Improving Task-Incremental Human Activity Recognition with Plasticity Techniques

F. G. Brabes da S.[1,2] and W. G. S. Lima[1,2] and A. Anjos da S.[1,2] and E. L. Colombini[1,2] and C. G. Ghedini[1,2]

[1] UNICAMP - Universidade Estadual de Campinas, Campinas SP, Brazil
[2] H.IAAC - Hub de Inteligência Artificial e Arquiteturas Cognitivas, Campinas SP, Brazil
{f247085, w195512}@dac.unicamp.br, {ansoanjos, estherlc, cghedini}@unicamp.br

Abstract. Sensor-based Human Activity Recognition (HAR) has been applied across various domains, including healthcare monitoring, fitness tracking, and smart home systems. These applications require the ability to accurately detect and respond to a wide range of human activities, each with varying distributions, which imposes a significant challenge. The task-incremental learning paradigm can address this problem by enabling HAR systems to adapt to changes in distribution and learn new activities over time without forgetting the previously learned ones. Continual adaptation is essential for maintaining high performance, as it allows the system to effectively respond to these changes. Although several strategies in the continual learning literature have been evaluated for task-incremental scenarios in HAR, there is still room for improvement, as the results are not as good as those achieved with conventional approaches. This work proposes two new neuroplasticity-inspired techniques that can be integrated with any learning strategy. Inspired by the brain's ability to reorganize and strengthen connections over time, these methods focus on enhancing the model's flexibility and long-term knowledge retention. The proposed techniques were evaluated alongside the WA-ADB and WA-MDF strategies on well-known HAR datasets. Experimental results demonstrated that the new techniques significantly enhanced the models' ability to retain knowledge, which holds significant potential for improving the robustness and longevity of HAR systems in real-world applications.

CCS Concepts: • **Computing methodologies** → *Supervised learning by classification*; **Lifelong machine learning**; *Neural networks*; **Online learning settings**.

Keywords: continual learning, human activity recognition (HAR), neuroplasticity, task-incremental learning

1. INTRODUCTION

In recent years, Human Activity Recognition (HAR) has emerged as a prominent research focus, driven by advancements in sensing technologies and the proliferation of sensor-equipped devices. These systems identify human activities by analyzing data collected from various sensors embedded within these devices, allowing for a range of applications.

The primary challenges in HAR include ensuring consistent performance across diverse distributions for the same activity and recognizing new activities in real-time. Traditional machine learning models face difficulties when activities are updated: they either require complete retraining when new activities are introduced or risk degrading performance on previously well-classified activities due to biases in the new data.

To address these challenges, task-incremental learning scenarios can be introduced. This approach encompasses strategies that enable systems to incrementally learn new tasks while retaining previously acquired knowledge. It involves presenting processed data from a sequence of class groups forming a task. Figure 1 illustrates the structure of a sequence of class groups, each forming a distinct task.

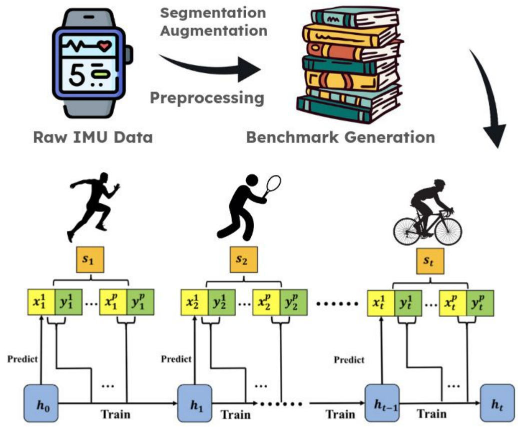

Fig. 1: Task-incremental learning scenario for HAR.

Specifically in this figure, each task, denoted as h_i, is composed of a single class s_i. In this setup, the data reflects the necessary updates and adjustments commonly required in real-world applications.

On other other hand, the plasticity-stability dilemma is a major challenge in task-incremental learning, requiring a balance between adapting to new information (plasticity) and preserving previously acquired knowledge (stability). This often leads to the interference, where enhancing performance on new tasks results in forgetting previous learned tasks.

Additionally, research on deep learning models, such as that by Lee et al. [Lee et al. 2011], indicate that these models learn hierarchical representations of input data, progressing towards the target output. Inspired by neuroplasticity - the brain's remarkable ability to change and adapt - this study aims to replicate these effects in neural networks. When visualizing a deep neural network as an intelligent mind, fundamental concepts are learned in the layers closer to the input, while more complex representations emerge in layers closer to the output.

Several continual learning strategies have already been studied and applied to HAR. In this work, we build on the protocol described by Jha et al. [Jha et al. 2021] to propose neuroplasticity-based techniques that aim to enhance the the stability and generalization of existing continual learning strategies in HAR.

For evaluation, we consider two continual learning strategies: WA-MDF (Weight Alignment for Maintaining Discrimination and Fairness) [Zhao et al. 2019] and WA-ADB (Weight Alignment for Adjusting Decision Boundary) [Kim and Kim 2020]. To ensure a comprehensive evaluation, we utilized some of the most used datasets in HAR studies: UCIHAR (Human Activity Recognition Using Smartphones Dataset) [Reyes-Ortiz et al. 2012], PAMAP2 (Physical Activity Monitoring) [Reiss et al. 2012], HAPT (Smartphone-Based Recognition of Human Activities and Postural Transitions) [Anguita et al. 2013] and DSADS (Daily and Sports Activities) [Barshan and Yüksek 2013]. The experimental protocol incorporates these datasets and the new techniques into Avalanche [Carta et al. 2023], a continual learning framework.

1.1 Contributions

In this work, we contribute to the area by:

—Proposing Plasticity and Metaplasticity strategies for improving continual learning strategies.

—Evaluating their performance with WA-ADB and WA-MDF strategies on widely adopted HAR datasets: UCIHAR, PAMAP2, HAPT, and DSADS.

—Development of new Avalanche plugins to generate HAR benchmarks, techniques and strategies.

2. RELATED WORK

The field of continual learning is actively expanding, with a variety of approaches being proposed in the literature, most of which focus on image-based applications. However, continual learning setups are essential in the HAR domain because users will not consistently perform the same set of activities in the same way after the model is initially trained. Despite that, there are considerably fewer works exploring continual learning in time series data, even though it is crucial for HAR.

In this direction, Jha et al. [Jha et al. 2021] conducted a comprehensive review of state-of-the-art continual learning strategies for task-incremental, sensor-based HAR. They benchmarked these strategies through extensive empirical evaluation, assessing their ability to recognize new activities without forgetting old ones, their computational cost, including memory usage and training time, and their feasibility on resource-constrained devices. The evaluation encompassed 10 techniques on 8 HAR datasets using regularization and rehearsal-based methods. Findings reveal that rehearsal techniques outperform regularization methods, which struggle to retain old knowledge. Notably, rehearsal techniques require minimal memory and often benefit from random sampling. Furthermore, these techniques are not highly sensitive to training data size, achieving good accuracy with 30% of data, and their low computation cost, making them suitable for resource-limited devices.

Since the Plasticity-based and Metaplasticity-based techniques are not restricted to any particular approach and have the potential for broad applicability across various strategies, our primary focus in this work is not to compare different continual learning strategies, but to evaluate the impact of integrating these techniques with existing ones. Therefore, we build on the findings of Jha et al. [Jha et al. 2021] to select two potential strategies for evaluating our proposed techniques: WA-MDF (Weight Alignment for Maintaining Discrimination and Fairness) [Zhao et al. 2019] and WA-ADB (Weight Alignment for Adjusting Decision Boundary) [Kim and Kim 2020]. Both strategies address the issue of biased predictions toward newly learned classes in neural networks that are trained without continual learning strategies. According to these studies, this bias occurs because the weight vectors of new classes in the last, or classification, layer tend to have greater norms, which results in a preference for predicting these new classes. Consequently, the model's performance may degrade on previously learned classes, leading to a less balanced and less accurate overall classification.

2.1 WA-MDF

WA-MDF [Zhao et al. 2019] addresses classification layer imbalance by correcting the weights according to the norms of the classification vectors. Let the output of the neural network be $o(x) = W^T\phi(x)$, where $\phi(x)$ is the feature extractor of the input x and W are the classification layers' weights. W can be split in $W = (W_o, W_n)$, where W_o and W_n refer to the weights for old and new classes, such that $W_o = (w_1, w_2, w_{C_o^b})$, $W_n = (w_{C_o^b+1}, \ldots, w_{C_o^b+C^b})$, where C^b is the total number of new classes and C_o^b is the total number of old classes.

Therefore, the alignment proposed is such that the weights for new classes is $\hat{W} = \gamma W_n$ where $\gamma = \frac{Mean(Norm_o)}{Mean(Norm_n)}$. In this manner, the mean norm of weights is the same for new and old classes.

2.2 WA-ADB

WA-ADB [Kim and Kim 2020] was also proposed to adjust the dimensionality or scaling of classification vectors to better handle unbalanced classes. Unlike WA-MDF, which adjusts weights uniformly, WA-ADB modifies the weights based on the proportion between new and old classes.

For a dataset D with K classes, given that n_i is the number of samples from class i ($n_1 \geqslant \ldots n_i \cdots \geqslant$

n_K), the norm of classification vectors is adjusted by a factor n_1/n_i, i.e., $w_i = (\frac{n_1}{n_t})^\gamma w_i$. During training, weight vectors are updated with each gradient descent step. A greater γ increases the prediction frequency for infrequent classes. Additionally, normalization ensures that the class's conditional probability has the same variance, regardless of sample size.

3. PLASTICITY TECHNIQUES

As highlighted in Section 1, a crucial capability in HAR applications is not only the model's adaptability but also its capacity to retain knowledge from previous data without the need for retraining.

Several strategies have already been evaluated in the HAR domain. Our main goal is to leverage these well-established strategies by combining them with techniques that make models less plastic during training. This approach ensures that the model is less biased towards new classes compared to a more plastic model after training on new data. In this context, Plasticity refers to a network's ability to update its internal representations in response to various learning signals within its training distribution. By addressing the challenge of model bias towards new classes, we aim to enhance the effectiveness and robustness of continual learning strategies.

We developed a technique to reduce model plasticity by stabilizing internal representations during training. This is achieved by constraining the update for the internal layers, in such manner that the most internal layers change less.

The layer updating, using the Adam optimizer, can be defined as

$$\Delta\theta = -lr * \frac{m_t}{\sqrt{s_t} + \epsilon}, \tag{1}$$

where lr is the learning rate, m_t is the moving average of the gradient, s_t the moving average of the squared gradient, and ϵ a small value to prevent division from zero.

Therefore, since the norm of the parameter update is proportional to the learning rate, reducing the learning rate will consequently reduce the norm. From this perspective, the desired behavior of reducing the model's plasticity can be accomplished by employing a scalar, β, as a multiplicative factor to modulate the learning rate of each layer. For a network with n layers, the adjusted learning rate for layer x, denoted as $lr(x)$, follows Equation 2 and is illustrated in Figure 2, where lr_{base} is the unaltered learning rate. By setting β in the range $0 < \beta < 1$ we ensure that the internal representations undergo minimal changes, thereby enhancing the model's robustness, where

$$lr(x) = lr_{base} * \beta^{n-x}. \tag{2}$$

In a task-incremental scenario, the objective is to reduce the plasticity of the model's internal representations over time, rather than merely achieving a more robust internal interpretation. To accomplish this, we introduce a Metaplasticity factor (γ) to dynamically adjust the model's plasticity during training. Similar to learning rate schedulers, the Plasticity factor β is updated for each experience using $\beta = \gamma^{n_{\text{prev}}}$, where n_{prev} denotes the number of previous experiences. By setting γ within the range $(0, 1)$, we ensure that plasticity decreases over time. Consequently, the adjusted learning rate of layer x in the context of the Metaplasticity technique is defined by:

$$lr(x) = lr_{base} * \gamma^{(n-x)*n_{\text{prev}}}. \tag{3}$$

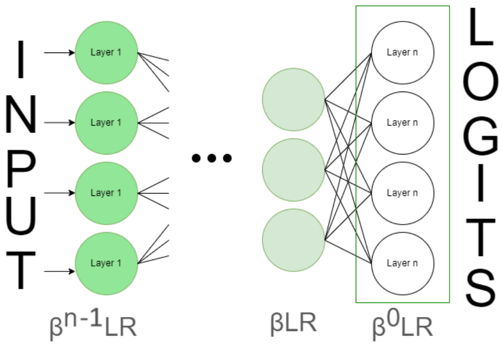

Fig. 2: The Plasticity technique adapted learning rates with plasticity factor β.

Table I: Datasets main characteristics.

Datasets	**Activity**	**# Sensors**	**#Classes**	**Balanced**
DSADS [Barshan and Yüksek 2013]	Daily Activities and Sports	45	19	Yes
HAPT [Anguita et al. 2013]	Daily Activities & Postural Transitions	6	12	No
PAMAP2 [Reiss et al. 2012]	Daily Activities	52	12	No
UCIHAR [Reyes-Ortiz et al. 2012]	Daily Activities	9	6	No

4. EXPERIMENTAL SETUP

This Section details the components of the experimental setup, including the datasets, evaluation metrics, and protocol.

4.1 Datasets

We selected the publicly available datasets commonly used in HAR studies. Table I presents the datasets and summarizes their main characteristics.

4.2 Baselines

Two baselines were used for evaluating the Plasticity-based techniques performance on the task-incremental scenario: 1) **finetuning**, which consists of training the model without applying any strategy, and 2) **offline**, which uses data from all classes for supervised training. Therefore, **finetuning** establishes the lower performance bound for strategies, while **offline** represents the upper bound.

4.3 Evaluation Metrics

We use $F1_{Macro}$ and $F1_{Micro}$ on the testing set for evaluation. Specifically, $F1_{Macro}$ is useful for imbalanced datasets, as it gives equal weight to each class. In contrast, $F1_{Micro}$ provides an overall performance view by accounting for the total number of true positives, false negatives, and false positives across all classes.

$F1_{Macro}$ calculates the class average of F1 scores that is the harmonic mean between precision and recall regarding each class. This ensures that the model's performance on minority classes is not masked by its performance on majority classes. It is given by $F1_{Macro} = \frac{1}{C}\sum_{i=1}^{C}\left(\frac{2*Precision_i*Recall_i}{Precision_i+Recall_i}\right)$.

On the other hand, $F1_{Micro}$ aggregates the contributions of all classes for its calculation. It is the harmonic mean computed by using the total of True Positives (TP), False Negatives (FN), and False

Positives (FP), given by $F1_{Micro} = \frac{2*TP}{2*TP+FP+FN}$.

In the task-incremental scenario, scores are calculated only for classes the model has encountered, while classes it hasn't seen are excluded from evaluation. This approach ensures that the model is only evaluated on tasks it has been trained on. At the experiment's end, the mean and standard deviation of these metrics are computed across all tasks except the first, which does not reflect the incremental nature of the problem.

4.4 Protocol

To address the Human Activity Recognition (HAR) problem, we ensure a subject-based train/test split to avoid data leakage, randomly sampling 80% of the subjects for training and using the remainder for testing. For the task-incremental scenario, each task is defined by randomly sampling two new classes from the training set, balancing reduced training time with increased task difficulty.

We performed a grid search to optimize hyperparameters, including learning rates, weight decay (L2 penalty), and the plasticity or metaPlasticity factor (when applicable). The best results were selected based on the mean $F1_{Macro}$ score across three experiments for each combination. The network architecture is a Multilayer Perceptron (MLP) with layer parameters derived from the work of Jha et al. [Jha et al. 2021]. The experiments employ rehearsal memory with six samples per class and apply Knowledge Distillation (KD) on previous tasks. For the evaluated datasets, tests indicated that the optimal values for the plasticity factors are $0.85 < \beta < 0.95$ and $0.925 < \gamma < 0.975$.

The results are averaged over 10 experiment runs after identifying best hyperparameters for each dataset and strategy permutation. The source code for the avalanche plugins described in this article are available on GitHub at `https://github.com/H-IAAC/META2`.

5. RESULTS

Figure 3 presents the mean $F1_{macro}$ scores per current task across tasks (experiences) for WA-ADB and WA-MDF strategies, both individually and in combination with Plasticity and Metaplasticity techniques. The Plasticity technique, represented by the red and purple curves, occasionally outperformed the pure strategies shown in green and orange. However, it did not consistently improve the overall performance. Thus, addressing decreased Plasticity alone does not necessarily enhance the model's ability to retain information about previous classes.

Conversely, the Metaplasticity technique, represented by the brown and pink curves, consistently outperforms the other strategies. This flexible Plasticity across tasks significantly enhances knowledge retention and achieves the best performance, as highlighted in Table II, which presents the mean and standard deviation of $F1_{Macro}$ and $F1_{Micro}$ scores. Additionally, it is noteworthy that the proposed techniques do not significantly increase training time, as changes are made at most once per experience.

To support this claim, we applied the Mann-Whitney U test [Mann and Whitney 1947]. The null hypothesis (H_0) is that the technique does not result in a statistically significant difference in performance under the $F1_{Macro}$ metric, whether the Metaplasticity technique is applied or not. The alternative hypothesis (H_A) is that the applying this technique increases performance in comparison with the plain strategy. The Mann-Whitney U test was selected because the data is not normally distributed due to forgetting between experiences, and the small sample size makes this non-parametric test more appropriate.

Table III shows the p-values for each dataset and strategy combination. Using a significance level of 0.05, which is commonly used in the literature, we find that in six out of the eight cases, we can reject the null hypothesis, indicating that the Metaplasticity technique significantly improves performance over the plain strategy. These results supports our claim, as only in two cases we cannot reject the null hypothesis.

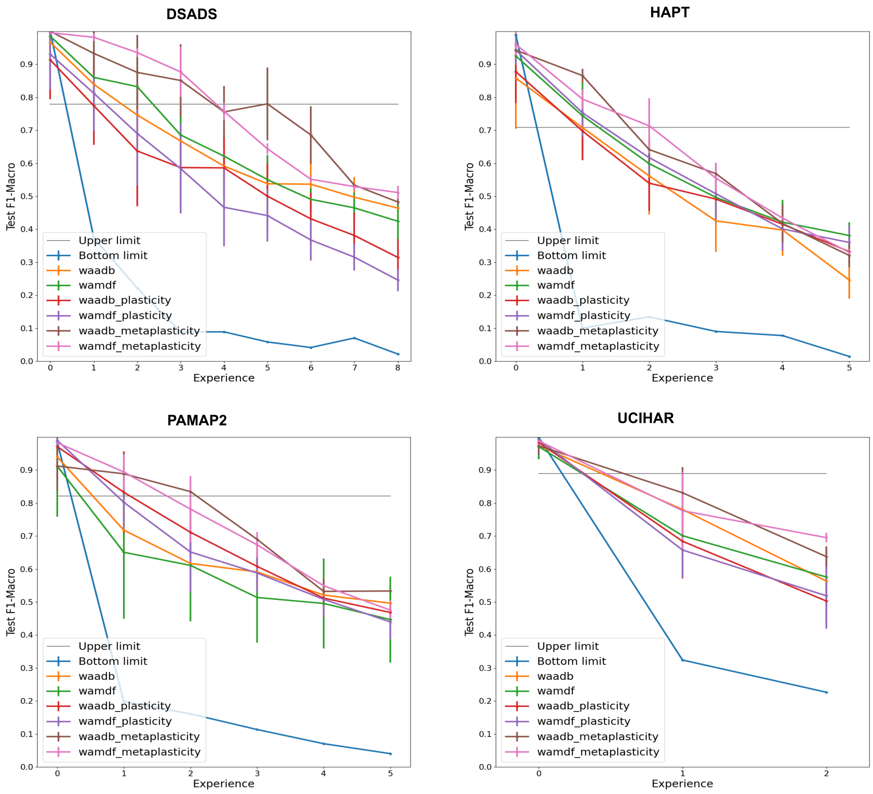

Fig. 3: $F1_{Macro}$ scores for task-incremental learning.

Table II: $F1_{Macro}$ and $F1_{Micro}$ scores across tasks.

	Dataset			
Strategy	**DSADS**	**HAPT**	**PAMAP2**	**UCIHAR**
$F1_{Macro}$ Scores				
WA-ADB	0.610 (0.148)	0.468 (0.180)	0.589 (0.106)	0.672 (0.142)
WA-MDF	0.616 (0.175)	0.528 (0.149)	0.544 (0.174)	0.639 (0.121)
Plasticity WA-ADB	0.527 (0.175)	0.495 (0.140)	0.626 (0.154)	0.594 (0.131)
Plasticity WA-MDF	0.490 (0.207)	0.528 (0.159)	0.598 (0.148)	0.588 (0.115)
Metaplasticity WA-ADB	**0.737 (0.172)**	0.563 (0.192)	**0.696 (0.150)**	0.735 (0.113)
Metaplasticity WA-MDF	0.724 (0.181)	**0.565 (0.178)**	0.675 (0.162)	**0.736 (0.093)**
$F1_{Micro}$ Scores				
WA-ADB	0.641 (0.134)	0.599 (0.177)	0.699 (0.096)	0.730 (0.108)
WA-MDF	0.614 (0.159)	0.675 (0.142)	0.657 (0.184)	0.692 (0.103)
Plasticity WA-ADB	0.557 (0.158)	0.652 (0.118)	0.626 (0.154)	0.714 (0.125)
Plasticity WA-MDF	0.532 (0.190)	0.674 (0.152)	0.598 (0.148)	0.700 (0.134)
Metaplasticity WA-ADB	**0.755 (0.159)**	**0.719 (0.122)**	0.755 (0.123)	0.769 (0.084)
Metaplasticity WA-MDF	0.743 (0.166)	0.709 (0.109)	**0.763 (0.131)**	**0.804 (0.105)**

Therefore, Metaplasticity-based strategies, particularly Metaplasticity WA-MDF, demonstrate superior performance in both $F1_{Macro}$ and $F1_{Micro}$ scores across all evaluated tasks. This suggests that incorporating Metaplasticity principles enhances the effectiveness of the model across the diverse evaluated datasets.

6. CONCLUSION

Human Activity Recognition presents many challenges, including potential changes in the class of interest or distribution shifts over time.

Table III: p-values for the Mann-Whitney U test supporting the improvements of the Metaplasticity techniques under $F1_{Macro}$ score, comparing with plain strategies.

	Dataset			
Strategy	**DSADS**	**HAPT**	**PAMAP2**	**UCIHAR**
WA-ADB	**7e-7**	**0.011**	**0.001**	0.086
WA-MDF	**4e-5**	0.251	**0.002**	**0.002**

Our research addresses these challenges by proposing neuroplasticity-inspired techniques that can be integrated with any continual learning strategy.

Our findings indicate that the Plasticity technique did not improve the evaluated strategies. On the other hand, the Metaplasticity technique demonstrated promising results by maintaining knowledge of previously seen classes and mitigating the problem of catastrophic forgetting. This represents a significant advancement in the field of HAR, offering potential for more effective and efficient systems.

Future research will focus on developing a model that incorporates state-of-the-art deep learning techniques, such as vision transformers. We will also explore whether well-established strategies and our proposed techniques for task-incremental scenarios in multilayer perceptrons maintain their performance when applied to these advanced models.

REFERENCES

Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. L. A public domain dataset for human activity recognition using smartphones. In *The European Symposium on Artificial Neural Networks*, 2013.

Barshan, B. and Yüksek, M. C. Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-Worn Sensor Units. *The Computer Journal* 57 (11): 1649–1667, 07, 2013.

Carta, A., Pellegrini, L., Cossu, A., Hemati, H., and Lomonaco, V. Avalanche: A pytorch library for deep continual learning. *Journal of Machine Learning Research* 24 (363): 1–6, 2023.

Jha, S., Schiemer, M., Zambonelli, F., and Ye, J. Continual learning in sensor-based human activity recognition: An empirical benchmark analysis. *Information Sciences* vol. 575, pp. 1–21, 2021.

Kim, B. and Kim, J. Adjusting decision boundary for class imbalanced learning. *IEEE Access* vol. PP, pp. 1–1, 04, 2020.

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Commun. ACM* vol. 54, pp. 95–103, 10, 2011.

Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18 (1): 50–60, 1947.

Reiss, A., Schmidt, A., Lorenz, L., Diewald, S., Hirche, S., and Urban, B. Introducing a new benchmarked dataset for physical activity monitoring. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*. ACM, pp. 1–8, 2012.

Reyes-Ortiz, J., Anguita, D., Ghio, A., Oneto, L., and Parra, X. Human activity recognition using smartphones. UCI Machine Learning Repository, 2012.

Zhao, B., Xiao, X., Gan, G., Zhang, B., and Xia, S. Maintaining discrimination and fairness in class incremental learning. *CoRR* vol. abs/1911.07053, 2019.

This project was supported by the Brazilian Ministry of Science, Technology and Innovations, with resources from Law nº 8,248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published Arquitetura Cognitiva (Phase 3), DOU 01245.003479/2024-10.