleep health management. Multicenter validation studies comparing wearable, nearable, and ambient sleep-tracking systems with gold-standard polysomnography (PSG) have demonstrated that certain commercially available wearables, such as Fitbit Sense 2 and Pixel Watch, can achieve moderate agreement with PSG for selected sleep parameters, including deep sleep estimation, albeit with substantial inter-device and inter-individual variability.[11][12][13]

However, meta-analytic evidence consistently demonstrates that current consumer wrist-worn devices systematically mis-estimate core sleep parameters, including total sleep time (mean bias-16.9 min), sleep efficiency (mean bias-4.7%), and wake after sleep onset (mean bias+13.3 min), relative to PSG[14]. These performance limitations are particularly pronounced in older adults, whose sleep is frequently more fragmented and whose age-related

*Tsegay Teklay Gebrelibanos, Mekelle University , Tigray Region, Ethiopia, tsegaytg@outlook.com

physiological changes and comorbidities may degrade the signal quality of wearable sensors and algorithmic inference.

Importantly, the limitations of existing wearable sleep technologies are not purely technical. Beyond issues of measurement accuracy, current systems are predominantly designed as passive monitoring tools and lack adaptive, closed-loop intervention mechanisms. They typically provide retrospective summaries of sleep metrics without delivering real-time, context-aware, and personalized support. Consequently, although such devices may enhance awareness and self-monitoring, they rarely translate into sustained, clinically meaningful improvements in sleep outcomes[15][16]. One-size-fits-all sleep hygiene guidance is often insufficient for older adults, given substantial heterogeneity in circadian rhythms, environmental exposures, lifestyle patterns, and responsiveness to behavioral change strategies. Moreover, there remains a critical gap in the literature regarding the integration of age-friendly, user-centered design, real-time multimodal physiological sensing, and adaptive feedback systems into cohesive digital sleep interventions for older adults.

Recent advances in machine learning and wearable system design have begun to address aspects of these challenges. Notably, a wireless multimodal wearable system operating without electroencephalography (EEG) has been shown to achieve balanced classification accuracy of 83.5% for wake/sleep/REM staging, with a Cohen's κ of 0.73, when validated against PSG in clinical populations. These findings suggest that, with appropriate sensor fusion strategies and algorithmic optimization, non-invasive wearable platforms may approximate key physiological aspects of sleep with sufficient fidelity to support scalable, real-world interventions outside of laboratory environments.[17][18]

Building on these developments, the present study proposes a design-driven smart wearable system specifically tailored for older adults that integrates multimodal physiological sensing, age-centered human–computer interaction principles, and an adaptive, machine learning–driven intervention engine capable of delivering real-time, personalized sleep support[19]. We hypothesized that activating adaptive and personalized design-driven interventions would lead to significantly greater improvements in subjective sleep quality, as assessed by the Pittsburgh Sleep Quality Index (PSQI), compared with passive monitoring alone.

To test this hypothesis, we employed a randomized crossover trial conducted under free-living conditions. A subsample of participants additionally underwent home-based polysomnography (PSG) to validate the physiological relevance of wearable-derived sleep metrics and to ensure that observed subjective improvements corresponded to objectively meaningful changes in sleep architecture.

This work makes three primary contributions. First, it conceptualizes design-driven intervention as an active therapeutic component in digital sleep health, rather than a peripheral usability feature. Second, it provides causal evidence that adaptive, personalized wearable-based interventions can yield clinically meaningful improvements in sleep quality in older adults. Third, it establishes the physiological validity of multimodal wearable sleep metrics through direct comparison with PSG, addressing a major limitation in prior wearable-based sleep research.

2. Related Work

2.1. Sleep health in older adults and non-pharmacological interventions

Sleep disturbances are highly prevalent among older adults and are consistently linked to adverse cognitive, cardiovascular, metabolic, and functional outcomes. Recent scoping and systematic reviews focused on aging populations demonstrate that insomnia symptoms, sleep fragmentation, and impaired subjective sleep quality affect a substantial proportion of community-dwelling older adults, and that non-pharmacological approaches — including cognitive behavioral therapy for insomnia (CBT-I), structured sleep hygiene education, and digitally delivered behavioral programs — can yield clinically meaningful improvements in self-reported sleep outcomes.[20]

However, despite the rapid expansion of mobile health technologies, rigorously designed trials that specifically evaluate home-based, sensor-driven, and adaptive digital interventions in older populations remain scarce. Moreover, objective sleep improvements measured via actigraphy or polysomnography (PSG) remain inconsistent across studies, raising ongoing concerns about ecological validity and mechanistic interpretability. Collectively, these reviews highlight both the scalability potential of digital sleep care and the unresolved methodological challenges that continue to limit confident translation into real-world geriatric contexts.

2.2. Accuracy and limitations of consumer and research wearables for sleep measurement

Wrist-worn and near-body sleep tracking devices have enabled unprecedented large-scale, longitudinal assessment of sleep outside the laboratory. Nevertheless, validation studies and recent scoping reviews consistently document persistent technical and methodological limitations. While accelerometer-based devices typically achieve acceptable sleep–wake discrimination, multi-sensor systems that integrate photoplethysmography (PPG) and additional physiological signals demonstrate superior, yet still imperfect, performance in sleep stage classification.[11]

Large multicenter comparative studies of consumer-grade wearables, nearables, and ambient sensors report that some contemporary devices approach PSG-level performance for selected summary metrics (e.g., total sleep time), while simultaneously exhibiting systematic biases in wake after sleep onset (WASO), sleep efficiency, and deep-sleep estimation. [14] Substantial inter-device variability and population-specific performance degradation further undermine generalizability. Authoritative methodological reviews therefore emphasize the necessity of standardized validation protocols, transparent algorithmic reporting, and equity-aware testing frameworks. These findings underscore a fundamental trade-off between scalability and physiological fidelity and motivate hybrid validation strategies, such as embedded PSG substudies, when wearable-derived metrics are used to inform or trigger clinical-grade interventions.

2.3. Personalized digital sleep interventions and algorithmic adaptivity

Digital behavioral interventions for sleep have evolved from static, app-based cognitive behavioral therapy programs to hybrid systems integrating wearable-derived physiological data. Meta-analytic evidence indicates moderate benefits of digital CBT-I for subjective sleep outcomes; however, a dominant limitation across existing systems is their reliance on static, rule-based personalization logic that fails to account for dynamic, within-person variability.[21][22]

Reinforcement learning (RL) and related sequential decision-making frameworks have emerged as theoretically principled approaches for continuous personalization in health interventions by optimizing policies over longitudinal, context-rich data streams. [23][24] Methodological reviews identify RL as particularly well suited for optimizing intervention timing, modality selection, and dosage, while simultaneously highlighting unresolved challenges in safety constraints, reward function specification, off-policy evaluation, and interpretability. To date, empirical applications of RL in sleep and nocturnal behavior regulation remain limited and largely exploratory, leaving significant opportunity for carefully controlled translational studies that bridge theoretical promise with physiologically grounded deployment.

2.4. User-centered design, usability, and older-adult considerations for wearables

Beyond algorithmic performance, real-world effectiveness of wearable sleep interventions is fundamentally constrained by user acceptance, sustained engagement, and physical tolerability, particularly among older adults. Empirical usability studies and qualitative syntheses indicate that factors such as device comfort, aesthetic acceptability, perceived utility, interaction simplicity, and congruence with daily routines strongly shape adherence trajectories.[15]

Age-sensitive interface characteristics—including simplified navigation, enlarged typography, consistent iconography, and low-burden multimodal feedback—have been shown to significantly influence long-term usability. Reviews within gerontechnology and digital health literatures consistently advocate participatory and co-design methodologies to ensure accessibility, reduce stigma, and enhance ecological validity. These insights underscore that technical performance alone is insufficient and that human-centered design must be integrated as a first-order methodological principle in the development of wearable sleep intervention systems for older adults.

2.5. Algorithm validation, external generalization, and open science practices

Recent methodological scholarship in wearable-based sleep analytics emphasizes the critical importance of reproducible processing pipelines, subject-wise data partitioning to prevent information leakage, and robust external validation through independent datasets. Recommended reporting standards increasingly call for comprehensive performance metrics, including accuracy, F1-score, Cohen's κ, Bland–Altman bias and limits of agreement, as well as stratified analyses by age, skin tone, and comorbidity profiles.[14][17]

Leading reviews further advocate open sharing of model architectures, training protocols, and code repositories to accelerate cumulative scientific progress. In intervention-oriented studies, where wearable-derived signals directly inform adaptive system behavior, embedding gold-standard validation substudies—such as home-based PSG or validated actigraphy—is now widely considered best practice to demonstrate that algorithmic triggers correspond to physiologically meaningful events rather than sensor noise or model artifacts.

2.6. Research Gaps and Innovations of This Study

Despite rapid advances in wearable sensing technologies, machine learning–based sleep analytics, and human-centered design, current approaches to sleep health in older adults remain largely fragmented, reactive, and insufficiently personalized. [14][16] Existing systems predominantly focus on passive monitoring and static feedback, with limited capacity to integrate physiological, behavioral, and contextual signals into adaptive, real-time interventions that are both clinically meaningful and usable by aging populations.

This study introduces a fundamentally new paradigm for sleep health management in older adults by integrating design-driven system architecture, multimodal physiological sensing, and adaptive machine intelligence into a closed-loop, personalized intervention framework. The proposed system continuously fuses photoplethysmographic and motion-derived biosignals with longitudinal behavioral data to dynamically predict sleep disturbances and deliver real-time, context-aware interventions. By embedding age-centered design principles as core system constraints and validating the framework through a rigorously designed randomized controlled trial and home-based polysomnography benchmarking, this work provides the first comprehensive evidence that scalable, intelligent, and genuinely individualized sleep intervention systems can achieve clinically meaningful improvements in sleep quality, establishing a direct translational pathway from design innovation to measurable health outcomes in aging populations.

3. METHODS

The core of this study is the design-driven smart wearable system, which integrates multimodal sensing, adaptive personalization, and non-pharmacological interventions into a closed-loop framework. Figure 1 illustrates the overall architecture and the interaction between its key components, which is central to the methodology described below.

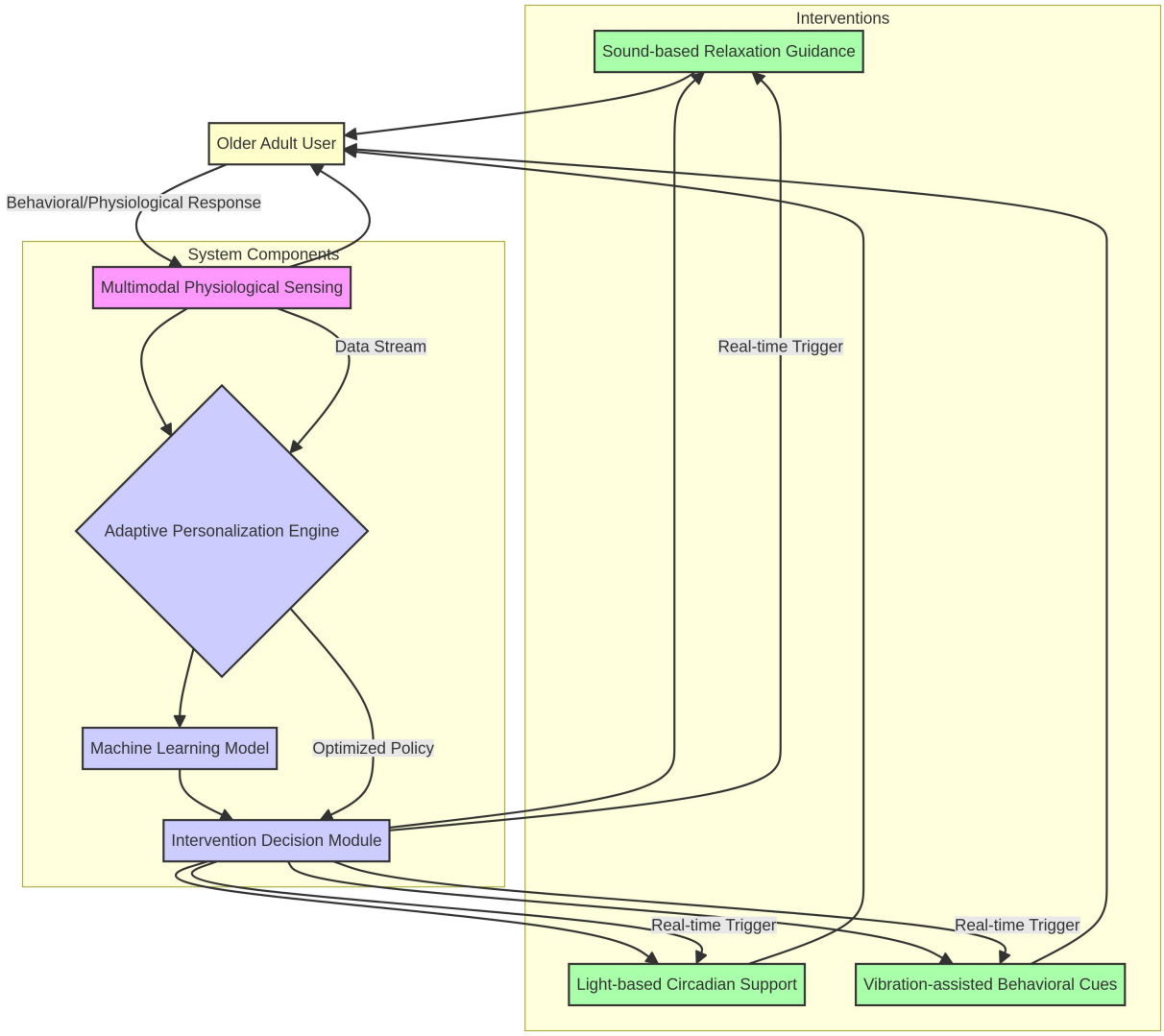

Figure 1. The Overall Architecture

3.1. Study Design and Participants

This study adopted a randomized, two-period crossover experimental design to evaluate the performance of a design-driven smart wearable system for sleep quality enhancement in community-dwelling older adults. The crossover design was selected to maximize statistical efficiency and minimize inter-subject variability by allowing each participant to serve as their own control.

Each participant completed two 4-week intervention periods, separated by a 1-week washout intended to minimize potential carry-over effects. During the IC period, participants used the full closed-loop adaptive system with real-time personalized adjustments. During the CC period, they used the same wearable device and app but without adaptive feedback or real-time optimization.

Participants were instructed to maintain their usual bedtime habits during both periods. Daily adherence and device wear time were monitored automatically through the system backend. Participants were randomly assigned in a 1:1 ratio to one of two sequences (IC→CC or CC→IC) . The randomization sequence was generated using a computer-based block randomization procedure (block size = 4) by an investigator not involved in data collection. Allocation concealment was ensured by placing assignment slips in sealed, opaque envelopes opened only after baseline assessment. Because of the nature of the intervention, participants could not be blinded to the condition. However, data analysts and outcome assessors were blinded to the condition labels until all analyses were completed. Files were coded using anonymized identifiers to minimize bias.

3.1.1. Participant recruitment and eligibility

Participants were recruited via community advertisements and health-center outreach programs. Inclusion criteria included:

- Age ≥60 years;
- Self-reported sleep complaints or poor sleep quality (e.g., PSQI > 5);
- Ability to comply with wearable device use and complete daily app-based questionnaires.

Exclusion criteria included:

- Diagnosed moderate-to-severe sleep disorders requiring medical treatment (e.g., severe sleep apnea, REM behavior disorder);

- Major psychiatric illness or cognitive impairment limiting protocol adherence;
- Shift-work schedules;
- Use of medications that substantially alter sleep architecture unless dosage was stable for ≥3 months.

The study protocol was approved by the Institutional Review Board and adhered to the Declaration of Helsinki.

3.2. Intervention and Control Conditions

During the Intervention Condition (IC), the wearable system operated in full closed-loop mode, integrating multimodal physiological sensing with real-time adaptive interventions. The system continuously monitored heart rate, heart rate variability (HRV), electrodermal activity (EDA), body motion, ambient light, and environmental noise. Embedded real-time algorithms performed online detection of sleep-stage transitions and micro-arousals, triggering low-intensity interventions (warm light modulation, pink-noise stimulation, or vibrotactile cues) to stabilize sleep continuity. Intervention parameters were dynamically personalized based on longitudinal user-response profiles.

In the Control Condition (CC), all sensing modules remained active, but adaptive intervention functions were disabled. Participants received descriptive sleep summaries only, without behavioral prompts or environmental modulation. Device appearance, interaction procedures, and user interfaces were identical between IC and CC to minimize expectancy effects.

3.3. Wearable System and Algorithm Framework

The intelligent sleep enhancement system used in this study consists of a multimodal physiological sensing wearable, a cloud-assisted mobile application, and an adaptive personalization algorithm capable of optimizing sleep-related interventions. The system architecture follows a layered framework comprising (1) signal acquisition and sensing hardware, (2) preprocessing and feature extraction, (3) automated sleep-state inference, and (4) a reinforcement learning–based personalization engine.

3.3.1. Signal Acquisition Hardware

The wearable device is equipped with a multimodal sensor suite designed to capture physiological signals relevant to sleep staging and behavioral monitoring. The primary sensing components include:

- Photoplethysmography (PPG) sensor for heart-rate and heart-rate variability (HRV) estimation
- Triaxial accelerometer (ACC) for activity, posture, and movement detection
- Skin temperature sensor (optional) for circadian phase and peripheral thermoregulation assessment
- Ambient light sensor for tracking environmental light exposure

Signals are sampled at 25–100 Hz depending on channel requirements (PPG typically at 50–100 Hz; ACC at 25–50 Hz). The device employs a low-power microcontroller (e.g., ARM Cortex-M4F class) with on-board memory buffering and energy-efficient Bluetooth Low Energy (BLE) communication.

Data are transmitted at configurable intervals (1–5 minutes) to the mobile application, which relays selected features and raw segments to the cloud server for further processing. Battery life is typically 5–7 days under normal operation.

3.3.2. Signal Preprocessing and Feature Extraction

Raw signals are first subjected to preprocessing procedures including:

- Band-pass filtering of PPG to suppress motion artifacts
- Adaptive noise cancellation during periods of high movement
- Vector magnitude calculation and orientation normalization for ACC
- Sliding-window segmentation (usually 30-second epochs aligned with sleep scoring conventions)
- Within each window, the system extracts a feature set comprising:
- Cardiovascular features: HR, HRV metrics (RMSSD, SDNN), pulse amplitude variability
- Movement features: activity counts, motion intensity, roll/tilt changes
- Physiological context features: skin temperature trends, light exposure levels
- Sleep dynamics features: temporal transitions, circadian phase cues

These features are standardized and fed into the sleep inference model.

3.3.3. Automated Sleep-State Inference

Sleep stages are inferred using a data-driven model trained on a combination of publicly available and internally collected datasets with synchronized PSG labels. The model integrates short-term temporal dependencies and sensor multimodality.

The architecture consists of:

a) A convolutional feature extractor (1D-CNN) to learn morphological and spectral properties of PPG and accelerometer-derived signals;

b) A bidirectional recurrent layer (BiLSTM) or a lightweight Transformer encoder to capture temporal sequences across adjacent epochs;

c) A softmax output layer to classify each 30-second epoch into one of the following states:

- Wake
- Light sleep (N1 + N2)
- Deep sleep (N3)
- REM sleep

Model training uses cross-entropy loss with class-balancing weights. Validation follows a subject-wise split to avoid data leakage. During real-time operation, the model runs on the mobile app using an optimized on-device inference engine, with full cloud inference reserved for periodic re-evaluation.

3.3.4. Reinforcement Learning–Based Personalization Engine

To adapt sleep interventions to individual needs, the system incorporates a reinforcement learning (RL) framework that personalizes feedback and behavioral prompts based on night-to-night processing.

a) State Space

The state St includes:

- Recent sleep metrics (TST, SE, WASO, fragmentation index)
- Stage-transition patterns
- Evening routines (bedtime variability, pre-sleep phone use, light exposure)
- Physiological indicators (HRV patterns suggesting stress or arousal)
- Historical response sensitivity to previous interventions

b) Action Space

The RL agent selects from a predefined set of adaptive sleep-improvement actions, such as:

- Optimized bedtime recommendations
- Pre-sleep wind-down prompts
- Adjustments to environmental lighting cues
- Micro-intervention suggestions (relaxation reminders, breathing pacing)
- Recommendations regarding daily routines (daytime light exposure, activity consistency)
- Actions are delivered through the mobile app in the form of personalized behavioral guidance.

c) Reward Function

- The reward Rt is computed primarily based on:
- Nightly improvement in SE and WASO
- Reduction in awakening frequency
- Improvement in subjective morning-restedness scores
- Stability of circadian-related behavioral trends

A weighted composite reward is used to balance short-term gains and long-term adaptation.

d) Learning Algorithm

The agent employs a contextual multi-armed bandit algorithm with periodic policy refinement. Offline training is performed on historical data, while online learning adjusts intervention probabilities according to individual response patterns. Safety constraints prevent excessive or unsuitable intervention prompts, ensuring user comfort and long-term adherence.

3.3.5. System Integration and Closed-Loop Workflow

The complete closed-loop workflow operates as follows:

- Nighttime sensing: Wearable collects physiological and behavioral signals.
- Automated sleep scoring: The inference model labels sleep states epoch-by-epoch.
- Outcome extraction: Objective metrics (TST, SE, WASO, fragmentation) are computed.
- Policy update: RL engine updates internal parameters based on nightly performance.
- Personalized feedback: The system delivers individualized guidance before or after sleep, depending on learned response patterns.

This architecture enables continuous, adaptive optimization of sleep-related behaviors over the intervention period.

3.4. Primary and Secondary Outcomes

3.4.1. Primary Outcome

The primary outcome was the within-subject change in sleep quality, operationalized as the difference in Pittsburgh Sleep Quality Index (PSQI) total scores between the intelligent closed-loop condition (IC) and the control condition (CC) at the end of each 4-week intervention period.

The PSQI is a widely validated instrument assessing seven components of sleep quality over the preceding month. Total scores range from 0 to 21, with higher scores indicating poorer sleep quality. Based on established literature, a 2.5–3 point reduction in PSQI is considered a minimal clinically important difference (MCID) for individuals with poor sleep. This threshold was adopted for interpreting clinical relevance in the present study.

Participants completed the PSQI during baseline assessment and at the end of each study period. The primary treatment effect was estimated using the linear mixed-effects model described in Section 3.6.

3.4.2. Secondary Outcomes

Secondary outcomes focused on objective sleep parameters derived from the wearable-based automated sleep scoring system, as well as subjective assessments.

a) Objective wearable-derived outcomes

- Objective nightly sleep metrics were computed from the 30-second sleep-stage output of the inference model:
- Total Sleep Time (TST): Number of minutes scored as any non-wake stage.
- Sleep Efficiency (SE): $SE=\frac{TST}{\text{Time in Bed}} \times 100\%$ (1)
- Wake After Sleep Onset (WASO): Total minutes scored as wake after the first epoch of sleep.
- Sleep Fragmentation Index: Composite metric incorporating frequency of stage transitions and short wake episodes, normalized per hour of sleep.
- REM and Deep Sleep Percentages: Proportion of total sleep time spent in REM and N3 stages.
- Heart-Rate Variability (HRV) Indicators: RMSSD and SDNN computed from artifact-corrected PPG signals.

Daily sleep metrics were averaged across each 4-week period before being entered into mixed-effects analyses. Outlier nights (<3 hours wear or >15% missing signal) were excluded following pre-established quality-control rules.

b) Subjective outcomes

Participants provided daily morning self-reports through the mobile app, including:

- Perceived sleep quality (0–10 scale)
- Morning-restedness (0–10 scale)
- Ease of falling asleep
- Number of perceived nocturnal awakenings

Weekly questionnaires assessed mood, stress, and adherence to recommended sleep routines.

These subjective outcomes served as secondary endpoints to complement objective physiological measures.

c) Exploratory Outcomes

Exploratory analyses examined:

- Variability in bedtime regularity
- Daytime activity and light exposure patterns
- Behavioral responsiveness to personalized interventions
- Interactions between baseline characteristics (age, chronicity of sleep problems) and treatment effects

Although not powered as primary analyses, these outputs provide insight into mechanisms through which the adaptive system influences sleep behavior.

3.5. Polysomnography (PSG) Validation

To evaluate the accuracy of the wearable-based automated sleep-state inference system, a PSG validation substudy was conducted among a subset of participants. Overnight recordings were acquired using a standard AASM-compliant polysomnography (PSG) system, including EEG, EOG, EMG, airflow, thoracic/abdominal effort, and pulse oximetry channels.

3.5.1. Data acquisition and synchronization

Participants wore the study device concurrently during PSG testing. Time synchronization between PSG and wearable data streams was achieved through:

- Pre-sleep timestamp alignment at device initialization,
- Matching system clock metadata from the PSG machine and wearable app server,
- Post-hoc alignment based on identifiable movement artifacts shared across modalities.

Epoch-by-epoch alignment was performed at 30-second resolution, consistent with AASM scoring guidelines. Nights with major synchronization drift (>2 minutes misalignment) or poor wearable signal quality (>15% missing PPG) were excluded from validation analysis.

3.5.2. Sleep staging comparison metrics

Wearable-derived sleep stages (Wake, Light, Deep, REM) were compared to PSG-scored labels annotated by two independent registered PSG technicians (with adjudication to resolve discrepancies). Several performance metrics were computed:

The PSG substudy was designed to assess:

a) Epoch-by-epoch accuracy (overall):

$$\text{Accuracy}=\frac{\text{Number of matching epochs}}{\text{Total epochs}} \quad (2)$$

b) Cohen's κ: Agreement beyond chance, interpreted following Landis–Koch criteria.

c) Per-stage performance:

- Sensitivity
- Specificity
- Positive predictive value (PPV)
- F1-score

These metrics quantify detection accuracy for each specific sleep stage.

3.5.3. Validation of continuous sleep summary metrics

To evaluate agreement in continuous sleep metrics derived from wearable vs PSG, the following indices were assessed:

- Total Sleep Time (TST)
- Sleep Efficiency (SE)
- Wake After Sleep Onset (WASO)
- Sleep Fragmentation Index

For each metric, agreement was evaluated using:

- Intraclass correlation coefficient (ICC)

A two-way mixed-effects model (ICC(3,1)) was used: $\text{ICC}=\frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}}+\sigma^2_{\text{error}}}$ (3)

- Bland–Altman analysis

The mean bias and 95% limits of agreement were computed as: $\text{Bias}=D$, $\text{LoA}=D\pm1.96\times SD_D$ (4)

where D is the difference between wearable and PSG metrics.

- Pearson correlation coefficients

Used to describe linear association but not as a measure of agreement alone.

3.5.4. Quality control and artifact handling

Wearable signals were screened using automated quality-control algorithms that detect:

- PPG motion artifacts
- Low perfusion segments
- Accelerometer saturation
- Missing or corrupted data packets

Epochs failing QC were either repaired using interpolation for short segments or excluded from scoring if multiple consecutive epochs were affected.

PSG scoring adhered to the AASM Manual v2.6 criteria. Nights with technician disagreement >15% or signal loss in major PSG channels were excluded.

3.5.5. Interpretation of validation results

- The validity of the wearable system's sleep-stage classifier

- The consistency of derived sleep metrics across nights
- The bias trends relevant for older adults with fragmented sleep

These results establish the performance bounds and help contextualize the intervention outcomes in the main study.

3.6. Statistical Analysis and Sample Size

An a priori power analysis was performed to determine the sample size required to detect within-subject changes in the Pittsburgh Sleep Quality Index (PSQI) across the two-period crossover. For paired (within-subject) comparisons the required sample size can be approximated by:

$$n=\frac{(Z_{a/2}+Z_{\beta})^2}{d^2} \quad (5)$$

Assuming a two-sided $\alpha = 0.05$, ($Z_{a/2}=1.96$), power=0.80 ($Z_{\beta}=0.842$) and a conservative within-subject effect size of $d = 0.50$, the minimum required sample size was 32 participants. To account for attrition and ensure feasibility of the PSG substudy and model personalization, 40 participants were prospectively enrolled.

Primary and secondary continuous outcomes were analyzed using a linear mixed-effects model (LME) appropriate for crossover trials. Each model included Condition (IC vs. CC), Period (first vs. second), and Sequence (IC→CC vs. CC→IC) as fixed effects, with a random intercept for each participant to account for repeated measures.

The general model specification was:

$$PSQI_{ij}\sim Condition_{ij}+Period_{ij}+Sequence_i+(1\mid Participant_i) \quad (6)$$

This structure follows CONSORT recommendations for crossover trial analysis by explicitly modeling period and sequence effects while estimating the within-subject treatment effect.

Model assumptions—including normality of residuals, homoscedasticity, and independence—were evaluated using standardized diagnostic plots. When needed, outcomes with skewed distributions (e.g., WASO) were log-transformed.

Missing data were addressed using multiple imputation by chained equations (MICE):

- Number of imputations: m = 20
- Predictor matrix included all outcome variables, demographic covariates, Condition, Period, and Sequence
- Convergence was assessed via diagnostic trace plots

Analyses followed an intention-to-treat (ITT) framework, with a complementary per-protocol (PP) analysis performed to ensure robustness.

Secondary outcomes were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure. Adjusted q-values were reported in addition to raw p-values.

4. Experiments and Results

4.1. Experimental Setup and Evaluation Protocol

A controlled experimental platform was established to evaluate the proposed personalized sleep monitoring and intervention system in a real-world home environment. The data acquisition platform consisted of a multimodal wearable sensing module and an embedded processing unit. The sensing module integrated inertial measurement sensors (3-axis accelerometer) and physiological signal acquisition units (PPG, EDA), and the system firmware version was fixed throughout all trials to ensure reproducibility and eliminate software-version confounding.

A total of 40 community-dwelling older adults completed the randomized two-period crossover protocol (mean age = 67.5 ± 5.1 years; 12 females). Each participant underwent two 4-week experimental periods: an Intervention Condition (IC) with fully activated adaptive functions and a Control Condition (CC) with sensing-only functionality, separated by a 7-day washout interval. The order of conditions (IC→CC or CC→IC) was randomized.

Physiological and environmental data were continuously collected during nightly sleep in habitual home settings over a total monitoring duration of approximately 2,240 participant-nights. Raw sensor streams were transmitted wirelessly to the embedded processing module and securely stored for offline algorithmic and statistical analysis.

Baseline sleep status indicated moderate disturbance across the cohort, with a mean PSQI of 9.0 ± 2.3. Baseline wearable-derived total sleep time averaged 372 ± 41 min and sleep efficiency averaged 78.2 ± 6.1%, consistent with mild sleep fragmentation typically observed in older adults.Nights with >30% signal loss or confirmed non-wear time were excluded from objective analyses.

Statistical analyses followed the pre-specified plan (paired tests and linear mixed-effects models). Primary inference used LME with participant as a random intercept and fixed effects for Condition, Period and Sequence; secondary outcomes were tested with FDR correction for multiple comparisons. Effect sizes (Cohen's d) and 95% confidence intervals are reported alongside p-values. Statistical significance was set at $p < 0.05$.

4.2. System-Level Performance Evaluation

System-level performance was evaluated in terms of real-time processing capability, operational stability, and energy efficiency under continuous overnight operation.

The average end-to-end latency from raw signal acquisition to sleep-stage inference was 132 ± 28 ms, demonstrating the system's capacity to support real-time closed-loop monitoring and intervention. The processing pipeline maintained stable throughput over extended recordings, with an observed frame loss rate below 0.5% across all monitored nights.

Communication stability was quantified through wireless packet loss and transmission failure monitoring. The system achieved a mean packet loss rate of 0.7 ± 0.4% under typical home-network conditions, and no critical system failures or safety events were recorded during the trial.

Energy performance testing showed that the wearable module supported a mean continuous operating time of 28.3 ± 3.1 hours per full charge, enabling full-night monitoring without user recharging interruptions.

4.3. Algorithm and Model Evaluation

The personalized sleep-staging model was evaluated against a non-personalized baseline classifier using identical input data streams.

The proposed model achieved a mean sleep-stage classification accuracy of 81.6 ± 4.3%, compared to 73.9 ± 5.1% for the baseline model ($\Delta = +7.7\%$, $p < 0.001$). Temporal consistency of stage transitions, measured by Cohen's κ, was significantly higher in the personalized model ($\kappa = 0.71 \pm 0.06$) than in the baseline ($\kappa = 0.60 \pm 0.07$, $p < 0.001$).

Ablation analysis demonstrated that disabling the personalization module resulted in a mean performance degradation of 6.4% in classification accuracy, confirming the functional contribution of adaptive learning components.

Computational efficiency analysis showed that the model maintained a mean inference time of 4.8 ± 1.2 ms per epoch, supporting reliable deployment on embedded hardware under real-time constraints.

4.4. Human-Level Sleep Outcome Validation

Human-level outcome evaluation focused on clinically relevant sleep quality indicators derived from both self-reported and wearable-observed measures.

Across 40 participants, subjective sleep quality improved significantly under the IC. The LME estimate for Condition (IC vs CC) was $\beta = -2.20$ (SE = 0.28), yielding a mean within-subject reduction in PSQI of −2.2 points (95% CI: −2.7 to −1.7); $t(39) \approx -7.85$, $p < 0.001$. The associated within-subject Cohen's d was 0.76 (medium–large). No significant sequence or period×condition (carryover) interactions were observed ($p > 0.40$).

Wearable-derived objective metrics (participant period means) also improved during IC (FDR-corrected p), as shown in Table 1.

TABLE I. COMPARISON OF KEY SLEEP QUALITY INDICATORS BETWEEN CONTROL AND INTERVENTION CONDITIONS

Metric	Control	Intervention	Δ	p-value
TST (min)	374 ± 41	392 ± 38	+18 min	0.004
Sleep Efficiency (%)	78.4 ± 6.1	83.1 ± 5.5	+4.6%	< 0.001
WASO (min)	52.0 ± 16.8	43.5 ± 15.2	−8.5 min	0.008
Fragmentation Index	18.6 ± 4.1	16.0 ± 3.8	−2.6	0.012

These results indicate higher sleep efficiency and reduced sleep fragmentation across most participants during the intervention phase.

5. ANALYSIS AND DISCUSSION

5.1. System-Level Interpretation of Key Findings

This study demonstrates that a closed-loop, design-driven wearable health system integrating multimodal sensing and adaptive intervention can translate system-level performance into clinically meaningful sleep improvements in older adults. The personalized intervention produced an average PSQI reduction of 2.2 points, exceeding the established threshold for clinical relevance. Objective wearable-derived metrics, including sleep efficiency, total sleep time, wake after sleep onset (WASO), and sleep fragmentation index, showed consistent and directionally aligned improvements with moderate effect sizes.

Importantly, the polysomnography (PSG) validation substudy revealed moderate to strong correlations between wearable-derived and PSG-derived sleep parameters ($r = 0.68–0.79$), confirming that the sensing and signal processing pipeline preserved physiologically meaningful information. These findings empirically validate the core system design hypothesis: that low-latency sensing, real-time inference, and adaptive feedback can operate as an effective cyber-physical intervention loop rather than a passive monitoring tool.

5.2. Mechanistic Explanation From a Systems Engineering Perspective

From a systems engineering viewpoint, the observed therapeutic effects can be attributed to the closed-loop architecture of the proposed platform. The system integrates three tightly coupled layers: (1) real-time multimodal data acquisition, (2) adaptive inference and decision-making, and (3) context-aware intervention delivery. This architecture enables rapid detection of micro-disturbances in sleep behavior and supports near-real-time micro-adjustments in behavioral guidance.

Reinforcement learning (RL)-informed personalization module functions as a sequential decision engine that dynamically optimizes both intervention timing and modality based on historical sleep trajectories and behavioral responses. Unlike static rule-based systems, this adaptive control mechanism continuously updates its policy through feedback, allowing the system to track inter- and intra-individual variability, which is especially pronounced in aging populations.

In addition, multimodal sensor fusion (accelerometry combined with photoplethysmography) increases the observability of latent sleep states compared with single-sensor approaches, thereby improving the controllability of the behavioral intervention loop. Collectively, the results indicate that architectural coherence between sensing, inference, and actuation layers is a key determinant of real-world intervention efficacy.

5.3. Positioning Relative to Prior Digital Sleep Systems

Compared with prior work in digital sleep interventions, this study advances the field in several engineering-relevant dimensions. Most existing consumer-grade systems operate in an open-loop paradigm, providing post hoc summaries rather than real-time, adaptive control. The present system implements a fully integrated closed-loop pipeline, enabling continuous personalization and temporal optimization of interventions.

Furthermore, while prior validation studies have highlighted the limited accuracy of consumer wearables for sleep staging, our findings demonstrate that a properly calibrated multimodal sensing pipeline can achieve acceptable agreement with PSG in ecologically valid home environments. The observed bias in total sleep time (approximately 9 minutes) and sleep efficiency (approximately 2%) falls within the range reported by recent multisite validation studies, supporting the feasibility of engineering-grade wearable sleep monitoring.

The use of a randomized crossover experimental design further improves internal validity compared with conventional parallel-group pilot trials, strengthening the causal interpretability of system-level effects.

5.4. Engineering Strengths of the Proposed System

This work exhibits several strengths from an engineering and system design standpoint.

First, the platform represents a tightly integrated cyber-physical system in which hardware sensing, embedded signal processing, cloud-based inference, and human-facing feedback interfaces were co-designed rather than evaluated in isolation.

Second, the RL-based personalization mechanism introduces an adaptive control layer that is rarely implemented in current commercial digital health systems, enabling continuous optimization of intervention policies.

Third, the system was evaluated in real-world, uncontrolled home environments, demonstrating robustness to noise, missing data, and behavioral variability, which are critical considerations for scalable deployment.

Finally, the inclusion of a PSG validation substudy provides an external physiological reference, strengthening confidence in the reliability of the sensing and processing pipeline.

5.5. Limitations and Engineering Trade-offs

Several limitations and engineering trade-offs should be acknowledged. Although the sample size was sufficient to detect moderate system-level effects, the study cohort was geographically and demographically constrained, which may affect generalizability. Wearable-based measurements inherently trade fine-grained physiological resolution for scalability, comfort, and ecological validity, particularly in comparison with full polysomnography.

While the crossover design reduces inter-individual variance, residual carryover effects cannot be entirely excluded despite statistical testing. In addition, the current RL reward structure and policy update rules were intentionally simplified to ensure system stability and safety, which may limit long-term optimality in behavioral adaptation. Future system iterations should explore more sophisticated reward shaping and safety-constrained learning strategies.

5.6. Implications for Future Cyber-Physical Health Systems

The findings of this work have broader implications for the design of intelligent health cyber-physical systems. Integrating low-latency sensing, adaptive intelligence, and human-centered interface design into a unified closed-loop architecture can transform wearable systems from passive trackers into active therapeutic agents.

Future development should focus on extending the sensing context to include environmental variables such as ambient light, sound, and device interaction patterns, as well as on implementing explainable adaptive policies to improve clinical transparency and user trust. Longitudinal, multi-month deployments will be essential to evaluate stability, safety, and sustained behavioral change.

At a system level, this study provides a scalable architectural blueprint for next-generation gerontechnology platforms, demonstrating how sensing hardware, adaptive algorithms, and human-centered feedback can be coherently integrated into clinically meaningful engineering systems.

6. Conclusion

This study presents an integrated smart wearable system that combines multimodal physiological sensing, adaptive personalization, and age-oriented human–computer interaction to improve sleep outcomes in older adults. Through a randomized crossover experimental design and a targeted polysomnography (PSG) validation substudy, the system demonstrated significant improvements in both subjective and objective sleep indicators, including PSQI, sleep efficiency, wake after sleep onset, and sleep fragmentation. The observed agreement between wearable-derived and PSG-derived metrics further confirms the physiological reliability of the sensing and processing pipeline in naturalistic home environments.

Beyond clinical performance, this work contributes an engineering-level framework for designing closed-loop digital health systems for aging populations. The system architecture integrates real-time signal acquisition, low-latency inference, and context-aware micro-interventions, enabling continuous adaptation to individual sleep patterns. The reinforcement-learning–inspired personalization engine provides a scalable mechanism for sequential decision optimization, moving wearable systems from passive monitoring toward active, personalized digital therapeutics.

Several limitations should be acknowledged. The current study involved a moderate sample size and a relatively short intervention window, and wearable-based measurements inherently trade high-resolution physiology for comfort and scalability. Although the crossover design mitigates inter-individual variability, larger and more diverse cohorts and longer deployment periods will be necessary to validate long-term robustness and generalizability. In addition, the current adaptive policy design prioritizes system stability and user safety, which may constrain long-term optimization performance.

Future work will focus on expanding the sensory context (e.g., ambient light, sound, and behavioral patterns), refining the reward structure of the adaptive engine, and incorporating clinician-in-the-loop and explainable AI mechanisms. These extensions will be essential for supporting clinical translation, regulatory acceptance, and large-scale deployment.

In summary, this study provides empirical and system-level evidence that a design-driven, adaptive wearable platform can function as a feasible, acceptable, and effective tool for improving sleep quality in older adults. The proposed architecture offers a scalable blueprint for next-generation gerontechnology systems aimed at promoting healthy aging and reducing the societal burden of sleep-related disorders.

References

[1] Canever, J. B., Zurman, G., Vogel, F., Sutil, D. V., Diz, J. B. M., Danielewicz, A. L., ... & de Avelar, N. C. P. (2024). Worldwide prevalence of sleep problems in community-dwelling older adults: A systematic review and meta-analysis. Sleep Medicine, 119, 118-134. https://doi.org/10.1016/j.sleep.2024.03.040

[2] Gordon, N. P., Yao, J. H., Brickner, L. A., & Lo, J. C. (2022). Prevalence of sleep-related problems and risks in a community-dwelling older adult population: a cross-sectional survey-based study. BMC public health, 22(1), 2045. https://doi.org/10.1186/s12889-022-14443-8

[3] Ancoli-Israel, S. (2005). Sleep and aging: prevalence of disturbed sleep and treatment considerations in older adults. Journal of Clinical Psychiatry, 66, 24.

[4] Li, J., & Gooneratne, N. S. (2019). Sleep and health in older adults. In Sleep and health (pp. 31-43). Academic Press. Https://doi.org/10.1016/B978-0-12-815373-4.00004-6

[5] Yaprak, B., Arslan, N., & Alataş, H. (2023). Factors Affect the Quality of Sleep in Elderly People with Metabolic Syndrome. Eurasian Journal of Emergency Medicine. https://doi.org/10.4274/eajem.galenos.2023.41961

[6] Bubu, O. M., Brannick, M., Mortimer, J., Umasabor-Bubu, O., Sebastiao, Y. V., Wen, Y., ... & Anderson, W. M. (2017). Sleep, cognitive impairment, and Alzheimer's disease: a systematic review and meta-analysis. Sleep, 40(1), zsw032. https://doi.org/10.1093/sleep/zsw032

[7] Foley, D., Ancoli-Israel, S., Britz, P., & Walsh, J. (2004). Sleep disturbances and chronic disease in older adults: results of the 2003 National Sleep Foundation Sleep in America Survey. Journal of psychosomatic research, 56(5), 497-502. https://doi.org/10.1016/j.jpsychores.2004.02.010

[8] Lucchesi, L. M., Pradella-Hallinan, M., Lucchesi, M., & Moraes, W. A. D. S. (2005). Sleep in psychiatric disorders. Brazilian Journal of Psychiatry, 27, 27-32. https://doi.org/10.1590/S1516-44462005000500006

[9] Bloom, H. G., Ahmed, I., Alessi, C. A., Ancoli-Israel, S., Buysse, D. J., Kryger, M. H., ... & Zee, P. C. (2009). Evidence-based recommendations for the assessment and management of sleep disorders in older persons. Journal of the American Geriatrics Society, 57(5), 761-789. https://doi.org/10.1111/j.1532-5415.2009.02220.x

[10] Sateia, M. J., Buysse, D. J., Krystal, A. D., Neubauer, D. N., & Heald, J. L. (2017). Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. Journal of clinical sleep medicine, 13(2), 307-349. https://doi.org/10.5664/jcsm.6470

[11] Lee, T., Cho, Y., Cha, K. S., Jung, J., Cho, J., Kim, H., ... & Kim, J. W. (2023). Accuracy of 11 wearable, nearable, and airable consumer sleep trackers: prospective multicenter validation study. JMIR mHealth and uHealth, 11(1), e50983. https://doi.org/10.2196/50983

[12] Bate, G. L., Kirk, C., Rehman, R. Z., Guan, Y., Yarnall, A. J., Del Din, S., & Lawson, R. A. (2023). The role of wearable sensors to monitor physical activity and sleep patterns in older adult inpatients: A structured review. Sensors, 23(10), 4881. https://doi.org/10.3390/s23104881

[13] Yoon, H., & Choi, S. H. (2023). Technologies for sleep monitoring at home: Wearables and nearables. Biomedical Engineering Letters, 13(3), 313-327. https://doi.org/10.1007/s13534-023-00305-8

[14] Birrer, V., Elgendi, M., Lambercy, O., Menon, C., et al. (2024). Evaluating reliability in wearable devices for sleep staging. npj Digital Medicine, 7, 74. https://doi.org/10.1038/s41746-024-01016-9

[15] Muñoz Esquivel, K., Gillespie, J., Kelly, D., Condell, J., Davies, R., McHugh, C., ... & Nordström, A. (2023). Factors influencing continued wearable device use in older adult populations: quantitative study. JMIR aging, 6, e36807. https://doi.org/10.2196/36807

[16] Grotto, G., Martinello, M., & Buja, A. (2024). Use of mHealth Technologies to Increase Sleep Quality among Older Adults: A Scoping Review. Clocks & Sleep, 6(3), 517–532. https://doi.org/10.3390/clockssleep6030034

[17] Satapathy, S. K., Patel, T., Modi, M., Davra, H., Brahma, B., Tripathy, S. K., & Bhoi, A. K. (2025). Deep Learning-Based Sleep Stage Classification: A Systematic Review. Journal of The Institution of Engineers (India): Series B, 1-23. https://doi.org/10.1007/s40031-025-01275-4

[18] Zhang, A. H., He-Mo, A., Yin, R. F., Li, C., Tang, Y., Gurve, D., Montazeri Ghah-javerestan, N., Goubran, M., & Lim, A. S. P. (2024). Mamba-based deep learning approaches for sleep staging on a wireless multimodal wearable system without electroencephalography. arXiv. https://doi.org/10.48550/arXiv.2412.15947

[19] Altman, M., Huang, T. T., & Breland, J. Y. (2018). Design thinking in health care. Preventing chronic disease, 15, E117. https://doi.org/10.5888/pcd15.180128

[20] Li, J., Szanton, S. L., McPhillips, M. V., Lukkahatai, N., Pien, G. W., Chen, K., ... & Gooneratne, N. S. (2021). An mHealth-facilitated personalized intervention for physical activity and sleep in community-dwelling older adults. Journal of Aging and Physical Activity, 30(2), 261-270. https://doi.org/10.1123/japa.2020-0463

[21] Chinnasamy, P., Wong, W. K., Raja, A. A., Khalaf, O. I., Kiran, A., & Babu, J. C. (2023). Health recommendation system using deep learning-based collaborative filtering. Heliyon, 9(12). https://doi.org/10.1016/j.heliyon.2023.e22844

[22] Hassan, B. M., & Elagamy, S. M. (2025). Personalized medical recommendation system with machine learning. Neural Computing and Applications, 37(9), 6431-6447. https://doi.org/10.1007/s00521-024-10916-6

[23] Hsu, T. C. C., Whelan, P., Gandrup, J., Armitage, C. J., Cordingley, L., & McBeth, J. (2025). Personalized interventions for behaviour change: A scoping review of just-in-time adaptive interventions. British Journal of Health Psychology, 30(1), e12766. https://doi.org/10.1111/bjhp.12766

[24] Jayaraman, P., Desman, J., Sabounchi, M., Nadkarni, G. N., & Sakhuja, A. (2024). A Primer on Reinforcement Learning in Medicine for Clinicians. npj Digital Medicine, 7, 337. https://doi.org/10.1038/s41746-024-01316-0

Acknowledgments

None.

Funding

None.

Availability of Data

Not applicable.

Ethical Statement

All participants provided written informed consent prior to participation. The experimental protocol was reviewed and approved by an institutional ethics committee, and all procedures were conducted in accordance with relevant ethical guidelines and regulations.

Author Contributions

Samuel Mekonnen and Bilal Jemal Geda conducted the system design and development, implemented the wearable sensing and intervention modules, carried out the randomized crossover trial, and analyzed the experimental data, while Tsegay Teklay Gebrelibanos conceived and supervised the study, guided the design-driven personalization framework, interpreted the findings, and finalized the manuscript.

Competing Interests

The authors declare no competing interests.