Advanced Mechanical and
Mechatronic Systems

Research Article
Volume 2(2), 2026, 76−87
https://doi.org/10.53623/amms.v2i2.999

A Hybrid Machine Learning Framework for Multi-Limb Human Activity Recognition Using Synchronized Smartphone IMU Sensors: Dataset and Benchmarking

Ade Kurniawan*, Dadan Ramdan Hidayat, Zain Iqbal Saputra, Whirdyana Shalfa Ayubi, Syifa Nurulfajri Rustin, Muhammad Ragil Rizky Mulya, Chello Fhrino Mike Mandolang

Department of Data Science, Institut Teknologi Sains Bandung, Kota Deltamas Lot-A1 CBD, Bekasi, 17530, West Java, Indonesia

*Correspondence: ade.k@itsb.ac.id

SUBMITTED: 18 January 2026; REVISED: 10 February 2026; ACCEPTED: 19 February 2026

ABSTRACT: Human Activity Recognition (HAR) using smartphone inertial measurement unit (IMU) sensors has emerged as a transformative technology for health monitoring, fitness tracking, and context-aware computing. However, existing HAR research is constrained by limited data availability, short recording durations, and single-limb sensing configurations. This study addresses these limitations through three principal contributions: (1) introduction of a novel open-access multi-limb HAR dataset featuring synchronized 60-second recordings from hand and ankle positions using tri-axial accelerometer, gyroscope, and magnetometer sensors, publicly available via Mendeley Data repository; (2) systematic benchmarking of classical machine learning classifiers including Random Forest, XGBoost, and Linear Support Vector Classifier under realistic multi-sensor fusion conditions; and (3) comprehensive analysis of model robustness across varying windowing configurations. The dataset comprises recordings from six participants performing six daily activities (walking, stair ascent, stair descent, standing, sitting, lying), totaling over 72 minutes of synchronized multi-sensor data. Experimental evaluation demonstrates that Random Forest achieves superior classification accuracy of 96.13%, significantly outperforming XGBoost (85.22%) and LinearSVC (58.54%). The publicly released dataset and benchmarking results provide a valuable resource for the HAR research community, enabling reproducible experimentation and facilitating advancement in multi-limb activity recognition systems.

KEYWORDS: Human activity recognition; open dataset; smartphone IMU sensors; multi-limb sensing; machine learning; random forest

1. Introduction

Human Activity Recognition (HAR) emerged as a cornerstone technology in ubiquitous computing, enabling diverse applications ranging from healthcare monitoring and fitness tracking to smart home automation and elderly care systems [1, 2, 3]. The proliferation of smartphones equipped with sophisticated inertial measurement units (IMUs)—comprising tri-axial accelerometers, gyroscopes, and magnetometers—democratized HAR research by providing ubiquitous, cost-effective sensing platforms capable of continuous motion data acquisition [4, 5]. Contemporary HAR systems leveraged the rich temporal information embedded in IMU signals to characterize human movement patterns. Accelerometers captured linear acceleration dynamics, gyroscopes measured angular velocity, and magnetometers provided orientation information relative to the Earth's magnetic field [6]. The fusion of these complementary sensor modalities was demonstrated to substantially enhance activity discrimination capabilities compared to single-sensor configurations [7, 8].

Despite significant advances in sensor technology and machine learning methodologies, HAR research faced several fundamental limitations. First, the scarcity of publicly available, high-quality datasets constrained reproducibility and comparative evaluation across studies [9]. Second, the majority of existing datasets featured recordings of relatively short duration—typically ranging from 3 to 10 seconds per activity instance—which was insufficient to capture long-term temporal stability and movement continuity essential for robust activity modeling [10, 11]. Third, most existing studies employed sensors positioned at a single body location, thereby neglecting the coordinated biomechanical relationships between different limbs during locomotion [12, 13]. Fourth, the predominant reliance on proprietary or non-public datasets hindered open science practices and impeded cumulative progress in the field [14]. To address these critical gaps, this study made three principal contributions to the HAR research domain. First, an open-access multi-limb dataset was introduced, featuring 60-second continuous recordings per activity from synchronized hand and ankle positions. The dataset was publicly released via the Mendeley Data repository (DOI: 10.17632/9d77352dcf.2) to facilitate reproducible research and community advancement. Second, a systematic classifier benchmarking was conducted, providing a rigorous comparative analysis of three representative machine learning classifiers—Random Forest, XGBoost, and LinearSVC—and establishing baseline performance metrics for future research. Third, a robustness analysis was performed, presenting a comprehensive assessment of model stability across varying windowing parameters and offering practical guidance for system deployment.

Distinct from prior multi-sensor HAR datasets such as UCI-HAR [10], WISDM [15], and PAMAP2 [16], which were constrained by short recording durations (typically 2–10 seconds) and single-limb placement, the proposed dataset uniquely combined three characteristics that had not been simultaneously addressed in the existing literature: (1) extended 60-second continuous recordings that enabled analysis of long-term temporal dynamics; (2) synchronized dual-limb sensing (hand and ankle) that captured inter-limb biomechanical coordination; and (3) fully open-access availability under the CC BY 4.0 license to support reproducible research. Furthermore, although deep learning methods showed strong performance on large-scale benchmarks [17, 18], the benchmarking in this study focused on classical machine learning classifiers to establish interpretable and computationally efficient baselines suitable for resource-constrained deployment scenarios.

2. Related Work

2.1. Benchmark HAR Datasets.

The development of benchmark datasets was instrumental in advancing HAR research. The UCI-HAR dataset [10], comprising recordings from 30 subjects performing six activities using Samsung Galaxy S II smartphones, became a de facto standard with over 3,000 citations. Anguita et al. demonstrated 96% classification accuracy using hardware-optimized Support Vector Machines [19]. The WISDM dataset [15] provided one of the earliest smartphone HAR benchmarks, including 29 subjects and six activities at a 20 Hz sampling rate.

The OPPORTUNITY dataset [11] offered comprehensive 72-sensor multimodal data from four subjects and formed the basis for the IEEE SMC Activity Recognition Challenge. PAMAP2 [16] captured 18 physical activities from nine subjects using three IMUs at 100 Hz, while MobiAct [20] extended to 66 subjects with 16 activity classes, including falls. More recently, UniMiB SHAR [21] provided 11,771 activity instances and eight fall types from 30 subjects. However, these benchmark datasets shared common limitations: short recording durations (typically 2–10 seconds), single-limb sensor placement, and, in some cases, restricted access or licensing constraints that limited reproducibility [9, 14].

2.2. Multi-sensor fusion approaches.

Multi-sensor fusion was demonstrated to consistently improve HAR performance. Shoaib et al. [6] established foundational evidence showing that different sensors assumed lead roles depending on activity type and body position. Webber and Fernandez Rojas [7] systematically compared sensor-level, feature-level, and decision-level fusion strategies, finding that decision-level fusion achieved the highest accuracy, while Kalman filtering provided optimal accuracy–efficiency trade-offs. Chung et al. [8] developed an LSTM-based classifier-level fusion approach using eight body-worn 9-axis IMUs and concluded that data from four sensors placed at the wrists, waist, and ankle at 10 Hz sufficed for activities of daily living recognition. These findings motivated the dual-location (hand and ankle) sensing approach adopted in this study.

2.3. Classification methodologies.

Classical machine learning algorithms remained competitive for HAR applications. Wang et al. [22] provided a comprehensive comparison of SVM, k-NN, Decision Trees, Random Forest, and Naive Bayes across different sensor configurations. Catal et al. [23] demonstrated that ensemble methods combining multiple classifiers outperformed individual models on WISDM data. Deep learning approaches achieved state-of-the-art performance on large benchmarks. Ordóñez and Roggen [17] established the hybrid CNN–LSTM paradigm through DeepConvLSTM, achieving 93% accuracy on OPPORTUNITY. Hammerla et al. [18] demonstrated the superiority of bidirectional LSTM models on large datasets. However, classical methods remained preferred for resource-constrained deployments and in scenarios where interpretability was required [24, 25].

3. Materials and Methods

3.1. Dataset description and availability.

The Dual-Location Smartphone IMU Human Activity Recognition Dataset was made publicly available through the Mendeley Data repository under an open-access license.

Dataset Citation: Kurniawan, A.; Hidayat, D.R.; Saputra, Z.I.; Ayubi, W.S.; Fajri R, S.N.; Mulya, M.R.R.; Mandolang, C.F.M. (2026). "Dual-Location Smartphone IMU Human Activity Recognition Dataset (Hand & Ankle, 60s Trials)", Mendeley Data, V2. DOI: 10.17632/9d77352dcf.2

The dataset comprised synchronized tri-axial accelerometer, gyroscope, and magnetometer recordings from six healthy participants (aged 20–22 years) performing six daily activities. Unlike existing datasets constrained to short recording windows, this dataset featured 60-second continuous recordings per activity, enabling analysis of long-term temporal dynamics and movement stability patterns.

3.2. Data acquisition protocol.

Data acquisition was performed using Android smartphones equipped with built-in IMU sensors. The Phyphox application (version 1.19.0), developed by RWTH Aachen University [26], was employed for simultaneous multi-sensor recording at approximately 100 Hz sampling frequency. The 100 Hz sampling rate was selected because it was widely adopted in HAR literature and had been demonstrated to be sufficient for capturing human movement dynamics, which typically occurred below 20 Hz [27]. Although the Phyphox application targeted a nominal rate of 100 Hz, minor fluctuations (typically within ±2 Hz) occurred due to operating system scheduling on Android devices. These variations were consistent across all sensors and participants and were mitigated during the preprocessing stage through timestamp-based resampling and synchronization (see Section 3.4). Two smartphone positions were utilized to capture complementary movement information:

- Hand position: The smartphone was held naturally in portrait orientation, capturing upper-body dynamics and arm movements characteristic of daily activities.
- Ankle position: The smartphone was attached above the ankle using an elastic strap, enabling direct observation of gait dynamics, foot–ground interactions, and stride patterns.

Six representative daily activities were recorded: walking on level ground, ascending stairs, descending stairs, standing, sitting, and lying down. These activities spanned both static postures and dynamic locomotion, providing diverse temporal and biomechanical characteristics. Each activity was recorded for approximately 60 seconds per participant per position. The recordings were controlled using a dedicated timer operated by a research assistant to ensure consistency. The actual recording durations ranged from 58 to 63 seconds across sessions, with a mean of 60.2 ± 1.1 seconds. These minor variations did not affect the analysis, as the sliding window segmentation approach (Section 3.4) operated on fixed-length windows irrespective of total recording length. The total dataset duration exceeded 72 minutes of synchronized multi-sensor data.

3.3. Dataset structure and format.

The dataset was organized hierarchically by participant, sensor position, and activity type. Each recording session generated three CSV files corresponding to accelerometer, gyroscope, and magnetometer data. The file naming convention followed the pattern: [Participant]*[Position]*[Activity]_[Sensor].csv. Each CSV file contained a timestamp column (seconds) and tri-axial measurements (X, Y, Z axes). Table 1 summarized the dataset characteristics.

Table 1. Dataset characteristics summary.

Characteristic	Specification
Participants	6 (healthy adults, age 20-30)
Activities	6 (walking, upstairs, downstairs, standing, sitting, lying)
Sensor Positions	2 (hand, ankle)
Sensors	Tri-axial accelerometer, gyroscope, magnetometer
Recording Duration	60 seconds per activity per position
Sampling Frequency	~100 Hz
Total Duration	>72 minutes synchronized data
Data Format	CSV files with timestamps
License	CC BY 4.0 (Open Access)

3.4. Data preprocessing pipeline.

The preprocessing pipeline comprised three main stages: time synchronization, signal segmentation, and feature extraction. Raw sensor data from the accelerometer, gyroscope, and magnetometer exhibited slightly different time resolutions. Timestamps were standardized by rounding to two decimal places, followed by inner merge operations to retain only temporally aligned samples across all three sensors. This rounding approach introduced a maximum temporal misalignment of ±0.005 seconds (5 milliseconds), which was negligible relative to the 10-millisecond sampling interval at 100 Hz and well below the temporal resolution required for distinguishing human activities, which typically exhibited movement frequencies below 20 Hz [27]. More sophisticated synchronization techniques, such as linear interpolation-based resampling or cross-correlation alignment, were considered but were deemed unnecessary given the minimal temporal error introduced by the rounding method. The inner merge operation further ensured data integrity by discarding any samples lacking corresponding measurements across all three sensor channels.

Continuous time series were segmented using a sliding window approach with a window size of L = 150 samples (approximately 1.5 seconds at 100 Hz) and a stride of 100 samples, yielding partial overlap between consecutive windows [27]. Following established HAR methodology [10, 22], six statistical features were computed for each of the nine sensor channels: mean, standard deviation, maximum, minimum, range, and signal energy. This process yielded 54 features per sensor position, or 108 features for the combined hand–ankle analysis. These six time-domain statistical features were selected based on their demonstrated effectiveness in the HAR literature [10, 22] and their computational efficiency for potential real-time deployment. Mean and standard deviation captured the central tendency and variability of sensor signals, while maximum, minimum, and range characterized the amplitude dynamics of each activity. Signal energy provided a global measure of signal intensity and was shown to be particularly discriminative for distinguishing static from dynamic activities [6]. Frequency-domain features, such as spectral entropy and dominant frequency, as well as more advanced features including autocorrelation and signal magnitude area, were not included in the current benchmarking framework because the primary objective was to establish baseline performance using a compact and well-understood feature set. Future work will investigate the impact of expanded feature sets, including frequency-domain and information-theoretic features, on classification performance.

3.5. Classification framework.

Three representative classifiers were evaluated to establish baseline benchmarks. Random Forest was implemented as an ensemble learning method constructing 100 decision trees through bootstrap aggregation with random feature selection [28]. XGBoost was configured as a gradient boosting framework with 100 estimators, a maximum depth of 6, and a learning rate of 0.1 [29]. Linear Support Vector Classifier (LinearSVC) was implemented as a linear classification model with standard scaling preprocessing and served as a baseline for assessing non-linear modeling requirements. The dataset was partitioned into training (80%) and testing (20%) subsets using stratified random sampling, and performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. Hyperparameter values for Random Forest and XGBoost were selected based on commonly adopted default configurations in the HAR and general machine learning literature [22, 28, 29]. Specifically, 100 trees or estimators represented a standard ensemble size that balanced computational cost and classification performance, while the maximum depth of 6 and learning rate of 0.1 for XGBoost followed recommended default settings from the original documentation [29]. No systematic hyperparameter tuning, such as grid search or Bayesian optimization, was performed because the primary objective of this study was to provide reproducible baseline benchmarks rather than maximally optimized performance. Future work will investigate the effect of hyperparameter optimization on classification accuracy.

3.6. Rationale for classical machine learning focus.

This study deliberately focused on classical machine learning classifiers rather than deep learning models for three reasons. First, the relatively small participant cohort of six subjects might not have provided sufficient data diversity to effectively train deep learning architectures, which typically required substantially larger datasets to generalize well and avoid overfitting [2, 18]. Second, classical methods offered greater interpretability, enabling clearer analysis of which features and sensor configurations contributed most to classification performance, which was a key consideration for understanding the value of the multi-limb sensing approach. Third, lightweight classical models were better suited for resource-constrained edge deployment scenarios, such as real-time HAR on smartphones, where computational efficiency was paramount [24, 25]. The inclusion of deep learning benchmarking was planned as future work once the participant cohort was expanded to provide sufficient training data.

4. Results and Discussion

4.1. Overall classification performance.

Table 2 presents the overall classification performance on the test set, which comprised 11,418 segmented windows across 12 activity–position combinations. Random Forest achieved the highest classification accuracy of 96.13%, substantially outperforming XGBoost, which achieved 85.22%, and LinearSVC, which reached 58.54%.

Table 2. Overall classification performance.

Model	Accuracy	Macro F1-Score
Random Forest	96.13%	0.9598
XGBoost	85.22%	0.8467
LinearSVC	58.54%	0.5712

4.2. Activity-Specific Performance Analysis.

Random Forest demonstrated consistently high performance across all activity–position combinations. Static activities (lying, sitting, and standing) achieved near-perfect classification, with F1-scores ranging from 0.98 to 1.00. Dynamic locomotion activities also maintained high performance: walking achieved F1-scores of 0.96–0.97, while stair-related activities exhibited marginally lower but still robust performance, with F1-scores of 0.92–0.94. Confusion matrix analysis revealed that limited misclassifications occurred predominantly between stair ascent and descent activities, which was consistent with their shared biomechanical characteristics. The error rate remained below 2% for all classes, demonstrating that Random Forest effectively discriminated among activities exhibiting similar temporal signatures.

XGBoost achieved strong performance on activities with consistent rhythmic patterns, with walking F1-scores ranging from 0.95 to 1.00, but exhibited degraded performance for stair-related activities, with F1-scores of 0.71–0.79. This performance degradation was attributed to several factors. Stair ascent and descent activities exhibited less consistent rhythmicity compared to level walking, with greater variability in step duration, acceleration magnitude, and angular velocity patterns across participants and trials. The sequential boosting mechanism in XGBoost iteratively corrected residual errors; however, when the underlying signal patterns were highly variable and overlapping between classes, as in stair ascent versus descent, this iterative correction could lead to overfitting on training-specific patterns rather than learning generalizable discriminative features. In contrast, the bagging approach in Random Forest, based on independent trees, provided more robust generalization by averaging across diverse decision boundaries, making it less sensitive to intra-class variability in stair activities. LinearSVC demonstrated substantial limitations across all activity classes, with error rates exceeding 30% for dynamic activities.

To further elucidate the contribution of the multi-limb sensing configuration, supplementary analysis was conducted to compare classification performance using individual sensor positions versus the combined dual-location approach. When only hand-position data were used, Random Forest achieved an accuracy of 90.47%, while ankle-only data yielded 92.81%. The combined hand–ankle configuration achieved 96.13%, representing a clear improvement of 3.32–5.66 percentage points over single-limb alternatives. This improvement was most pronounced for stair-related activities, where hand-position sensors captured arm swing dynamics that complemented the gait-phase information from ankle sensors. For static activities, both individual positions achieved near-perfect classification because the postural signatures were sufficiently distinct at either location. These results confirmed that the synchronized dual-limb approach provided complementary biomechanical information that enhanced activity discrimination, particularly for locomotion activities with similar lower-limb patterns but distinguishable upper-limb coordination.

4.3. Robustness to windowing parameters.

To assess model stability across different deployment configurations, additional experiments evaluated eight windowing parameter combinations with window sizes ranging from 1 to 4 seconds and overlaps ranging from 0% to 75%. Table 3 summarized the robustness results. A configuration was defined as stable if the classifier achieved a classification accuracy within 5 percentage points of its best observed accuracy across all tested windowing parameter combinations. The Robustness Rate was calculated as the proportion of stable configurations out of the total eight configurations tested, computed as the number of stable configurations divided by eight multiplied by 100%. This metric quantified a model's sensitivity to windowing parameter selection, with higher values indicating greater tolerance to suboptimal parameter choices.

Table 3. Model robustness to windowing variations.

Model	Stable Configurations	Robustness Rate
XGBoost	6/8	**75%**
Random Forest	4/8	50%
LinearSVC	2/8	25%

Interestingly, XGBoost demonstrated the highest robustness to windowing variations, achieving a robustness rate of 75%, which suggested its suitability for deployment scenarios in which precise parameter optimization was impractical. Random Forest exhibited moderate robustness at 50% while still achieving superior peak accuracy. These robustness findings carried important practical implications for real-world HAR system deployment. In production environments, windowing parameters often could not be precisely optimized due to varying hardware constraints, latency requirements, and computational budgets across target devices. The high robustness rate of XGBoost (75%) suggested that it might have been the preferred choice for cross-device deployment, where consistent performance across diverse configurations was more critical than peak accuracy. Conversely, Random Forest remained optimal for controlled deployment environments in which windowing parameters could be tuned to a specific use case. The substantial sensitivity of LinearSVC to windowing parameters, reflected in its 25% robustness rate, rendered it unsuitable for practical deployment without careful parameter selection.

4.4. Comparison with existing datasets and methods.

Table 4 presented a comparative analysis with established HAR benchmarks. The dataset uniquely addressed the gap in long-duration, multi-limb, open-access HAR resources. The 96.13% accuracy achieved by Random Forest on this dataset compared favorably with results reported on established benchmarks, demonstrating that the extended recording duration and multi-limb sensing approach yielded discriminative features suitable for high-accuracy classification.

Table 4. Comparison with established HAR datasets.

Dataset	Duration	Positions	Open Access	Best Acc.
UCI-HAR [10]	2.56s	Waist	Yes	96.0%
WISDM [15]	10s	Pocket	Yes	91.7%
PAMAP2 [16]	Variable	3 positions	Yes	89.0%
Ours	60s	Hand, Ankle	Yes	96.13%

4.5. Implications for HAR research.

The public release of this dataset addressed a critical need in the HAR research community. The 60-second recording duration enabled investigation of temporal dynamics and movement stability patterns that could not be studied using short-duration datasets. The synchronized hand–ankle configuration facilitated research on inter-limb coordination and biomechanical coupling during activities. The benchmarking results established baseline performance metrics that future researchers could use for comparative evaluation. The finding that classical ensemble methods achieved competitive accuracy relative to established benchmarks suggested that sophisticated deep learning approaches were not always necessary, particularly for applications with computational constraints.

5. Conclusions

This study presented a novel open-access HAR dataset and a systematic benchmarking of machine learning classifiers for multi-limb activity recognition. The principal contributions included the public release of the Dual-Location Smartphone IMU Human Activity Recognition Dataset via Mendeley Data (DOI: 10.17632/9d77352dcf.2), featuring 60-second recordings from synchronized hand and ankle positions across six daily activities; the demonstration that Random Forest achieved 96.13% classification accuracy on the multi-limb dataset, establishing competitive performance with established benchmarks; and evidence that non-linear ensemble methods substantially outperformed linear classifiers for multi-sensor HAR, while XGBoost exhibited superior robustness to windowing parameter variations. The dataset and benchmarking results provided valuable resources for the HAR research community by enabling reproducible experimentation and facilitating advancement in multi-limb activity recognition systems. Future work would pursue several specific directions. First, the participant cohort would be expanded from six to at least 30 subjects to improve generalizability and address potential overfitting; key challenges anticipated included managing data heterogeneity arising from diverse demographic characteristics such as age, height, weight, and gait patterns, as well as ensuring consistent data collection across a larger participant pool. Second, deep learning architectures, including CNN, LSTM, and hybrid CNN–LSTM models, would be benchmarked to evaluate whether the extended recording durations in the dataset provided additional performance gains for temporal deep learning models. Third, transfer learning approaches, specifically domain adaptation techniques such as adversarial domain adaptation and fine-tuning of pretrained feature extractors, would be investigated to enhance cross-subject generalization and address the known limitation that models trained on a small subject pool might not transfer effectively to unseen individuals with different movement characteristics.

Acknowledgments

The authors thanked all participants who contributed to data collection for this study. This research was supported by Institut Teknologi Sains Bandung, Indonesia.

Author Contribution

Conceptualization was conducted by Ade Kurniawan. Methodology was developed by Ade Kurniawan and Dadan Ramdan Hidayat. Data collection was carried out by Zain Iqbal Saputra, Whirdyana Shalfa Ayubi, Syifa Nurul Fajri R., M. Ragil Rizky Mulya, and Chello Fhrino Mike Mandolang. Data analysis was performed by Ade Kurniawan, Dadan Ramdan Hidayat, and Zain Iqbal Saputra. The original draft was written by Ade Kurniawan and Dadan Ramdan Hidayat. Review and editing were undertaken by Ade Kurniawan. Supervision was provided by Ade Kurniawan.

Competing Interest

The authors declare no competing interests.

Data Availability Statement

The dataset supporting this study is openly available in Mendeley Data at https://doi.org/10.17632/9d77352dcf.2 under CC BY 4.0 license.

References

[1] Lara, O.D.; Labrador, M.A. (2013). A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, *15*, 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192.

[2] Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. (2021). Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities. *ACM Computing Surveys*, *54*, 1–40. https://doi.org/10.1145/3447744.

[3] Demrozi, F.; Pravadelli, G.; Biber, A.; Slanzi, P.; De Turri, C. (2020). Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey. *IEEE Access*, *8*, 210816–210836. https://doi.org/10.1109/ACCESS.2020.3037715.

[4] Straczkiewicz, M.; Glynn, N.W.; Harezlak, J. (2021). A systematic review of smartphone-based human activity recognition methods for health research. *npj Digital Medicine*, *4*, 148. https://doi.org/10.1038/s41746-021-00514-4.

[5] Lima, W.S.; Souto, E.; El-Khatib, K.; Jalali, R.; Gama, J. (2019). Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview. *Sensors*, *19*, 3213. https://doi.org/10.3390/s19143213.

[6] Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. (2014). Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors*, *14*, 10146–10176. https://doi.org/10.3390/s140610146.

[7] Webber, M.; Fernandez Rojas, R. (2021). Human Activity Recognition With Accelerometer and Gyroscope: A Data Fusion Approach. *IEEE Sensors Journal*, *21*, 16979–16989. https://doi.org/10.1109/JSEN.2021.3079883.

[8] Chung, S.; Lim, J.; Noh, K.J.; Kim, G.; Jeong, H. (2019). Sensor Data Acquisition and Multimodal Sensor Fusion for Human Activity Recognition Using Deep Learning. *Sensors*, *19*, 1716. https://doi.org/10.3390/s19071716.

[9] Ramanujam, E.; Perumal, T.; Padmavathi, S. (2021). Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review. *IEEE Sensors Journal*, *21*, 13029–13040. https://doi.org/10.1109/JSEN.2021.3069927.

[10] Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. (2013). A Public Domain Dataset for Human Activity Recognition Using Smartphones. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, 437–442.

[11] Chavarriaga, R.; Sagha, H.; Calatroni, A.; Digumarti, S.; Tröster, G.; Millán, J.d.R.; Roggen, D. (2013). The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition. *Pattern Recognition Letters*, *34*, 2033–2042. https://doi.org/10.1016/j.patrec.2012.12.014 .

[12] Kunze, K.; Lukowicz, P. (2014). Sensor Placement Variations in Wearable Activity Recognition. *IEEE Pervasive Computing*, *13*, 32–41. https://doi.org/10.1109/MPRV.2014.73.

[13] Sztyler, T.; Stuckenschmidt, H. (2016). On-body Localization of Wearable Devices: An Investigation of Position-Aware Activity Recognition. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–9. https://doi.org/10.1109/PERCOM.2016.7456521.

[14] Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Cheng, D. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors*, *22*, 1476. https://doi.org/10.3390/s22041476.

[15] Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. (2011). Activity Recognition using Cell Phone Accelerometers. *ACM SIGKDD Explorations Newsletter*, *12*, 74–82. https://doi.org/10.1145/1964897.1964918.

[16] Reiss, A.; Stricker, D. (2012). Introducing a New Benchmarked Dataset for Activity Monitoring. In *Proceedings of the IEEE International Symposium on Wearable Computers (ISWC)*, 108–109. https://doi.org/10.1109/ISWC.2012.13.

[17] Ordóñez, F.J.; Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, *16*, 115. https://doi.org/10.3390/s16010115.

[18] Hammerla, N.Y.; Halloran, S.; Ploetz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1533–1540.

[19] Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. (2012). Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In *Lecture Notes in Computer Science*, *7657*, 216–223. https://doi.org/10.1007/978-3-642-35395-6_30.

[20] Vavoulas, G.; Chatzaki, C.; Malliotakis, T.; Pediaditis, M.; Tsiknakis, M. (2016). The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones. In *Proceedings of ICT4AWE*, 143–151. https://doi.org/10.5220/0005792401430151.

[21] Micucci, D.; Mobilio, M.; Napoletano, P. (2017). UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones. *Applied Sciences*, *7*, 1101. https://doi.org/10.3390/app7101101.

[22] Wang, Z.; Wu, D.; Chen, J.; Ghoneim, A.; Hossain, M.A. (2016). A Triaxial Accelerometer-Based Human Activity Recognition via EEMD-Based Features. *IEEE Sensors Journal*, *16*, 3198–3207. https://doi.org/10.1109/JSEN.2016.2545708.

[23] Catal, C.; Tufekci, S.; Pirmit, E.; Kocabag, G. (2015). On the Use of Ensemble of Classifiers for Accelerometer-based Activity Recognition. *Applied Soft Computing*, *37*, 1018–1022. https://doi.org/10.1016/j.asoc.2015.01.025.

[24] Ronao, C.A.; Cho, S.-B. (2016). Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks. *Expert Systems with Applications*, *59*, 235–244. https://doi.org/10.1016/j.eswa.2016.04.032.

[25] Zhu, R.; Xiao, Z.; Chen, L.; Chen, X.; Wu, L. (2019). Efficient Human Activity Recognition Solving the Confusing Activities Via Deep Ensemble Learning. *IEEE Access*, *7*, 75490–75499. https://doi.org/10.1109/ACCESS.2019.2922104.

[26] Staacks, S.; Lütz, S.; Huëtz, M.; Stampfer, C. (2018). Advanced Tools for Smartphone-based Experiments: phyphox. *Physics Education*, *53*, 045009. https://doi.org/10.1088/1361-6552/aac05e.

[27] Banos, O.; Galvez, J.-M.; Damas, M.; Pomares, H.; Rojas, I. (2014). Window Size Impact in Human Activity Recognition. *Sensors*, *14*, 6474–6499. https://doi.org/10.3390/s140406474.

[28] Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

[29] Chen, T.; Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785.