Recognizing Drinking Gestures with Wrist-Worn Inertial Sensors: Public Dataset and Lightweight CNN Approach

Pedro D. Gohl[1], Laura I. Queiroz[1], Eduardo Souto[1], Eulanda M. Santos[1]

[1]Institute of Computing – Federal University of Amazonas (UFAM)
Manaus – AM – Brazil

{pedro.gohl, isabelle.queiroz, esouto, emsantos}@icomp.ufam.edu.br

Abstract. *Monitoring fluid intake via wearable sensors is challenging due to gesture variability and overlap with similar hand movements. We propose a deep learning approach for liquid intake detection using wrist-worn accelerometer and gyroscope data, segmented into 4-second windows at 50 Hz. We introduce a Multi-Scale 1D Convolutional Neural Network (MS-Conv1D) that extracts spatiotemporal features at multiple temporal resolutions. Evaluated via 50-fold Leave-One-Subject-Out cross-validation, the model achieved F1-scores of 93.33% (binary) and 91.50% (multiclass), outperforming traditional classifiers and most state-of-the-art baselines, while remaining efficient enough for real-time execution on embedded hardware.*

1. Introduction

Water constitutes the primary component of the human body, playing a vital role in sustaining life by supporting nutrient transport, temperature regulation, maintenance of cellular structure, and joint lubrication [Lorenzo et al. 2019]. Consequently, maintaining adequate water balance is imperative for overall health and well-being.

Dehydration, when fluid loss exceeds intake, impairs physical and cognitive performance and negatively affects gastrointestinal, renal, cardiac, and hepatic functions [Liaqat et al. 2022, Likhtenshtein 2021]. Despite its importance, water intake is often overlooked, particularly among vulnerable populations such as the elderly, who may experience decreased thirst sensation [Begg 2017]. Reliable detection of drinking gestures is therefore the cornerstone of any automatic system that estimates the volume consumed and, ultimately, hydration status.

Traditional monitoring still relies largely on self-report questionnaires, which are error-prone and impractical for continuous use. Accurate detection of a drinking gesture is the first indispensable step toward automatic estimation of consumed volume and, ultimately, hydration status. Wearable devices equipped with inertial sensors offer a privacy-preserving, container-agnostic alternative that blends naturally into daily routines. Although commercial smartwatches allow manual water logs (e.g., Apple Health, Garmin Hydration), no current model performs passive recognition of the drinking gesture, an unmet need that this study begins to address.

Environment-based solutions using cameras [Iosifidis et al. 2012, Tham et al. 2014] and smart bottles with embedded sensors [Griffith and Biswas 2019, Liu et al. 2020, Roy et al. 2022] improve objectivity, but they raise privacy concerns or require specific containers [Cohen et al. 2021]. Consequently, wrist-worn inertial-sensor systems have emerged as a practical and privacy-preserving alternative [Amft et al. 2010, Anderez et al. 2018, Gomes and Sousa 2019].

Despite encouraging results, two intertwined challenges persist: (i) the scarcity of sizeable, publicly available datasets that enable fair benchmarking, and (ii) the need for models with modest computational footprints that can run on resource-constrained wearables.

To tackle these gaps, we collected InGesture[1], a dataset comprising 50 participants, more than 5,000 labeled drinking events, and 39,000 seconds of inertial sensor data, surpassing previous datasets in both size and diversity. We also propose a tailored CNN architecture for recognizing fluid intake events from inertial sensor data. Our streamlined design aimed to ensure strong generalization across subjects. Validated via Leave-One-Subject-Out cross-validation, it achieved strong results with an F1-score of 93.33% in binary and 91.50% in multiclass scenarios. The full set of preprocessing, training and evaluation scripts can be accessed at code-link (link concealed for double-blind review), enabling complete reproducibility.

The remainder of this article is organized as follows: Section 2 reviews related work; Section 3 details the dataset, processing pipeline, model, and validation protocol; Section 4 presents results; and Section 5 discusses conclusions and outlines future directions.

2. Related Work

Research on automatic fluid-intake monitoring spans three broad approaches: environment-based systems, container-attached sensors, and wearable devices.

Environment-based solutions use external sensors, such as cameras and sensors embedded in furniture or objects, to identify drinking episodes. Previous studies have proposed the use of cameras attached to intelligent systems to automatically capture these events [Iosifidis et al. 2012, Tham et al. 2014]. While this approach provides high accuracy, it poses challenges related to user privacy and monitoring limitations in controlled environments, reducing its applicability for continuous use.

Another research direction explores container-attached sensors, such as smart cups and bottles, to monitor fluid intake. Studies such as those by [Griffith and Biswas 2019], [Liu et al. 2020], and [Roy et al. 2022] developed systems that use accelerometers and flow sensors embedded in containers to detect drinking episodes and estimate the ingested volume. However, these approaches require users to use specific containers, which may hinder adherence and limit their applicability in various settings.

Advances in wearable technology led inertial sensors embedded in smartwatches and wristbands to be extensively explored for the automatic detection of fluid intake. These solutions offer continuous, non-intrusive monitoring, independent of the container used. Previous studies have investigated this approach, using both traditional machine learning models and deep neural networks to detect drinking patterns from inertial sensor data. [Amft et al. 2010] explored the use of wrist and shoulder sensors to detect drinking behavior and estimate the ingested volume. [Hamatani et al. 2018] proposed

[1]The dataset has been made publicly available at `https://data.mendeley.com/datasets/fdxst56tcj/5`

a method based on analyzing arm posture during drinking, achieving 84% accuracy in detecting drinking events and a 15% average error in volume estimation. Additionally, [Anderez et al. 2018] and [Gomes and Sousa 2019] developed real-time detection systems using wearable sensors, relying on machine learning models and pulse signals.

More recent studies further highlight the potential of this approach. [Moccia et al. 2022] compared the performance of various deep and shallow machine learning methods for automatic drinking detection, using a private dataset with hand gesture signals. [Martínez et al. 2023] proposed an algorithm for detecting fluid intake in older adults, focusing on detection robustness even in cases of uncontrolled movements, achieving 92% accuracy. [Li et al. 2024] developed a multimodal system, combining inertial sensors on the wrist, container-embedded sensors, and an in-ear microphone to detect drinking events. Their study compared different machine learning models and demonstrated that sensor fusion significantly improved detection, achieving F1-score of 96.5% in event-based evaluation. Another relevant advancement was the study by [Cergolj et al. 2024], which developed an optimized model for edge devices, prioritizing low energy consumption. The solution only activates inference when there is a high probability of drinking, reducing battery consumption by a factor of 5.8 during inactivity periods. Their tests demonstrated F1-score of 89.4% in offline evaluations and 74.5% in real-world conditions, highlighting the feasibility of efficient deployment in wearable devices.

Despite these advances, most existing studies still face challenges related to robustness in real-world scenarios and the availability of public datasets to train generalizable models. Therefore, this study proposes two main contributions: (i) a deep learning model for fluid intake detection, based on a convolutional architecture, specifically trained to handle inertial sensor signals and validated using Leave-One-Subject-Out (LOSO) cross-validation to ensure robustness; and (ii) a new dataset, incorporating drinking-like gestures to enhance the model's reliability in real-world scenarios. Unlike previous datasets, which have limited participant numbers and standardized movements, our dataset considers natural variations in user movements, making the model more adaptable to different profiles and contexts.

3. Materials and Methods

This section details the wrist-worn device used for data capture (Section 3.1), the acquisition protocol (Section 3.2), the data processing methods (Section 3.3), the gesture classification pipeline (Section 3.4), and the experimental protocol for evaluation and testing (Section 3.5).

3.1. Witmotion Wrist Monitoring Device

Data were collected using the WitMotion wrist wearable device, model WT901BLECL5.0 [WitMotion 2024], worn on the dominant wrist of each participant. The device is light, comfortable, and equipped with a USB rechargeable battery, providing up to 10 hours of activity collection during daily use. This module (Figure 1) integrates high-precision gyroscopes, accelerometers, and geomagnetic field sensors, and leverages high-performance microprocessors, advanced dynamics calculation, and Kalman dynamic filtering algorithms to swiftly determine the current real-time motion posture of the module. Notably, the output content can be customized, with a data output frequency of 200 Hz. For this study, only accelerometer and gyroscope data were utilized, with a sampling frequency of 200 Hz. These sensor streams form the raw multivariate time series used for gesture recognition. For data collection, the device was connected via Bluetooth to an Android smartphone, which utilized an application for real-time data labeling and exported it to CSV files.

Figure 1. Wrist monitoring device used for data collection.

3.2. Study Population and Acquisition Protocol

The study was approved by the Ethical Committee of Federal University of Amazonas under the number 70951823.0.0000.5020. Fifty (50) healthy volunteers, both men and women (mean $\pm$ SD, 26 $\pm$ 11 years, age range 18 to 67), willingly enrolled in the study upon signing a written informed consent form.

The protocol investigated liquid intake recognition in a controlled environment. Participants wore a monitor on their dominant wrist and were guided through a series of daily hand gestures (Figure 2), with fluid intake gestures providing visual context in Figure 3. A researcher performed real-time ground-truth labeling. Participants used varied containers (personal bottles, 100 ml, or 200 ml cups) for drinking and moved freely between tasks. Uncatalogued intermediate movements were labeled as “other”. The seven targeted gestures were: (1) fluid intake; (2) answering a cell phone; (3) scratching the head; (4) passing a hand over the face; (5) holding the chin; (6) adjusting glasses or touching temples; and (7) stretching the back with hands behind the neck.

The resulting dataset includes accelerometer, gyroscope, and magnetometer signals, alongside roll, pitch, and heading angles. Data is provided in two versions. The raw data uses datetime timestamps and three categories: “other” (0), “fluid intake” (1), and “activity” (2). The processed data converts timestamps to milliseconds and applies an 8-class system: "other" (0) alongside the seven specific gestures numbered 1 through 7, respectively.

The dataset comprises a total of 5265 labeled events, where each event represents a continuous interval of a specific label. During each 10-minute collection period, there were 20 fluid intake events, 10 answering cell phone events, and 2 events for each of the 5 similar activities, totaling 40 events per collection, along with 41 intermediate events related to other activities.

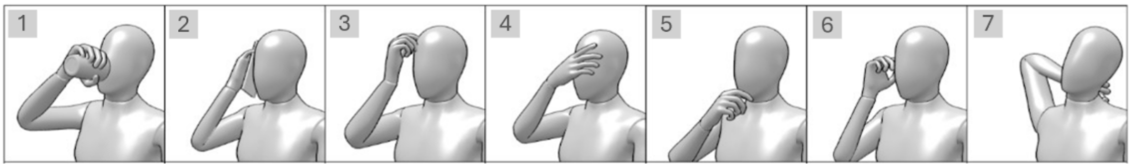

Figure 2. Hand gestures investigated in this study.

Figure 3. Participants demonstrating fluid intake gestures.

3.3. Data Pre-processing

Although the device streams at 200 Hz, we conducted resampling to standardize the frequency to 50 Hz (as in [Wang et al. 2022]) using linear interpolation. Prior to segmentation, the continuous data was z-score normalized; to strictly prevent data leakage, the scalers were fitted and applied separately for the training and testing datasets. We then segmented the data into fixed 4-second intervals—a size chosen because it reflects the mean duration of the fluid intake activity—with a 50% overlap.

3.4. Network Architecture

The preprocessed 4-second signal windows (comprising 6 channels from the accelerometer and gyroscope axes) serve as input to our proposed Multi-Scale Convolutional Neural Network (MS-Conv1D). The network begins with a 1D CNN layer utilizing 64 filters and a kernel size of 3, followed by batch normalization (BN) and max-pooling (pool size of 2). To capture diverse temporal dependencies simultaneously, a multi-scale convolutional block then processes the data through three parallel branches. Each branch utilizes 128 filters with varying kernel sizes (3, 5, and 7). The outputs of these branches are concatenated along the last axis, normalized, and pooled.

Next, a final convolutional layer containing 256 filters with a kernel size of 3 and Batch Normalization is applied. A Global Average Pooling (GAP) layer replaces the traditional flatten layer to further reduce spatial dimensions and mitigate overfitting. The extracted features then feed directly into a dense output layer utilizing a softmax activation function to generate class probability distributions.

Models were trained using the Adam optimizer (initial learning rate of 0.001 with a scheduler) and either binary or sparse categorical cross-entropy loss functions, depending on the classification task. To ensure a robust assessment of the model's generalization across unseen subjects, performance was evaluated using a 50-fold Leave-One-Subject-Out (LOSO) cross-validation strategy, with final metrics reported as the mean across all folds. Figure 4 illustrates the proposed architecture.

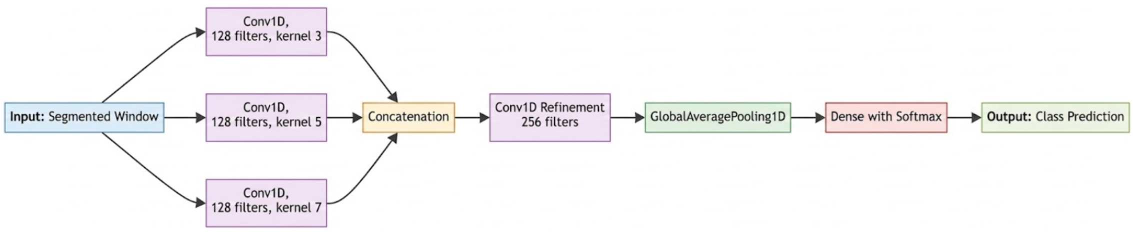

Figure 4. The overall pipeline of the proposed model for activity recognition.

3.5. Validation and Performance Metrics

To ensure robustness and generalization across individuals, all proposed and benchmark models were evaluated using Leave-One-Subject-Out (LOSO) cross-validation on the collected dataset, ensuring a fair comparison. Given the dataset's class imbalance, traditional accuracy is insufficient. Instead, performance was evaluated based on True Positives (TP), False Positives (FP), and False Negatives (FN) using precision, recall, and the F1-score:

$$\text{precision} = \frac{TP}{TP + FP}, \quad \text{recall} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{F1-score} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} \tag{2}$$

4. Results and Discussion

We conducted two series of experiments employing binary and multiclass classification models. This section provides a comprehensive analysis of the results obtained from our proposed methodology in comparison with baseline classifiers.

In our experiments, all models, including our proposed approach and benchmark classifiers were trained and tested using the same dataset collected in this study. The methodologies provided in [Senyurek et al. 2019], [Huang et al. 2020], [Moccia et al. 2022] and [Wang et al. 2024] were re-implemented and adapted to our dataset under identical experimental conditions, following the Leave-One-Subject-Out (LOSO) cross-validation protocol. Table 1 outlines the benchmark algorithms and their corresponding hyperparameters.

Table 1. Benchmark algorithms hyperparameters.

Author	Method	Hyperparameters
Huang et al. (2020)	AdaBoost, Decision Tree, Random Forest, Naïve Bayes, k-NN, SVM	AdaBoost: CART base estimators; Decision Tree: Impurity-based (CART); Random Forest: Min leaf=1, min parent=2; Naïve Bayes: Normal distribution; k-NN: k=3; SVM: Linear kernel; Sliding window: 50Hz, 50% overlap; Input: Statistical features (mean, std, max, min, kurtosis, skewness).
Senyurek et al. (2019)	CNN-LSTM	Conv1D K: input/2, input/4, input/8. F: 128, 64, 32. MaxPool: 2. Dense: 32. LSTM: 64, 64. Dropout: 0.5. Opt: Adam (lr: 0.001).
Moccia et al. (2022)	CNN-LSTM	TD-Conv1D F: 100, 150, 150 (K: 3). MaxPool: 3. LSTM: 150. Dense: 1000, 500. Dropout: 0.5. Opt: Adam (lr: 0.001).
Moccia et al. (2022)	CNN	Conv1D F: 100, 150, 150 (K: 3). MaxPool: 3. Dense: 1000, 500 (multi) / 200, 100 (binary). Dropout: 0.5. Opt: Adam (lr: 0.001).
Wang et al. (2024)	MHA-TCN	TCN: 9 layers, 64 F (K: 3, dilation: 2^i). MHA: 8 heads (dim: 16). FCN: 64. Dropout: 0.3. Opt: Adam (lr: 0.0005).
Proposed Model	MS-Conv1D	Init Conv1D: 64 F (K: 3). MS-Branches: 3×128 F (K: 3, 5, 7). Final Conv1D: 256 F (K: 3). Pool: Max 2, GAP. Opt: Adam (lr: 0.001).

4.1. Binary Classification Results

The task of detecting liquid intake activity presents significant challenges due to gesture variability and potential overlaps with similar activities. Table 2 presents the results obtained for binary classification. In terms of recall, the Naïve Bayes model achieved the highest value (92.32%), but at the cost of a very low precision (32.97%), indicating a high false positive rate. Our proposed model exhibited a better tradeoff, providing 92.12% recall and 96.01% precision, resulting in the highest F1-score of 93.33%. The SVM model demonstrated strong precision performance (84.05%), but its lower recall (69.20%) reduced its overall performance.

It is important to note that the superiority of our model can be attributed to the deep feature extraction capability of deep learning networks, which allowed it to differentiate between fluid intake gestures and other hand movements. In contrast, shallow classifiers like Decision Trees and k-NN exhibited inferior performance, not necessarily due to their inherent limitations, but because they relied on manually extracted features that may not have been sufficiently representative of the complex temporal patterns in inertial sensor data.

Table 2. Results for drinking activity detection (binary classification).

Approach	Classifier	Recall (%)	Precision (%)	F1-score (%)
Shallow Learning	KNN	74.48	79.94	75.38
	SVM	69.20	84.05	73.90
	Random Forest	66.66	89.23	73.54
	Decision Tree	69.30	65.76	66.39
	Naïve Bayes	92.32	32.97	47.27
	AdaBoost	76.23	84.72	77.85
Deep Learning	CNN Moccia et al. (2022)	82.49	94.34	86.76
	CNN-LSTM Moccia et al. (2022)	72.14	83.85	75.17
	CNN-LSTM Senyurek et al. (2019)	72.43	80.45	74.28
	TCN-MHA Wang et al. (2024)	95.24	92.64	93.45
	Proposed Model	**92.12**	**96.01**	**93.33**

4.2. Multiclass Classification Results

Table 3 presents the results for multiclass classification, where each activity was treated as a separate class. When compared to the results attained in the scenario of binary classification, most models exhibited an increase in F1-score, reflecting their improved ability to distinguish between drinking and non-drinking activities. The SVM model achieved an F1-score of 82.99%, demonstrating balanced performance across all classes. In contrast, Naïve Bayes, despite achieving high recall (86.78%), suffered from lower precision, leading to a moderate F1-score of 66.11%.

Our proposed CNN model, however, outperformed the vast majority of baselines, achieving an F1-score of 91.50%, confirming its robustness in recognizing fluid intake events amid similar activities. The key factor behind this performance lies in the ability of deep learning architectures to leverage temporal dependencies in the sensor data, leading to more accurate classification of ambiguous activities, such as answering the phone or adjusting glasses.

Table 3. Results for drinking activity detection (multiclass classification).

Approach	Classifier	Recall (%)	Precision (%)	F1-score (%)
Shallow Learning	KNN	83.89	77.06	79.38
	SVM	85.14	82.71	82.99
	Random Forest	85.23	86.26	84.28
	Decision Tree	76.76	78.02	76.20
	Naïve Bayes	86.78	54.96	66.11
	AdaBoost	79.46	76.44	76.05
Deep Learning	CNN Moccia et al. (2022)	90.19	90.77	88.31
	CNN-LSTM Moccia et al. (2022)	86.32	81.51	82.54
	CNN-LSTM Senyurek et al. (2019)	90.07	87.02	87.84
	TCN-MHA Wang et al. (2024)	95.59	92.93	93.90
	Proposed Model	**90.69**	**93.01**	**91.50**

4.3. Comparative Discussion of Related Work

Our findings are consistent with recent literature but offer distinct advantages. For example: [Li et al. 2024] explored a multi-sensor fusion approach combining wrist motion, container movement, and acoustic swallowing signals, achieving 96.5% F1-score in drinking detection. While their multimodal approach improves robustness, it requires additional sensor modalities (e.g., microphones), increasing hardware complexity. Our single-device approach offers a more practical solution for real-world applications.

In their turn, [Martínez et al. 2023] focused on fluid intake quantification for elderly using wearable sensors. Their model achieved high accuracy of 92%, but their dataset was limited to controlled settings. In contrast, our model generalizes better to diverse hand gestures and real-life conditions. Finally, [Cergolj et al. 2024] developed an energy-efficient drinking detection system optimized for edge computing on smartwatches. Their model reduced power consumption by 5.8×, but achieved lower real-world precision (74.5%). Our deep learning approach, while computationally more demanding, offers higher accuracy and adaptability.

5. Conclusion and Future Work

This study introduced a deep-learning model for fluid-intake gesture recognition using inertial sensors embedded in wearable devices. To support this research, we developed InGesture, a new dataset containing 5,265 labeled drinking events from 50 participants, ensuring diversity in the collected gestures and enabling a robust evaluation of the models.

Our proposed CNN demonstrated superior performance, achieving F1-scores of 93.3% (binary) and 91.5% (multiclass), surpassing traditional shallow-learning baselines.

In contrast to previous studies that rely on private data, we release both the dataset and the complete set of preprocessing, training and evaluation scripts at dataset-link and code-link (links concealed for double-blind review), fostering reproducible progress in hydration-monitoring research.

However, several limitations must be acknowledged. First, data collection occurred in a controlled environment, which raises concerns regarding generalisation to free-living scenarios where daily routines, distractions and environmental variability may affect performance. Second, although participants varied in age and gender, the sample may not fully represent the broader population—especially older adults or individuals with motor impairments—introducing potential bias. Cultural and habitual differences in drinking behaviour (e.g., container choice or posture) add further challenges. In terms of internal validity, real-time manual labelling is susceptible to human error, particularly during ambiguous transitions between gestures. Finally, the current approach detects drinking events but does not yet estimate the volume consumed, a key factor for precise hydration assessment.

Preliminary on-device tests on a Samsung Galaxy Watch 8 confirm that real-time inference is feasible. Ongoing work focuses on model-compression strategies (quantisation and pruning) to reduce memory and compute demands, and on sensor-fusion approaches that combine inertial data with smart-container or physiological signals to estimate intake volume.

Future research will prioritise large-scale free-living evaluations, broader demographic coverage and regression models for liquid volume, ultimately closing the remaining gap toward fully passive, wrist-based hydration monitoring.

Acknowledgements

This research was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES-PROEX) – Finance Code 001, and the National Council for Scientific and Technological Development – CNPq. This work was also partially supported by Amazonas State Research Support Foundation – FAPEAM, through the POSGRAD 2024–2025 project.

References

Amft, O., Bannach, D., Pirkl, G., Kreil, M., and Lukowicz, P. (2010). Towards wearable sensing-based assessment of fluid intake. In *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 298–303.

Anderez, D. O., Lotfi, A., and Langensiepen, C. (2018). A hierarchical approach in food and drink intake recognition using wearable inertial sensors. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, pages 552–557.

Begg, D. P. (2017). Disturbances of thirst and fluid balance associated with aging. *Physiology & Behavior*, 178:28–34.

Cergolj, V. et al. (2024). Drinking event detection on a sensing wristband using machine learning. *Journal of Ambient Intelligence and Smart Environments*, pages 1–20. Preprint.

Cohen, R., Fernie, G., and Roshan Fekr, A. (2021). Fluid intake monitoring systems for the elderly: a review of the literature. *Nutrients*, 13(6):2092.

Gomes, D. and Sousa, I. (2019). Real-time drink trigger detection in free-living conditions using inertial sensors. *Sensors*.

Griffith, H. and Biswas, S. (2019). Improving water consumption estimates from a bottle-attachable sensor using heuristic fusion. In *2019 IEEE 20th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–3.

Hamatani, T., Elhamshary, M., Uchiyama, A., and Higashino, T. (2018). Fluidmeter: Gauging the human daily fluid intake using smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, pages 1–25.

Huang, H.-Y. et al. (2020). Fluid intake monitoring system using a wearable inertial sensor for fluid intake management. *Sensors*, 20(22):6682.

Iosifidis, A., Marami, E., Tefas, A., and Pitas, I. (2012). Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2201–2204.

Li, J.-H. et al. (2024). Multi-sensor fusion approach to drinking activity identification for improving fluid intake monitoring. *Applied Sciences*, 14(11):4480.

Liaqat, S., Dashtipour, K., Rizwan, A., Usman, M., Shah, S. A., Arshad, K., Assaleh, K., and Ramzan, N. (2022). Personalized wearable electrodermal sensing-based human skin hydration level detection for sports, health and wellbeing. *Scientific Reports*.

Likhtenshtein, G. I. (2021). Water: Clinical aspects. In *Biological Water: Physicochemical Aspects*, pages 481–512. unknown.

Liu, K.-C., Hsieh, C.-Y., Huang, H.-Y., Chiu, L.-T., Hsu, S. J.-P., and Chan, C.-T. (2020). Drinking event detection and episode identification using 3d-printed smart cup. *IEEE Sensors Journal*, pages 13743–13751.

Lorenzo, I., Serra-Prat, M., and Yébenes, J. C. (2019). The role of water homeostasis in muscle function and frailty: A review. *Nutrients*.

Martínez, P., Gordillo-Castillo, N., and Cortés Sáenz, D. (2023). Towards fluid intake quantification in older adults: An algorithm for movement detection using accelerometry and gyroscope sensors. In *Congreso Nacional de Ingeniería Biomédica*, pages 222–231, Cham. Springer Nature Switzerland.

Moccia, S., Solbiati, S., Khornegah, M., Bossi, F. F., and Caiani, E. G. (2022). Automated classification of hand gestures using a wristband and machine learning for possible application in pill intake monitoring. *Computer Methods and Programs in Biomedicine*, 219:106753.

Roy, A., Dutta, H., Griffith, H., and Biswas, S. (2022). An on-device learning system for estimating liquid consumption from consumer-grade water bottles and its evaluation. *Sensors*, 22(7):2514.

Senyurek, V., Imtiaz, M., and Hassan, N. (2019). Detection of drinking via a wrist-worn inertial sensor.

Tham, J. S., Chang, Y. C., and Fauzi, M. F. A. (2014). Automatic identification of drinking activities at home using depth data from rgb-d camera. In *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*, pages 153–158. IEEE.

Wang, C., Kumar, T. S., De Raedt, W., Camps, G., Hallez, H., and Vanrumste, B. (2022). Drinking gesture detection using wrist-worn imu sensors with multi-stage temporal convolutional network in free-living environments. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1778–1782.

Wang, C., Kumar, T. S., De Raedt, W., Camps, G., Hallez, H., and Vanrumste, B. (2024). Eating speed measurement using wrist-worn imu sensors towards free-living environments. *IEEE journal of biomedical and health informatics*, 28(10):5816–5828.

WitMotion (2024). WT901BLECL 9-axis BLE accelerometer inclinometer. `https://www.wit-motion.com/BLE/52.html`.