University of Kent

Kent Academic Repository

Liu, Zhao, Soria, Daniele, Ang, Chee Siang and Shergill, Sukhi (2026) *Self-supervised transfer learning for IMU-based upper-limb action detection and motion quality analysis in an immersive VR functional task.* Journal of Sensor and Actuator Networks, 15 (3).

Downloaded from
https://kar.kent.ac.uk/115613/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/doi:10.3390/jsan15030046

This document version
Publisher pdf

DOI for this version

Licence for this version
CC BY (Attribution)

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Journal of
Sensor and Actuator Networks

MDPI

Article

Self-Supervised Transfer Learning for IMU-Based Upper-Limb Action Detection and Motion Quality Analysis in an Immersive VR Functional Task

Zhao Liu [1,*,†], Daniele Soria [1,†], Chee Siang Ang [1,2,3,†] and Sukhi Shergill [2,4,*,†]

[1] School of Computing, University of Kent, Canterbury CT2 7NZ, UK; d.soria@kent.ac.uk (D.S.); c.s.ang@kent.ac.uk (C.S.A.)
[2] Kent and Medway Medical School, Canterbury CT2 7FS, UK
[3] National Heart & Lung Institute, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK
[4] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London WC2R 2LS, UK
* Correspondence: zl239@kent.ac.uk (Z.L.); sukhi.shergill@kcl.ac.uk (S.S.)
† These authors contributed equally to this work.

Abstract

Wearable inertial sensing has considerable potential for process-level analysis of upper-limb function, but further evidence is needed to understand how it can be applied within ecologically structured immersive virtual reality (VR) tasks. Most VR-based functional assessments rely primarily on outcome-level indicators, such as task completion time, success rate, or error count, which may not fully capture how a task is executed. This exploratory study investigated whether wearable IMU signals collected during an immersive VR sushi-making task could support binary detection of a core upper-limb manipulation phase and provide additional information about task execution beyond global performance outcomes. A total of 45 participants contributed usable motion recordings for this study, with five Xsens DOT sensors placed on the hands, forearms, and waist. Three signal modalities were analysed, including acceleration (ACC), gyroscope angular velocity (GYR), and Euler angles. The downstream recognition problem was formulated as a binary classification task (*Placing* vs. *Non-Placing*), and a self-supervised learning (SSL) pretrain–fine-tune strategy was evaluated against conventional machine learning and from-scratch deep learning baselines using five subject-wise validation splits. The strongest overall performance was achieved with hand-mounted accelerometer signals, with LeftHand–ACC achieving a Macro-F1 of 0.712 ± 0.128 and RightHand–ACC achieving 0.679 ± 0.118. Under both hand-ACC settings, SSL fine-tuning showed higher mean Macro-F1 than the Balanced Random Forest baseline and the same deep architecture trained from scratch. Recognition performance varied substantially across sensor locations, signal modalities, and task segments, with distal upper-limb sensors generally outperforming waist-based configurations. Cross-age analyses further showed that within-cohort and cross-cohort performance did not fully align, indicating sensitivity to age-related distribution shift. Beyond classification, Log Dimensionless Jerk (LDLJ) derived from the *Placing* action showed a significant positive association with Cognitron motor control time cost ($r = 0.636$, $p < 0.001$). These findings suggest that wearable IMU sensing can provide preliminary process-level information during immersive VR functional tasks, including task-phase detection, sensing-configuration comparison, cross-cohort generalisation assessment, and exploratory motion-quality analysis. The results should be interpreted as evidence of feasibility rather than as a mature biomechanical or clinical assessment model.

Check for updates

Academic Editor: Jordi Mongay Batalla

Received: 23 April 2026
Revised: 8 June 2026
Accepted: 9 June 2026
Published: 12 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: wearable inertial sensors; IMU; virtual reality; upper-limb movement; binary action detection; transfer learning; self-supervised learning; human activity recognition; motion smoothness; Log Dimensionless Jerk

1. Introduction

Wearable sensing technologies, particularly inertial measurement units (IMUs), have become an important approach for capturing human motion in health monitoring, rehabilitation, and activity recognition research [1,2]. Compared with camera-based systems, IMU-based sensing is relatively low-cost, portable, and suitable for repeated data collection in naturalistic or semi-structured task settings. These sensors can record acceleration, angular velocity, and orientation-related signals from multiple body locations, thereby providing multi-dimensional motion signals relevant to movement execution across the upper limbs, trunk, and other body segments [2]. When combined with machine learning methods, wearable motion signals can support automated analysis of behavioural patterns and functional performance.

However, a large proportion of widely used publicly available Human Action Recognition (HAR) datasets focus on relatively coarse locomotion activities such as walking, sitting, standing, or climbing stairs [3–5]. In comparison, evidence for IMU-based recognition of subtle upper-limb manipulation phases remains much more limited [6]. This gap is particularly relevant in rehabilitation and health-related assessment contexts, where clinically meaningful performance often depends on subtle manipulation-based actions such as grasping, placing, bilateral coordination, and object handling [7]. Such task-relevant upper-limb movements are often characterised by small amplitudes, lower signal-to-noise ratios, and greater inter-subject variability than coarse whole-body activities, which can make robust recognition more challenging [8,9]. In addition, labelled datasets for these actions are often small because data collection and annotation are time-consuming and resource-intensive, especially in health-related and ageing populations [9,10].

Virtual reality (VR), particularly immersive VR, offers a useful and increasingly adopted environment for studying upper-limb performance and rehabilitation because it enables standardised and repeatable task delivery while preserving a degree of ecological relevance through immersive and interactive task simulation. Within VR, participants can perform structured, function-led activities that resemble real-world daily tasks, thereby supporting more controlled investigation of upper-limb behaviour under comparable task demands [11]. This is particularly valuable for functional assessment and rehabilitation research, where interest lies not only in whether a task is completed successfully, but also in how the task is executed. Recent studies have used VR-based grasping, reaching, object transfer, manual dexterity, and other functional tasks to assess upper-limb ability, task performance, and rehabilitation effects [12]. For example, Everard et al. contrasted the classic Box and Block Test with immersive VR, hand-tracking, and mixed-reality versions for post-stroke manual dexterity assessment; Juan et al. compared hand interaction and controller interaction in immersive VR upper-limb rehabilitation [13]; Ham et al. evaluated a mixed-reality system with individual finger-movement tracking for post-stroke hand and finger rehabilitation [14]; and Chen et al. reported the effectiveness, safety, and patient perceptions of an immersive VR-based exercise system for post-stroke upper-limb motor rehabilitation [15]. More broadly, recent review evidence also suggests that immersive VR is becoming an increasingly relevant technological approach for upper-extremity rehabilitation in neurological disorders [16]. Nevertheless, wearable sensor-based analysis of upper-limb manipulation phases in immersive VR settings has received comparatively lim-

ited attention [17]. However, in many existing studies, evaluation is still frequently centred on outcome-level indicators, such as task completion time, success rate, error count, clinical scale mapping, usability, or global digital health metrics [18–20]. Although these measures are useful for characterising overall performance differences, they do not fully reveal how the task is executed. Kanzler et al., for instance, showed that both VR and task complexity substantially influenced digital health metrics and their measurement properties in upper-limb assessment, with intra-participant variability increasing markedly as task complexity increased [21]. This suggests that participants with comparable overall outcomes may still differ in movement organisation, segment transitions, local control strategies, and coordination patterns. Likewise, studies comparing different extended-reality task implementations indicate that the interaction mode and task medium themselves can shape the observed functional results [12,13]. Therefore, summary behavioural outcomes alone may be insufficient to capture process-level differences in functional task execution.

From this perspective, incorporating wearable motion analysis into immersive functional tasks may provide important additional value. Compared with focusing only on final task outcomes, process-level motion analysis can provide a more detailed account of how upper-limb action phases are organised during task execution and can offer complementary evidence for understanding the relationship between movement characteristics and broader functional performance [22]. Existing work has shown that movement smoothness and related kinematic quality measures can reflect upper-limb motor control and are meaningfully related to activity-level function after stroke. In addition, recent work has examined the measurement properties of smoothness metrics in upper-limb reaching, supporting their potential use as interpretable process-level indicators [23]. Motivated by this perspective, the present study employs an immersive, multi-step, daily-activity-like VR task to examine the feasibility of IMU-based upper-limb action detection and process-level motion analysis. It further explores the potential value of process-level motion features for cognitively relevant functional task assessment.

From a methodological perspective, several challenges need to be addressed in this context. First, deep learning models trained from scratch often perform poorly when only limited labelled data are available. Recent work has shown that transfer learning, especially self-supervised pre-training followed by task-specific fine-tuning, can substantially improve wearable activity recognition under low-label conditions [24]. However, evidence for its application to subtle upper-limb manipulation tasks remains comparatively limited [10]. Second, practical performance and deployment feasibility are influenced by sensor placement and signal modality, such as acceleration (ACC), gyroscope (GYR), and orientation-related signals [25]. While hand- and wrist-based acceleration signals are often informative, their relative usefulness for small-amplitude manipulation actions in immersive functional tasks remains insufficiently established [26]. Third, age-related dataset shift is a major practical concern. Target populations for rehabilitation and cognitive screening are often older adults, whereas data collection is frequently easier in younger participants. Previous studies suggest that models trained only on younger cohorts can generalise less well to older adults, whereas mixed-age training may improve transferability in some settings [27,28]. Whether this pattern also holds in immersive VR task settings requires further evaluation.

Beyond classification performance, motion-derived features may also have interpretive value for functional and cognitive assessment. In particular, Log Dimensionless Jerk (LDLJ) has been widely used as a kinematic smoothness metric for quantifying movement intermittency and execution quality [29–31]. LDLJ and related jerk-based measures have been applied in upper-limb assessment and IMU-based movement analysis [30,32,33]. Prior work has also suggested that movement smoothness may be sensitive to cognitive-motor

decline in conditions such as mild cognitive impairment and Alzheimer's disease [34,35]. In this study, we therefore consider motion analysis not only as a classification problem, but also as a potential source of behaviourally meaningful indicators that may enrich functional assessment.

Against this background, the present study examined wearable IMU-based motion analysis as a process-level extension of immersive VR functional assessment. We used windowed IMU signals from the continuous VR sushi-making workflow to examine whether a core manipulation phase could be detected from upper-limb motion patterns. Specifically, we formulated a binary *Placing* versus *Non-Placing* recognition task as an initial and pragmatic modelling step, where *Placing* represented the repeated transport-and-release actions embedded throughout the task and *Non-Placing* represented the remaining manipulation phases. This binary formulation was used to test whether wearable sensing could recover meaningful task-phase information under small-sample, class-imbalanced, and continuous VR task conditions.

This study had three aims. First, we tested whether self-supervised transfer learning could improve binary upper-limb action detection from wearable IMU signals compared with from-scratch deep learning and random-forest baselines. Second, we examined recognition performance across sensor locations, signal modalities, task segments, and age-cohort settings. Third, we explored whether LDLJ, a movement-smoothness metric derived from the *Placing* action, was associated with Cognitron motor-control performance. Together, these analyses examined whether wearable IMU sensing could provide process-level motion information during an immersive VR functional task.

This manuscript builds on our previous vKitchen preprint, which focused mainly on the VR task framework and behavioural validation across VR, 2D tablet, and real-world formats. The present study is distinct because it focuses on wearable IMU-based process-level motion analysis, including binary upper-limb action detection, SSL transfer learning, sensor-location and modality comparison, age-cohort generalisation, and LDLJ-based movement-smoothness analysis.

2. Methods

Kitchen tasks represent instrumental activities of daily living (IADLs) that require coordinated, goal-directed upper-limb actions and are therefore commonly used in rehabilitation and functional assessment contexts [36–38]. Compared to isolated laboratory movements, kitchen-based tasks provide a more functionally meaningful context in which reaching, grasping, transporting, pressing, and object manipulation are embedded within an everyday goal [38,39]. Among the many possible kitchen tasks, sushi making was selected as the target task because, relative to a full real-kitchen workflow, it is comparatively safe, highly standardisable, and can be naturally decomposed into repeatable manipulation steps. This makes it suitable for studying upper-limb manipulation patterns while retaining the ecological logic of a real cooking activity.

2.1. VR Task and Action Definitions

VR Task: *vKitchen* [40] is a fully immersive VR sushi-making simulation developed in Unity3D [41] and deployed on the Oculus Quest 2 headset (Meta Platforms Technologies, LLC, Menlo Park, CA, USA). Participants completed the task while wearing a head-mounted display (HMD) and using handheld controllers. The task was designed as a multi-step functional activity involving sequential object manipulation, visuomotor control, bilateral coordination, and action transitions (Figure 1a). The full workflow comprised an initial *Remember* phase and a *Selection* phase, followed by the sequential task steps *Placing*

Bamboo, *Placing Seaweed*, *Placing Rice*, *Flatten Rice*, *Placing Ingredients*, *Rolling*, *Cutting*, and *Plating* (Figure 1c).

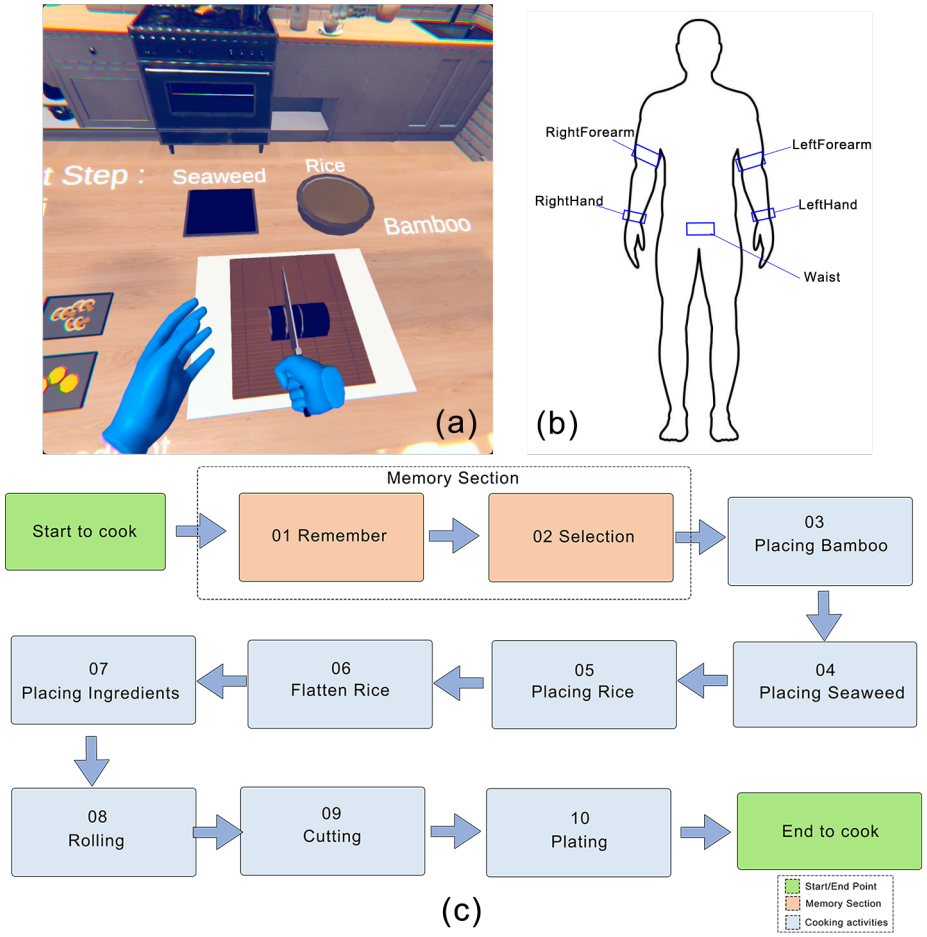


Figure 1. Overview of the *vKitchen* VR sushi-making task and sensor setup. (**a**) Example first-person view in the VR task. (**b**) Xsens DOT sensor placement on the body (hands, forearms, waist). (**c**) Task workflow and key manipulation steps.

To reduce extraneous cognitive load, the VR interaction scheme was simplified to two main functions. During the ingredient recall phase, participants used the "A" button to select ingredients from a visual panel. During the cooking phase, object manipulation was controlled by the side "Trigger" button. Participants extended their virtual hand toward an object, pressed the Trigger to grasp it, and released the button to place it. This simplified dual-button mapping was intended to keep the interaction demands transparent and predictable, thereby preserving attention for the task itself rather than for interface learning [42,43].

For the VR motion-analysis condition, each participant completed one formal VR sushi-making trial after standardised task instructions and a brief VR tutorial. The tutorial was used only to familiarise participants with the virtual environment and controller-based object interaction, did not involve the formal sushi-making task, and was not included in the motion-analysis dataset.

2.2. Cognitron Assessment

To provide an external cognitive reference for interpreting task-related behavioural and motion-derived measures, participants also completed the Cognitron cognitive assessment battery after the three cooking tasks. Cognitron is an online cognitive testing platform

derived from the Great British Intelligence Test and has been used in research involving participants with a range of cognitive and motor abilities [44].

In the present study, the Cognitron battery comprised 10 subtests, each targeting a specific cognitive domain relevant to functional cognition. These included reaction speed, attention, emotional discrimination, motor coordination, mental rotation, visual and verbal memory, visual search, and cognitive flexibility. A summary of the subtests is shown in Table 1.

Table 1. Cognitron subtests and associated cognitive domains.

No.	Test Name	Cognitive Domain
1	SRT	Reaction speed and attention
2	ED	Emotional discrimination
3	MC	Reaction time and motor coordination
4	M2Ds	Mental rotation and pattern recognition
5	PMOi	Immediate visual memory and attention
6	PMOd	Delayed visual memory and attention
7	PWMi	Immediate verbal memory and recognition
8	PWMd	Delayed verbal memory and recognition
9	TD	Visual search, attention, and reaction speed
10	SS	Attention and cognitive flexibility

Note: Additional information on the Cognitron platform is available from the official website.

All participants completed the Cognitron battery after finishing the VR task. Completion of the Cognitron assessment required approximately 25–30 min. The Cognitron measures were used as external reference variables to explore whether behavioural and motion-derived indicators from *vKitchen* showed interpretable associations with standardised cognitive domains.

Action definitions: For motion analysis, action classes were defined according to shared movement mechanics rather than treating each task step as a separate semantic label. Steps involving similar pick-and-place kinematics were grouped into a single class, *Placing*, comprising *Placing Bamboo*, *Placing Seaweed*, *Placing Rice*, *Placing Ingredients*, and *Plating*. These segments all involve a similar distal movement pattern: grasping an object, transporting it to a target location, and releasing it. Because this common kinematic structure is likely to dominate the inertial signal, separating these steps reliably using IMU data alone would be difficult and would also lead to very small per-class sample sizes.

By contrast, *Flatten Rice*, *Rolling*, and *Cutting* involve distinct interaction mechanics and were therefore retained as separate classes. *Flatten Rice* involves repeated downward tapping or pressing actions; *Rolling* requires a coordinated bimanual forward movement; and *Cutting* consists of repeated slicing actions with a more regular oscillatory pattern. Accordingly, four action classes were defined for motion analysis: *Placing*, *Flatten Rice*, *Rolling*, and *Cutting*.

Label source: Action labels were generated by synchronising the IMU recordings with the VR task timeline. Specifically, the task start time was aligned with the sensor recording start time, after which action start and end timestamps were extracted from the VR system logs. These log-derived labels were then verified against first-person (egocentric) video recordings of each VR session and corrected when needed. This combination of system logs and video inspection was used to improve the temporal fidelity of labels in the continuous task setting, where action boundaries may not always be perfectly separable.

Task formulation: In the present study, the downstream recognition problem was first formulated as a binary classification task: *Placing* versus *Non-Placing*, where *Non-Placing* consisted of *Flatten Rice*, *Rolling*, and *Cutting*. This design choice was made for both practical and methodological reasons. Within the sushi-making workflow, *Placing* is

the most frequent, relatively homogeneous, and mechanically stable manipulation phase, whereas the remaining actions are more heterogeneous and less frequent. Under limited sample size and label availability, a coarse-to-fine modelling strategy provides a more stable starting point for benchmarking recognition performance [8,9,24]. The binary setup was therefore adopted to establish a stable and reproducible baseline for the present study. We formulated the task as a binary classification problem because *Placing* was the dominant and most consistently repeated action in the task, while the other action types contained fewer windows and were combined as *Non-Placing*. This binary setting provided a more stable first-step evaluation of whether IMU signals could distinguish the main manipulation phase from the remaining task phases.

2.3. Participants and Data Acquisition

For the age-cohort analyses, participants aged below 60 years were grouped as younger adults, whereas those aged 60 years or above were grouped as older adults, following the commonly used 60-year threshold for older populations [45]. Of these, six were excluded from the primary motion-analysis dataset because of substantial sensor dropout or incomplete motion recordings, leaving 45 participants (young = 22, older = 23) with complete usable motion data for the main classification analyses. For body-location-specific analyses, participants with missing data at a given location were excluded from that location-specific comparison only, rather than being removed entirely from all analyses. For the LDLJ–Cognitron association analysis, two further participants were excluded because the relevant Cognitron data were unavailable, resulting in an effective sample of 43 participants for that analysis.

Five Xsens DOT sensors [46] were used to record movement signals from the left hand, right hand, left forearm, right forearm, and waist (Figure 1b). The sensors were selected because of their compact wearable form, wireless data transmission, and strap-based fixation, which made them suitable for VR-based movement recording. Because participants operated VR controllers with both hands, the hand sensors were fixed on the back of the wrist area to minimise interference with controller use while remaining close to distal manipulation-related movements. The forearm sensors were fixed on the left and right forearms, approximately aligned with the corresponding hand sensors, to capture forearm movement and rotation. The waist sensor was fixed at the posterior midline of the lower back of the waist region as a proximal reference for trunk-level movement during the seated task. These five sensors were intended to capture a broad representation of upper-limb and trunk movement during the VR task while minimising interference with controller-based interaction.

From the raw sensor streams, three signal modalities were extracted: tri-axial ACC, GYR, and Euler angles. The original sampling rate was 60 Hz. All preprocessing, model training, evaluation, and statistical analyses were implemented in Python (version 3.8).

This study was approved by the Central Research Ethics Advisory Group at the University of Kent (reference number CREAG51-03-23).

2.4. Pre-Processing Pipeline and Windowing

Before model training, the IMU recordings were screened for data completeness and basic signal usability using custom scripts. Sensor streams were first checked for sufficient data length and obvious recording failure, such as prolonged unchanged values or values remaining at zero. Time-series plots were then generated for each recording and visually inspected to identify missing segments or abnormal signal patterns. Recordings with missing, substantially incomplete, or clearly unusable sensor data were excluded from the

corresponding sensor-specific analysis. Therefore, the effective sample size varied across sensor configurations.

The continuous raw IMU recordings were transformed into clean fixed-length windows for downstream modelling.

- **Cleaning:** Action names were first normalised across the log files and motion records to ensure consistent label assignment. Participants with missing recordings for a specific body location were excluded from analyses involving that location. Missing-location recordings were identified primarily through all-zero channels or absent sensor files.
- **Resampling:** Signals were downsampled from 60 Hz to 30 Hz to match the input specification of the SSL foundation model and its official implementation, thereby ensuring consistency between the present downstream task and the original pre-training framework [24]. This step also reduced computational load while preserving the temporal structure needed for upper-limb action analysis.
- **Windowing and label purity**: After timestamp-level action labels had been assigned to the continuous IMU time series, a sliding-window scheme was applied using a window length of 10 s and a stride of 1 s [2,10]. Each window corresponded to 300 time points after resampling to 30 Hz. Because windows near action boundaries could contain samples from more than one action, the label of each window was assigned according to the dominant action within that window. Label purity was defined as the proportion of samples in the window belonging to the dominant action. Windows were retained only when dominant-label purity was at least 0.60. This threshold was used as a lower inclusion boundary rather than as an indication that most retained windows had low label purity. Because the VR task was performed as a continuous sequence of actions, some sliding windows naturally contained transitions between adjacent task phases. This criterion allowed transition-overlapping windows to be retained while excluding windows with highly ambiguous labels, thereby preserving sufficient labelled data for exploratory binary task-phase detection.
- **Normalisation:** Each retained window was z-normalised per channel along the time axis during dataset construction. This step reduced the effect of absolute scale differences across participants, sensors, and signal modalities, and made ACC, GYR, and Euler inputs more comparable across folds and model settings.
- **Binary label mapping:** After the purity-based filtering step, retained windows were mapped to the binary recognition task. Windows dominated by *Placing* were labelled as *Placing*, whereas windows dominated by *Flatten Rice*, *Rolling*, or *Cutting* were labelled as *Non-Placing*.

2.5. SSL Foundation Model

A wearable self-supervised learning (SSL) foundation model proposed by Yuan et al. [24] was adopted, using the official implementation released by the authors for transfer learning fine-tuning. This model was pre-trained in a multi-task self-supervised manner on large-scale wrist accelerometer data from the UK Biobank (approximately 700,000 person-days of unlabelled data), with the aim of learning general temporal representations that could transfer to downstream HAR tasks under limited-label conditions [24].

The model backbone is a 1D convolutional ResNet-V2 with 18 layers and approximately 10 million parameters. The backbone outputs 1024-dimensional feature vectors. For downstream tasks, a 512-dimensional fully connected layer is attached after the representation layer, followed by a classification head. In the present study, all inertial signals were resampled to 30 Hz and fed to the model in the form of fixed 10 s windows, consistent with the official implementation.

2.6. Baseline Models

To evaluate the benefit of SSL-based transfer learning under low-label conditions, the fine-tuned SSL model was compared against conventional supervised baselines.

Balanced Random Forest (BRF): A Balanced Random Forest classifier was used as a classical machine learning baseline because it is specifically designed for imbalanced classification settings by rebalancing the data at the tree level [47]. BRF was trained on the same window-level samples as the deep models. For each retained window, the multichannel time-series input was flattened into a one-dimensional feature vector before classification. The BRF model used 600 trees, no manually restricted maximum tree depth, a minimum leaf size of 3, and the square root of the number of features as the maximum feature set considered at each split. Balanced sampling was applied using the default "auto" sampling strategy, with replacement disabled. A fixed random state of 42 was used. A participant-level validation split was used to select the classification threshold before final testing. The BRF baseline provided a non-deep reference for evaluating whether SSL transfer yielded gains beyond an imbalance-aware ensemble method.

Deep model trained from scratch: To isolate the effect of pre-training from that of the model architecture itself, the same deep backbone and downstream classification head used in the SSL setting were also trained with random parameter initialisation and without loading pre-trained weights. This *from-scratch* setting used the same data splits, input representation, and downstream optimisation strategy as the SSL fine-tuning experiments. Comparison between this model and the fine-tuned SSL model therefore allowed the contribution of transfer learning to be evaluated more directly [24].

2.7. Evaluation Protocol

Model performance was evaluated using five subject-wise group-based evaluation splits, with participant ID used as the grouping variable. Within each split, all windows from the same participant were assigned entirely to the training, validation, or test subset, preventing participant-level leakage across subsets. A group-based validation split was constructed within the training data and used for early stopping and model selection. This subject-wise design is important in wearable human activity recognition because random window-level splitting can inflate performance when windows from the same participant appear in multiple subsets.

Macro-F1 was used as the primary performance metric because it gives equal weight to both classes and is therefore less dominated by the more frequent *Placing* class. Macro-precision, macro-recall, accuracy, and Cohen's κ were also reported as supplementary metrics.

To examine sensor design choices, a systematic comparison was conducted across sensor locations and signal modalities. Specifically, single-input settings were evaluated for five body locations (left hand, right hand, left forearm, right forearm, and waist) and three signal modalities (ACC, GYR, and Euler angles), training and testing each location–modality pair independently. This design allowed for the identification of the configurations that were most informative for the binary upper-limb action-detection task.

To assess generalisation across age cohorts, within-cohort and cross-cohort performance were further compared. Within-cohort analyses were conducted separately for the younger and older groups using the same subject-wise evaluation logic. Cross-cohort experiments examined Young $\rightarrow$ Older and Older $\rightarrow$ Young transfer, in which a model trained on all subjects in one age group was tested on all subjects in the other.

To evaluate the contribution of transfer learning, the SSL fine-tuning model was compared directly against the Balanced Random Forest baseline and the deep model

trained from scratch. All models were evaluated using the same subject-wise split logic to ensure fair comparison across methods.

2.8. Model Training Settings

For SSL fine-tuning and from-scratch deep learning training, the input size was set to 300 time points, corresponding to 10 s windows resampled at 30 Hz. For ACC, GYR, and Euler inputs, the same three-channel network input format was used, with the same window length and resampling rate. No modality-specific adaptation layer was introduced for GYR or Euler inputs. Models were trained using the Adam optimiser with AMSGrad enabled, a learning rate of 1×10^{-4}, a batch size of 256, and a maximum of 80 epochs. Early stopping was applied with a patience of 10 epochs. For the binary classification task, cross-entropy loss was used. Because the retained windows were class-imbalanced, a weighted cross-entropy loss was applied, with class weights calculated from the inverse class frequency in the training set. Weighted sampling was not used. Random seeds were fixed at 42 for NumPy (version 1.26.4; NumPy Developers, Austin, TX, USA) and PyTorch (version 2.10.0; Meta AI, Menlo Park, CA, USA).

For SSL fine-tuning, pretrained weights from the wearable SSL foundation model were loaded, and the full network was fine-tuned rather than frozen. The from-scratch model used the same architecture, input representation, data splits, and optimisation settings, but was trained with random initialisation and without loading pretrained weights. Hyperparameters were not tuned on the test data; the main settings, including window length, stride, learning rate, batch size, maximum epoch number, and early-stopping patience, were fixed before the final evaluation based on the pretrained SSL framework and preliminary feasibility runs.

2.9. Task-Segment Analysis

To examine whether recognition performance varied across different stages of the sushi-making workflow, an additional task-segment analysis was conducted. The rationale was that the full task includes action phases with different levels of movement consistency and complexity, which may affect classification difficulty.

Based on the task structure, three segment settings were defined: *Up to Action 1*, *Up to Action 2*, and *Full*. The *Up to Action 1* setting included the earliest action sequence used in the classification analysis, namely *Placing* and *Flatten Rice*. The *Up to Action 2* setting extended this by including *Rolling*. The *Full* setting included the complete action sequence used in the binary recognition task: *Placing*, *Flatten Rice*, *Rolling*, and *Cutting*. For reference, the corresponding mean action-only durations were approximately 60.05 s for *Up to Action 1*, 145.41 s for *Up to Action 2*, and 159.08 s for *Full*.

All task-segment analyses used the same preprocessing pipeline, model configurations, and subject-wise evaluation strategy as the full-task analysis. The purpose of this additional analysis was to determine whether earlier, shorter, and more functionally coherent task phases produced cleaner or more discriminative motion patterns than the full multi-stage task.

2.10. LDLJ-Based Motion Smoothness Analysis

In addition to action classification, movement smoothness was explored as a process-level feature using LDLJ. As mentioned before, LDLJ is a kinematic smoothness metric derived from acceleration signals and is widely used to quantify movement intermittency or irregularity in upper-limb motion [29–31]. For the present analysis, LDLJ was computed from tri-axial acceleration data recorded from the left-hand sensor during the *Placing* action segment. This action was selected because it was frequent, mechanically simple, and represented the core transport-and-release movement pattern in the task.

Acceleration signals along the x, y, and z axes were first filtered using a fourth-order zero-phase Butterworth low-pass filter with a cutoff frequency of 6 Hz to suppress high-frequency noise while preserving the main movement-related components [30]. Jerk was then calculated as the time derivative of the filtered acceleration signal along each axis. The squared jerk magnitude was obtained by summing the squared jerk values across the three axes and integrating the result over the movement duration using trapezoidal numerical integration.

Let $\mathbf{a}(t) = [a_x(t), a_y(t), a_z(t)]^\top$ denote the tri-axial acceleration signal after filtering, and let $\mathbf{j}(t)$ denote the corresponding jerk signal:

$$\mathbf{j}(t) = \frac{d\mathbf{a}(t)}{dt}. \tag{1}$$

The integrated

squared jerk was then computed as follows:

$$J = \int_0^T \|\mathbf{j}(t)\|^2 \, dt, \tag{2}$$

where T is the movement duration. Following the implementation used in this study, a time-normalised dimensionless jerk score was obtained as follows:

$$\mathrm{DJ} = \sqrt{J\,T^5}, \tag{3}$$

and the final LDLJ score was calculated as follows:

$$\mathrm{LDLJ} = \ln(\mathrm{DJ} + 10^{-6}). \tag{4}$$

Here, ln denotes the natural logarithm. Higher LDLJ values indicate less smooth and more irregular movement execution [29,30].

LDLJ was used here as an exploratory motion-derived indicator of movement quality. Previous studies have suggested that jerk-related characteristics and fine motor smoothness may be altered in cognitive ageing, mild cognitive impairment (MCI), and Alzheimer's disease (AD) [34,35]. To examine its cognitive-motor relevance in the present dataset, participant-level LDLJ scores were correlated with Cognitron motor control performance (v_motorControl). Before correlation analysis, outliers were removed using the interquartile range (IQR) rule. Pearson's correlation was then applied to the cleaned data to assess the association between movement smoothness and cognitive-motor performance.

3. Results

3.1. Dataset Summary and Label Purity After Preprocessing

A total of 51 participants completed the VR condition. After excluding unusable or incomplete IMU recordings caused by sensor dropout or location-specific signal-quality issues, up to 45 participants contributed usable IMU data to the motion-analysis dataset. Because data availability varied by sensor location, the main hand-ACC analyses included 44 participants. Participant characteristics and VR task-duration statistics are summarised in Table 2.

After preprocessing, each main hand-ACC dataset contained 8097 retained windows from 44 participants, including 5547 *Placing* windows and 2550 *Non-Placing* windows. This corresponded to a moderate class imbalance of approximately 2.18:1. Although the minimum label-purity threshold was 0.60, retained windows were generally of high purity,

with a mean purity of 0.946, a median of 1.000, and 79.3% of windows in the 0.90–1.00 purity range (Table 3).

Table 2. Participant characteristics and VR task duration summary by age group.

Measure	Younger Group	Older Group	Total
Participants, n	25	26	51
Age, mean ± SD	26.64 ± 5.10	71.35 ± 5.30	49.43 ± 23.15
Age range	18–43	60–80	18–80
Sex	11 F, 14 M	18 F, 8 M	29 F, 22 M
Total duration, mean ± SD	205.67 ± 90.88	372.91 ± 147.27	290.93 ± 148.11
Total duration range	77.33–448.17	172.38–735.87	77.33–735.87
Action duration, mean ± SD	136.72 ± 52.94	266.13 ± 121.61	202.70 ± 114.06
Action duration range	60.99–313.54	117.53–602.83	60.99–602.83

Note: Age is reported in years, and task duration is reported in seconds. F = female; M = male. Younger adults were defined as participants aged below 60 years, and older adults were defined as participants aged 60 years or above. Total duration includes the *Remember*, *Selection*, and action-execution phases; action duration excludes the *Remember* and *Selection* phases.

Table 3. Window-level class distribution and label-purity summary for the main hand-ACC datasets.

Measure	LeftHand–ACC	RightHand–ACC
Participants	44	44
Total retained windows	8097	8097
Placing windows	5547 (68.5%)	5547 (68.5%)
Non-Placing windows	2550 (31.5%)	2550 (31.5%)
Mean label purity	0.946	0.946
Median label purity	1.000	1.000
Purity 0.60–0.70	892 (11.0%)	892 (11.0%)
Purity 0.70–0.80	20 (0.2%)	20 (0.2%)
Purity 0.80–0.90	766 (9.5%)	766 (9.5%)
Purity 0.90–1.00	6419 (79.3%)	6419 (79.3%)

3.2. Overall Performance Against Baselines

Table 4 compares SSL fine-tuning with the BRF baseline and the same deep architecture trained from scratch using hand acceleration signals. Across both hand-based ACC configurations, SSL fine-tuning achieved the highest Macro-F1. For LeftHand–ACC, SSL fine-tuning reached 0.712 ± 0.128, compared with 0.599 ± 0.036 for BRF and 0.476 ± 0.147 for training from scratch. For RightHand–ACC, the corresponding values were 0.679 ± 0.118, 0.588 ± 0.045, and 0.536 ± 0.180, respectively.

SSL fine-tuning showed a higher mean Macro-F1 than the two baselines across both hand-ACC configurations. These descriptive results suggest that SSL-based transfer learning may provide a useful initialisation strategy in the present low-label, class-imbalanced upper-limb action-detection setting. However, because formal paired significance tests were not conducted, these differences should be interpreted as exploratory rather than confirmatory evidence of model superiority.

Table 4. Descriptive baseline comparison for the main hand-ACC configurations (Macro-F1, mean ± SD).

Method	LeftHand–ACC	RightHand–ACC
Balanced Random Forest (BRF)	0.599 ± 0.036	0.588 ± 0.045
Trained From Scratch	0.476 ± 0.147	0.536 ± 0.180
SSL Fine-Tune (All Layers)	**0.712 ± 0.128**	**0.679 ± 0.118**

Note: Values are reported descriptively as mean ± SD. Bold values indicate the highest Macro-F1 within each sensor configuration. Formal paired significance tests were not conducted; therefore, differences between models should be interpreted as exploratory performance comparisons.

Per-class precision, recall, F1 scores, and confusion-matrix summaries for the main hand-ACC models are provided in Appendix Tables A2 and A3.

3.3. Effects of Sensor Location and Modality

Figure 2 summarises recognition performance across 15 single-input settings obtained from five body locations and three signal modalities. The full numerical results are presented in Table 5.

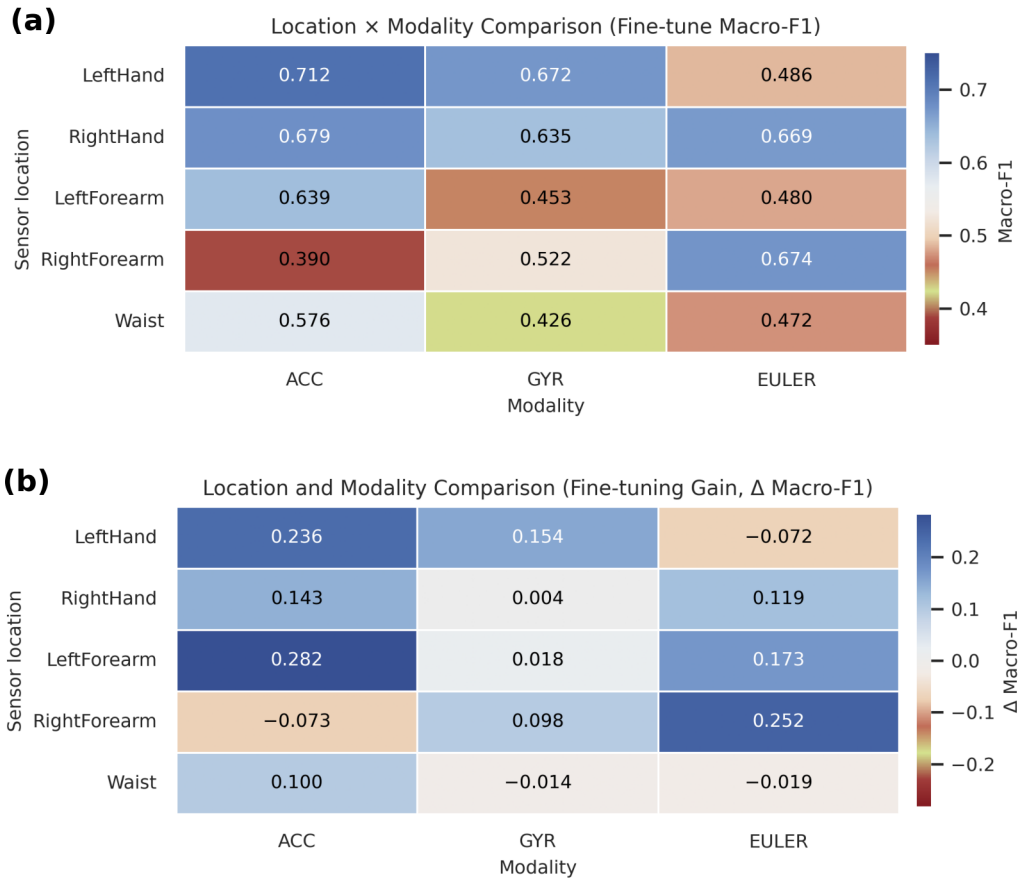


Figure 2. Location and modality comparison. (**a**) Fine-tune Macro-F1 across single-input settings. (**b**) Fine-tuning gain (Δ Macro-F1 = fine-tune − scratch) across the same settings.

Table 5. Location and modality comparison results.

Location–Modality	Scratch	Fine-Tune	ΔF1
RightHand–ACC	0.536 ± 0.180	0.679 ± 0.118	+0.143
RightHand-GYR	$\mathbf{0.631 \pm 0.097}$	0.635 ± 0.128	+0.004
RightHand–Euler	0.550 ± 0.186	0.669 ± 0.114	+0.119
Waist–ACC	0.476 ± 0.137	0.576 ± 0.119	+0.100
Waist–GYR	0.440 ± 0.131	0.426 ± 0.136	−0.014
Waist–Euler	0.491 ± 0.160	0.472 ± 0.148	−0.019
LeftHand-ACC	0.476 ± 0.147	$\mathbf{0.712 \pm 0.128}$	+0.236
LeftHand–GYR	0.518 ± 0.171	0.672 ± 0.123	+0.154
LeftHand–Euler	0.558 ± 0.160	0.486 ± 0.144	−0.072
LeftForearm–ACC	0.357 ± 0.131	0.639 ± 0.136	+0.282
LeftForearm–GYR	0.435 ± 0.138	0.453 ± 0.238	+0.018
LeftForearm–Euler	0.307 ± 0.121	0.480 ± 0.136	+0.173
RightForearm–ACC	0.463 ± 0.130	0.390 ± 0.111	−0.073
RightForearm–GYR	0.424 ± 0.146	0.522 ± 0.200	+0.098
RightForearm–Euler	0.422 ± 0.162	0.674 ± 0.123	+0.252

Note: Results are subject-wise 5-fold GroupKFold Macro-F1 (mean ± SD). ΔF1 denotes the descriptive mean difference between SSL fine-tuning and training from scratch (Fine-tune – Scratch). The best result in each metric column is highlighted in bold.

Overall, recognition performance varied substantially by both sensor location and signal modality. Hand- and forearm-based inputs generally outperformed waist-based inputs, with LeftHand–ACC showing the strongest overall performance after fine-tuning. Accelerometer-based hand configurations were among the most stable and informative settings, whereas waist-based GYR and Euler configurations were generally weaker.

Fine-tuning gains were also configuration-dependent. Several configurations showed higher Macro-F1 after SSL fine-tuning, particularly LeftForearm–ACC, RightForearm–Euler, and LeftHand–ACC. However, not all inputs benefited from fine-tuning, and some waist- or orientation-based configurations showed little improvement or a decline. These results indicate that transfer learning performance depended strongly on whether the selected sensor location and modality captured task-relevant movement information.

3.4. Within-Age Versus Cross-Age Results

To examine age-related distribution shift, we compared within-age subject-wise validation with cross-age train and test evaluation using the two best-performing hand acceleration configurations. The full results are reported in Table 6.

Within-age validation showed different performance profiles between the younger and older cohorts. LeftHand–ACC performed more consistently across age settings, whereas RightHand–ACC showed greater variability between cohorts. In the cross-age setting, performance differed from the corresponding within-age results in both transfer directions, indicating that models trained in one age group did not generalise identically to the other.

Across the two sensor settings, LeftHand–ACC remained above 0.58 in all within-age and cross-age evaluations, while RightHand–ACC showed a wider range of values. These results suggest that age-group composition influenced classification performance and that within-cohort validation may not fully reflect cross-cohort generalisation.

Table 6. Within-age and cross-age generalisation.

Setting	LeftHand–ACC	RightHand–ACC
Young only (CV)	0.687 ± 0.153	0.507 ± 0.200
Older only (CV)	0.642 ± 0.109	0.627 ± 0.114
Young $\rightarrow$ Older	0.6389	0.5999
Older $\rightarrow$ Young	0.5895	0.5547

Note: CV = subject-wise 5-fold GroupKFold within the same cohort (Macro-F1, mean±SD). Cross-age settings train on all subjects in the source cohort and test on all subjects in the target cohort (single run, no SD).

3.5. Task-Segment Analysis Results

To examine whether recognition performance varied across different stages of the sushi-making workflow, we further compared SSL fine-tuning performance across three segment settings: *Up to Action 1*, *Up to Action 2*, and *Full*. Here, *Up to Action 1* included *Placing* and *Flatten Rice*; *Up to Action 2* included *Placing*, *Flatten Rice*, and *Rolling*; and *Full* included *Placing*, *Flatten Rice*, *Rolling*, and *Cutting*. The corresponding mean action-only durations were approximately 60.05 s, 145.41 s, and 159.08 s, respectively. The full results are shown in Table 7.

Task-segment restriction influenced recognition performance. For LeftHand–ACC, the earliest segment produced the highest Macro-F1 and Cohen's κ, suggesting that the shorter and more homogeneous task segment was easier to classify. For RightHand–ACC, Macro-F1 was highest in the full-task setting, while κ was highest in the earliest segment.

These results indicate that recognition behaviour changed as the analysed sequence became longer and included a broader range of movement types. They also suggest that segment-level analysis can reveal performance patterns that may be obscured when the entire task is treated as a single undifferentiated recognition stream.

Table 7. Fine-tuning performance across task segments for the two best sensor configurations.

Task Segment	Sensor	Macro-F1	Cohen's κ
Up to Action 1 (60.05 s)	LeftHand–ACC	0.767	0.552
	RightHand–ACC	0.595	0.307
Up to Action 2 (145.41 s)	LeftHand–ACC	0.708	0.464
	RightHand–ACC	0.616	0.280
Full (159.08 s)	LeftHand–ACC	0.712	0.446
	RightHand–ACC	0.679	0.225

Note: Up to Action 1 includes *picking and Placing* and *Flatten*; *Up to Action 2* includes *picking and Placing*, *Flatten*, and *rolling*; *Full* includes *picking and Placing*, *Flatten*, *rolling*, and *cutting*. Durations are mean action-only durations, excluding memory and instruction phases.

3.6. LDLJ and Cognitive-Motor Association

In addition to action classification, we explored whether the motion-derived smoothness indicator Log LDLJ was associated with cognitive-motor performance. Figure 3 displays the relationship between LeftHand LDLJ during the *Placing* action and the Cognitron motor control score after outlier removal.

As summarised in Table 8, in this exploratory bivariate analysis, LDLJ showed a positive correlation with Cognitron motor control performance ($r = 0.636$, $p < 0.001$, $N = 43$). The corresponding coefficient of determination was $R^2 = 0.405$. Since higher LDLJ values indicate less smooth and more irregular movement execution, participants with higher LDLJ tended to show higher motor-control time-cost values in the Cognitron task.

This exploratory result provides a quantitative description of the bivariate association between motion smoothness in the VR task and motor-control performance measured outside the task framework.

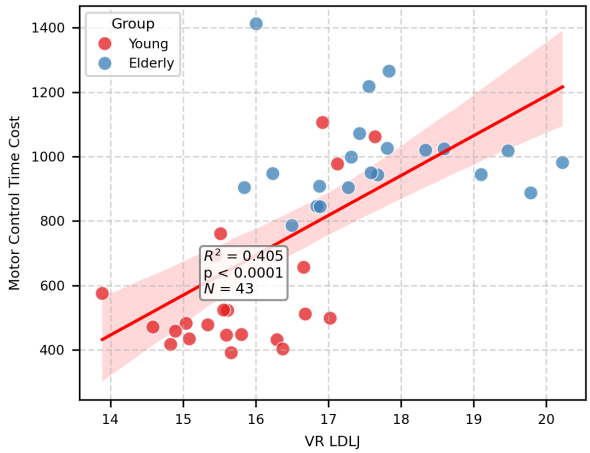


Figure 3. Association between LeftHand LDLJ during the *Placing* action and Cognitron motor control score after outlier removal. Higher LDLJ values indicate less smooth movement execution.

Table 8. Summary statistics for the Cognitron motor control outcome and its correlation with LeftHand LDLJ.

Measure	N	Mean±SD	Range	Correlation (*r*)	*p*-Value
v_motorControl	43	782.13 ± 278.28	390.87–1412.33	0.636	<0.001 ***

Note: The Cognitron motor control outcome was used as an external cognitive-motor benchmark in the LDLJ association analysis. Higher values indicate longer completion time and poorer motor control performance. *** indicates $p < 0.001$.

4. Discussion

This study showed that wearable IMU signals can provide process-level information during an immersive VR functional task, but also highlighted several constraints on how such signals should be interpreted. SSL fine-tuning showed higher mean Macro-F1 than the baseline approaches in the main hand-ACC settings, yet these differences should be interpreted descriptively because formal paired significance tests were not conducted. Performance also varied substantially by sensor location, signal modality, task segment, and age-cohort setting. These findings suggest that the value of transfer learning in VR-based motion analysis depends not only on the model architecture, but also on whether the sensing configuration and analysed task segment capture movement information that is relevant to the target action.

In addition to classification performance, the LDLJ analysis suggested that movement smoothness during the *Placing* action was associated with an external cognitive-motor benchmark. This indicates that IMU-derived movement features can provide additional objective data for interpreting the behavioural process of VR task execution, rather than relying only on task-completion outcomes. As a small-sample exploratory study, the present work provides an initial step toward examining whether immersive VR tasks combined with wearable sensing could be developed into a clinically meaningful assessment model in future research.

4.1. SSL Transfer Under Low-Label Binary Upper-Limb Action Detection

Under the hand-ACC setting, end-to-end fine-tuning of the wearable SSL foundation model showed higher mean Macro-F1 than both the traditional BRF baseline and the same deep architecture trained from scratch (Table 4). This pattern is consistent with previous evidence showing that large-scale self-supervised pre-training on wearable inertial data can learn transferable temporal representations that improve downstream HAR performance, particularly when labelled data are limited [24]. In the present study, the descriptive performance pattern suggests that pre-trained temporal features may have remained useful even when the downstream task involved binary detection of upper-limb action phases embedded in a complex, ecologically structured VR task.

This result is important because the downstream task in *vKitchen* differs substantially from the types of activities that dominate many public wearable HAR datasets. Most benchmark datasets emphasise coarse locomotor activities such as walking, sitting, standing, or stair climbing, where movement cycles are relatively long, repetitive, and well separated in the signal space [48,49]. By contrast, the upper-limb action phases analysed in *vKitchen* are shorter in duration, uneven in frequency, and embedded within continuous goal-directed behaviour. Action boundaries are therefore less distinct, transitions are more common, and movement patterns may overlap across neighbouring segments. The fact that SSL still improved performance under these conditions suggests that pre-training helped the model capture general short-timescale motion patterns and temporal dependencies, which would have been difficult to learn robustly from the downstream task data alone.

At the same time, the absolute recognition levels remained lower than those typically reported for coarse locomotor activity recognition. This is not unexpected. Compared with activities such as walking, sitting, or stair climbing, the action classes in *vKitchen* were shorter in duration, more heterogeneous in execution, and more likely to exhibit partial overlap in their motion patterns [48,49]. The task also involved realistic transitions between action segments rather than artificially isolated movement trials. Taken together, these characteristics make the recognition problem intrinsically more difficult. The present results should therefore be interpreted not against the standards of coarse activity recognition alone, but in relation to the more challenging problem of detecting subtle, functionally embedded upper-limb action phases within continuous task performance.

From an applied perspective, the key point is that SSL transfer showed a favourable descriptive performance pattern within a small, imbalanced, and ecologically motivated dataset. This supports the practical value of transfer learning for wearable motion analysis in function-led VR assessment, where large labelled datasets are unlikely to be routinely available. In clinical or rehabilitation-oriented research, collecting hundreds or thousands of labelled trials from target populations is often unrealistic. A strategy that can make better use of limited data is therefore especially relevant for translation.

4.2. Sensing Configuration: Location and Modality

The comparison across locations and modalities showed that sensing configuration was a major determinant of recognition performance. Across the 15 single-input settings, the strongest overall results were obtained from hand-mounted accelerometer signals, particularly LeftHand–ACC and RightHand–ACC, while several forearm-based configurations also performed well after fine-tuning. In contrast, waist-based configurations were generally weaker, and in some cases did not improve after fine-tuning at all.

These patterns are biomechanically plausible and consistent with the movement characteristics of the *vKitchen* task. The task primarily required participants to manipulate virtual objects using their hands and forearms, including repeated reaching, grasping, transporting, pressing, rolling, and cutting actions. Therefore, wrist- and forearm-mounted

sensors captured richer and more task-relevant motion variations than the waist sensor, which mainly reflected more indirect trunk-level movement. This signal distribution supports the ecological consistency of the task design, as the recorded IMU patterns reflected the expected movement structure of an upper-limb object-manipulation activity. It also reinforces the importance of aligning sensor placement with the movement characteristics of the target VR task [50].

Signal quality and environmental magnetic disturbance should also be considered when interpreting modality-specific results. Magnetic disturbance is particularly relevant to orientation-derived estimates, such as Euler angles, because these signals depend on sensor fusion procedures and may be affected by the stability of magnetometer-informed heading estimates in magnetically complex environments [51–53]. By contrast, the strongest results in the present study were obtained from accelerometer signals, particularly the hand-mounted ACC configurations, which are less directly dependent on magnetometer-derived heading estimates. This may partly explain why ACC provided a more stable and informative input modality than Euler angles in several configurations.

The modality comparison further suggests that ACC was particularly informative in this setting. From a signal perspective, acceleration captures linear movement dynamics that are closely tied to initiation, transport, deceleration, release, and abrupt changes in movement, all of which are relevant to the action phases examined here. This is especially plausible for the *Placing* class, which is largely defined by object transport and release, and for *Flatten Rice*, which involves brief, repeated downward motions. GYR and Euler signals also contained useful information in some configurations, but the pattern of gains was less consistent. This suggests that the relative contribution of different modalities was task-dependent rather than uniform across all body locations [54]. Taken together, these findings indicate that for manipulation-centric VR tasks, hand-mounted ACC offers a strong and comparatively stable sensing choice.

An additional point is that fine-tuning gains were not uniformly positive. Several configurations showed little benefit or even a decline after fine-tuning, including Waist–GYR, Waist–Euler, LeftHand–Euler, and RightForearm–ACC. This is also informative. It suggests that the advantage of transfer learning depends not only on the model itself, but also on whether the selected input channel contains sufficiently stable and relevant information for the target task. If the signal source is weak, noisy, or only indirectly related to the action of interest, fine-tuning may provide little benefit and may even amplify unstable patterns. In that sense, the present ablation results show that transfer learning is not a substitute for thoughtful sensor design. Instead, the two need to be considered together.

4.3. Task-Segment Performance

The task-segment analysis showed that recognition performance was not constant across the workflow. For the strongest configuration, LeftHand–ACC, the highest Macro-F1 and the highest Cohen's κ were both observed in the earliest segment, *Up to Action 1*. This suggests that the first stage of the task contained motion patterns that were more internally consistent and more readily separable than those seen in the longer multi-stage sequences.

One likely explanation is that earlier task phases involved a narrower and more homogeneous set of movements. By contrast, later stages introduced additional movement diversity and more complex transitions between action types, increasing within-class variability and reducing separability. This interpretation is also consistent with the structure of the binary task itself: the *Placing* class is relatively frequent and mechanically consistent, whereas later combinations of actions bring in a broader range of manipulation styles.

As the analysed sequence becomes longer, the model is asked to accommodate more heterogeneous movement patterns within the same recognition framework.

The pattern for RightHand–ACC was less uniform, with the full-task setting yielding the highest Macro-F1 but lower κ values than the earliest segment. This suggests that the effect of task segmentation may depend on both sensor location and performance metric. The divergence between F1 and κ is worth noting. Macro-F1 is driven by class-wise balance in predictive performance, whereas κ additionally reflects agreement beyond chance given the overall prediction structure. The fact that these two metrics did not always move in the same direction suggests that segment restriction influenced not only accuracy balance, but also the structure of class assignments.

More broadly, the segment analysis indicates that restricting analysis to shorter, functionally coherent parts of the task can meaningfully change recognition behaviour. This has methodological implications for future model design, particularly if certain task phases prove more stable or more informative than the full workflow. In ecologically structured tasks, it may not always be optimal to treat the entire task as a single undifferentiated recognition stream. Instead, phase-aware modelling or segment-aware feature extraction may offer a useful direction for future work.

4.4. Within-Age and Cross-Age Performance

The within-age and cross-age analyses showed that age-group composition influenced model performance. Within-age validation indicated different performance profiles between the younger and older cohorts, and cross-age transfer further demonstrated that models did not generalise identically across the two groups. These findings are consistent with broader concerns in wearable HAR that movement distributions may shift across age groups because of differences in execution style, variability, amplitude control, or compensatory movement patterns [27,28].

Because the notation may not be immediately intuitive, it is useful to clarify what the different settings mean in practical terms. In the *young-only* setting, the model was trained and evaluated only within the younger cohort using subject-wise five-fold cross-validation. In the *older-only* setting, the same procedure was applied only within the older cohort. These two conditions therefore quantify how well the model recognises actions when both training and testing are drawn from the same age group. By contrast, *Young* $\rightarrow$ *Older* means that the model was trained on all younger participants and then tested on all older participants, whereas *Older* $\rightarrow$ *Young* means the reverse. These latter two settings are therefore not standard validation settings, but cross-cohort transfer tests designed to evaluate sensitivity to age-related distribution shift.

In the present study, cross-age transfer did not reproduce within-cohort performance exactly in either direction. This suggests that subject-wise cross-validation conducted within a single cohort may not fully represent deployment conditions when the target population differs from the training sample. For rehabilitation or screening applications, this is an important practical issue, because the eventual target users are often older adults, while data collection is frequently easier in younger participants. A model that performs well in a young-only validation setting should not automatically be assumed to generalise to older adults.

At the same time, the results do not imply that age-specific modelling must always be treated as a binary choice between "young-only" and "older-only" training. Rather, they point to the importance of explicitly considering cohort composition during model development. A mixed-age or adaptation-based strategy may provide a more realistic direction for future work, especially where collecting very large older-adult datasets is

difficult. The main implication here is that age-related distribution shift should be treated as a substantive modelling issue rather than as a minor implementation detail.

Another point is that the two hand configurations did not behave identically. LeftHand–ACC showed a comparatively stable pattern across the within-age and cross-age settings, whereas RightHand–ACC exhibited greater fluctuation. This may reflect asymmetric task execution, individual hand dominance, or differences in how the two hands contributed to the continuous manipulation sequence. Although the present data are not sufficient to definitively resolve this issue, the result indicates that sensor-side asymmetry may matter in upper-limb action recognition and should not automatically be treated as negligible.

To provide context for the external benchmark used in the correlation analysis, the Cognitron motor control outcome for the 43 participants included in this analysis had a mean value of 782.13 $\pm$ 278.28, with a range from 390.87 to 1412.33 (see Table 8; participant-level values are provided in Appendix Table A1).

Beyond action classification, the LDLJ analysis showed that movement smoothness in the VR task was significantly associated with this external cognitive-motor benchmark, namely the Cognitron motor control outcome summarised in Table 8. Higher LDLJ values, indicating less smooth and more irregular movement execution, were associated with poorer motor-control performance. This result suggests that motion-derived features in *vKitchen* are not only technically useful for action recognition, but may also capture behaviourally meaningful variation relevant to cognitive-motor function.

This finding is broadly in line with previous work suggesting that jerk-related characteristics and fine motor smoothness may be altered in cognitive ageing, MCI, and AD [34,35]. In the present context, the observed association does not by itself establish diagnostic utility, but it does indicate that process-level movement quality may be meaningfully related to performance measured outside the VR task framework. This supports the broader rationale of this study: wearable motion data can extend the value of an immersive functional task beyond summary classification metrics alone.

It is also notable that this association was observed in an action segment as simple as *Placing*. This suggests that even relatively basic upper-limb manipulation phases may contain measurable information about movement efficiency and smoothness. This is methodologically relevant because simple action segments are easier to isolate, more frequent, and potentially more reproducible than highly compound task stages. If such simple segments already contain meaningful process-level variation, then they may be particularly useful candidates for future feature extraction and cross-task comparison.

More broadly, this result strengthens the case for combining behavioural outcomes with process-level motion features when analysing functional task performance. Summary behavioural indicators capture whether and how successfully a task is completed, whereas motion-derived measures provide additional information about how the task is executed. The present findings suggest that these two levels of analysis can be complementary rather than competing.

4.5. Limitations and Future Work

This study has several limitations. First, the overall dataset size was modest, and some participants were excluded from location-specific analyses because of sensor dropout or incomplete recordings. This may have affected the stability of certain configuration-level estimates. Although SSL fine-tuning achieved higher mean Macro-F1 than the baseline models, formal paired statistical testing of model-performance differences was not conducted; future studies with larger samples and prespecified comparisons could further assess the robustness of these observed differences. Second, handedness was not collected.

Although the task was designed as a bimanual activity with objects arranged around the participant to encourage use of both hands, handedness may still be relevant when interpreting side-specific differences between left- and right-hand sensor configurations. Third, transfer learning was evaluated using a single SSL foundation model. Although the model performed well relative to the baselines used here, benchmarking against alternative pre-training objectives and architectures would strengthen the generalisability of the conclusions. Fourth, the downstream recognition problem was intentionally formulated as a binary task in order to establish a stable baseline under low-label conditions. While this was methodologically appropriate for the current dataset, it does not yet address the full fine-grained action recognition problem. Fifth, action labels were derived from task logs and verified against egocentric video, but the task remained a continuous workflow rather than a set of isolated action clips. Although label-purity analysis showed that most retained windows were dominated by a single action label, windows close to action boundaries could still include transition movements. The 0.60 purity threshold was therefore used as a pragmatic balance between retaining sufficient training data and excluding highly ambiguous windows. Sixth, this study did not include independent validation of IMU signal accuracy or environmental magnetic disturbance. This limitation is particularly relevant for orientation-derived features such as Euler angles.

Future work should therefore proceed in several directions. One priority is to extend the current binary formulation to four-class action recognition using the same task framework. A second is to test whether the present findings generalise across additional VR manipulation tasks beyond sushi making. A third is to examine age-aware modelling strategies, such as mixed-cohort training or explicit domain adaptation, to improve cross-cohort transfer performance. A fourth is to investigate whether segment-aware modelling can further improve action recognition in continuous functional tasks. More broadly, future studies may explore whether combining action classification with process-level indicators such as LDLJ can improve the interpretability and practical value of VR-based functional assessment systems.

5. Conclusions

This study examined whether wearable motion data collected during the immersive *vKitchen* task could extend the assessment value of the system beyond outcome-level behavioural scores. Using multi-location IMU recordings from the hands, forearms, and waist, a self-supervised pretrain–fine-tune strategy was evaluated for binary upper-limb action detection under low-label, class-imbalanced conditions.

Across the main classification analyses, SSL fine-tuning showed higher mean Macro-F1 than the conventional machine learning and from-scratch deep learning baselines in the main hand-ACC settings, with the strongest descriptive performance observed for hand-mounted accelerometer signals. The sensor ablation results further showed that recognition performance depended strongly on sensing configuration, with distal upper-limb locations generally providing more informative signals than the waist. Task-segment analysis demonstrated that recognition performance also varied across stages of the workflow, with earlier and shorter segments yielding the strongest performance for the best sensor configuration. In addition, cross-age evaluation showed that model behaviour was sensitive to cohort composition, indicating that within-cohort validation alone does not fully capture generalisation across age groups.

Beyond classification, the exploratory LDLJ analysis showed a positive bivariate association between motion smoothness in the VR task and Cognitron motor control performance. This suggests that wearable motion signals in *vKitchen* may provide not

only technical input for action recognition, but also process-level information relevant to cognitive-motor performance.

Taken together, the findings indicate that immersive function-led tasks can support meaningful wearable motion analysis at multiple levels: action recognition, sensor-design evaluation, cross-cohort generalisation assessment, and exploratory motion-quality analysis. This extends the value of the *vKitchen* framework beyond final performance scores alone and supports the broader use of process-level motion analysis in VR-based functional assessment.

Author Contributions: Conceptualisation, D.S., S.S. and C.S.A.; methodology, Z.L., D.S., S.S. and C.S.A.; software, Z.L.; formal analysis, Z.L.; investigation, Z.L.; resources, D.S., S.S. and C.S.A.; data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, Z.L., D.S., S.S. and C.S.A.; visualisation, Z.L.; supervision, D.S., S.S. and C.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: The China Scholarship Council at the University of Kent funded this project. No. 202208150024.

Institutional Review Board Statement: This study has been approved by the Central Research Ethics Advisory Group at the University of Kent (reference number CREAG51-03-23).

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

Data Availability Statement: The aggregated results supporting the conclusions of this study are reported in this manuscript and Appendix A. The raw IMU time-series data, video-derived action labels, and participant-level split files are not publicly available because they contain detailed participant-level behavioural recordings and are subject to institutional data-governance and ethics restrictions. De-identified derived datasets and analysis code may be made available from the corresponding author upon reasonable request, subject to ethics approval and data-sharing agreements.

Acknowledgments: This study was funded by the University of Kent-China Scholarship Council.

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

Appendix A

Table A1. Participant-level Cognitron motor control values for the 43 participants included in the LDLJ correlation analysis.

Participant ID	Cognitron Motor-Control Time Cost (ms)
1	390.87
2	431.20
3	760.43
4	1061.33
5	511.07
6	522.37
7	523.50
8	434.00
9	1105.63
10	402.53
11	482.03
12	445.53
13	656.13
14	416.63
15	477.47
16	470.73
17	498.07
18	692.20
19	458.03
20	447.37
21	574.97
25	976.90
30	1412.33
33	903.23
34	943.90
36	887.03
37	981.30
38	1023.73
39	1017.67
41	1019.80
42	785.87
44	1025.47
45	907.77
47	845.53
49	942.90
50	949.27
51	947.07
52	903.57
53	1071.30
55	1217.43
56	998.10
57	1265.10
58	844.37

Note: The Cognitron Motor Control task requires participants to click accurately on the screen in response to a target. The value reported here is the mean reaction time for the Motor Control task, expressed in milliseconds. Higher values indicate longer response time during target-clicking performance.

Table A2. Split-level per-class performance for the main hand-ACC models.

Dataset	Class	Precision	Recall	F1-Score
LeftHand–ACC	Placing	0.821 ± 0.048	0.772 ± 0.119	0.789 ± 0.053
LeftHand–ACC	Non-Placing	0.586 ± 0.096	0.634 ± 0.125	0.594 ± 0.020
LeftHand–ACC	Macro average	0.704 ± 0.032	0.703 ± 0.016	0.691 ± 0.031
RightHand–ACC	Placing	0.809 ± 0.056	0.777 ± 0.103	0.786 ± 0.035
RightHand–ACC	Non-Placing	0.569 ± 0.099	0.598 ± 0.150	0.567 ± 0.053
RightHand–ACC	Macro average	0.689 ± 0.034	0.687 ± 0.036	0.676 ± 0.029

Note: Values are reported as mean±SD across the five subject-wise evaluation splits. Class labels were mapped as 0 = *Placing* and 1 = *Non-Placing*.

Table A3. Split-level confusion-matrix summary for the main hand-ACC models.

Dataset	True Class	Predicted Placing	Predicted Non-Placing
LeftHand–ACC	Placing	900.4 ± 170.3	263.4 ± 136.3
LeftHand–ACC	Non-Placing	205.0 ± 85.9	341.4 ± 45.2
RightHand–ACC	Placing	905.8 ± 145.0	258.0 ± 112.9
RightHand–ACC	Non-Placing	226.8 ± 106.8	319.6 ± 50.9

Note: Values are mean±SD across the five subject-wise evaluation splits. The table summarises confusion matrices calculated separately within each split.

References

1. Gastaldi, L.; Digo, E. Recent Advance and Application of Wearable Inertial Sensors in Motion Analysis. *Sensors* **2025**, *25*, 818.
2. 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.
3. 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; IEEE: New York, NY, USA, 2012; pp. 108–109.
4. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explor. Newsl.* **2011**, *12*, 74–82.
5. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. In Proceedings of the ESANN 2013 Proceedings, Bruges, Belgium, 24–26 April 2013; pp. 3–4.
6. Alam, G.; McChesney, I.; Nicholl, P.; Rafferty, J. Open datasets in human activity recognition research—Issues and challenges: A review. *IEEE Sens. J.* **2023**, *23*, 26952–26980.
7. Turk, R.; Whitall, J.; Meagher, C.; Stokes, M.; Roberts, S.; Woodham, S.; Clatworthy, P.; Burridge, J. Task selection for a sensor-based, wearable, upper limb training device for stroke survivors: A multi-stage approach. *Disabil. Rehabil.* **2023**, *45*, 1480–1487.
8. Yin, Y.; Xie, L.; Jiang, Z.; Xiao, F.; Cao, J.; Lu, S. A systematic review of human activity recognition based on mobile devices: Overview, progress and trends. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 890–929.
9. Kim, G. J.; Parnandi, A.; Eva, S.; Schambra, H. The use of wearable sensors to assess and treat the upper extremity after stroke: A scoping review. *Disabil. Rehabil.* **2022**, *44*, 6119–6138.
10. Demrozi, F.; Pravadelli, G.; Bihorac, A.; Rashidi, P. Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access* **2020**, *8*, 210816–210836.
11. Soleimani, M.; Ghazisaeedi, M.; Heydari, S. The efficacy of virtual reality for upper limb rehabilitation in stroke patients: A systematic review and meta-analysis. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 135.
12. Everard, G.; Burton, Q.; Van de Sype, V.; Bibentyo, T.N.; Auvinet, E.; Edwards, M.G.; Batcho, C.S.; Lejeune, T. Extended reality to assess post-stroke manual dexterity: Contrasts between the classic box and block test, immersive virtual reality with controllers, with hand-tracking, and mixed-reality tests. *J. Neuroeng. Rehabil.* **2024**, *21*, 36.
13. Juan, M.-C.; Elexpuru, J.; Dias, P.; Sousa Santos, B.; Amorim, P. Immersive Virtual Reality for Upper Limb Rehabilitation: Comparing Hand and Controller Interaction. *Virtual Real.* **2023**, *27*, 1157–1171.
14. Ham, Y.; Yang, D.-S.; Choi, Y.; Shin, J.-H. Effectiveness of Mixed Reality-Based Rehabilitation on Hands and Fingers by Individual Finger-Movement Tracking in Patients with Stroke. *J. Neuroeng. Rehabil.* **2024**, *21*, 140.
15. Chen, J.; Or, C.K.; Li, Z.; Yeung, E.H.K.; Zhou, Y.; Hao, T. Effectiveness, Safety and Patients' Perceptions of an Immersive Virtual Reality-Based Exercise System for Poststroke Upper Limb Motor Rehabilitation: A Proof-of-Concept and Feasibility Randomized Controlled Trial. *Digit. Health* **2023**, *9*, 20552076231203599.

16. Ceradini, M.; Losanno, E.; Micera, S.; Bandini, A.; Orlandi, S. Immersive VR for Upper-Extremity Rehabilitation in Patients with Neurological Disorders: A Scoping Review. *J. Neuroeng. Rehabil.* **2024**, *21*, 75.
17. Zhang, X.; Liang, J.; Chen, M.; Xu, R.; Meng, L. A gamified virtual reality and inertial measurement unit-based framework for fine-grained upper limb motor assessment in stroke patients. *J. Intell. Med.* **2025**, *2*, 148–161.
18. Mancuso, V.; Sarcinella, E.D.; Bruni, F.; Arlati, S.; Di Santo, S.G.; Cavallo, M.; Cipresso, P.; Pedroli, E. Systematic review of memory assessment in virtual reality: Evaluating convergent and divergent validity with traditional neuropsychological measures. *Front. Hum. Neurosci.* **2024**, *18*, 1380575.
19. Kirkham, R.; Kooijman, L.; Albertella, L.; Myles, D.; Yücel, M.; Rotaru, K. Immersive virtual reality-based methods for assessing executive functioning: Systematic review. *JMIR Serious Games* **2024**, *12*, e50282.
20. Cubillos, C.; Rienzo, A. Digital Cognitive Assessment Tests for Older Adults: Systematic Literature Review. *JMIR Ment. Health* **2023**, *10*, e47487.
21. Kanzler, C.M.; Armand, T.; Simovic, L.; Sylvester, R.; Domnik, N.; Eilfort, A.M.; Rohner, C.; Gassert, R.; Gonzenbach, R.; Lambercy, O. Influence of Virtual Reality and Task Complexity on Digital Health Metrics Assessing Upper Limb Function. *J. Neuroeng. Rehabil.* **2024**, *21*, 125.
22. Schiefelbein, M.L.; Salazar, A.P.; Marchese, R.R.; Rech, K.D.; Schifino, G.P.; Figueiredo, C.S.; Cimolin, V.; Pagnussat, A.S. Upper-Limb Movement Smoothness after Stroke and Its Relationship with Measures of Body Function/Structure and Activity: A Cross-Sectional Study. *J. Neurol. Sci.* **2019**, *401*, 75–78.
23. Cornec, G.; Lempereur, M.; Mensah-Gourmel, J.; Robertson, J.; Miramand, L.; Medee, B.; Bellaiche, S.; Gross, R.; Gracies, J.-M.; Remy-Neris, O.; et al. Measurement Properties of Movement Smoothness Metrics for Upper Limb Reaching Movements in People with Moderate to Severe Subacute Stroke. *J. Neuroeng. Rehabil.* **2024**, *21*, 90.
24. Yuan, H.; Chan, S.; Creagh, A.P.; Tong, C.; Acquah, A.; Clifton, D.A.; Doherty, A. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *npj Digit. Med.* **2024**, *7*, 91.
25. Davoudi, A.; Mardini, M.T.; Nelson, D.; Albinali, F.; Ranka, S.; Rashidi, P.; Manini, T.M. The effect of sensor placement and number on physical activity recognition and energy expenditure estimation in older adults: Validation study. *JMIR mHealth uHealth* **2021**, *9*, e23681.
26. Paraschiakos, S.; Cachucho, R.; Moed, M.; van Heemst, D.; Mooijaart, S.; Slagboom, E.P.; Knobbe, A.; Beekman, M. Activity recognition using wearable sensors for tracking the elderly. *User Model. User-Adapt. Interact.* **2020**, *30*, 567–605.
27. Fatima, S. Activity recognition in older adults with training data from younger adults: Preliminary results on in vivo smartwatch sensor data. In Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual, 18–22 October 2021; pp. 1–4.
28. Allahbakhshi, H.; Röcke, C.; Weibel, R. Assessing the transferability of physical activity type detection models: Influence of age group is underappreciated. *Front. Physiol.* **2021**, *12*, 738939.
29. Balasubramanian, S.; Melendez-Calderon, A.; Roby-Brami, A.; Burdet, E. On the analysis of movement smoothness. *J. Neuroeng. Rehabil.* **2015**, *12*, 112.
30. Melendez-Calderon, A.; Shirota, C.; Balasubramanian, S. Estimating Movement Smoothness From Inertial Measurement Units. *Front. Bioeng. Biotechnol.* **2021**, *8*, 558771.
31. Mohamed Refai, M.I.; Saes, M.; Scheltinga, B.L.; van Kordelaar, J.; Bussmann, J.B.; Veltink, P.H.; Buurke, J.H.; Meskers, C.G.M.; van Wegen, E.E.H.; Kwakkel, G.; et al. Smoothness metrics for reaching performance after stroke: A systematic review. *J. Neuroeng. Rehabil.* **2021**, *18*, 154.
32. Bayle, N.; Lempereur, M.; Hutin, E.; Motavasseli, D.; Remy-Neris, O.; Gracies, J.-M.; Cornec, G. Comparison of Various Smoothness Metrics for Upper Limb Movements. *Sensors* **2023**, *23*, 1337.
33. Roren, A.; Mazarguil, A.; Vaquero-Ramos, D.; Deloose, J.B.; Vidal, P.P.; Nguyen, C.; Rannou, F.; Wang, D.; Oudre, L.; Lefèvre-Colau, M.M. Assessing Smoothness of Arm Movements with Jerk. *Front. Bioeng. Biotechnol.* **2022**, *9*, 770935.
34. Yan, J.H.; Dick, M.B. Practice effects on motor control in healthy seniors and patients with mild cognitive impairment and Alzheimer's disease. *Aging Neuropsychol. Cogn.* **2006**, *13*, 385–410.
35. Yan, J.H.; Rountree, S.; Massman, P.; Doody, R.S.; Li, H. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *J. Psychiatr. Res.* **2008**, *42*, 1203–1212.
36. Winstein, C.J.; Wolf, S.L.; Dromerick, A.W.; Lane, C.J.; Nelsen, M.A.; Lewthwaite, R.; Cen, S.Y.; Azen, S.P. Effect of a Task-Oriented Rehabilitation Program on Upper Extremity Recovery Following Motor Stroke: The ICARE Randomized Clinical Trial. *JAMA* **2016**, *315*, 571–581.
37. Hendry, K.; Ownsworth, T.; Beadle, E.; Chevignard, M.P.; Fleming, J.; Griffin, J.; Shum, D.H.K. Cognitive Deficits Underlying Error Behavior on a Naturalistic Task after Severe Traumatic Brain Injury. *Front. Behav. Neurosci.* **2016**, *10*, 190.
38. Manera, V.; Petit, P.-D.; Derreumaux, A.; Orvieto, I.; Romagnoli, M.; Lyttle, G.; David, R.; Robert, P.H. 'Kitchen and Cooking', a Serious Game for Mild Cognitive Impairment and Alzheimer's Disease: A Pilot Study. *Front. Aging Neurosci.* **2015**, *7*, 24.

39. Giovannetti, T.; Yamaguchi, T.; Roll, E.; Harada, T.; Rycroft, S.S.; Divers, R.; Hulswit, J.; Tan, C.C.; Matchanova, A.; Ham, L.; et al. The Virtual Kitchen Challenge: Preliminary Data from a Novel Virtual Reality Test of Mild Difficulties in Everyday Functioning. *Aging Neuropsychol. Cogn.* **2019**, *26*, 823–841.
40. Liu, Z.; Soria, D.; Shergill, S.; Ang, C.S. A fully immersive gamified virtual reality assessment of functional cognition. *Res. Sq.* **2025**, *preprint*. https://doi.org/10.21203/rs.3.rs-7251163/v1.
41. Unity Technologies. *Unity 3D*, version 2021.3.24f1 LTS; Unity Technologies: San Francisco, CA, USA, 2021.
42. Kourtesis, P.; Collina, S.; Doumas, L.A.A.; MacPherson, S.E. Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An Immersive Virtual Reality Neuropsychological Battery with Enhanced Ecological Validity. *J. Int. Neuropsychol. Soc.* **2021**, *27*, 181–196.
43. Jang, S.; Choi, S.-W.; Son, S.J.; Oh, J.; Ha, J.; Kim, W.J.; Roh, H.W.; Kim, K.Y.; Lee, S.; Jung, E.; et al. Virtual Reality-Based Monitoring Test for MCI: A Multicenter Feasibility Study. *Front. Psychiatry* **2023**, *13*, 1057513.
44. Hampshire, A.; Trender, W.; Chamberlain, S.R.; Jolly, A.E.; Grant, J.E.; Patrick, F.; Mazibuko, N.; Williams, S.C.R.; Barnby, J.M.; Hellyer, P.; et al. Cognitive Deficits in People Who Have Recovered from COVID-19. *EClinicalMedicine* **2021**, *39*, 101044.
45. World Health Organization. Ageing and Health. Available online: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed on 26 May 2026).
46. Beange, K.H.E.; Chan, A.D.C.; Graham, R.B. Investigating Concurrent Validity of Inertial Sensors to Evaluate Multiplanar Spine Movement. *J. Biomech.* **2024**, *164*, 111939.
47. Chen, C.; Liaw, A.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; Technical Report 666; University of California: Berkeley, CA, USA, 2004.
48. Haresamudram, H.; Essa, I.; Plötz, T. Towards Learning Discrete Representations via Self-Supervision for Wearables-Based Human Activity Recognition. *Sensors* **2024**, *24*, 1238.
49. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. Deep Learning for Sensor-Based Human Activity Recognition: Overview, Challenges, and Opportunities. *ACM Comput. Surv.* **2021**, *54*, 77.
50. Atallah, L.; Lo, B.; King, R.; Yang, G.-Z. Sensor Placement for Activity Detection Using Wearable Accelerometers. In Proceedings of the 2010 International Conference on Body Sensor Networks, Biopolis, Singapore, 7–9 June 2010; pp. 184–187.
51. Fan, B.; Li, Q.; Liu, T. How Magnetic Disturbance Influences the Attitude and Heading in Magnetic and Inertial Sensor-Based Orientation Estimation. *Sensors* **2018**, *18*, 76.
52. Chen, H.; Schall, M.C., Jr.; Fethke, N.B. Effects of Movement Speed and Magnetic Disturbance on the Accuracy of Spatial Orientation Measurements Using Inertial Measurement Units. *Sensors* **2017**, *17*, 1470.
53. Kok, M.; Hol, J. D.; Schön, T.B. Using Inertial Sensors for Position and Orientation Estimation. *Found. Trends Signal Process.* **2017**, *11*, 1–153.
54. Woodman, O.J. *An Introduction to Inertial Navigation*; Technical Report UCAM-CL-TR-696; University of Cambridge, Computer Laboratory: Cambridge, UK, 2007.

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.