

Article

To Transfer or Not to Transfer and Why? Meta-Transfer Learning for Explainable and Controllable Cross-Individual Activity Recognition

Qiang Shen ¹, Stefano Teso ², Fausto Giunchiglia ^{1,2} and Hao Xu ^{1,3,4,*}

¹ College of Computer Science and Technology, Jilin University, Changchun 130012, China; shenqiang19@mails.jlu.edu.cn (Q.S.)

² Department of Information Engineering and Computer Science (DISI), University of Trento, 38123 Trento, Italy; stefano.teso@gmail.com

³ Chongqing Research Institute, Jilin University, Chongqing 401123, China

⁴ School of Artificial Intelligence, Jilin University, Changchun 130012, China

* Correspondence: xuhao@jlu.edu.cn

Abstract: Human activity recognition (HAR) plays a central role in ubiquitous computing applications such as health monitoring. In the real world, it is impractical to perform reliably and consistently over time across a population of individuals due to the cross-individual variation in human behavior. Existing transfer learning algorithms suffer the challenge of “negative transfer”. Moreover, these strategies are entirely black-box. To tackle these issues, we propose X-WRAP (eXplain, Weight and Rank Activity Prediction), a simple but effective approach for cross-individual HAR, which improves the performance, transparency, and ease of control for stakeholders in HAR. X-WRAP works by wrapping transfer learning into a meta-learning loop that identifies the approximately optimal source individuals. The candidate source domains are ranked using a linear scoring function based on interpretable meta-features capturing the properties of the source domains. X-WRAP is optimized using Bayesian optimization. Experiments conducted on a publicly available dataset show that the model can effectively improve the performance of transfer learning models consistently. In addition, X-WRAP can provide interpretable analysis according to the meta-features, making it possible for stakeholders to get a high-level understanding of selective transfer. In addition, an extensive empirical analysis demonstrates the promise of the approach to outperform in data-sparse situations.

Keywords: human activity recognition; transfer learning; meta-learning; domain adaptation



Citation: Shen, Q.; Teso, S.; Giunchiglia, F.; Xu, H. To Transfer or Not to Transfer and Why? Meta-Transfer Learning for Explainable and Controllable Cross-Individual Activity Recognition. *Electronics* **2023**, *12*, 2275. <https://doi.org/10.3390/electronics12102275>

Academic Editors: Ana Kuzmanic Skelin, Maya Dimitrova, Mirjana Bonkovic and Daniele Riboni

Received: 28 March 2023

Revised: 7 May 2023

Accepted: 12 May 2023

Published: 18 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human activity recognition from wearable sensors is a key element of many human-centric applications, such as smart personal assistants [1], healthcare assessment [2–6], sports monitoring [7], and aging care [8]. In HAR, typically one first collects a training set of examples of rich, multi-modal sensor observations, such as gravitational acceleration and Global Positioning System (GPS), labeled with corresponding activity annotations, and then uses this data to learn a machine learning classifier that predicts activities from sensor measurements [9,10].

Many state-of-the-art approaches assume the training and testing data to be independent and identically distributed. This assumption, however, usually does not hold in practice: sensor data for HAR are collected from a diverse pool of individuals, and behavior patterns are person-dependent [11] owing to biological and environmental factors, meaning that the same activity can be performed differently by different individuals [1]. In practice, while a certain number of participants' data can be collected and annotated for training, the target users are usually not available at the training time [9]. This is what defines cross-individual (or cross-subject) HAR.

The key challenge in cross-individual HAR is how to train a model on known users' data and achieve good recognition performance on new-coming target individuals. A standard solution is to generalize information from well-known training individuals to the target using techniques from transfer learning and domain adaptation [12]. For instance, Zhao et al. [13] tackled cross-individual HAR using a transfer learning approach based on decision trees and k -means clustering. Similarly, Wang et al. [14] explored intra-class knowledge transfer and proposed a transfer learning model for cross-domain activity recognition tasks, while a previous study [15] focused on the problem of cross-dataset activity recognition using an approach that extracts both spatial and temporal features. A previous study [16] proposed another algorithm that adapts to the characteristics and behaviors of different individuals with reduced training data.

However, the success of domain adaptation approaches is not always guaranteed. Existing domain adaptation approaches assume that source and target domains share an identical label space, ignoring the fact that different behavior patterns can contribute to differently distributed sensory data across multiple individuals. If the source and target domains are not sufficiently similar or the source domains have low-quality labeled data, transferring from such a weakly related source may hinder the performance of the target, which is known as negative transfer [12,17]. The negative transfer phenomenon indicates that the source domain data and task contribute to the reduced performance of learning in the target domain, which happens in the real-world scenario where people perform diversely [12]. Despite the fact that avoiding negative transfer is an important issue, little research work has been published to analyze or predict negative transfer for cross-individual HAR tasks. Another issue with these strategies is that they are entirely black-box: it is difficult to extract the reasons why information from a particular source individual was transferred to the target. This is problematic, partly because it makes it hard to identify bugs in the transfer learning step and partly because it prevents stakeholders from controlling and debugging the transfer process.

Motivated by these observations, we propose X-WRAP (eXplain, Weight and Rank Activity Prediction), a novel approach for cross-individual HAR designed for achieving high transfer performance, interpretability, and controllability and specifically tailored for realistic settings where training data for the target individual are scarce. X-WRAP takes an existing baseline transfer learning algorithm and learns a source selection model that identifies the (approximately) optimal source individuals for the given algorithm. The possible sources are first ranked using a linear scoring function using interpretable meta-features. The latter encodes properties of candidate sources as well as their relation to the target individual, enabling the scoring function to properly disambiguate between promising and unpromising candidates. Then, X-WRAP selects the higher-scoring candidates using a learned threshold and applies the baseline transfer algorithm to them to obtain a predictor for the target. The parameters of the scoring function and the threshold are specifically learned so that they generalize across different sources and targets, enabling applications to previously unobserved subjects about whom little is known. The learning problem itself involves repeatedly invoking and evaluating the baseline transfer algorithm, which can be computationally expensive. In order to cope with this, X-WRAP leverages state-of-the-art Bayesian optimization (BO) algorithms [18]. Thus, this architecture offers several major technical novelties. First, X-WRAP facilitates the introspection of the transfer process by selecting candidate sources to improve the performance of cross-individual HAR tasks and avoid the negative transfer. Second, it is completely model- and transfer-algorithm agnostic, making it possible to use any state-of-the-art transfer learning. Finally, BO enables X-WRAP to find high-quality parameters while keeping the number of evaluations of the learning objective (and hence the number of calls to the costly baseline transfer algorithm) at a minimum.

Summarizing, our contributions are as follows:

- We propose X-WRAP, a simple but effective approach for cross-individual HAR that optimizes source selection for transfer learning algorithms in a way that generalizes across individuals with heterogeneous sensor data.
- We propose to use a Bayesian optimization strategy for training the meta-transfer learning framework in an efficient and model-agnostic fashion. The training process includes a data masking strategy implemented in the meta-learning loop.
- X-WRAP builds on an interpretable ranking step and thus enables stakeholders to obtain a high-level understanding of the reasons behind the selective transfer and control (or debug) of the HAR system.
- We report an extensive empirical evaluation of X-WRAP on a real-world dataset and several baseline transfer learning algorithms. Our results indicate that X-WRAP improves the post-transfer performance of cross-individual HAR. In addition, the label-level results indicate the consistent superiority of our approach for almost all activities. Furthermore, we set up experiments to prove that X-WRAP is explainable and controllable.

The remainder of the paper is structured as follows: Section 2 positions X-WRAP with respect to existing approaches. Section 4.3.1 defines the motivation of selective transfer learning for cross-individual HAR. Section 3 introduces X-WRAP, the proposed explainable meta-transfer algorithm for cross-individual HAR, and Section 4 describes and discusses our experimental evaluation of X-WRAP on real-world data. Finally, Section 6 presents some concluding remarks and illustrates promising directions for future work.

2. Related Work

2.1. Human Activity Recognition

A wide range of HAR (human activity recognition) approaches have been developed. Some exploit shallow machine learning models and manually constructed features [19–22], both statistical [23] and distribution-based [24], while others rely on deep learning and automatically learned representations [9,10,25]. Building on this insight, models such as DeepConvLSTM [26] and DeepSense [27] leverage a hybrid architecture that combines CNNs and RNNs. AttenSense [28] implemented an attention-based model into a multi-modal neural network, which is well suited for capturing both spatial and temporal correlations. However, the attention weights are identical across individuals, which does not work well in the real world with heterogeneous data.

The issue with “pure” HAR is that it assumes the training and test data to be independent and identically distributed (IID), which—as we mentioned—is often not realistic. This restricts the applicability of HAR approaches to real-world tasks involving diverse and changing individuals [9]. X-WRAP does not make any such assumption. Rather, it selectively transfers knowledge about known individuals to new or changed ones, which is defined as the cross-individual activity recognition task.

2.2. Multi-Source Unsupervised Domain Adaptation

Cross-individual activity recognition is a more challenging and realistic task where a HAR predictor has access to annotations from a whole pool of individuals and is applied to potentially different target individuals [29,30]. Essentially, approaches for cross-individual activity recognition use the UDA (unsupervised domain adaptation) approach, which is an actively studied area of research in machine learning and computer vision. UDA aims to train the model with labeled source domain data and test on unlabeled target domain data. Traditionally, UDA methods use maximum mean discrepancy (MMD) as metrics and minimize the distance between the source and target domains [31]. In addition, adversarial learning [32–35] and contrastive learning [36] are also applied for domain adaptation without knowing information from target domains.

Considering there are multiple domains in the real-world task, multi-source unsupervised domain adaptation (MS-UDA) has been raised as a novel research area, which is more practical and valuable [37]. Guo et al. [38] apply different distance-based metrics

to measure the correlations between the source and target domain in order to choose data samples dynamically during the training process. Recently, adversarial and GAN-based models have been applied as trending strategies for MSDA tasks [39]. In addition, latent space learning and domain generation have also been applied [40].

However, existing approaches on MS-UDA ignore that some of the source features may not exist in the target domain, which may lead to negative transfer. In addition, sometimes the target domain only shares a part of features with certain source domains.

2.3. Partial Domain Adaptation and Domain Selection

Traditional domain adaptation approaches assume that source and target domains share an identical label space. In practice, however, source features may not exist in the target domain, which can contribute to negative transfer. To tackle this issue, partial domain adaptation (PDA) and domain selection approaches have been proposed. Some previous methods employed sample selection and sample weighing techniques for domain adaptation. Bhatt et al. [41] proposed to adapt iteratively by selecting the best sources that learn shared representations faster. Chen et al. [42] used a hand-crafted re-weighting vector so that the source domain label distribution is similar to the unknown target label distribution. Mancini et al. [43] modeled the domain dependency using a graph and utilizes auxiliary metadata for predictive domain adaptation.

Recently, partial domain adaptation (PDA) approaches have been proposed to solve the challenge where the target domain only contains a part of the data and labels from source domains. A previous study [44] applies a selective weighting mechanism to multiple adversarial networks. After that, Cao et al. [45] use one adversarial network and class-level weight to judge source samples. Zhang et al. [46] propose an auxiliary domain classifier be utilized to derive the possibility that a source sample is contained in the target label space.

2.4. Transfer Learning for Activity Recognition

Transfer learning leverages information about a well-labeled source domain to learn higher-quality models for a target domain with few annotations, while domain adaptation does the same without assuming the target domain has any annotations [12,47,48]. In both tasks, the domains may differ in terms of distribution, representation, or both. Methods used on cross-individual HAR tasks include the following approaches: transfer component analysis (TCA) [49], which acquires a kernel in the reproducing kernel Hilbert space to minimize the MMD between domains; stratified transfer learning (STL) [14], which is designed specifically for exploiting the intra-affinity of classes to perform intra-class knowledge transfer; joint distribution adaptation (JDA) [50], which is based on minimizing joint distribution between domains; balanced distribution adaptation (BDA) [51,52], which extends JDA to adaptively adjust the importance of marginal distribution and conditional distribution; and local domain adaptation (LDA) [53], which offers a balance between domain- and class-level matching and utilizes high-level abstract clusters to organize data.

A previous study [14] explored the intra-class knowledge transfer and proposed a transfer learning model for cross-domain activity recognition tasks. Qin et al. [15] focused on the source domain selection problem to avoid negative transfer and proposed an adaptive transfer learning model to extract both spatial and temporal features for the cross-dataset activity recognition task. Garcia et al. [16] proposed a user-adaptive model in order to adapt to each user's characteristics and behaviors with reduced training data for the task of human activity recognition.

The work in [54] proposed a generalizable independent latent excitation for multi-individual HAR tasks, which can enhance the generalization ability of the cross-individual model. In addition, deep learning models for transfer learning are also applied to domain adaptation tasks, such as cross-individual HAR. ContrasGAN [34] tackles the domain information transferring problem by adding contrastive learning during the adaption with the goal to minimize the intra-class discrepancy and maximize the inter-class margin.

Existing transfer learning methods assume that the transfer of any information from users in the source domain to the target domain is always beneficial. However, this can contribute to negative transfer [17] in the real-world scenario because people are diverse in that they display different behaviors, and the distributions of features in the source and target domains are different. Our proposed meta-transfer model is different from these methods. It leverages pre-learned knowledge to transfer selectively from the source domain, which helps filter out useless information that leads to negative transfer. In addition, the previously described approaches are completely black-box and ignore the fact that source features may not exist in the target domain. X-WRAP builds on existing transfer algorithms by smartly predicting source individuals that are (approximately) optimal for a given target and does so in a transparent fashion, thus facilitating the understanding and control of opaque approaches like those listed above.

2.5. Explainability and Explanatory Debugging

Most work on explainable AI is concerned with extracting explanations from black-box predictors [55,56], as doing so can reveal bugs and biases in the model's logic [57]. The explanations output by X-WRAP achieves the same effect. These works, however, do not give any guidance as to how to debug the learned model. Our approach is inspired by recent approaches that improve the model's behavior by acquiring and learning from corrective supervision on the model's explanations [58–60]. X-WRAP builds on these ideas to correct the transfer algorithm whenever it selects the wrong sources but applies them to the transfer learning step rather than to the prediction step. To the best of our knowledge, these approaches have never been used in HAR nor to control the behavior of a transfer learning algorithm. Moreover, X-WRAP can make use of any predictive model and transfer learning algorithms, making it possible to build on state-of-the-art methods.

3. Method

3.1. Problem Formulation

In the simplest case of HAR, examples annotated by some individuals are used to learn a machine learning classifier that generalizes to unseen inputs from the same individual. In this work, we tackle cross-individual human activity recognition, a more challenging and realistic setting where a HAR predictor has access to annotations from a whole pool of individuals and is applied to potentially different target individuals. Cross-individual HAR is very common in real-world applications.

In the following, sensor measurements (such as acceleration, GPS coordinates, etc.) are encoded as a vector $\mathbf{x} \in \mathbb{R}^d$ and activities (e.g., “running”, “walking”, “swimming”) as a one-hot vector $\mathbf{y} \in \{0, 1\}^A$, where A is the number of alternative activities and y_a is the annotation of activity $a \in \{1, \dots, A\}$. We assume access is given to N training sets $u = \{(\mathbf{x}_{u,i}, \mathbf{y}_{u,i}) : i = 1, \dots, M_u\}$, one for each individual $u \in \mathcal{U} = \{1, \dots, N\}$, and a target user $t \in \mathcal{U}$ for which training data are scarce or absent. The goal is to compute a predictor that performs well for user t by leveraging training data available for one or more of the other users.

In cross-individual human activity recognition settings, each individual u has unique characteristics and behaviors. Different individuals also have different priors over activities. Formally, this means that both the prior over observations $p_u(\mathbf{y})$ and the conditional distribution of sensor observations given priors $p_u(\mathbf{x} | \mathbf{y})$ depend strongly on the individual u .

3.2. Overall Architecture of X-WRAP

Based on the above motivation mentioned in Section 4.3.1, we propose X-WRAP (eXplain, Weight and Rank Activity Prediction), a meta-transfer learning approach for cross-individual human activity recognition tasks, which is also specifically designed for interpretability and control. As shown in Figure 1, our proposed model has three main modules to selectively transfer from the multiple source domains for cross-individual human activity recognition:

- Firstly, a meta-training procedure is trained to get a meta-model for selecting proper source individuals according to various statistical meta-features.
- Then, the meta-model is applied to the transfer learning model to selectively transfer knowledge from existing individuals. A transfer learning model is applied to adapt the model trained on selected source domains, which uses a domain classifier to minimize the distribution distance between source and target domains.
- Based on these two procedures, we can then deploy our system to achieve selective and explainable HAR according to the parameters of the meta-model. In addition, the function of empowering the coordinator with control over the meta-model is designed, which can provide an explainable report and tool for adjusting the parameters in order to gain a better performance.

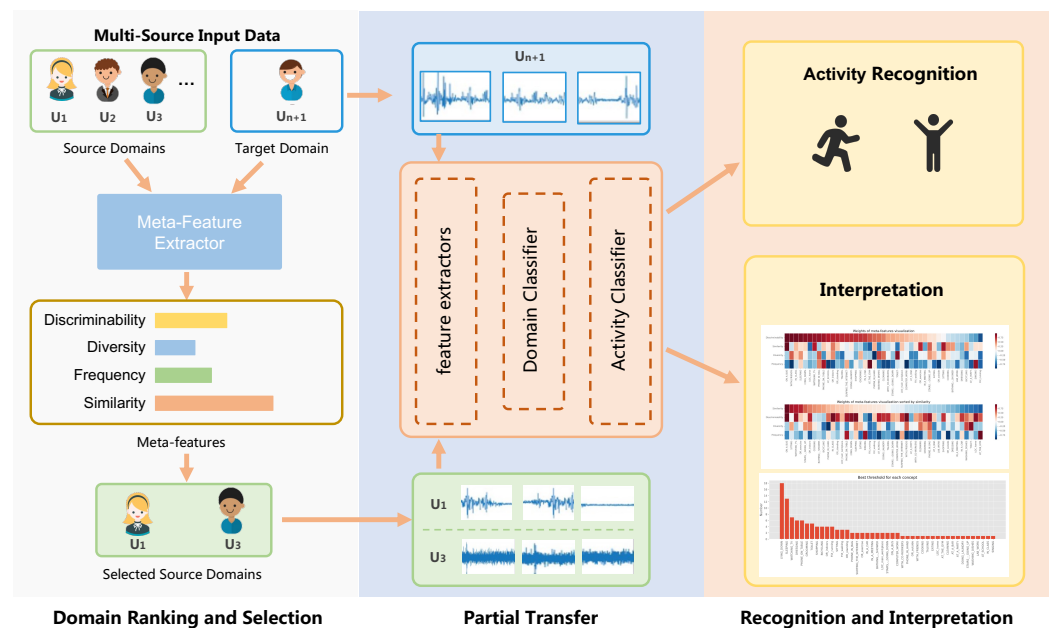


Figure 1. Schematic illustration of the overall architecture of X-WRAP.

3.3. Meta-Transfer Mechanism

Specifically, X-WRAP takes a baseline transfer learning algorithm “Transfer”, such as CORAL [61], a set of candidate sources $\mathcal{S} \subseteq \mathcal{U}$, a target individual $t \in \mathcal{U} \setminus \mathcal{S}$, and an activity $a \in \{1, \dots, A\}$ and outputs a subset of sources $\mathcal{S}^* \subseteq \mathcal{S}$ that is approximately optimal for predicting activity a with Transfer.

The core of the function Transfer is the ranking step responsible for identifying the source individuals $\mathcal{S}^* \subseteq \mathcal{S}$. X-WRAP sorts potential sources using a linear scoring function Score, defined as follows:

$$\text{Score}(s, a, t; \mathbf{w}^{(a)}) = \sum_i w_i^{(a)} \phi_i(s, a, t) \tag{1}$$

Here, the ϕ_i 's are meta-features that capture salient information about the source s , action a , and target t , while the w_i 's are learned weights chosen specifically so as to score more beneficial source individuals higher than the others using the training procedure described below. Notice that the weights are shared by all sources and targets but differ across activities a ; however, the score itself does depend on all three elements because of the meta-features.

After ranking the candidates, X-WRAP applies the Select function to choose those that score above a certain threshold $\tau^{(a)}$, which is also a parameter to be learned. Specifically, our approach returns a selected subset \mathcal{S}^* , defined as follows:

$$\mathcal{S}^* = \text{Select}(\mathcal{S}, a, t; \theta^{(a)}) = \{s \in \mathcal{S} : \text{Score}(s, a, t; \mathbf{w}^{(a)}) \geq \tau^{(a)}\} \tag{2}$$

where $\theta^{(a)} = (\mathbf{w}^{(a)}, \tau^{(a)})$ are learnable parameters, \mathcal{S} is the set of source domains, and $\tau^{(a)}$ is the threshold to select proper source users for activity a . Thus, Equation (2) can select the source domains that have scores higher than the threshold. X-WRAP then simply applies the transfer learning model to the selected individuals, obtaining a predictor for activity a that leverages the information of sources in \mathcal{S}^* . The transfer learning model here is a deep domain adaptation approach as shown in Figure 1. The transfer model contains three parts, which is similar to the previous work [62], which can maximize the similarity between data classes across domains. The first part is an LSTM-based feature extractor. Then, we use a domain classifier that discriminates between the source and the target domains during training. Finally, fully connected layers are applied for classifying activities. In addition, the proposed meta-learning architecture is model-agnostic, which means it can also be applied to any transfer learning approach to select proper source domains.

3.4. The Meta-Features

The meta-features $\phi_i(s, a, t)$ are designed for data valuation to select the best source domains, which need to satisfy three desiderata. First, they should enable the ranking function to identify source individuals that are likely to provide useful information. For instance, candidates on which the HAR predictor performs poorly may be inadequate as sources for the transfer. Second, they should highlight similarities and dissimilarities between the distribution of activities and sensor measurements of the candidate source and target individuals. This further facilitates discriminating between promising and unpromising individuals for a given target. Third, they should be understandable to human stakeholders. This is essential for enabling sufficiently expert users to understand why selected individuals have achieved a high enough score and to provide corrective feedback on the behavior of X-WRAP.

Previous study [63] surveyed the principles for selecting valuable data, which contain model-driven principles and data-driven principles. Model-driven principles refer to evaluating the data by training a model and measuring the data according to the performance on a validation set intuitively. Previous studies compute the metrics, such as accuracy and negated loss functions, as the validation of the performance to evaluate the data [64,65]. Data-driven principles indicate computing metrics, such as frequency, diversity, and similarity, to reference distribution directly on the dataset instead of the model. Frequency or monotonicity indicates the amount of the data and having more data is more valuable [66]. Diversity indicates that the data points cover a larger region of the input space. The dataset can improve the model's predictive performance if it contains more intra-diversity [67,68]. In transfer learning, similarity to the reference distribution is crucial for transfer performance. Tay et al. [67] measured similarity using the translated negative maximum mean discrepancy (MMD) between the two distributions to evaluate the quality of the dataset.

Motivated by the previous studies, we implement two groups of meta-features to satisfy the desideratum, which are then concatenated by X-WRAP and used together in Equation (1). The meta-features in the first group capture information about a single individual, while the second group of meta-features is concerned with how dissimilar the source and target individuals are. The computations of the meta-features are shown in Table 1. The meta-features can be categorized as follows:

- Meta-features based on predictability: By predictability (or discriminability), we refer to the easiness of classifying a particular activity a using a supervised classifier, 1-nearest neighbor (1-NN), fit on the training set of an individual u . In order to estimate predictability, for each candidate source u , we take 20% of the training set as a validation set to compute multiple metrics, including accuracy and F_1 score, and the rest as the training set.
- Meta-features based on diversity: By diversity, we mean the intrinsic heterogeneity in its activity patterns. For each individual u , we measure this by computing the number of distinct activities that they perform and the Shannon entropy of the activity

annotations $\{y_{i,a}\}$ available in the training set. These meta-features model the intrinsic difficulty of predicting the behavior of an individual u and are useful for preventing unpredictable individuals being used as sources.

- Meta-features based on frequency: We also account for the frequency of the context labels by computing the number of times a certain activity occurs in the training annotations.
- Meta-features based on dissimilarity: X-WRAP measures diversity using the maximum mean discrepancy (MMD), a well known statistic that estimates how “different” two (empirical) distributions are and that has found ample application in hypothesis testing [69] and domain adaptation [70].

Table 1. List of meta-features. In the definition column, TP refers to the number of true positive samples; FN indicates the number of false negative samples; TN refers to the number of true negative samples; and FP refers to number of false positive samples. Pr indicates the *precision* of the machine learning model, which is formalized as $\frac{TP}{TP+FP}$ and Rc indicates the *recall* of the machine learning model, which is formalized as $\frac{TP}{TP+FN}$. $p_{u,a}$ indicates the probability of activity a . For the definition of MMD, $\phi(\cdot)$ is the feature map that maps the original instances into the reproducing kernel Hilbert space (RKHS) \mathcal{H} .

Meta-Feature Type	Meta-Feature	Measurement	Definitions
Single individual	Predictability	Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
	Predictability	F_1 score	$\frac{2*Pr*Rc}{Pr+Rc}$
	Diversity	Number of activity types	$\#types$
	Diversity	Shannon entropy	$-\sum_a p_{u,a} \ln(p_{u,a})$
	Frequency	Number of instances	$\#y_{i,u}$
Paired individuals	Dissimilarity	MMD	$\left\ \sum_{i=1}^{n_1} \phi(\mathbf{x}_i) - \sum_{j=1}^{n_2} \phi(\mathbf{y}_j) \right\ _{\mathcal{H}}^2$

3.5. Fitting X-WRAP Using Meta-Learning

The only remaining element is the learning objective. Recall that our aim is to acquire the parameters $\theta^{(a)} = (\mathbf{w}^{(a)}, \tau^{(a)})$ for each activity a that ensure that Select function in Equation (2) selects source users that are maximally beneficial for the baseline transfer learning algorithm. Moreover, and crucially, we want these parameters to work well regardless of the choice of source and target individuals.

X-WRAP achieves this using a simple but effective meta-learning strategy. At a high level, the idea is to directly optimize the performance of the predictor’s output by the baseline transfer mechanism Transfer when applied to the known users \mathcal{U} . Formally, let $L_{a,t}(f)$ be a loss function that measures the quality of the output of the prediction by a classifier f for activity a and individual t , e.g., the number of mistakes or the binary cross-entropy. In addition, let $f_{S,a,t}$ be the predictor output by running the baseline transfer learning algorithm Transfer on sources $S \subseteq \mathcal{U}$. Then, the improvement in prediction performance due to transferring from S compared to not transferring at all is as follows:

$$L_{a,t}(f_{\emptyset,a,t}) - L_{a,t}(f_{S,a,t}) \tag{3}$$

and the average improvement due to parameters $\theta^{(a)}$ is given by the following:

$$\mathbb{E}_{\mathcal{U}}[L_{a,t}(f_{\emptyset,a,t}) - L_{a,t}(f_{S,a,t})], \quad \text{where } S = \text{Select}(\mathcal{U}, a, t; \theta^{(a)}). \tag{4}$$

Notice that the expectation runs over all possible choices of source individuals \mathcal{U} and the overall best parameters are those that maximize it. In practice, however, this expectation cannot be computed because the ground-truth distribution of \mathcal{U} is unknown.

X-WRAP works around this issue by optimizing a leave-one-out estimator of the expected benefit over the known users \mathcal{U} , namely,

$$\operatorname{argmin}_{\theta^{(a)}} \sum_{u \in \mathcal{U}} L_{a,u}(f_{\mathcal{S},a,u}), \quad \text{where } \mathcal{S} = \text{Select}(\mathcal{U} \setminus \{u\}, a, u; \theta^{(a)}). \quad (5)$$

In words, for each activity $a = 1, \dots, A$, X-WRAP seeks parameters $\theta^{(a)}$ that lead to a high post-transfer performance when applying the baseline transfer learning algorithm to the subset \mathcal{S} of individuals known to the system \mathcal{U} , minus one hold-out individual u used as the target. Naturally, both training and validation data are available for these individuals, making it possible to compute the performance and features in an unbiased manner.

3.6. Training X-WRAP Using Bayesian Optimization

The issue with minimizing Equation (5) is that computing the loss involves invoking the baseline transfer learning algorithm Transfer. The training and testing procedures are shown in Figure 2. X-WRAP is a meta-learning structure and each round of the training process involves the transfer model, which is not cheap to evaluate. Therefore, we treat Equation (5) as an expensive global optimization problem and solve it using effective Bayesian optimization (BO) algorithms [18,71]. This brings two advantages. First, BO is designed for expensive-to-evaluate problems and can effectively keep the number of evaluations of the loss function within an acceptable budget, which is 300 rounds in this work. Second, BO does not need gradient information in order to explore the space of the candidate parameters and therefore can also work in our model- and algorithm-agnostic setting. In short, BO algorithms seek a global maximum of a black-box function f by repeatedly sampling the value of the function at well-chosen inputs and learning a surrogate that matches the corresponding outputs. The query inputs are chosen so as to maximize the expected information about the structure of the function that they convey, estimated using the surrogate itself. The BO loop proceeds as in the following Algorithm 1.

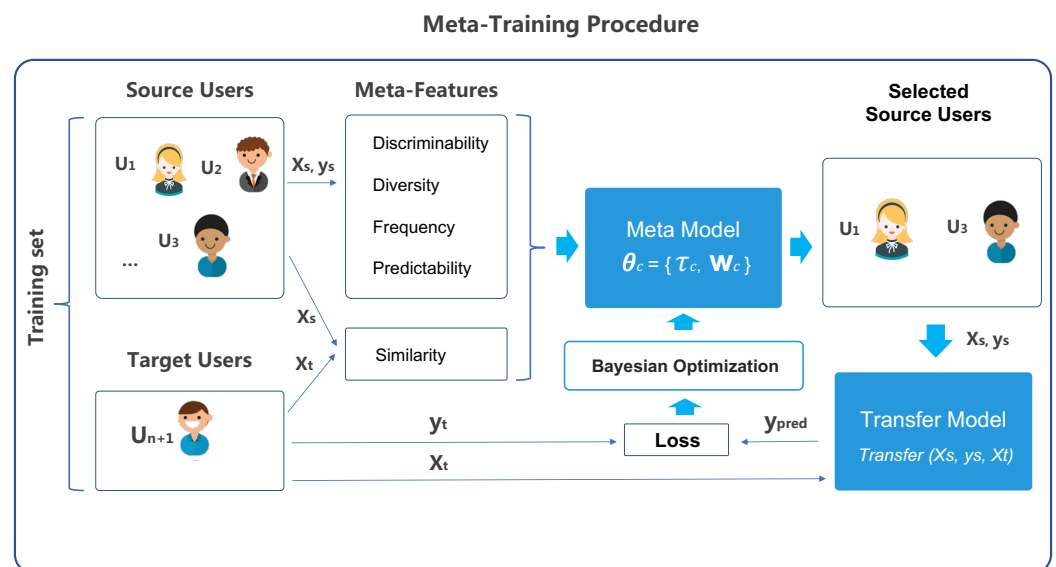


Figure 2. Cont.

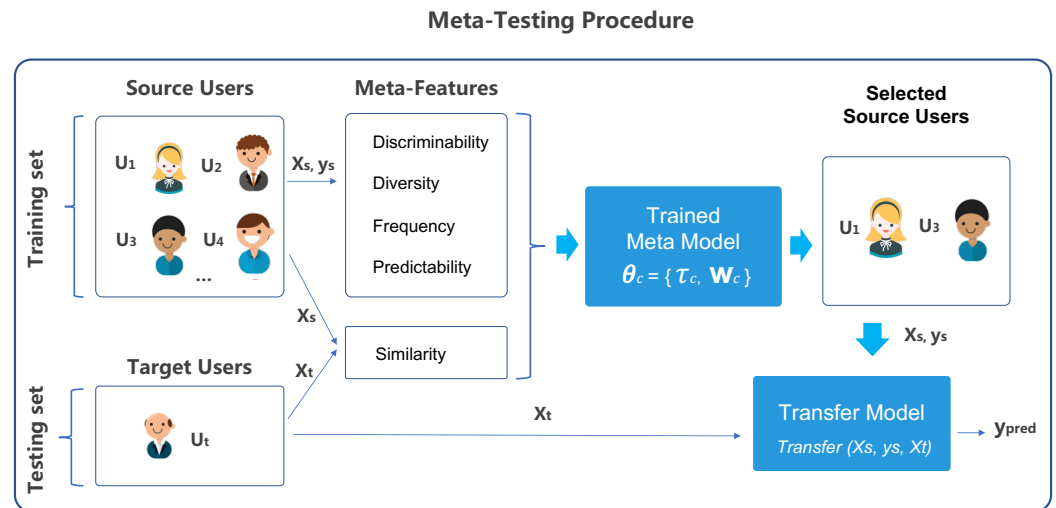


Figure 2. Schematic illustration of the training (top) and evaluation (bottom) procedures of X-WRAP.

Algorithm 1 X-WRAP's training procedure: u is the training set of user u ; t and a are the target user and activity, respectively; I is the max iteration budget; and Θ is the space of possible parameters.

```

1: function EvalTransferLoss( $\{1, \dots, N\}, a, \theta$ )
2:   for  $t = 1, 2, \dots, N$  do
3:     rank candidate sources  $s \in [N] \setminus \{t\}$  according to Score (Equation 1)
4:      $S' \leftarrow$  select  $\tau$  highest scoring candidates
5:      $f \leftarrow$  Transfer( $S', a, t$ )
6:      $L \leftarrow L + l(f, i)$ 
7:   return  $L$ 

8: function LearnToTransfer( $\{1, \dots, N\}, a, I, \Theta$ )
9:   # Initialize parameter-loss data
10:  sample  $\theta_0 = (\mathbf{w}_0, \tau_0)$  at random from  $\Theta$ 
11:   $L_0 \leftarrow$  EvalTransferLoss( $\{1, \dots, N\}, a, \theta_0$ )
12:   $\mathcal{H} \leftarrow \{(\theta_0, L_0)\}$ 
13:  # Optimize  $\theta$  using Bayesian optimization
14:  for  $i = 1, 2, \dots, I$  do
15:    fit surrogate  $g : \Theta \rightarrow \mathbb{R}$  on observations  $\mathcal{H}$ 
16:     $\theta_i \leftarrow \operatorname{argmax}_{\theta \in \Theta} \operatorname{acq}(g, \theta)$ 
17:     $L_i \leftarrow$  EvalTransferLoss( $\{1, \dots, N\}, a, \theta_i$ )
18:     $\mathcal{H} \leftarrow \mathcal{H} \cup \{(\theta_i, L_i)\}$ 
19:  # Return best parameters found
20:   $i^* \leftarrow \operatorname{argmin}_{i=1, \dots, I} L_i$ 
21:  return  $\theta_{i^*}$ 

```

X-WRAP follows the meta-learning manner, which means that the parameters are fitted with all training users and then evaluated on a distinct set of test users. The training set only participates in the training procedure and is used to train the meta-model, as in the procedure in Figure 2 (top figure). Specifically, in the training procedure, we split the training users into source users and target users with the leave-one-user-out setting. As for the testing procedure, all of the testing users are used for target users, and the selected best source users are chosen according to the trained parameters from training users (who work as source users), as shown in Figure 2 (bottom figure). Considering that the computations of the meta-features do not need labels from target users, the proposed model is unsupervised.

In this way, only the sensory data of testing users are used in the test procedure but the labels are not used.

3.7. Model Interpretation

Compared to traditional transfer learning, which is completely black-box, in X-WRAPit is straightforward to extract the reasons why a particular individual is selected (or not) from the scoring function. This follows from two aspects. First, like other interpretable approaches [72] and explainability techniques [73], X-WRAP relies on a linear model scoring function Score from which the relative contribution of each meta-feature is easy to read off from the associated weight. Second, the meta-features are themselves easy to interpret for sufficiently expert stakeholders. Indeed, prediction accuracy and entropy are easy to assign intuitive meaning to for trained statisticians, while cross-individual diversity (modeled by MMD) can be specifically broken down along different and intuitive axes, such as geographical similarity. This is precisely what we do in our last experiment. This formulation ignores information overlap between sources, which is not a huge deal if Transfer performs learning afterward. This could be solved by making use of submodular scoring functions or techniques from deep active learning, but this is not entirely straightforward.

4. Result

4.1. Dataset and Data Processing

We ran a set of experiments on the ExtraSensory [74] dataset, a well known context recognition dataset collected via a mobile application in the open world. The ExtraSensory dataset contains over 300,000 multi-labeled instances (with classes such as “outside”, “at a restaurant”, and “with friends” from a total of 51 labels) from 60 users. Each example associates readings from multiple sensors to ground-truth context annotations describing the activity, location, and social context (e.g., “walking”, “with friends”, “at home”). The sensors include motion-reactive sensors (e.g., accelerometer, watch accelerometer, gyroscope, magnetometer), location services (GPS), audio, watch compass, phone state indicators (e.g., WIFI), and low-frequency sensors (e.g., air pressure, humidity, temperature). As for the labels, a flexible user interface is provided for participants to self-report their context in terms of what they were doing, who they were with, where they were, where their phone was, and so on.

The dataset contains 60 subjects (34 female and 26 male users) with heterogeneous sensory data. Specifically, 34 of the subjects are iPhone users (iPhone 4 to iPhone 6; iOS versions 7, 8, and 9), while 26 subjects use android phones, with various devices (Samsung, Nexus, Motorola, Sony, HTC, Amazon Fire-Phone, and PlusOne). In addition, The participants have different nationalities and are from diverse ethnic backgrounds, including Chinese, Mexican, Indian, Caucasian, African-American, and more. A total of 93% of the subjects are right-handed and wear the smartwatch on their left wrist, and almost all were students or research assistants. Table 2 presents additional subject characteristics. As pointed out in the previous study [74], the ExtraSensory dataset with high heterogeneity needs to be considered with the domain adaptation for new users. The potential bias in the dataset can contribute to poor performance for context and activity recognition models and personalized models are required to recognize the activity for unique individuals [1,75]. These settings are applied to research question 2 to research question 5. Research question 1 applied the same dataset but has different settings and detailed settings will be introduced in Section 4.3.1.

In the following experiments (research question 2 to research question 5), we follow the settings of meta-learning. Specifically, X-WRAP is meta-trained on a set of training users, which means that the parameters are fitted with all training users and then evaluated on a distinct set of test users. We selected 30 users at random as the training set and the remaining 30 distinct users as the testing set. The training set only participates in the training procedure and is used to train the meta-model, as in the procedure discussed in Section 3.5, while the testing users instead play the role of target users and are only used

to test the generalization ability of the learned models. In the training procedure, we split the training users into source users and target users with the leave-one-user-out setting. Specifically, we take every single user from the 30 training users as the target user and the remaining 29 users as source users. As for the testing procedure, all of the 30 testing users are used for target users, and one or multiple selected best source users are chosen according to the trained parameters from 30 training users (who work as source users). Considering that the computations of the meta-features do not need labels from target users, the proposed model is unsupervised. In this way, only the sensory data of testing users are used in the test procedure but the labels are not used.

Table 2. Statistics for the 60 users in the ExtraSensory.

	Range	Mean (Standard Deviation)
Age (years)	18–42	24.7 (5.6)
Height (cm)	145–188	171 (9)
Weight (kg)	50–93	66 (11)
Body mass index (kg/m ²)	18–32	23 (3)
Participation duration (days)	2.9–28.1	7.6 (3.2)

4.2. Baseline Models and Transfer Mechanisms

In order to probe the effect of the transfer mechanisms, we apply X-WRAP to two diverse sets of well-known baseline transfer learning approaches used in cross-domain HAR. The baseline deep learning models are as follows:

- The **domain adversarial neural network (DANN)** [62] is a deep domain adaptation approach, which uses a domain classifier that discriminates between the source and the target domains during training.
- The **distribution-embedded deep neural network (DDNN)** [76] is a state-of-the-art network featuring learning approaches for activity recognition.
- **Triple-DARE** [77] is a neural network method that combines three unique loss functions to enhance intra-class compactness and inter-class separation within the embedding space of multi-labeled datasets. In this experiment, we slightly modify the model from lab-to-field transfer to cross-individual transfer to make the settings the same as the proposed model.
- **HDCNN** [78] is a cross-domain transfer learning model that uses KL divergence loss on the acquired feature vectors.
- The **convolutional deep domain adaptation model for time series data (CoDATS)** [54] is the latest domain adaptation model for time series data.

Notice that HDCNN and Triple-DARE are applied on the ExtraSensory dataset and handle a lab-to-field transfer. For a fair comparison, we change them slightly to cross-individual transfer tasks. In addition, we compare some existing shallow transfer learning models:

- **Transfer component analysis (TCA)** [49] is a feature extraction approach for domain adaptation. It aims to learn a feature subspace shared by different individuals that minimizes the discrepancy in distribution, measured using MMD. The transfer is then achieved by projecting inputs onto this shared subspace.
- **Correlation alignment (CORAL)** [61] is another unsupervised approach for domain adaptation in HAR. It works by minimizing cross-domain differences by aligning the second-order statistics of source and target distributions.
- **Balanced distribution adaptation (BDA)** [51] tackles domain adaptation by minimizing the marginal and conditional distribution discrepancy between domains. It leverages marginal and conditional distributions between domains and re-weights the importance of those two distributions.
- **Manifold embedded distribution alignment (MEDA)** [79] performs dynamic marginal and conditional distribution alignment for unsupervised domain adap-

tation. It learns a domain-invariant classifier in Grassmann manifold with structural risk minimization and performs dynamic distribution alignment to account for the importance of marginal and conditional distributions.

- **Easy transfer learning (EasyTL)** [80] learns both non-parametric transfer features and classifiers by learning intra-domain structures for unsupervised domain adaptation.

Note that all the approaches are fully unsupervised and therefore do not require annotations for the target individual to be available at any time. As for the underlying predictor of shallow transfer learning models, we follow other work on transfer learning for HAR and other tasks [14,15,51,79,80] and use 1-nearest neighbor (1-NN) as the underlying classifier.

4.3. Experimental Results

We answer empirically the following research questions:

- **RQ1:** Is the selective meta-transfer mechanism in X-WRAP needed for cross-individual HAR?
- **RQ2:** Does X-WRAP improve the performance of cross-individual HAR?
- **RQ3:** Does X-WRAP perform better when data for the target user are sparse?
- **RQ4:** Does X-WRAP enable experts to control the system whenever sensors go awry?
- **RQ5:** Does X-WRAP provides a reliable explanation for the partial transfer?

To this end, we apply X-WRAP to a well-known context recognition dataset and study the conditions under which cross-individual transfer across users is beneficial, focusing on the characteristic of the users, activities, and prediction tasks. All methods were implemented in Python 3 using scikit-learn and the Hyperopt Bayesian optimization package [81].

4.3.1. Motivation for Selective Meta-Transfer

In this experiment, an empirical experiment is conducted to prove that heterogeneous sensory data decrease the performance of the HAR models. Specifically, we assess whether there is a correlation between distribution discrepancy and the change in performance after transferring the model. For this experiment, we use 30 training users and for each pair of users consider one the source user and the other the target user. Then, the data of each source user is split into training data and testing data with a ratio of 8:2. For each pair of users and activity, a binary KNN (1-NN) classifier is trained on the training data (80%) of source users and tested on both the testing data (20%) of the source user and 20% of the data of the target user for fairness. We then compute the difference in the performance of the testing on the source user themselves and the other target user. Then, for each activity and each pair of users, the dissimilarity (MMD) is computed according to Section 3.4. Finally, the correlations between dissimilarity (MMD) and the difference in performance are visualized and computed to study the impact of heterogeneity existing in sensory data for cross-individual HAR.

As shown in Figure 3, the results visualize the correlations between similarity and change in performance after transfer. Specifically, each point in this figure indicates one activities of one pair of users. Generally speaking, from this figure, a positive correlation between similarity and change in performance between the source domain and target domain can be seen, such as the activity. Among these subplots, the activity “At home” and “Indoors” show strong positive correlations between the MMD and change in performance, where the performance of the model will drop after transferring information between two dissimilar users. Interestingly, there is another kind of distribution in the subplots of “Eating”, “Standing”, “Walking”, and “Phone in hand”. When the value of MMD is close to zero, the change in performance is small. Along with the increase in MMD, the performances of some pairs start to decrease while the performances of other pairs start to increase. Therefore, we would infer that a large distance between two sub-datasets of users indicates notable changes in performance after transferring the information in the task of activity recognition.

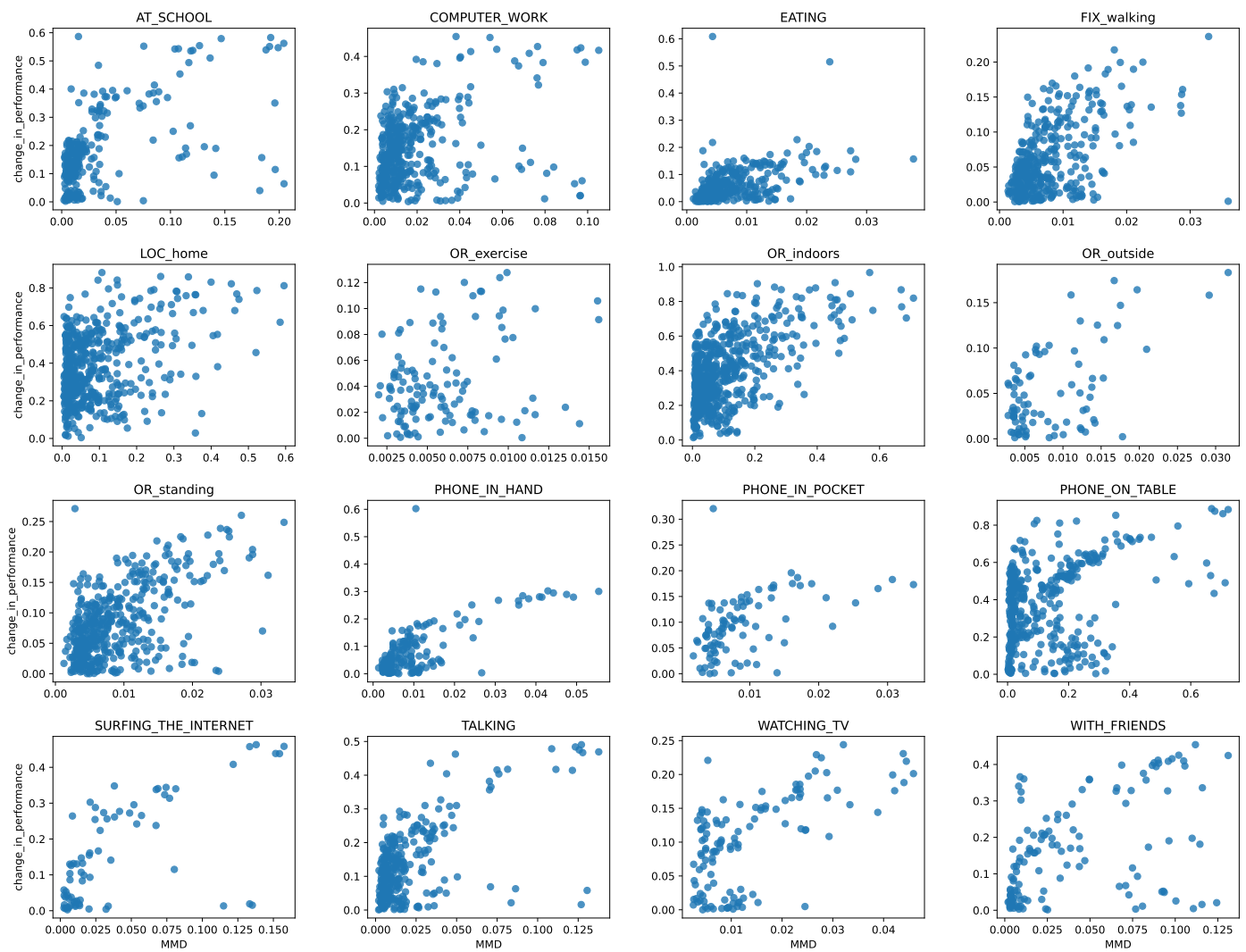


Figure 3. Correlation between MMD and change in performance for different activities.

To take one step further, we inspected the correlations between similarity and change in performance quantitatively by computing the correlation coefficient. We randomly selected activities with more than 100 examples to filter out the rare activities from 20 users. The correlation results are presented in Table 3. Generally speaking, most of the activities show positive correlations. We find strong correlations for activities with complex and diverse behavior patterns, such as “Phone in hand” and “Surf the Internet”. Thus, the discrepancy of sensor readings can contribute to negative transfer. In contrast, the activities with a constant pattern such as “Phone on table” show weak correlation since the sensor features are similar across individuals. The consistent positive correlations among all activities suggest that the similarity between the features of the source and target users should be considered in the meta-transfer procedure.

Table 3. Spearman correlations between MMD and change in performance.

Activity	Corr.	<i>p</i> -Value	Activity	Corr.	<i>p</i> -Value
Surfing the Internet	0.658	<10 ⁻⁹	Standing	0.488	<10 ⁻²⁷
Phone in Hand	0.570	<10 ⁻¹⁴	Walking	0.451	<10 ⁻¹⁸
Phone in Pocket	0.561	<10 ⁻⁸	Talking	0.424	<10 ⁻¹³
Watching Tv	0.556	<10 ⁻¹⁰	Drive	0.381	0.003
With Friends	0.532	<10 ⁻¹⁰	Outside	0.341	0.001
At School	0.525	<10 ⁻¹⁸	Phone on Table	0.315	<10 ⁻⁹
Eating	0.498	<10 ⁻²⁰	Computer Work	0.293	<10 ⁻⁸
Indoors	0.493	<10 ⁻³⁹	Home	0.290	<10 ⁻¹⁰

4.3.2. X-WRAP Helps Improve Transfer Performance

For starters, we evaluate the impact of meta-transfer learning on transfer performance by applying X-WRAP to each of the above transfer mechanisms. For each approach, we evaluated X-WRAP using the following procedure. For each activity c in the activity hierarchy and for each user t in the dataset, we selected the other users as candidate sources and then recorded the performance achieved by applying the transfer mechanism by itself and that obtained by combining it together with X-WRAP. The performances averaged over all activities are shown in Tables 4 and 5. Generally, X-WRAP works effectively on all of the candidate transfer algorithms by selecting proper source users. Note that selective transfer using CoDATS outperforms slightly transferring from all domains. We assume the reason is CoDATS is considered a multi-source domain transfer itself. In addition, applying X-WRAP to the DANN can achieve the highest performance. Thus, in the following experiments, we apply the DANN as the transfer learning module.

To take one step further, the performances of all activities averaged over 30 test users are reported in Table 6. Generally, X-WRAP outperforms transfers from all users consistently for most of the activities with an average margin of 2.63%. X-WRAP works extremely well on activities such as “With Friends”, “At Workplace”, and “Computer Work” compared to “Transfer from All” and “Transfer at Random”. We infer that these activities have large distances of sensory features across different individuals, and those source users who are dissimilar to the target user can contribute to negative transfer. Therefore, X-WRAP can improve the performance by filtering out useless users.

Table 4. Comparison between random transfer and transfer with and without X-WRAP on shallow transfer learning.

	TCA	BDA	EasyTL	MEDA	Coral
Transfer from all	81.24 \pm 0.16	89.12 \pm 0.35	81.08 \pm 0.23	85.98 \pm 0.11	89.87 \pm 0.06
Transfer at random	81.97 \pm 1.27	89.68 \pm 4.21	82.37 \pm 2.98	86.10 \pm 1.19	90.60 \pm 0.83
x-wrap	83.01 * \pm 0.41	90.83 * \pm 0.19	84.29 * \pm 0.25	88.78 * \pm 0.61	92.50 * \pm 0.13

* indicates significance based on Student t -test ($p < 0.05$).

Table 5. Comparison between random transfer and transfer with and without X-WRAP on deep transfer learning.

	DANN	DDNN	HDCNN	Triple-DARE	CoDATS
Transfer from all	88.92 \pm 0.82	89.20 \pm 0.97	87.23 \pm 2.01	89.15 \pm 1.29	92.23 \pm 1.62
Transfer at random	88.97 \pm 3.81	89.01 \pm 3.74	84.01 \pm 9.06	89.26 \pm 4.29	90.22 \pm 2.10
x-wrap	92.62 * \pm 0.59	91.48 * \pm 0.67	88.23 \pm 4.27	91.27 \pm 0.83	92.25 * \pm 1.02

* indicates significance based on Student t -test ($p < 0.05$).

Table 6. Comparison of F_1 scores achieved using DANN as baseline transfer learning algorithm for each activity label.

Activity	Transfer from All	Transfer at Random	X-WRAP
Bathing Shower	98.07 \pm 1.18	98.94 \pm 2.91 (+0.87)	98.99 \pm 1.18 (+0.92)
Bicycling	96.65 \pm 0.38	97.47 \pm 1.82 (+0.82)	97.51 * \pm 0.42 (+0.86)
Cleaning	96.45 \pm 1.32	97.17 \pm 3.57 (+0.72)	97.59 * \pm 0.85 (+1.15)
Computer Work	80.24 \pm 0.41	80.51 \pm 0.93 (+0.26)	87.48 * \pm 0.61 (+7.23)
Cooking	96.56 \pm 0.82	96.37 \pm 1.82 (−0.19)	98.19 \pm 4.12 (+1.63)
Doing Laundry	98.82 \pm 2.01	99.34 \pm 3.86 (+0.52)	99.35 \pm 1.82 (+0.53)
Dressing	99.08 \pm 2.60	99.17 \pm 2.81 (+0.09)	98.56 \pm 1.92 (−0.51)
Grooming	97.66 \pm 1.82	98.17 \pm 3.18 (+0.51)	98.64 \pm 1.92 (+0.98)
Eating	91.52 \pm 0.31	92.69 \pm 0.98 (+1.17)	96.23 * \pm 0.21 (+4.72)
In a Car	96.36 \pm 0.71	97.44 \pm 0.98 (+1.08)	97.65 * \pm 0.52 (+1.3)
In a Meeting	96.10 \pm 0.91	95.29 \pm 1.62 (−0.81)	97.72 * \pm 0.65 (+1.63)
In Class	96.39 \pm 0.31	95.52 \pm 0.84 (−0.87)	98.7 * \pm 0.59 (+2.31)
At Home	60.75 \pm 0.46	61.59 \pm 0.90 (+0.84)	64.48 * \pm 0.36 (+3.73)
At Main Workplace	81.98 \pm 0.68	81.14 \pm 1.83 (−0.83)	90.63 * \pm 0.51 (+8.65)
Lying Down	76.90 \pm 0.27	76.8 \pm 0.49 (−0.1)	78.28 * \pm 0.31 (+1.38)
On A Bus	98.20 \pm 0.81	98.87 \pm 1.64 (+0.67)	98.94 * \pm 0.71 (+0.74)
Exercise	95.86 \pm 0.93	95.69 \pm 1.72 (−0.17)	97.17 * \pm 0.84 (+1.31)
Indoor	56.89 \pm 0.62	57.44 \pm 0.85 (+0.55)	62.9 * \pm 0.39 (+6.01)
Outside	93.79 \pm 1.09	95.01 \pm 2.90 (+1.22)	95.99 * \pm 0.62 (+2.21)
Standing	84.59 \pm 0.82	86.43 \pm 0.90 (+1.84)	90.44 * \pm 0.41 (+5.85)
Running	98.97 \pm 0.27	99.24 \pm 0.36 (+0.26)	99.58 * \pm 0.18 (+0.61)
Shopping	97.75 \pm 0.42	98.39 \pm 0.26 (+0.65)	98.41 * \pm 0.93 (+0.66)
Sleeping	79.89 \pm 0.31	79.49 \pm 0.61 (−0.4)	80.63 * \pm 0.03 (+0.73)
Stairs Going Down	98.84 \pm 0.50	99.04 \pm 0.36 (+0.2)	99.84 * \pm 0.17 (+1.0)
Stairs Going Up	99.20 \pm 0.12	99.37 \pm 0.63 (+0.17)	99.82 * \pm 0.08 (+0.62)
Surfing the Internet	83.10 \pm 0.81	88.43 \pm 2.33 (+5.33)	89.11 * \pm 1.03 (+6.01)
Talking	81.62 \pm 1.03	79.96 \pm 0.92 (−1.67)	88.75 * \pm 0.61 (+7.12)
Toilet	97.04 \pm 0.89	98.63 \pm 1.58 (+1.58)	98.76 * \pm 0.73 (+1.72)
Walking	90.16 \pm 0.51	90.99 \pm 0.42 (+0.83)	91.84 * \pm 0.48 (+1.69)
Washing Dishes	97.90 \pm 2.01	98.29 \pm 3.61 (+0.38)	98.93 \pm 2.83 (+1.02)
Watching TV	89.19 \pm 1.23	92.42 \pm 1.97 (+3.23)	90.67 * \pm 1.32 (+1.48)
Average	89.87	90.60 (+0.73)	92.50 (+2.63)

1. The results (mean and standard deviation) are averaged over 30 individuals, on different activities (unit: %), and each activity is run 10 times to compute the mean and standard deviation. Transfer from all uses $S = \mathcal{U}$, while transfer at random transfers from a randomly chosen subset of individuals \mathcal{S} (of the same size as the set used by X-WRAP, for fairness). 2. **Bold** indicates the best-in-row, the numbers in parentheses indicate relative change compared to transfer from all, **red** highlights negative transfer, and **blue** highlights positive transfer larger than 1%. 3. * indicates significance based on Student t -test ($p < 0.05$).

4.3.3. X-WRAP Works Well for Data-Scarce Individuals

Considering that collecting labeled data for the target user is expensive in the real-world scenario, in this experiment, we evaluate the performance of X-WRAP applied to Coral when data for the target user are sparse. Specifically, we subsample the training data for testing users and set the rate of training data from 1% to 80%. As shown in Figure 4, X-WRAP can not only work well with large datasets but also on few data, especially on the activity in Figure 4.

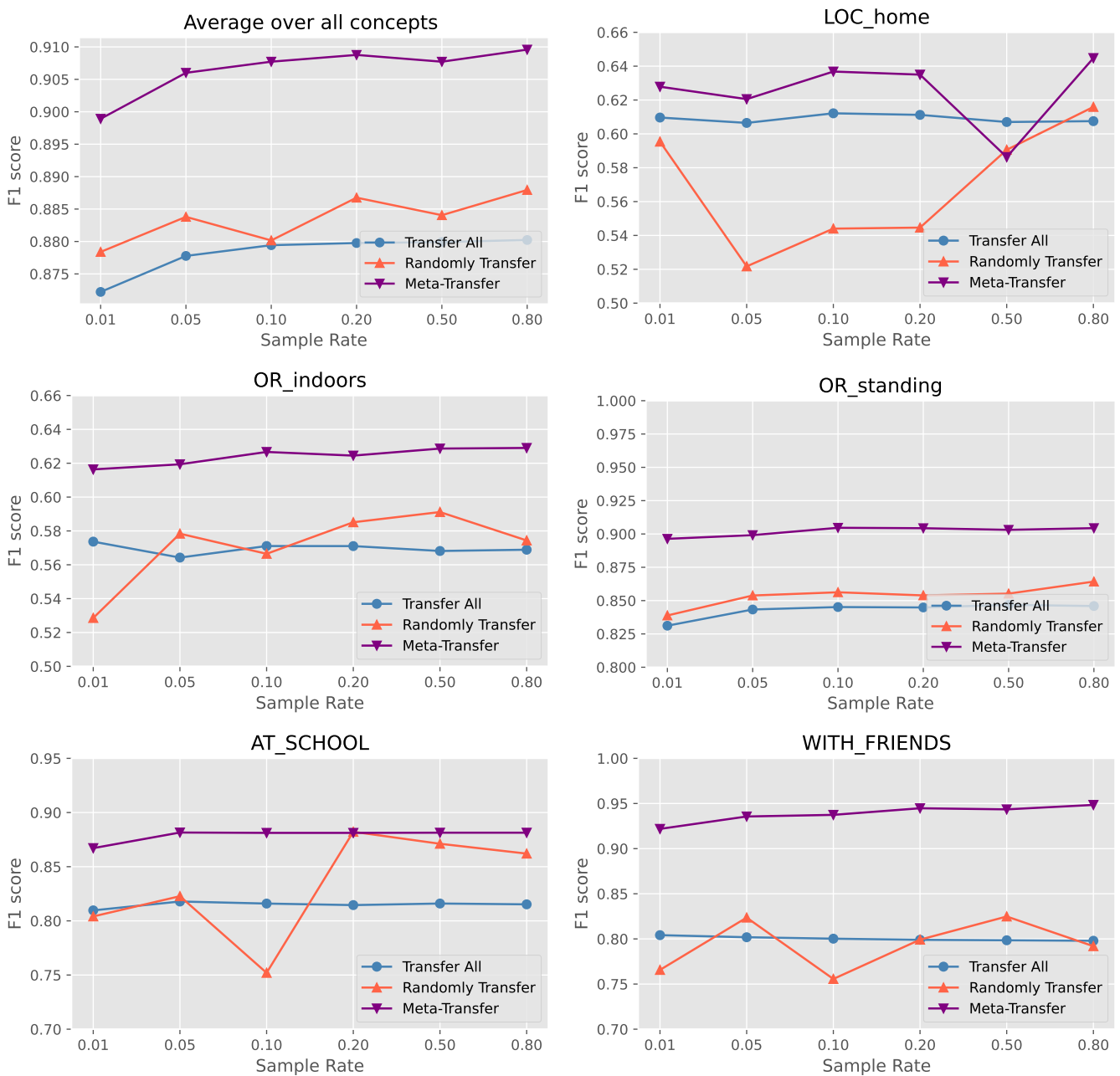


Figure 4. Performance of X-WRAP for increasing sample rate for the target individual (from 1% to 80%).

4.3.4. X-WRAP Enables Control in HAR System

In this final experiment, we study the possibility of empowering the system expert to control X-WRAP. We simulate the situation when certain sensors (i.e., GPS, accelerometer, watch accelerometer, and phone-state sensors) go awry by masking the data columns collected by these sensors. Then, we reduce the weights of similarity to simulate the operation of manipulating the parameter of X-WRAP. As shown in Figure 5, the performance of most activities can be improved after reducing the weight of similarity. On average, the improvement in performance ranges from 0.6% to 1.2%. In addition, there is a notable improvement in some activities. For instance, the activity “With Co-worker” can be improved more than 8% after changing the weights when the GPS sensor goes awry. In addition, “exercise” can be improved by around 3% after manipulating the weights when

the GPS sensor breaks. Thus, X-WRAP can empower the HAR system expert to obtain control when sensors go awry.

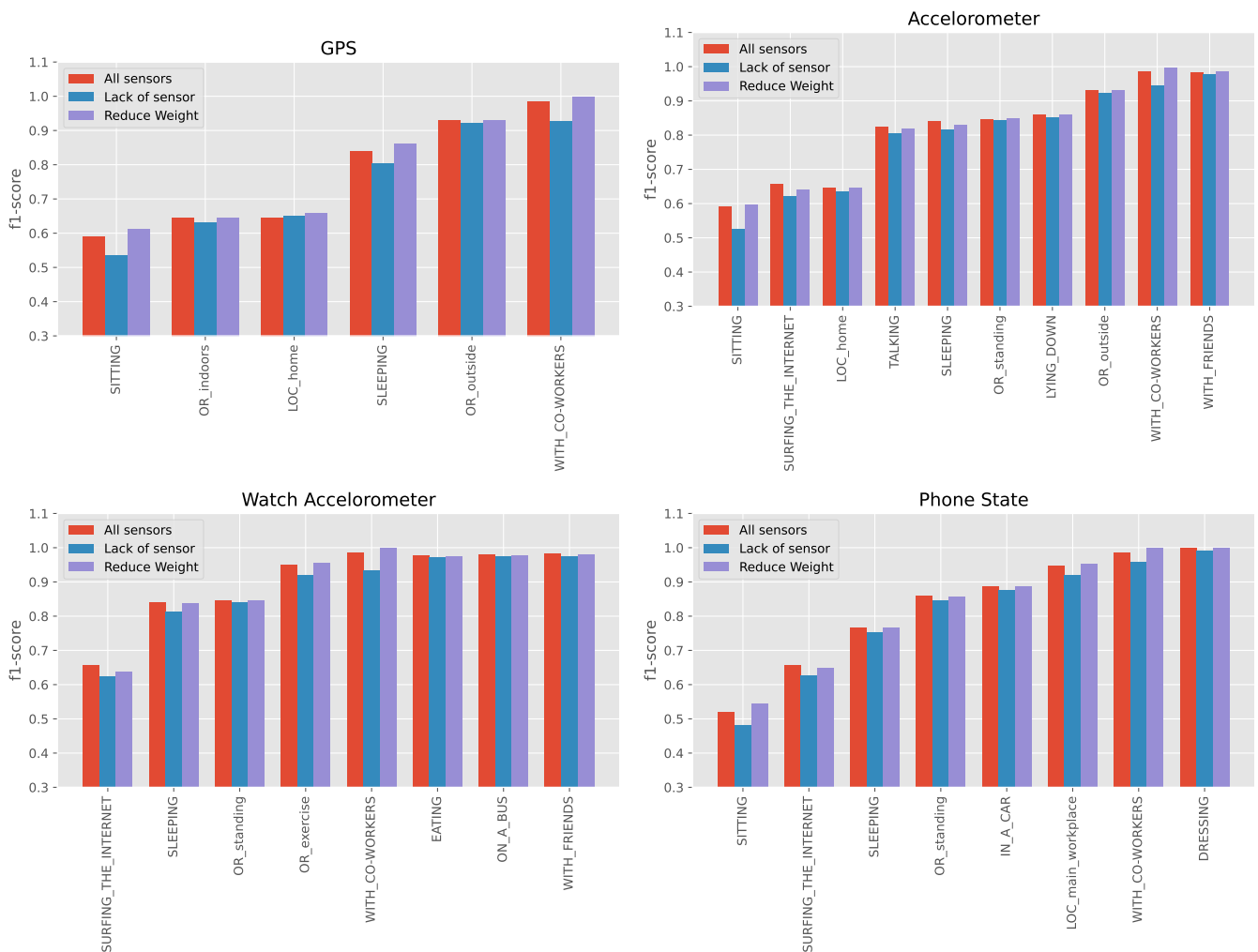


Figure 5. Performance of X-WRAP for reducing the weight of similarity when various sensors are removed.

4.3.5. X-WRAP Provides a Reliable Explanation for the Partial Transfer

In this subsection, we inspect the differences of meta-features on an activity level by visualizing the weights of meta-features in order to enhance the explainability of X-WRAP. In Figure 6, we visualize the weights of meta-features for each activity. For better visualization and comparison, we show four subplots, and the activities of each subplot are sorted according to the weights of certain meta-features. By exploring activity-specific weights of meta-features, some interesting insights into how the prediction of health was delivered in different individuals can be analyzed. It can be observed that predictability has the most positive contribution to source user selection, while similarity is the second most useful meta-feature and frequency contributes the least.

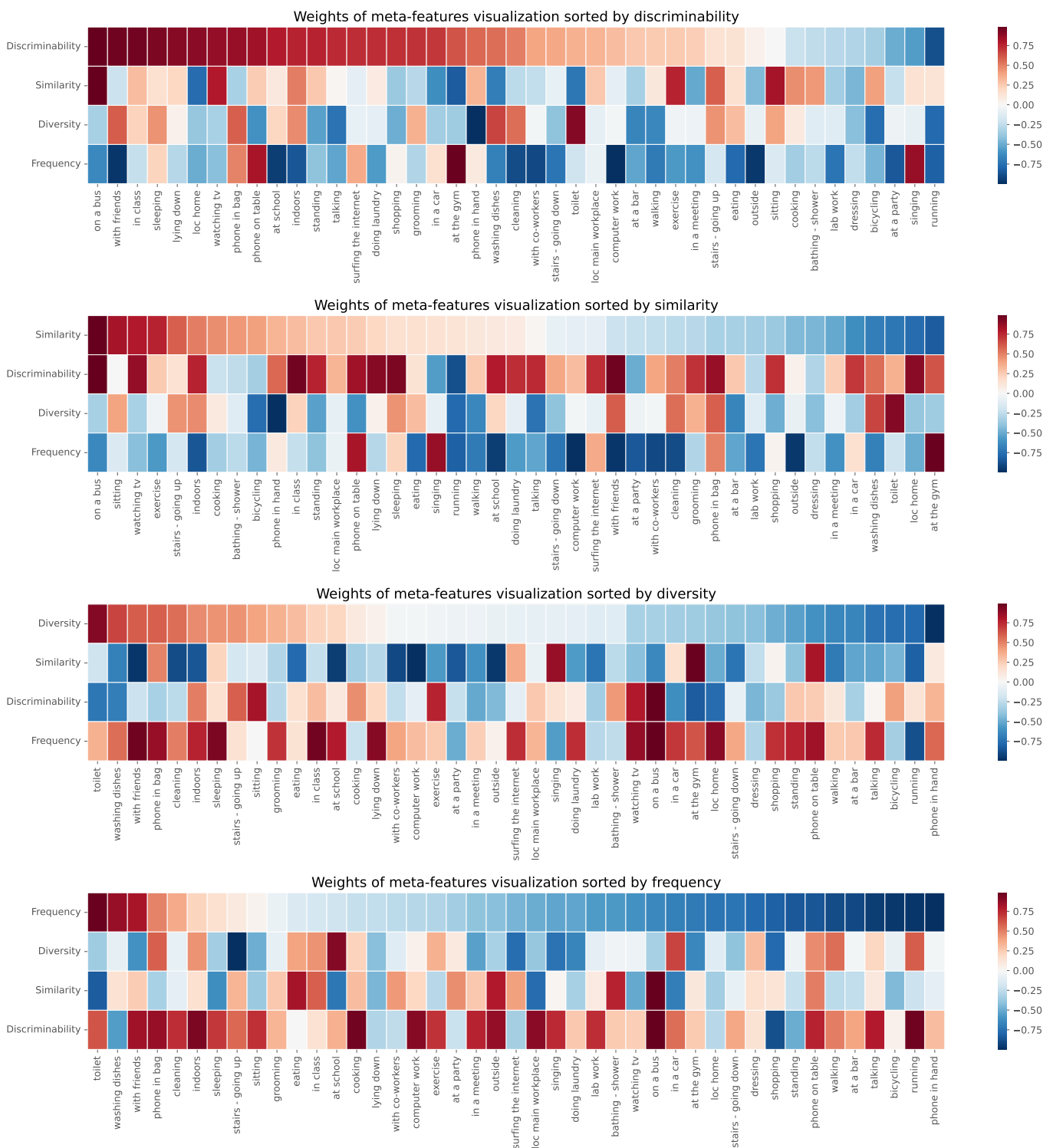


Figure 6. Weights of meta-features visualization.

In addition, we visualize the best parameter τ learned by X-WRAP, which indicates the number of selected source individuals. In Figure 7, the best numbers of source individuals for each activity are sorted. Generally, to achieve the best performance, the recognition of most activities needs to transfer only 1 or 2 source individuals out of 30. This indicates that there is sensory feature discrepancy between individuals for most of the activities and the individuals who have large differences with the target user would contribute to negative transfer. It can also be observed that those activities with simple and similar behavior

patterns require more source individuals. For instance, the activity “Lying Down” needs to transfer 18 users to achieve the best performance, and the recognition of “Sleeping” requires 13 individuals. Contrastingly, the activities that can be performed diversely require only one similar user to transfer, such as “AT School” and “Outsides”.

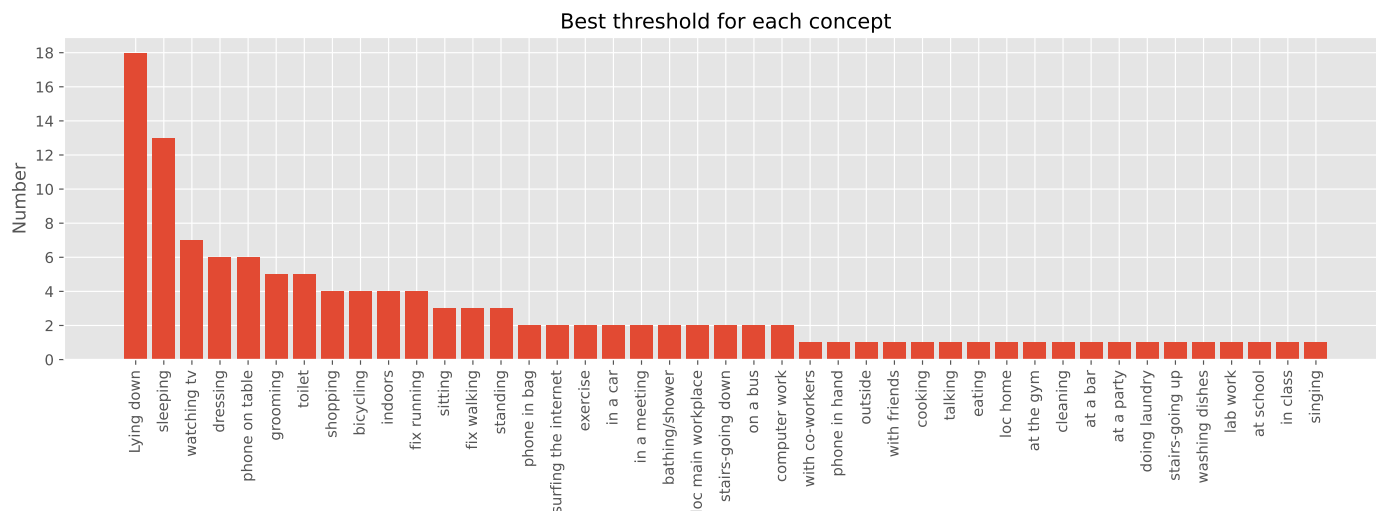


Figure 7. Parameter evaluation.

5. Discussion

X-WRAP is specifically designed to optimize the performance of a baseline transfer algorithm so as to generalize across individuals. X-WRAP builds on an interpretable linear function to rank the source domains and thus it enables stakeholders to understand and control (i.e., debug) the HAR system. However, there are some limitations in this work. Firstly, the importance of a feature is relative to the set of features that the model can use and may not reflect the actual importance of that feature according to the data-generating distribution. This is an intrinsic limitation of all statistical and causal explanation techniques [55]. Secondly, it may be tempting to use non-linear scoring functions (e.g., a multi-layer perceptron with one hidden layer); however, doing so can substantially hinder interpretability. Thirdly, training the X-WRAP with the DANN as the transfer module is expensive because of the deep learning structure. Thus, it is necessary to propose a more efficient way of training the model. As for the limitations caused by the dataset, using a larger dataset with more diverse characteristics can avoid the bias caused by the dataset. Finally, we point out that the interpretability could be further improved by encouraging X-WRAP to acquire sparse weights using, e.g., a sparsifying prior [72,82], but we leave this interesting direction to future work.

6. Conclusions

We introduced X-WRAP, a transfer learning approach for HAR that combines favorable accuracy with explainability and controllability. X-WRAP leverages meta-learning to acquire an interpretable source selection strategy that is (approximately) optimal for any given baseline transfer learning algorithm. Our experiments show that X-WRAP often improves the performance of the baseline while enabling human control.

Our approach can be extended in several directions. First and foremost, extending X-WRAP to sequential prediction would bring performance gains and only require minimal changes to the meta-learning procedure. Moreover, in many settings activities are organized hierarchically, implying that distinct activities may be statistically and logically related to each other. X-WRAP could be easily generalized to this setting, and the relations between activities could be leveraged to design more fine-grained meta-features. Finally, an interactive version of X-WRAP tailored for wearable personal assistants would also be

useful, as it would enable the machine to acquire supervision from the target individual that is most informative in terms of transfer learning.

Author Contributions: Conceptualization is done by Q.S., S.T. and H.X.; methodology is done by Q.S. and S.T.; software is done by Q.S. and S.T.; validation is done by Q.S., F.G. and S.T.; formal analysis is done by Q.S. and S.T.; investigation is done by F.G. and H.X.; resources is done by F.G. and H.X.; data curation is done by Q.S., S.T., F.G. and H.X.; writing—original draft preparation is done by Q.S. and S.T.; writing—review and editing is done by F.G. and H.X.; visualization is done by Q.S. and S.T.; supervision is done by F.G. and H.X.; project administration is done by F.G. and H.X.; funding acquisition is done by F.G. and H.X. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the National Natural Science Foundation of China: 62077027; European Union’s Horizon 2020 FET Proactive project: 823783.

Data Availability Statement: We use a public dataset ExtraSensory (<http://extrasensory.ucsd.edu/>).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, W.; Shen, Q.; Teso, S.; Lepri, B.; Passerini, A.; Bison, I.; Giunchiglia, F. Putting human behavior predictability in context. *EPJ Data Sci.* **2021**, *10*, 42. [[CrossRef](#)]
2. Hammerla, N.Y.; Fisher, J.; andras, P.; Rochester, L.; Walker, R.; Plötz, T. PD disease state assessment in naturalistic environments using deep learning. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
3. Intille, S. The Precision Medicine Initiative and Pervasive Health Research. *IEEE Pervasive Comput.* **2016**, *15*, 88–91. [[CrossRef](#)]
4. Gao, Y.; Long, Y.; Guan, Y.; Basu, A.; Baggaley, J.; Ploetz, T. Towards reliable, automated general movement assessment for perinatal stroke screening in infants using wearable accelerometers. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, New York, NY, USA, 29 March 2019; Volume 3, pp. 1–22.
5. O’Brien, J.; Gallagher, P.; Stow, D.; Hammerla, N.; Ploetz, T.; Firbank, M.; Ladha, C.; Ladha, K.; Jackson, D.; McNaney, R.; et al. A study of wrist-worn activity measurement as a potential real-world biomarker for late-life depression. *Psychol. Med.* **2017**, *47*, 93–102. [[CrossRef](#)] [[PubMed](#)]
6. Yao, X.; Plötz, T.; Johnson, M.; Barbaro, K.D. Automated detection of infant holding using wearable sensing: Implications for developmental science and intervention. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, New York, NY, USA, 21 June 2019; Volume 3, pp. 1–17.
7. Nguyen, L.N.N.; Rodríguez-Martín, D.; Català, A.; Pérez-López, C.; Samà, A.; Cavallaro, A. Basketball activity recognition using wearable inertial measurement units. In Proceedings of the XVI international conference on Human Computer Interaction, New York, NY, USA, 7–9 September 2015; pp. 1–6.
8. Lee, M.L.; Dey, A.K. Sensor-based observations of daily living for aging in place. *Pers. Ubiquitous Comput.* **2015**, *19*, 27–43. [[CrossRef](#)]
9. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–40. [[CrossRef](#)]
10. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [[CrossRef](#)]
11. Weiss, G.M.; Lockhart, J. The impact of personalization on smartphone-based activity recognition. In Proceedings of the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
12. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
13. Zhao, Z.; Chen, Y.; Liu, J.; Shen, Z.; Liu, M. Cross-people mobile-phone based activity recognition. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
14. Wang, J.; Chen, Y.; Hu, L.; Peng, X.; Philip, S.Y. Stratified transfer learning for cross-domain activity recognition. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–10.
15. Qin, X.; Chen, Y.; Wang, J.; Yu, C. Cross-dataset activity recognition via adaptive spatial-temporal transfer learning. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **2019**, *3*, 1–25. [[CrossRef](#)]
16. Garcia-Ceja, E.; Riegler, M.; Kvernberg, A.K.; Torresen, J. User-adaptive models for activity and emotion recognition using deep transfer learning and data augmentation. *User Model. User-Adapt. Interact.* **2020**, *30*, 365–393. [[CrossRef](#)]
17. Rosenstein, M.T.; Marx, Z.; Kaelbling, L.P.; Dietterich, T.G. To transfer or not to transfer. In Proceedings of the NIPS 2005 Workshop on Transfer Learning, Whistler, BC, Canada, 9 December 2005; Volume 898, pp. 1–4.
18. Brochu, E.; Cora, V.M.; de Freitas, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv* **2010**, arXiv:1012.2599.

19. Ling, B.; Intille, S. Activity Recognition from User-Annotated Acceleration Data. In Proceedings of the Pervasive Computing, Vienna, Austria, 21–23 April 2004.
20. Kim, E.; Helal, S.; Cook, D. Human activity recognition and pattern discovery. *IEEE Pervasive Comput.* **2009**, *9*, 48–53. [[CrossRef](#)]
21. Plötz, T.; Hammerla, N.Y.; Olivier, P. Feature Learning for Activity Recognition in Ubiquitous Computing. In Proceedings of the IJCAI 2011, 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
22. Shen, Q.; Teso, S.; Zhang, W.; Xu, H.; Giunchiglia, F. Multi-Modal Subjective Context Modelling and Recognition. In Proceedings of the 24th European Conference on Artificial Intelligence, Fourteenth International Workshop on Modelling and Representing Context, Santiago, Spain, 29 August 2020.
23. Figo, D.; Diniz, P.C.; Ferreira, D.R.; Cardoso, J.M. Preprocessing techniques for context recognition from accelerometer data. *Pers. Ubiquitous Comput.* **2010**, *14*, 645–662. [[CrossRef](#)]
24. Hammerla, N.Y.; Kirkham, R.; Andras, P.; Ploetz, T. On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. In Proceedings of the 2013 International Symposium on Wearable Computers, Zurich, Switzerland, 8–12 September 2013; pp. 65–68.
25. Zhao, Y.; Guo, S.; Chen, Z.; Shen, Q.; Meng, Z.; Xu, H. Marfusion: An Attention-Based Multimodal Fusion Model for Human Activity Recognition in Real-World Scenarios. *Appl. Sci.* **2022**, *12*, 5408. [[CrossRef](#)]
26. Francisco, O.; Daniel, R. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115.
27. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 351–360.
28. Ma, H.; Li, W.; Zhang, X.; Gao, S.; Lu, S. AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 3109–3115.
29. Shen, Q.; Feng, H.; Song, R.; Teso, S.; Giunchiglia, F.; Xu, H. Federated Multi-Task Attention for Cross-Individual Human Activity Recognition. In Proceedings of the 31st International Joint Conference on Artificial Intelligence, Vienna, Austria, 23–29 July 2022; pp. 3423–3429.
30. Shen, Q.; Feng, H.; Song, R.; Song, D.; Xu, H. Federated Meta-Learning with Attention for Diversity-Aware Human Activity Recognition. *Sensors* **2023**, *23*, 1083. [[CrossRef](#)]
31. Liu, Q.; Xue, H. Adversarial Spectral Kernel Matching for Unsupervised Time Series Domain Adaptation. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, ON, Canada, 19–26 August 2021.
32. Ozyurt, Y.; Feuerriegel, S.; Zhang, C. Contrastive Learning for Unsupervised Domain Adaptation of Time Series. *arXiv* **2022**, arXiv:2206.06243.
33. Wilson, G.; Doppa, J.R.; Cook, D.J. CALDA: Improving Multi-Source Time Series Domain Adaptation with Contrastive Adversarial Learning. *arXiv* **2021**, arXiv:2109.14778.
34. Sanabria, A.R.; Zambonelli, F.; Dobson, S.; Ye, J. ContrasGAN: Unsupervised domain adaptation in Human Activity Recognition via adversarial and contrastive learning. *Pervasive Mob. Comput.* **2021**, *78*, 101477. [[CrossRef](#)]
35. He, Q.Q.; Siu, S.W.I.; Si, Y.W. Attentive recurrent adversarial domain adaptation with Top-k pseudo-labeling for time series classification. *Appl. Intell.* **2022**, *2022*, 1–20. [[CrossRef](#)]
36. Eldele, E.; Ragab, M.; Chen, Z.; Wu, M.; Kwok, C.K.; Li, X. CoTMix: Contrastive Domain Adaptation for Time-Series via Temporal Mixup. *arXiv* **2022**, arXiv:2212.01555.
37. Hoffman, J.; Mohri, M.; Zhang, N. Algorithms and theory for multiple-source adaptation. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 237–270. [[CrossRef](#)]
38. Guo, H.; Pasunuru, R.; Bansal, M. Multi-source domain adaptation for text classification via distancenet-bandits. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 7830–7838.
39. Lin, C.; Zhao, S.; Meng, L.; Chua, T.S. Multi-source domain adaptation for visual sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 2661–2668.
40. Mancini, M.; Porzi, L.; Bulò, S.R.; Caputo, B.; Ricci, E. Boosting domain adaptation by discovering latent domains. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3771–3780.
41. Bhatt, H.S.; Rajkumar, A.; Roy, S. Multi-Source Iterative Adaptation for Cross-Domain Classification. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 3691–3697.
42. Chen, Q.; Liu, Y.; Wang, Z.; Wassell, I.; Chetty, K. Re-weighted adversarial adaptation network for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7976–7985.
43. Mancini, M.; Bulò, S.R.; Caputo, B.; Ricci, E. Adagraph: Unifying predictive and continuous domain adaptation through graphs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6568–6577.
44. Cao, Z.; Long, M.; Wang, J.; Jordan, M.I. Partial transfer learning with selective adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2724–2732.

45. Cao, Z.; Ma, L.; Long, M.; Wang, J. Partial adversarial domain adaptation. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8 September 2018; pp. 135–150.
46. Zhang, J.; Ding, Z.; Li, W.; Ogunbona, P. Importance weighted adversarial nets for partial domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8156–8164.
47. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [[CrossRef](#)]
48. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
49. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2010**, *22*, 199–210. [[CrossRef](#)]
50. Long, M.; Wang, J.; Ding, G.; Sun, J.; Yu, P.S. Transfer feature learning with joint distribution adaptation. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 1–8 December 2013; pp. 2200–2207.
51. Wang, J.; Chen, Y.; Hao, S.; Feng, W.; Shen, Z. Balanced Distribution Adaptation for Transfer Learning. In Proceedings of the IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; pp. 1129–1134.
52. Wang, J.; Chen, Y.; Feng, W.; Yu, H.; Huang, M.; Yang, Q. Transfer learning with dynamic distribution adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–25. [[CrossRef](#)]
53. Zhao, J.; Deng, F.; He, H.; Chen, J. Local Domain Adaptation for Cross-Domain Activity Recognition. *IEEE Trans. Hum.-Mach. Syst.* **2020**, *51*, 12–21. [[CrossRef](#)]
54. Qian, H.; Pan, S.J.; Miao, C. Latent independent excitation for generalizable sensor-based cross-person activity recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 2–9 February 2021; Volume 35, pp. 11921–11929.
55. Lipton, Z.C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **2018**, *16*, 31–57. [[CrossRef](#)]
56. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining blackbox models. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–42. [[CrossRef](#)]
57. Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W.; Müller, K.R. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.* **2019**, *10*, 1096. [[CrossRef](#)]
58. Kulesza, T.; Burnett, M.; Wong, W.K.; Stumpf, S. Principles of explanatory debugging to personalize interactive machine learning. In Proceedings of the 20th International Conference on Intelligent User Interfaces, New York, NY, USA, 29 March–1 April 2015; pp. 126–137.
59. Ross, A.S.; Hughes, M.C.; Doshi-Velez, F. Right for the right reasons: Training differentiable models by constraining their explanations. *arXiv* **2017**, arXiv:1703.03717.
60. Schramowski, P.; Stammer, W.; Teso, S.; Brugger, A.; Herbert, F.; Shao, X.; Luigs, H.G.; Mahlein, A.K.; Kersting, K. Making deep neural networks right for the right scientific reasons by interacting with their explanations. *Nat. Mach. Intell.* **2020**, *2*, 476–486. [[CrossRef](#)]
61. Sun, B.; Feng, J.; Saenko, K. *Return of Frustratingly Easy Domain Adaptation*; AAAI Press: Phoenix, AZ, USA, 2015.
62. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 2030–2096.
63. Sim, R.H.L.; Xu, X.; Low, B.K.H. Data valuation in machine learning: “ingredients”, strategies, and open challenges. In Proceedings of the 31st International Joint Conference on Artificial Intelligence, Vienna, Austria, 23–29 July 2022; pp. 5607–5614.
64. Jia, R.; Dao, D.; Wang, B.; Hubis, F.A.; Hynes, N.; Gürel, N.M.; Li, B.; Zhang, C.; Song, D.; Spanos, C.J. Towards efficient data valuation based on the shapley value. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, Naha, Japan, 16–18 April 2019; pp. 1167–1176.
65. Ghorbani, A.; Zou, J. Data shapley: Equitable valuation of data for machine learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 11–13 June 2019; pp. 2242–2251.
66. Xu, X.; Wu, Z.; Foo, C.S.; Low, B.K.H. Validation free and replication robust volume-based data valuation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 10837–10848.
67. Tay, S.S.; Xu, X.; Foo, C.S.; Low, B.K.H. Incentivizing collaboration in machine learning via synthetic data rewards. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 22 February–1 March 2022; Volume 36, pp. 9448–9456.
68. Xu, X.; Lyu, L.; Ma, X.; Miao, C.; Foo, C.S.; Low, B.K.H. Gradient driven rewards to guarantee fairness in collaborative machine learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 16104–16117.
69. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
70. Zhang, K.; Schölkopf, B.; Muandet, K.; Wang, Z. Domain adaptation under target and conditional shift. In Proceedings of the 30th International Conference on International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
71. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2015**, *104*, 148–175. [[CrossRef](#)]
72. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [[CrossRef](#)]
73. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should I trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.

74. Vaizman, Y.; Ellis, K.; Lanckriet, G. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Comput.* **2017**, *16*, 62–74. [[CrossRef](#)]
75. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782. [[CrossRef](#)]
76. Qian, H.; Pan, S.J.; Da, B.; Miao, C. A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 5614–5620.
77. Alajaji, A.; Gerych, W.; Buquicchio, L.; Chandrasekaran, K.; Mansoor, H.; Agu, E.; Rundensteiner, E. Domain Adaptation Methods for Lab-to-Field Human Context Recognition. *Sensors* **2023**, *23*, 3081. [[CrossRef](#)]
78. Khan, M.A.A.H.; Roy, N.; Misra, A. Scaling human activity recognition via deep learning-based domain adaptation. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–9.
79. Wang, J.; Feng, W.; Chen, Y.; Yu, H.; Huang, M.; Yu, P.S. Visual Domain Adaptation with Manifold Embedded Distribution Alignment. In Proceedings of the ACM Multimedia Conference (ACM MM), Seoul, Republic of Korea, 22–26 October 2018.
80. Wang, J.; Chen, Y.; Yu, H.; Huang, M.; Yang, Q. Easy Transfer Learning By Exploiting Intra-domain Structures. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019.
81. Komer, B.; Bergstra, J.; Eliasmith, C. Hyperopt-sklearn. In *Automated Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 97–111.
82. Ustun, B.; Rudin, C. Supersparse linear integer models for optimized medical scoring systems. *Mach. Learn.* **2016**, *102*, 349–391. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.