Trends in Health Informatics

www.thi.reapress.com

THI. Vol. 3, No. 1 (2026) 1–11.

Paper Type: Original Article

An Interpretable Unsupervised Framework for Exercise Pattern Discovery Using Wearable Sensor Data

Pooya Zeinali[1], **Amanna Ghanbari Talouki**[2,*]

[1] Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran; pooyazeinali@aut.ac.ir.

[2] Department of Technical and Engineering, Ayandegan University, Tonekabon, Iran; ghanbari@aihe.ac.ir.

Citation:

Received: 11 May 2025 *Revised: 05 August 2025* *Accepted: 24 October 2025*	Zeinali, P., & Ghanbari Talouki, A. (2026). An interpretable unsupervised framework for exercise pattern discovery using wearable sensor data. *Trends in Health Informatics, 3*(1), 1-11.

Abstract

The rapid growth of wearable sensing technologies has enabled continuous monitoring of physiological signals during physical activity, providing new opportunities for data-driven exercise analysis. This study uses the publicly available wearable sports health monitoring dataset, consisting of 500 records collected from athletes during various activities, to identify latent exercise patterns based on physiological responses. An unsupervised clustering framework is proposed to categorize athletes into three exercise groups: short-duration high-intensity activity, long-duration low-intensity activity, and moderate activity. A compact and interpretable set of physiological features (Heart Rate (HR), step count, Systolic Blood Pressure (SBP), and blood oxygen saturation (SpO_2)) was selected to reduce model complexity while preserving physiological relevance. Experimental results show that the identified clusters exhibit distinct and realistic physiological profiles, with average HR and step count differences of up to 18% and 70%, respectively, between high-intensity and long-duration exercise groups. Furthermore, activity status distributions across clusters remain balanced, indicating that grouping is driven by physiological workload rather than predefined activity labels. Overall, the proposed approach provides an interpretable and label-free framework for exercise pattern discovery using wearable data, with potential applications in personalized training design, performance monitoring, and health assessment.

Keywords: Unsupervised learning, Exercise pattern discovery, Wearable sensor data, Interpretable machine learning, Physiological signals.

1 | Introduction

The widespread adoption of wearable devices has significantly transformed health monitoring and sports analytics by enabling continuous and non-invasive collection of physiological data. Sensors embedded in

Corresponding Author: ghanbari@aihe.ac.ir

https://doi.org/10.22105/thi.vi.40

smartwatches, fitness trackers, and medical wearables can record Heart Rate (HR), movement patterns, blood oxygen saturation (SpO_2), and other vital signals in real time. These data streams provide valuable insights into an individual's physical condition, training intensity, and overall health status [1], [2].

In the context of sports and exercise science, understanding exercise intensity and duration is essential for optimizing performance, preventing injury, and designing personalized training programs. Traditional approaches often rely on predefined activity labels or subjective measures such as self-reported exertion levels. However, these methods suffer from several limitations, including labeling inaccuracies, inter-individual variability, and the inability to capture physiological responses that differ across athletes performing the same activity [3].

Recent advances in machine learning have enabled data-driven analysis of wearable sensor data, particularly through supervised classification techniques that attempt to map physiological signals to predefined activity types. While effective in controlled settings, supervised methods require large, accurately labeled datasets and tend to focus on recognizing activity names (e.g., walking or running) rather than underlying physiological workload. In real-world scenarios, the same activity can induce vastly different physiological responses depending on intensity, duration, fitness level, and environmental conditions [4].

Unsupervised learning provides a promising alternative by discovering latent patterns directly from physiological measurements without relying on activity labels. Clustering techniques, in particular, allow the identification of natural groupings in data that may correspond to meaningful exercise patterns such as low-intensity endurance activity or short-duration high-intensity exertion. Such approaches are especially valuable in wearable-based health monitoring, where labeled data are scarce or unreliable [5], [6].

Another challenge in wearable data analysis is model interpretability. Many studies employ high-dimensional feature sets or dimensionality reduction techniques such as Principal Component Analysis (PCA) to improve performance. While effective in certain contexts, these approaches may reduce physiological interpretability and complicate clinical or sports-related insights. Selecting a compact set of meaningful physiological features can reduce model complexity while preserving explanatory power [7].

Motivated by these challenges, this study proposes an interpretable unsupervised framework for categorizing exercise patterns using wearable sensor data. Athletes are grouped into three exercise categories (short-duration high-intensity activity, long-duration low-intensity activity, and moderate activity) based solely on physiological responses. A limited set of key physiological features is selected to maintain interpretability and reduce complexity. Clustering results are validated using activity status information that is not included during model training. Additionally, PCA is applied only for comparative visualization to assess whether dimensionality reduction enhances cluster separation.

The remainder of this paper is organized as follows. Section 2 reviews related work on wearable-based exercise analysis and unsupervised learning approaches. Section 3 describes the dataset, feature selection process, and clustering methodology. Section 4 presents experimental results and discussions. Section 5 outlines limitations and future research directions, and Section 6 concludes the paper.

2 | Related Works

The increasing availability of wearable sensor data has motivated extensive research in the fields of activity recognition, exercise intensity estimation, and health monitoring. Early studies primarily focused on supervised activity recognition, where physiological and motion-based features were used to classify predefined activity types such as walking, running, or cycling. These approaches demonstrated high accuracy in controlled environments but relied heavily on large, well-labeled datasets and often failed to generalize to real-world conditions [8], [9].

To address labeling challenges, several researchers explored machine learning techniques for estimating exercise intensity rather than activity type. HR–based models and Metabolic Equivalents (METs) have been widely used to quantify physical workload during exercise. While these methods provide useful approximations, they often assume population-level relationships and may not accurately reflect individual physiological responses, especially across different fitness levels [10], [11].

Unsupervised learning approaches have gained increasing attention as a means of discovering latent exercise patterns without relying on explicit labels. Clustering algorithms such as k-means, hierarchical clustering, and density-based methods have been applied to wearable sensor data to identify groups of similar physiological responses. These studies demonstrated that unsupervised methods could reveal meaningful exercise states and transitions that are not captured by traditional activity labels [12], [13].

Recent advances in unsupervised representation learning for time-series sensor data have increasingly adopted contrastive learning frameworks, which leverage carefully designed temporal augmentations and instance discrimination objectives to learn robust and invariant latent representations without requiring labeled supervision. However, these methods primarily optimize for representation quality and downstream predictive performance, while offering limited insight into interpretable and semantically meaningful activity pattern discovery [14].

Dimensionality reduction techniques, particularly PCA, are commonly employed in wearable data analysis to address high dimensionality and redundancy among features. Several studies have reported improved classification or clustering performance after applying PCA or similar transformations [15], [16]. However, the resulting components are often difficult to interpret physiologically, limiting their usefulness in clinical and sports science applications where explainability is essential.

Recent studies have further investigated weakly and self-supervised learning strategies for wearable sensor-based human activity recognition, focusing on reducing annotation requirements while maintaining competitive performance across standard recognition benchmarks. However, these approaches remain primarily centered on improving recognition accuracy under limited supervision, without explicitly addressing fully unsupervised and interpretable discovery of latent activity patterns [17].

More recent research has emphasized the importance of model interpretability and feature selection in health-related machine learning. Studies have shown that using a compact set of physiologically meaningful features can achieve comparable performance to high-dimensional models while improving transparency and trustworthiness [18], [19]. This shift aligns with the growing demand for Explainable Artificial Intelligence (XAI) in medical and wearable-based analytics.

Recent works have further advanced unsupervised representation learning for time-series sensor data through contrastive learning frameworks, which exploit temporal augmentations and instance-level discrimination to learn robust and invariant feature representations without labeled supervision. However, these approaches are primarily designed to enhance representation quality for downstream tasks such as classification or clustering, rather than enabling interpretable and semantically grounded pattern discovery [20].

Recent advances in wearable-based human activity recognition have increasingly leveraged large-scale self-supervised learning to overcome the limitations imposed by scarce labeled data. Notably, recent work has demonstrated that training on massive unlabeled datasets, spanning hundreds of thousands of person-days, can significantly enhance the generalizability and robustness of learned representations across diverse populations, devices, and activity domains. Despite these promising developments, such approaches primarily focus on improving downstream recognition performance, with limited attention to interpretable and fully unsupervised pattern discovery [21].

Recent advances in wearable sensor-based human activity recognition have increasingly shifted toward self-supervised and unsupervised representation learning to mitigate the reliance on large-scale labeled datasets. In particular, learning meaningful latent structures from raw time-series data has emerged as a promising direction for capturing intrinsic activity patterns. Within this line of research, recent studies have explored the use of self-supervised frameworks to learn compact and discrete latent representations, enabling more structured and semantically meaningful modeling of human activities without manual annotations. Despite these advances, existing studies often focus on either activity recognition or intensity estimation in isolation, and few works explicitly validate unsupervised exercise clusters using independent activity information. Moreover, the necessity of dimensionality reduction in low-dimensional, interpretable physiological feature spaces remains insufficiently explored [22].

In parallel, recent studies have leveraged hybrid deep architectures combining convolutional and transformer-based models within self-supervised frameworks to learn rich temporal representations from wearable sensor data, achieving strong recognition performance with limited labeled data. However, these approaches predominantly focus on optimizing predictive accuracy through complex model architectures, with limited attention to interpretable and structurally meaningful pattern discovery [23].

In addition, recent studies have explored semi-supervised activity detection frameworks that operate directly on wearable devices, emphasizing personalized modeling and privacy preservation while leveraging sparsely labeled data. However, such approaches typically remain focused on classification-oriented objectives and require partial supervision, limiting their applicability to fully unsupervised and interpretable pattern discovery settings [24].

More closely related to unsupervised pattern discovery, recent studies have investigated hybrid deep clustering frameworks that combine representation learning with clustering mechanisms to uncover latent activity structures from wearable sensor data, demonstrating improved capability in identifying complex and transitional activity patterns without labeled supervision. Nevertheless, these approaches often prioritize clustering performance and latent separability, with limited emphasis on interpretability and validation using independent physiological or activity-based evidence [25].

Recent research has systematically explored the trade-offs across supervised, unsupervised, and weakly/self-supervised learning paradigms in wearable-based human activity recognition, emphasizing that reducing reliance on labeled data remains a fundamental challenge, as unsupervised approaches often lack discriminative power while supervised methods demand costly annotations. This tension has motivated the development of hybrid and weakly supervised frameworks; however, the problem of fully unsupervised and interpretable pattern discovery remains largely unresolved [26].

In contrast to prior work, the present study adopts an unsupervised clustering framework based on a small set of key physiological features to categorize exercise patterns by intensity and duration. The proposed approach emphasizes interpretability, avoids reliance on activity labels during training, and includes an explicit validation step using withheld activity status information. Additionally, PCA is evaluated only as a comparative tool, providing insight into its practical necessity in this context.

3 | Methodology

3.1 | Dataset Description

This study utilizes the wearable sports health monitoring dataset, a publicly available dataset hosted on Kaggle that comprises physiological and activity-related measurements collected from athletes wearing health monitoring devices during various physical activities (e.g., walking, running, cycling) [27].

Each record in the dataset represents an individual observation captured at a specific timestamp and includes a total of 12 attributes. These attributes consist of: record ID, athlete ID, timestamp, HR, body temperature, blood pressure, SpO_2, step count, activity status, latitude, longitude, and secure transmission status.

To ensure data consistency and suitability for the present unsupervised analysis, records with missing or incomplete values were excluded. The final curated dataset consisted of 500 complete records. Activity status labels and contextual variables were retained solely for post-hoc validation and interpretation. They were not used during the clustering process, ensuring an unbiased evaluation of the model’s ability to identify latent physiological exercise signatures.

3.2|Feature Selection and Preprocessing

A key objective of this study is to maintain physiological interpretability while minimizing model complexity. Although the dataset provides 12 attributes, not all variables are directly relevant to exercise workload or physiological response. Therefore, a feature selection step was performed to identify a compact and meaningful subset of physiological parameters.

Based on established principles in exercise physiology, four core physiological features were selected for clustering: HR, step count, Systolic Blood Pressure (SBP), and SpO_2. HR serves as a primary indicator of cardiovascular response and exercise intensity, while step count reflects movement volume and exercise duration. SBP provides insight into cardiovascular workload, and SpO_2 represents oxygen utilization efficiency during physical exertion.

Other variables, including body temperature, diastolic blood pressure, geographic coordinates, timestamp, and secure transmission status, were excluded as they either introduce redundancy, lack direct relevance to exercise intensity, or primarily serve contextual or administrative purposes rather than physiological interpretation. Blood pressure values were parsed to extract systolic measurements prior to analysis.

Before clustering, all selected features were standardized using z-score normalization to eliminate scale disparities and ensure equal contribution during distance-based clustering.

3.3|Clustering Approach

3.3.1|K-means clustering algorithm

K-means is a widely used unsupervised clustering algorithm that aims to partition a dataset into K disjoint clusters by minimizing intra-cluster variance. Given a set of N data points $\{x_1, x_2, \ldots, x_n\}$, where each data point is represented in a d-dimensional feature space, the objective of K-means is to minimize the following cost function:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i}^{||x-\mu_i||^2}, \tag{1}$$

where

Ci denotes the i−th cluster, and μi represents the centroid of that cluster, computed as the mean of all data points assigned to Ci.

The algorithm iteratively alternates between two steps: 1) assignment of each data point to the nearest cluster centroid based on Euclidean distance, and 2) update of cluster centroids as the mean of assigned data points. This process continues until convergence, typically when cluster assignments no longer change, or the reduction in the cost function becomes negligible.

3.3.2|Application of k-means in this study

In this study, K-means clustering was applied to standardized physiological features to identify latent exercise patterns based solely on physiological responses. The number of clusters was set to K = 3, corresponding to three interpretable exercise categories commonly used in sports science: short-duration high-intensity exercise, long-duration low-intensity exercise, and moderate-intensity exercise.

Clustering was performed exclusively on four standardized physiological features: HR, step count, SBP, and SpO_2. Activity status labels were deliberately excluded during clustering to ensure that cluster formation was driven purely by physiological workload rather than predefined activity types.

The resulting clusters were visualized in the physiological feature space, as illustrated in *Fig. 1*, which demonstrates clear separation between exercise groups based on their physiological characteristics.

3.4 | Cluster Interpretation and Labeling

To interpret the physiological meaning of each cluster, the mean values of the selected features were computed for each cluster. These cluster-level statistics enabled the characterization of dominant exercise patterns in terms of intensity and duration.

Clusters exhibiting elevated HR and SBP combined with moderate step counts were interpreted as short-duration high-intensity exercise. Conversely, clusters characterized by high step counts and moderate cardiovascular response were associated with long-duration low-intensity exercise. The remaining cluster, displaying intermediate values across all features, was labeled as moderate-intensity exercise.

This post-hoc interpretation allowed the assignment of meaningful exercise labels while preserving the unsupervised nature of the clustering process.

3.5 | Validation Using Activity Status

To evaluate the validity of the inferred exercise clusters, activity status information, such as walking, running, cycling, and resting, was analyzed after clustering. A cross-tabulation and percentage-based analysis were conducted to examine the distribution of activity types within each exercise category.

Since activity status was not used during the clustering process, a balanced and physiologically plausible distribution of activities across clusters indicates that the model captures underlying physiological workload rather than merely reflecting activity labels. This analysis provides additional evidence that the proposed clustering approach successfully identifies meaningful exercise patterns from wearable sensor data.

3.6 | PCA for Comparative Analysis

PCA was applied solely for visualization and comparative purposes. The standardized physiological features were projected onto a two-dimensional space to evaluate whether dimensionality reduction improved cluster separation.

Results showed that PCA did not significantly enhance separability compared to the original feature space. Given the low dimensionality and high interpretability of the selected features, PCA was excluded from the final clustering pipeline to preserve physiological meaning.

4 | Results

4.1 | Relationship Between Heart Rate and Systolic Blood Pressure

The relationship between HR and SBP is illustrated in *Fig. 1*. As expected, a positive correlation is observed between these two physiological parameters, reflecting increased cardiovascular workload during higher-intensity exercise. Clusters associated with short-duration high-intensity activity exhibit elevated HR and SBP values, whereas moderate and long-duration exercise clusters demonstrate lower or more stable cardiovascular responses.

This observation is consistent with established exercise physiology principles, where increased cardiac output and arterial pressure accompany higher exercise intensity. The clear separation of clusters along these physiological dimensions further supports the validity of the proposed clustering approach.

4.2 | Relationship Between Step Count and Heart Rate

Fig. 2 illustrates the relationship between step count and HR for the identified exercise clusters. Step count primarily reflects movement volume and exercise duration, while HR represents exercise intensity. The long-duration exercise cluster is characterized by high step counts.

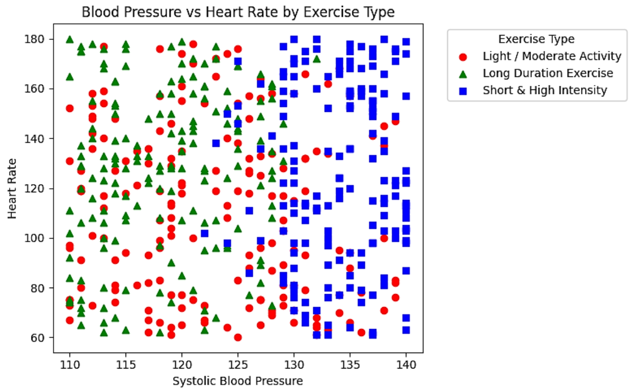

Fig. 1. Scatter plot showing the relationship between HR and SBP across clustered exercise patterns.

Combined with moderate HR values, indicating sustained but lower-intensity activity.

In contrast, the short-duration high-intensity cluster exhibits elevated HRs with comparatively lower step counts, reflecting intense but brief exercise sessions. The moderate activity cluster occupies an intermediate region, demonstrating balanced movement volume and cardiovascular response. This separation highlights the ability of the model to distinguish between intensity-driven and duration-driven exercise patterns.

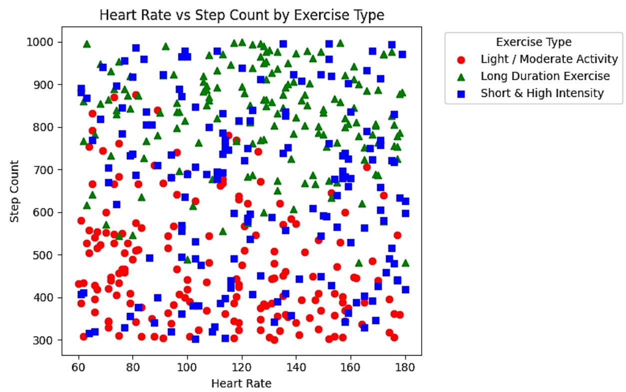

Fig. 2. Scatter plot of step count versus HR colored by cluster assignment.

4.3 | Distribution of Activity Status Within Exercise Types

Table 1 presents the percentage distribution of activity status within each inferred exercise type. Notably, no exercise category is dominated by a single activity label. For example, the light/moderate activity cluster includes a balanced mix of walking, running, cycling, and resting states, while the short-duration high-intensity cluster exhibits a nearly uniform distribution across activity types.

This result is particularly important because activity status information was not used during clustering. The balanced distributions indicate that the model identifies exercise patterns based on physiological response rather than activity names. Such behavior confirms the robustness of the clustering process and demonstrates that similar activities can induce different physiological workloads depending on the intensity and duration.

Table 1. Percentage distribution of activity status across inferred exercise types.

Activity Status/ExerciseType	Cycling	Resting	Running	Walking
Light / Moderate Activity	21.97%	22.54%	24.86%	30.64%
Long Duration Exercise	30.06%	26.38%	20.25%	23.31%
Short & High Intensity	26.83%	21.34%	25.00%	26.83%

4.4 | PCA-Based Visualization and Comparison

Fig. 3 shows the projection of standardized physiological features onto the first two principal components. While cluster structures remain observable, PCA does not significantly enhance separation compared to the original feature space. Furthermore, the principal components lack direct physiological interpretability, as they represent linear combinations of multiple features.

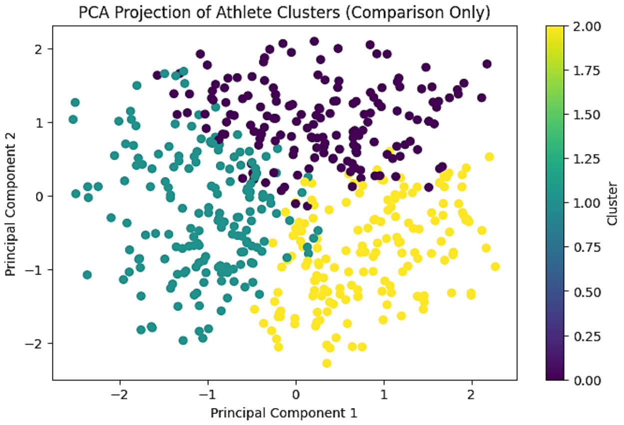

Fig. 3. Two-dimensional PCA projection of clustered physiological data.

Given the limited number of selected features and their clear physiological meaning, dimensionality reduction did not provide additional analytical benefits. This observation supports the decision to exclude PCA from the final clustering pipeline and reinforces the importance of interpretability in wearable-based health analytics.

5 | Discussion

Across all analyses, the proposed unsupervised framework successfully identifies meaningful exercise patterns that align with established physiological principles. The results demonstrate that clustering based on a small set of interpretable features can effectively differentiate exercise intensity and duration without relying on activity labels. Validation using independent activity status information and comparative PCA analysis further confirms the robustness and practicality of the proposed approach.

6 | Limitations and Future Work

6.1 | Limitations

Despite the promising results, this study has several limitations that should be acknowledged. First, the dataset represents a limited number of athletes and measurements, which may restrict the generalizability of the findings to broader populations with different fitness levels, age groups, or health conditions.

Second, the analysis is based on snapshot-level physiological measurements, and temporal dependencies between consecutive records were not explicitly modeled. Exercise intensity and physiological response often evolve over time, and ignoring temporal dynamics may limit the ability to capture transitions between exercise states.

Third, only a small set of physiological features was considered to maintain interpretability. While this design choice improves transparency, it may omit additional signals such as HR variability, respiration rate, or energy expenditure that could further enhance exercise characterization.

Finally, the clustering process relied on a fixed number of clusters. Although the selected number aligns with common exercise intensity categories, alternative cluster structures may exist and were not explored in this study.

6.2|Future Work

Future research can address these limitations in several directions. Incorporating temporal modeling techniques, such as Hidden Markov models or recurrent neural networks, could enable the detection of dynamic exercise patterns and transitions over time. Expanding the feature set to include additional physiological signals, while preserving interpretability, may further improve clustering robustness.

Moreover, adaptive or data-driven methods for determining the optimal number of clusters could be explored to capture more nuanced exercise patterns. Validation on larger and more diverse datasets, including athletes with different training backgrounds and health conditions, would also strengthen the applicability of the proposed framework.

7|Conclusion

This study presented an interpretable unsupervised learning framework for analyzing exercise patterns using wearable sensor data. By focusing on a compact set of physiologically meaningful features, the proposed approach successfully categorized athletes into three exercise groups based on intensity and duration without relying on predefined activity labels.

The clustering results demonstrated strong alignment with established physiological principles and were further validated using independent activity status information. Comparative analysis using PCA confirmed that dimensionality reduction did not enhance cluster separability, supporting the decision to preserve the original feature space for improved interpretability.

Overall, the findings highlight the effectiveness of unsupervised clustering in capturing latent exercise behaviors from wearable data while maintaining transparency and practical relevance. The proposed framework offers a scalable and data-driven foundation for personalized exercise analysis and wearable-based health monitoring applications.

Finally, integrating the clustering results into real-time wearable systems could enable personalized feedback and adaptive training recommendations, highlighting the practical potential of unsupervised learning in sports and health monitoring.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

All data are included in the text.

Conflict of Interest

The authors, being non-native English speakers, used AI tools to refine the English language of this manuscript.

References

[1] Dunn, J., Runge, R., & Snyder, M. (2018). Wearables and the medical revolution. *Personalized Medicine*, *15*(5), 429–448. https://doi.org/10.2217/pme-2018-0044

[2] Kristoffersson, A., & Lindén, M. (2022). A systematic review of wearable sensors for monitoring physical activity. *Sensors*, *22*(2), 573. https://doi.org/10.3390/s22020573

[3] Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: Recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise*, *44*(1 Suppl 1), S1. https://doi.org/10.1249/MSS.0b013e3182399b7e

[4] Li, F., Shirahama, K., Nisar, M. A., Köping, L., & Grzegorzek, M. (2018). Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, *18*(2), 679. https://doi.org/10.3390/s18020679

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer series in statistics New-York. https://doi.org/10.1007/978-0-387-84858-7

[6] Ige, A. O., & Noor, M. H. M. (2022). A survey on unsupervised learning for wearable sensor-based activity recognition. *Applied Soft Computing*, *127*, 109363. https://doi.org/10.1016/j.asoc.2022.109363

[7] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

[8] Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, *46*(3), 1–33. http://dx.doi.org/10.1145/2499621

[9] Wang, L., Gu, T., Tao, X., Chen, H., & Lu, J. (2011). Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive and Mobile Computing*, *7*(3), 287–298. https://doi.org/10.1016/j.pmcj.2010.11.008

[10] Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., … & Leon, A. S. (2011). 2011 compendium of physical activities: A second update of codes and MET values. *Medicine & Science in Sports & Exercise*, *43*(8), 1575–1581. https://doi.org/10.1249/MSS.0b013e31821ece12

[11] Silva, P., Dos Santos, E., Grishin, M., & Rocha, J. M. (2018). Validity of heart rate-based indices to measure training load and intensity in elite football players. *The Journal of Strength & Conditioning Research*, *32*(8), 2340–2347. https://doi.org/10.1519/JSC.0000000000002057

[12] Pham, S., Yeap, D., Escalera, G., Basu, R., Wu, X., Kenyon, N. J., … & Davis, C. E. (2020). Wearable sensor system to monitor physical activity and the physiological effects of heat exposure. *Sensors*, *20*(3), 855. https://doi.org/10.3390/s20030855

[13] Yan, Z., Subbaraju, V., Chakraborty, D., Misra, A., & Aberer, K. (2012). Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. *2012 16th International Symposium on Wearable Computers* (pp. 17–24). IEEE. https://doi.org/10.1109/ISWC.2012.23

[14] Tokas, P., Semwal, V. B., & Jain, S. (2025). A lightweight and explainable hybrid deep learning model for wearable sensor-based human activity recognition. *IEEE Sensors Journal*. https://doi.org/10.1109/JSEN.2025.3564045

[15] Van Der Maaten, L., Postma, E. O., & Van Den Herik, H. J. (2009). Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, *10*(1), 1–41. https://www.researchgate.net/publication/228657549_Dimensionality_Reduction_A_Comparative_Review?enrichId=rgreq-e4bc0ac4748b5adc9865c909d243caf9-XXX&enrichSource=Y292ZXJQYWdlOzIyODY1NzU0OTtBUzoxMTQ5MDI5ODMyNTQwMTZAMTQwNDQwNjQxNTExMg%3D%3D&el=1_x_2&_esc=publicationCoverPdf

[16] Thiemjarus, S., & Yang, G. Z. (2014). Context aware sensing. In *Body sensor networks* (pp. 355–404). Springer. https://doi.org/10.1007/978-1-4471-6374-9_9

[17] Olivia, S. (2025). Weakly and self-supervised learning strategies for human activity recognition using wearable sensor data. *International Journal of Engineering Technology Research & Management (IJETRM)*, *9*(12), 899–915. https://ijetrm.com/issues/files/Dec-2025-30-1767102638-DEC90.pdf

[18] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[19] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312. https://doi.org/10.1002/widm.1312

[20] Zhang, W. (2025). Robust self-supervised representation learning for multivariate industrial time series under sensor drift. *American Journal of Big Data*, *6*(05), 233–239. https://doi.org/10.71465/ajbd3454

[21] Yuan, H., Chan, S., Creagh, A. P., Tong, C., Acquah, A., Clifton, D. A., & Doherty, A. (2024). Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *NPJ Digital Medicine*, *7*(1), 91. https://doi.org/10.48550/arXiv.2206.02909

[22] Haresamudram, H., Essa, I., & Ploetz, T. (2024). Towards learning discrete representations via self-supervision for wearables-based human activity recognition. *Sensors*, *24*(4), 1238. https://doi.org/10.3390/s24041238

[23] Sun, Y., Xu, X., Tian, X., Zhou, L., & Li, Y. (2024). Efficient human activity recognition: A deep convolutional transformer-based contrastive self-supervised approach using wearable sensors. *Engineering Applications of Artificial Intelligence*, *135*, 108705. https://doi.org/10.1016/j.engappai.2024.108705

[24] Roy, A., Dutta, H., Bhuyan, A. K., & Biswas, S. (2024). On-device semi-supervised activity detection: A new privacy-aware personalized health monitoring approach. *Sensors*, *24*(14), 4444. https://doi.org/10.3390/s24144444

[25] Yang, W., Sui, Y., Zhang, Y., & Xia, S. (2025). *Unsupervised deep clustering for human behavior understanding* [presentation]. Proceedings of the 3rd international workshop on human-centered sensing, modeling, and intelligent systems (pp. 1–6). https://api.semanticscholar.org/CorpusID:278316494

[26] Sheng, T., & Huber, M. (2025). Reducing label dependency in human activity recognition with wearables: From supervised learning to novel weakly self-supervised approaches. *Sensors*, *25*(13), 4032. https://doi.org/10.3390/s25134032

[27] Ziya. (2023). *Wearable sports health monitoring dataset*. https://www.kaggle.com/datasets/ziya07/wearable-sports-health-monitoring-dataset