On Domain Generalization for Human Activity Recognition with Mix-Based Methods

Otávio Napoli and Edson Borin *

Institute of Computing - State University of Campinas
Av. Albert Einstein, 1251 - Cidade Universitária, Campinas - SP, 13083-889 - Brazil

Abstract. Domain generalization (DG) is a challenging problem that involves adapting a model trained on source domains to an unseen target domain. In human activity recognition (HAR), domain shifts often arise from differences in sensor placement, device specifications, or environmental factors, making generalization difficult. In this work, we investigate the effectiveness of mix-based methods like MixStyle and Exact Feature Distribution Mixing (EFDM) when integrated into state-of-the-art models like ResNet and TS2Vec for DG in HAR tasks, leveraging the DAGHAR benchmark. Our results demonstrate that MixStyle significantly outperforms both EFDM and Empirical Risk Minimization approaches, highlighting its effectiveness in addressing domain shifts.

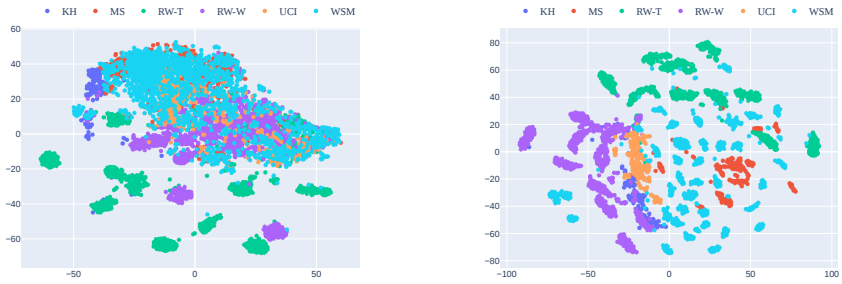
1 Introduction

Human activity recognition (HAR) aims to classify activities performed by individuals using data from inertial motion sensors, such as accelerometers and gyroscopes. Deep learning (DL) models have shown high effectiveness in HAR tasks [1]. However, these models are typically trained under the *i.i.d.* assumption, where training and test data are assumed to come from the same distribution. In practice, this assumption is often violated due to variations in sensor placement, user demographics, and data collection protocols, leading to domain shifts [2, 3].

For instance, Figure 1 presents a t-SNE visualization of samples from two activities (sitting and walking) [1] across six datasets from the DAGHAR benchmark [3], which comprises six smartphone-based datasets (KH, MS, RW-T, RW-W, UCI, and WISDM), each collected under different protocols, containing triaxial accelerometer and gyroscope data. In the sitting activity (Figure 1a), samples from RW-T and RW-W are partially separated from others. Similarly, in the walking activity (Figure 1b), nearly all datasets show distinct clusters. Although the activities remain the same, these separations highlight differences in data distributions across datasets, violating the *i.i.d.* assumption [3].

*This project was supported by the Ministry of Science, Technology, and Innovation of Brazil, with resources granted by the Federal Law 8.248 of October 23, 1991, under the PPI-Softex. The project was coordinated by Softex and published as Intelligent agents for mobile platforms based on Cognitive Architecture technology [01245.003479/2024-10]. The authors also thank CNPq (315399/2023-6 and 404087/2021-3) and Fapesp (2013/08293-7) for their financial support, and Discovery Laboratory for their computational resources.

[1] t-SNE reduces high-dimensional data to a lower-dimensional space while preserving local similarities, helping to reveal patterns and distribution differences in HAR.



(a) Sitting Activity (b) Walking Activity

Fig. 1: t-SNE projections of samples from two activities (sitting and walking) across six datasets in the DAGHAR benchmark. The projection is based on Fourier features extracted from the time-series data.

Generalizing DL models to new, unseen domains is challenging in HAR tasks [4]. In machine learning, this problem is known as domain generalization (DG) [2], which focuses on developing models capable of generalizing to unseen domains using only source domain data.

Wang *et al.* [5] categorize DG methods into three main groups: (i) data manipulation techniques, which enhance training data by augmenting it or simulating new scenarios in latent space; (ii) representation learning methods, which aim to extract domain-invariant features; and (iii) learning strategies, which include ensemble methods, meta-learning, and gradient-based techniques. This work focuses on data manipulation techniques due to their computational efficiency and ease of integration [6]. Specifically, we investigate mix-based methods, which improve generalization by perturbing the "style information" (representation embeddings) of training instances.

We adapt state-of-the-art DL models, including TS2Vec [7] and ResNet [1], for DG using mix-based methods such as MixStyle [6] and Exact Feature Distribution Mixing (EFDM) [8]. Our experiments reveal that MixStyle significantly enhances generalization performance in HAR tasks, outperforming the baseline Empirical Risk Minimization approach and EFDM in DAGHAR benchmark [3], with statistical significance, independently of the model or target domain.

2 Domain Generalization Problem and Methods

In this section, we introduce the notation and formalism used in DG, following the conventions of Zhang *et al.* [8]. Let $\mathcal{X}$ denote the input space and $\mathcal{Y}$ the output space. A *domain* is defined as a dataset S of n data points sampled from a joint distribution $P_{\mathcal{XY}}$. That is, $S = \{(x_i, y_i)\}|_{i=1}^{n} \sim P_{\mathcal{XY}}$, where x_i is the i-th input feature and y_i is the corresponding label. In DG, the training dataset, $\mathcal{S}_{\text{train}}$, comprises M source domains, each represented by a dataset S^i sampled from a unique joint distribution $P^i_{\mathcal{XY}}$, that is, $\mathcal{S}_{\text{train}} = \{\{(x^i_j, y^i_j)\}_{j=1}^{n_i} \sim$

$P^i_{\mathcal{XY}} \mid i = 1, \ldots, M\}$. The challenge in DG arises because the distributions of the source domains differ, *i.e.*, $P^i_{\mathcal{XY}} \neq P^j_{\mathcal{XY}}$ for $i \neq j$. Thus, the objective is to learn a function $h : \mathcal{X} \to \mathcal{Y}$ using $\mathcal{S}_{\text{train}}$ such that h generalizes well to an unseen test domain, $S_{\text{test}} = \{(x_i, y_i)\}|_{i=1}^n \sim P^u_{\mathcal{XY}}$, such that $P^i_{\mathcal{XY}} \neq P^u_{\mathcal{XY}}$ for all $i = 1, \ldots, M$. Since $P^u_{\mathcal{XY}}$ is inaccessible during training, DG methods aim to minimize the empirical risk, that is, $\min_h \mathbb{E}_{(x,y)\sim P^u_{\mathcal{XY}}} [\ell(h(x), y)]$ (where $\ell(\cdot, \cdot)$ is a loss function), by learning robust features that capture invariances across domains [2].

2.1 MixStyle

MixStyle [6] is a lightweight DG technique that simulates new styles by perturbing feature statistics (*e.g.*, mean and standard deviation) within a mini-batch. This method is applied at the feature map level (*e.g.*, outputs of convolutional layers) and is easily integrated into the traditional DL training pipeline.

Given a batch of feature maps $x = [x_1, x_2, \ldots, x_B]$ of size B, MixStyle generates a reference batch $\tilde{x}$ by randomly shuffling the batch dimension. If domain labels are available, $\tilde{x}$ is sampled to ensure x_i and $\tilde{x}_i$ come from different domains; otherwise, $\tilde{x}$ is shuffled randomly.

For each x, MixStyle computes mixed statistics as convex combinations of the original and reference statistics: $\gamma_{\text{mix}} = \lambda \cdot \sigma(x) + (1 - \lambda) \cdot \sigma(\tilde{x})$ and $\beta_{\text{mix}} = \lambda \cdot \mu(x) + (1-\lambda) \cdot \mu(\tilde{x})$, where $\lambda \sim \text{Beta}(\alpha, \alpha)$ is sampled from a Beta distribution. The mixed feature map is then computed as shown in Equation 1.

$$\text{MixStyle}(x) = \gamma_{\text{mix}} \cdot \frac{x - \mu(x)}{\sigma(x)} + \beta_{\text{mix}}. \tag{1}$$

It is worth noticing that MixStyle is applied during training with probability p and deactivated during testing to ensure stable inference. Also, gradients for $\mu(x)$ and $\sigma(x)$ are blocked to improve computational stability.

2.2 Exact Feature Distribution Mixing

Exact Feature Distribution Mixing (EFDM) [8] improves DG by aligning feature distributions across source domains. Unlike MixStyle, which introduces variability through random mixing, EFDM minimizes discrepancies between feature distributions, ensuring a shared statistical structure across domains. EFDM operates at the feature map level, where the feature map x is mixed with a reference map $\tilde{x}$ by interpolating sorted feature vectors.

2.3 Applicability to Time-Series Data

Existing methods like AFFAIR [4] and GILE [9] address DG in time-series data through complex frameworks involving domain-specific modules, custom loss

[2] We assume $P^u_{\mathcal{XY}}$ shares some underlying structure with the source domains. It should overlap with at least some $P^i_{\mathcal{X}}$, ensuring that learned features remain meaningful. We also assume the existence of invariant relationships between features and labels across domains.

functions, and specialized training strategies. Similarly, Lu *et al.* [10] leverage semantic information for generalization. However, these approaches often require extensive modifications or lack the flexibility to integrate with different models, limiting their applicability in real-world scenarios.

In contrast, MixStyle and EFDM offer computationally efficient and flexible alternatives. Initially developed for computer vision, they adapt seamlessly to time-series data by operating at the feature map level. This approach abstracts temporal dependencies and noise, enabling the capture of high-level representations. Such properties make them particularly effective in HAR, where domain shifts frequently arise from variations in sensor placement, device configurations, or environmental conditions [3].

3 Methodology

Datasets We conduct our experiments using the DAGHAR Domain Generalization dataset for HAR [3], a comprehensive benchmark consisting of six distinct datasets sourced from different collection protocols. Each dataset contains time-series data from accelerometers and gyroscopes, capturing six activities: walking, running, sitting, standing, and climbing stairs up and down.

Models We evaluate MixStyle and EFDM on two state-of-the-art models: ResNet [1] and TS2Vec [7]. ResNet uses residual blocks to extract features and we add MixStyle or EFDM layers after each residual block. TS2Vec is a self-supervised framework for time-series representation that captures contextual and temporal features. We adapt it for classification by attaching a multi-layer perceptron head to its encoder. MixStyle or EFDM layers are added after each dilated convolutional layer in the encoder.

Training and Evaluation Procedure We adopt a leave-one-dataset-out strategy [6, 3], where one dataset is designated as the target domain ($\mathcal{S}_{\text{test}}$) while the others serve as source domains for training ($\mathcal{S}_{\text{train}}$). D

As MixStyle and EFDM are applied probabilistically at the feature map level of each layer with predefined probability (p), we evaluate different values of p: 20%, 50%, 70%, and 100%. The configuration yielding the highest validation accuracy is selected and subsequently tested on the designated target domain [3].

Additionally, we investigate the impact of domain labels by exploring training under two scenarios: with and without domain labels (referred to as the "batch construction approach"). When domain labels are available, batches are constructed to sample data from two distinct domains. Else, data is randomly shuffled. In the first approach, the number of samples from each domain is downsampled to match that of the smallest domain.

Baseline comparisons are conducted against Empirical Risk Minimization (ERM), which trains models on combined source domain data using the Cross-

[3]DAGHAR provides predefined splits. We use the training set for model training, the validation set for hyperparameter tuning, and the test set for final evaluation.

Entropy loss function, similar to Napoli *et al.* [3]. ERM provides an evaluation of the model's inherent capacity to generalize to unseen domains.

Experimental Settings Each experiment is repeated three times, resulting in 1944 executions. Results are reported, and configurations are selected using the average and standard deviation of the accuracy of the three runs.

4 Experimental Results

Figure 2 compares the performance of models using MixStyle, EFDM, and ERM on the DAGHAR benchmark. MixStyle consistently outperforms EFDM and ERM across target domains, achieving higher accuracy in most cases. Additionally, in most cases, TS2Vec outperforms ResNet, with the best results observed when TS2Vec is combined with MixStyle. Exceptions occur in RW-Thigh and UCI datasets, where ResNet achieves higher accuracy than TS2Vec.

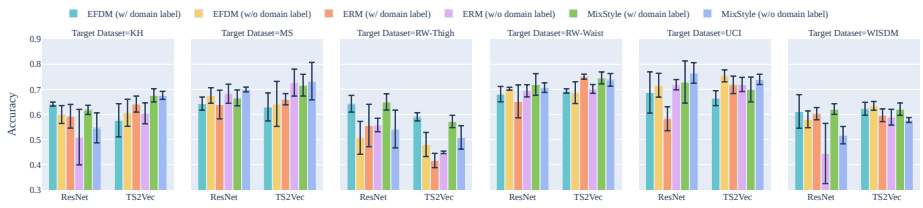


Fig. 2: Comparison of methods on DAGHAR. Bars represent mean accuracy, and error bars indicate standard deviation over three runs.

To assess statistical significance, a Wilcoxon signed-rank test was conducted at a p-value of 0.05. The graph in Figure 3 summarizes the results. Each node represents a mix method and batch construction approach, independent of the model and target domain, resulting in six nodes. Arrows denote statistically significant differences between methods; no arrow indicates no significance. The results confirm that MixStyle is statistically superior to both EFDM and ERM, independently of the model or target domain, demonstrating its robust generalization in HAR tasks. However, EFDM does not show significant improvement over ERM, and no statistical difference is found between the MixStyle batch construction approach, as indicated by the lack of arrows between these approaches.

Finally, these results highlight the importance of domain generalization in HAR, where sensor variability complicates real-world deployment. MixStyle's strong performance suggests its potential for robust HAR applications, extending to healthcare and fitness scenarios.

5 Conclusions

This work evaluated the impact of MixStyle and Exact Feature Distribution Mixing (EFDM) on the generalization of HAR models in a domain generalization

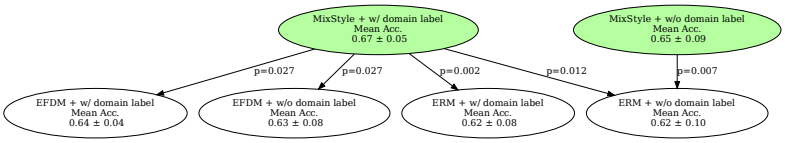


Fig. 3: Wilcoxon Precedence Graph for different mix-based and batch construction methods. Arrows indicate statistical significance; an arrow from node A to B signifies that A is statistically better than B, with weights showing p-value.

setting. We integrated these methods into state-of-the-art models, ResNet and TS2Vec, and tested them on the DAGHAR benchmark, designed for domain generalization in HAR.

The results show that MixStyle consistently outperforms EFDM and the baseline Empirical Risk Minimization (ERM), achieving superior accuracy across most target domains. TS2Vec also demonstrated better performance than ResNet, particularly when combined with MixStyle. These findings highlight MixStyle's effectiveness in enhancing generalization for HAR tasks.

References

[1] S. Mekruksavanich and A. Jitpattanakul. Deep residual network for smartwatch-based user identification through complex hand movements. *Sensors*, 22(8):3094, 2022.

[2] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[3] O. Napoli, D. Duarte, P. Alves, D. H. P. Soto, H. E. de Oliveira, A. Rocha, L. Boccato, and E. Borin. A benchmark for domain adaptation and generalization in smartphone-based human activity recognition. *Scientific Data*, 11(1):1192, 2024.

[4] X. Qin, J. Wang, Y. Chen, W. Lu, and X. Jiang. Domain generalization for activity recognition via adaptive feature fusion. *ACM Transactions on Intelligent Systems and Technology*, 14(1):1–21, 2022.

[5] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S Yu Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022.

[6] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021.

[7] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8980–8987, 2022.

[8] Y. Zhang, M. Li, R. Li, K. Jia, and L. Zhang. Exact feature distribution matching for arbitrary style transfer and domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8035–8045, 2022.

[9] Hangwei Qian, Sinno Jialin Pan, and Chunyan Miao. Latent independent excitation for generalizable sensor-based cross-person activity recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11921–11929, 2021.

[10] W. Lu, J. Wang, Y. Chen, S. Jialin Pan, C. Hu, and X. Qin. Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–19, 2022.