Generic semi-supervised adversarial subject translation for sensor-based activity recognition

Elnaz Soleimani, Ghazaleh Khodabandelou, Abdelghani Chibani, Yacine Amirat

Generic Semi-Supervised Adversarial Subject Translation for Sensor-Based Activity Recognition

Elnaz Soleimani[a], Ghazaleh Khodabandelou[a,*], Abdelghani Chibani[a], Yacine Amirat[a]

[a]*Laboratoire Images, Signaux et Systmes Intelligents (LISSI), University Paris-Est Crteil, France.*

Abstract

Performance of Human Activity Recognition (HAR) models, particularly deep neural networks, is highly contingent upon the availability of the massive amount of annotated training data. Though, data collection and manual labeling in the HAR domain are prohibitively expensive due to human resource dependence in both steps. Hence, domain adaptation techniques have been proposed to adapt the knowledge from the existing source of data. More recently, adversarial transfer learning methods have shown promising results for visual classification, yet limited for HAR problems, which are still prone to the unfavorable effects of the imbalanced distribution of samples. This paper presents a novel generic semi-supervised approach that takes advantage of the adversarial framework to tackle these shortcomings by leveraging knowledge from annotated samples exclusively from the source subject and unlabeled ones of the target subject. We conduct extensive subject translation experiments on three large, middle, and small-size datasets with different levels of imbalance to assess the robustness of the proposed model to the scale as well as imbalance in the data. The results demonstrate the effectiveness of our proposed algorithms over state-of-the-art methods, which led to up to 13%, 4%, and 13% improvement of our high-level activities recognition metrics for Opportunity, LISSI, and PAMAP2 datasets,

*Corresponding author
Email addresses: elnaz.soleimani@lissi.fr (Elnaz Soleimani), ghazaleh.khodabandelou@u-pec.fr (Ghazaleh Khodabandelou), chibani@u-pec.fr (Abdelghani Chibani), amirat@u-pec.fr (Yacine Amirat)

respectively.

Keywords: Human Activity Recognition, Transfer Learning, Adversarial Domain Adaptation, Generative Adversarial Network, Semi-Supervised Learning, Cross-Subject Transfer Learning

1. Introduction

Today, there is an unprecedented surge in artificial intelligence (AI) technologies in various areas such as image recognition, disease diagnosis, sentiment analysis and opinion mining, recommendation systems for products and services, automatic speech recognition, automated video surveillance, online support, ambient assisted living and robotics. The considerable progress witnessed in these areas has been made thanks to the contributions of the machine learning community and the rapid growth of processing units, which is stimulating big data collection and inspection. Compared with traditional machine learning approaches, the unparalleled performance of deep learning methods is made possible thanks to their capability of learning from a large number of samples. [1].

Nevertheless, there exist some fields of study which exceptionally face limitations on data acquisition and labeling due to practical constraints. Therefore, it is not feasible to provide enough labeled data in many cases. Taking the domain of sensors-based Human Activity Recognition (HAR) as an instance, the data collection and labeling tasks require the implication of human labelers and the use of pervasive and intrusive sensors such as video cameras, which makes it more challenging to preserve the privacy of human subjects. Furthermore, manual annotation is prohibitively expensive, especially for large-scale datasets. Most machine learning models work based on the primary condition that samples of training (source domain) and test (target domain) set must be drawn from the same distribution and feature space. However, in many real-world applications such as HAR, this assumption cannot be held and consequently causes a dramatic decrease in models performance [2]. In this context, how can a model trained on an initial amount of labeled data in a source domain be adapted to generalize on unlabeled data in a target domain? Today, it is considered that unlabeled data can give an indication of how a source domain and a target domain differ from each other. This information can be used by a classifier to modify its decisions in order to better generalize to the target domain [3]. Therefore, employing domain adaptation techniques could be beneficial for HAR in order to prevent models from suffering performance degradation when applied on new subjects or datasets.

Domain adaptation techniques have been considered as a way for automatic knowledge transfer from one domain to another to avoid the significant reduction of performance metrics, requiring as less amount of explicit training data as possible in the target domain [4]. It has been employed in almost every deep learning model when the target dataset does not contain enough labeled data [5]. Based on the availability of annotated data in the source or target domain, adaptation methods are categorized into supervised (inductive), semi-supervised (transductive), and unsupervised ones.

This paper develops a transductive method for HAR, which is based on the initial adversarial approach proposed in [6]. The latter outperforms only on datasets that are well-balanced and contain a huge amount of labeled samples. The proposed approach transfers knowledge from the source subject with distribution P_s to the target subject distributed with P_t using an adversarial framework. It has significantly enhanced the initial approach making it more generic and robust to prevent the decline in performance when there is not enough labeled data, and the imbalanced class distribution of the available samples imposes additional obstacles. The new proposed model uses annotated data exclusively from the source domain and unlabeled data from the target domain. This study focuses on datasets from body-worn sensors to cope with privacy. The main contributions of the present attempt lie in:

- Generic semi-supervised Adversarial Domain Adaptation technique that copes with cross-subject transfer learning problems in the Human Activity Recognition domain. The proposed model benefits from convolutional neural networks to perform more generalized automatic feature extraction, which is advantageous for classifying high-dimensional data of high-level activities. The proposed model is proven to be robust to the imbalance learning challenges by exploiting a micro-mini-batch learning strategy described in section 3.2.

- The proposed method has been extensively tested and evaluated by using cross-subject domain adaptation scenarios on three representative datasets; the two benchmark datasets, Opportunity [7] and PAMAP2 [8], and a dataset on rehabilitation self exercises that is collected for the purpose of the study at the LISSI laboratory [9]. The experimental results show that the final adapted classifier is always able to recognize with a reasonable rate the human activities using their high-dimensional feature vectors. Besides, the results illustrate the superiority of the proposed model over other state-of-the-art classification methods in terms of robustness w.r.t. weighted F1-score.

From this study, it has been concluded that the proposed approach can be applied to other transfer learning problems as well.

The overall structure of this paper takes the form of six sections, including this introductory one. Section 2 gives a brief overview of state-of-the-art concerning domain adaptation. The proposed model is presented in section 3. The fourth section lays out the experimental setup. The next section is dedicated to experiments, results and analysis of the evaluation. Finally, section 6 includes a discussion of the findings and future research into this area.

2. Related Work

Reducing the gap between source and target domains in machine learning in order to make machine learning algorithms able to self adapt their models to the target domain has been widely investigated these latest years. The approaches proposed in the state-of-the-art are classified according to three main categories, namely sample-based, feature-based, inference-based approaches [3]. The first one exploits the source distribution and targets the minimization of the target risk estimation without target labels. The second category uses techniques for matching the data distributions, shifting of data, sub-space mapping, etc., to make transformations that map source data into target data. The third category targets the adaptation of the inference procedure by including constraints related to the target domain in the source model, incorporating uncertainties through Bayesian inference, etc. Transfer learning, deep learning, adversarial learning are representative examples of techniques used in the HAR approaches. This classification is not mutually exclusive, but it provides a useful summary of domain adaptation methods. The proposed categorization demonstrates a modest number of conditions that allow domain adaptive classifiers to ensure performance. In this work, we organized our related work section, highlighting domain adaptation approaches from another perspective: The first part is mainly focused on the unsupervised and semi-supervised approaches. Next, the Adversarial methods are explored. In what follows, some of the most significant works related to our semi-supervised adversarial proposed approach are briefly explained.

Authors of [10], proposed a general unsupervised cross-domain learning framework that can exploit the intra-affinity of classes to perform intra-class knowledge transfer named Stratified Transfer Learning (STL). First, it obtains pseudo labels for the target domain by the majority voting technique. Then, it performs intra-class knowledge transfer iteratively to transform both domains into a common subspace. The model was extended to accomplish both source selection and knowledge transfer later in [11]. Although their research is dedicated to the HAR domain, the evaluation is limited to the adaptation of body parts on the same person or similar body parts on different person. Besides, the model utilized time domain and frequency domain feature extraction as their input which is not suitable for recognition of high-level activities with complicated patterns. Another unsupervised source selection algorithm was proposed in [12]. This algorithm is able to select the most similar K source domains from a list of available domains. Next, an effective Transfer Neural Network performs knowledge transfer for Activity Recognition (TNNAR) by capturing both the time and spatial relationship between activities during transfer. TNNAR was evaluated by body-part translation experiments, which provided more amount of samples for the model while the samples were limited to 4 common classes of activity on each dataset. Therefore, the effectiveness of TNNAR is not guaranteed on the imbalanced small-size datasets. The Geodesic Flow Kernel (GFK) presented in [13], models domain shift by integrating an infinite number of subspaces that characterize changes in geometric and statistical properties from the source to the target domain. GFK model learns feature representations that are invariant across domains. These new shallow representations possibly have limited transferability. On the contrary, the non-linearity of neural networks in our proposed model can avoid this problem.

Authors in [14] investigated different semi-supervised active learning strategies to scale activity recognition and proposed a dynamic k-means clustering-based active learning approach. Using active learning alleviates the labeling effort of data collection in the activity recognition pipeline. In spite of all improvements provided in the proposed model, such as computational complexity mitigation, the method is yet prone to under-fitting down to its limited generalization ability [12]. Pan et al. introduced Transfer Component Analysis (TCA), a semi-supervised approach for domain adaptation, which learns transfer components across domains in a reproducing kernel Hilbert space using maximum mean discrepancy [15]. The subspace spanned by these transfer components, preserves data properties and distributions of different domains close to each other. Therefore, using the new representations in this subspace, we can apply standard machine learning methods to train classifiers or regression models in the source domain and test them in the target domain. It means TCA applies representation transfer, and the classification task should be learned in another step. Besides, TCA only learns a global domain shift and does not fully consider the intra-class similarity [10]. Another semi-supervised domain adaptation method was introduced in [16] to remove the raining effect from the images using real unlabeled rainy images and paired synthetic (rainy, de-rained) images. However, the learning task is limited to learning the characteristics of 2 classes (rainy and de-rained), and the performance could not be generalized well to multi-class problems. Some approaches perform transfer learning by reweighing or taking samples of the source domain. Balanced Distribution Adaptation (BDA) is proposed in [17], which adaptively leverages the importance of the marginal and conditional distribution discrepancies. Based on BDA, a novel Weighted Balanced Distribution Adaptation (W-BDA) algorithm was proposed to tackle the class imbalance issue in transfer learning by considering not only the distribution adaptation between domains but also adaptively changing classes' weights. However, when there is a greater discrepancy between both distributions, this approach cannot evaluate their relative importance as it treats the two distributions equally. Besides, the method is designed only for the original space, where feature distortion will adversely affect the performance [18]. On the contrary, our proposed method could benefit from the alignment in the feature space of the convolutional neural networks that are highly competent for extracting discriminable features. TransAct is another transfer learning-enabled activity recognition model introduced in [19] that mitigates the degradation of recognition performance confront with activities with limited labeled samples. The model is introduced as a semi-supervised approach, though weak supervision is required in the target domain. It addressed the challenges by augmenting the Instance-based Transfer Boost algorithm with k-means clustering. This model is designed only to compensate for the domain shift in activity level.

Since the introduction of Generative Adversarial Networks (GANs) [20], adversarial machine learning is gaining increasing attention and achieving impressive performance in a wide variety of domains such as medicine [21, 22], text, and image processing [23, 24] and architecture [25, 26]. The idea behind GANs is to put generator and discriminator algorithms against each other in order to differentiate between the generated samples and real-world samples. Deep learning is used to build discriminators that continuously learn the best set of features, making it hard for the generator to pass the discriminator test [27]. The preliminary attempts of applying adversarial machine learning for HAR tackled the problem of producing synthetic data. In the latter, ground-truth annotations are generated automatically and the generator has complete knowledge about the target system [28, 29]. Nonetheless, enhancing the classification methods remains the main important challenge in this field.

With respect to knowledge transfer, some attempts applied adversarial machine learning for training a robust deep network by way of a common representation for knowledge transfer [30, 31]. Benefiting from the generalized adversarial framework, the generator component can be independently used as the target representation in order to reduce the difference between the training and test domain distributions and thus improve generalization performance. Bousmalis et al. proposed a generative adversarial network (GAN)-based method that adapts source domain images to appear as if drawn from the target domain [32]. Having transferred samples by this instance-based adaptation method, machine learning models can be trained over the target domain. He et al., proposed a semi-supervised generative adversarial framework for visual-audio transfer to enhance the Audio Emotion Recognition (AER) performance by exploring the visual-audio correlation [33]. The model takes the labeled videos (source domain) and partially labeled audio (target domain). While this setup lets the model benefit from the large labeled visual datasets available, it imposes a weak supervision requirement on the target domain that is not affordable in our problem setting. Besides, this model could be prone to an imbalanced distribution of the classes. Authors in [6] introduced SA-GAN, an adversarial transfer learning approach which provides enough data to train a classifier on the target domain. DANN is another semi-supervised domain adaptation approach in which a common embedding encoder is trained under an adversarial objective to generate domain invariant features between the source and target domain that have similar but different distributions [34]. VADA model [35] proposes further to minimize the conditional entropy of the target domain, based on the cluster assumption [36] and perform a virtual adversarial training on each domain. However, considering independent training constraints for each domain in both VADA and DANN models, leaves the important interplay between the source and target domains unexplored and may significantly limit the performance [37]. To overcome this limit, the IIMT model [37] uses the Mixup augmentation approach to compute inter-domain linearly interpolated examples from random pairs of source and target domains to boost the training set. On the other hand, such an inter-domain mixup would not preserve the class-specific features and may degrade the robustness of the model against the target domains of a higher distance from the source.

Despite all the efforts dedicated to developing Adversarial Domain Adaptation approaches for HAR, the majority of the attempts are exploiting only data produced by vision sensors. Due to the continuous nature of the sensor's data, the approaches exploiting wearable sensors such as accelerometers have a high tendency of being trapped by mode collapse problem in which the generator can only generate a single or few classes (modes) of the data. Besides, they are not immune to the imbalanced class distribution problem; thus, the classes with lower probability density values may have less chance of being generated. We summarized our state-of-the-art study in Table 1. The prior art could be categorized into 4 main groups according to the supervision level; Supervised, Weakly supervised, Semi-supervised, and Unsupervised. Supervised and weakly supervised approaches, despite their competitive performance, are not aligned with our problem setting, where there is no annotated sample available in the target domain. Despite categorizing themselves as unsupervised, some approaches such as [33, 38] take partial supervision in the target domain that does not fit the requirement of our problem. Other semi-supervised methods such as [6, 12, 14] are not designed to tackle the imbalanced class distribution problem. Unlike unsupervised approaches such as [13, 10], which are prone to under-fitting due to the limited learned representation, our proposed model benefits from convolutional architecture to extract high-level features to better recognize more abstract activities.

Table 1: State-of-the-art study summary

Reference	Supervision	Domain	Dataset/Modality
[10], [11]	Unsupervised	HAR	OPPORTUNITY, PAMAP2, and UCI DSADS
[12]	Unsupervised	HAR	OPPORTUNITY, PAMAP2, and UCI DSADS (Accelerometer, Gyroscope, Magnetometer)
[13]	Unsupervised	Vision	PASCAL, PASCAL, Caltech101, Amazon, Webcam, and DSLR
[22]	Unsupervised	Medical Image	Clinical high-resolution SD-OCT volumes of the retina with 49 B-scans
[6]	Semi-Supervised	HAR	Opportunity
[14]	Semi-Supervised	HAR	Custom Dataset (PIR motion and object sensors)
[15]	Semi-Supervised	Text Classification / Wifi Localization	Synthetic datasets, 2007 IEEE ICDM Contest, and 20-Newsgroups
[16]	Semi-Supervised	Vision	Rain800, Rain200H, DDN-SIRR
[17]	Semi-Supervised	Vision	USPS + MNIST, COIL20, and Office + Caltech
[28]	Semi-Supervised	HAR	Daily and Sports Activities Dataset (Inertial Sensors)
[30]	Semi-Supervised	Vision	MNIST, USPS, and SVHN
[32]	Semi-Supervised	Vision	MNIST, USPS, MNIST-M
[33]	Semi-Supervised	Vision / Audio	IFER, RAVESS, RAVDESS, and CREMA-D
[34]	Semi-Supervised	Vision / Text	Amazon Reviews, Office
[35]	Semi-Supervised	Vision / Wife Localization	MNIST, SVHN, SYN DIGITS, SYN SIGNS, GTSRB, CIFAR-10, STL-10.
[37]	Semi-Supervised	Vision / HAR	MNIST, SVHN, SYN DIGITS, CIFAR-10, STL-10, OPPORTUNITY, WiFi
[21]	Supervised	Medical Image	ADNI + Custom dataset
[23]	Supervised	Vision / Text	Caltech-UCSD Birds dataset, Oxford-102 Flowers
[24]	Supervised	Vision / Text	CUB, Oxford, MS COCO
[29]	Supervised	HAR	HASC2010corpus (Accelerometer)
[19]	Weakly Supervised(Few labeled in target domain)	HAR	HAR, MHealth, and DailyAndSports
[31]	Weakly Supervised	Vision	MNIST, UCF-101, ILSVRC2012, Google Street View House Numbers (SVHN)

3. Method

3.1. Objective

Sensor-based HAR can be formulated as predicting current activity according to a sequence of sensors outputs x_i [39]:

$$f : W \to Y \mid W = \{x_i : i = 1, ..., n\} \tag{1}$$

Whereas the correct activity sequence (ground truth) and input window of size n denoted as Y and W respectively. Any HAR related dataset has a finite amount of samples that are obtained from a limited number of human subjects. However, considering the requirements of applying HAR in real-world conditions, it is more interesting to evaluate the performance of an HAR model against many human subjects whose behaviors' data have not been included in the training dataset.

Let us define a learning domain D and a learning task T as bellow [2]:

$$D = \Big(X, P(X)\Big), T = \Big(Y, P(Y|X)\Big) \tag{2}$$

Where feature space X follows the distribution P and $P(Y|X)$ formulates conditional distribution of label space Y. The shift between the domains may root in the learning task, domain task, or both. Furthermore, the source and the target domains may be dissimilar also in terms of class distributions, typically known as *class imbalance* problem in machine learning. The conditional distributions of feature values could be the same in source and target domains, yet the labels may not follow the same distribution in both domains.

Subject level domain adaptation concentrates on the generalization of the knowledge a machine learning model. The latter is trained from a known subject and should be extended to unknown or unseen subject. Let us consider an HAR system that is supposed to recognise the activities of the inhabitants of smart homes. The inhabitants are considered as new subjects from the perspective of the HAR system. Even if the HAR system can be setup to temporarily collect data and learn inhabitants activities in the same time in a kind of a system initialisation mode, the annotation of the collected samples, by human experts or the inhabitant themselves to apply supervised machine learning will be infeasible. In this case the appropriate approach is a semi-supervised learning, where labeled data is provided in the source domain (subject) while target domains' samples are label excluded. Formally, the objective is to adapt the target domain $D_t = \big(X_t, P(X_t)\big)$ to the source domain $D_s = \big(X_s, P(X_s)\big)$ so as to have enough labeled data to train a HAR model on target domain. It has commonly been assumed that in the class imbalance domain adaptation problems, $P(X_s|Y_s = y_i) = P(X_t|Y_t = y_i)$ is held for all classes i, though the distribution of classes may not be the same in both domains [40]. Fig. 1 exemplifies this concept in the datasets used for this study. Taking any pair of subjects as the source and target domain, there exists a shift between the class probability distribution of both domains, which means $P(Y_s) \neq P(Y_t)$.

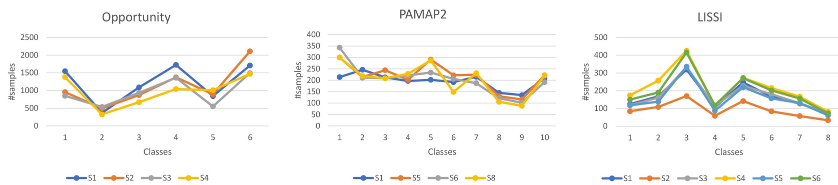


Figure 1: The class distribution of each subject P(Y) in Opportunity, PAMAP2 and LISSI dataset.

Classification of data with imbalanced class distribution along with absence of the labels in the target domain, may pose a significant drawback of the performance attainable by the adaptation process.

3.2. Adversarial Adaptation

Domain adaptation methods can be sort out into 4 categories based on the type of knowledge transferred: Instance, feature representation, parameter, and relational transfer [6]. The first two categories focus on drawing the samples of both domains closer, by direct transformation or finding a common representation respectively. Parameter and relational transfer methods transform prior knowledge and parameters and data relationship between domains. The proposed solution is kind of instance transfer except that it combines the data transformation and classifier training procedure.

More formally, let us consider $X^s = \{(x_s, y_s)^i \mid i = 0 \rightarrow n_s\}$ represents the set of n_s labeled samples from the source domain $D_s = (X_s, P(X_s))$ and $X^t = \{(x_t)^i \mid i = 0 \rightarrow n_t\}$ denotes the set of n_t unlabeled samples from the target domain $D_t = (X_t, P(X_t))$. The proposed adversarial adaptation model consists of a Generator (G), Discriminator (D), and Classifier (C). The generator $G(x, z; \theta_G)$ is a differentiable function represented by a Convolutional Neural Network (CNN) that generate synthetic data, called also fake samples, by using the input and noise vector, whereas z denotes a random noise vector. The discriminator $D(x; \theta_D)$ is defined as a CNN that outputs a single scalar indicating the probability that x came from the target domain rather than generator. The classifier $C(x; \theta_C)$ is also a CNN predicts the class of the input.

These elements are playing a min-max game together based on the cost function $\mathbb{J}$ which combines the loss functions of adversary and classification tasks as follows [6]:

$$\begin{aligned}&\mathbb{J}(G, C, D, X^s, X^t) = \\ &\min_{G,C} \max_{D} \mu \mathbb{J}_D(D, G, X^t, X^s) + \lambda \mathbb{J}_C(C, G, X^s)\end{aligned} \tag{3}$$

The impact of the classification $\mathbb{J}_C$ and adversary $\mathbb{J}_D$ task loss on the generation task are reflected and controlled by μ and λ coefficient respectively in $\mathbb{J}$. Generator supposed to generate artificial data which are similar to the samples of the target domain by descending on the gradient of $\mathbb{J}(G, C, D, X^s, X^t)$:

$$\begin{aligned}&\frac{\partial}{\partial \theta_G} \mathbb{J}(G, C, D, X^s, X^t) = \\ &\mu \nabla_{\theta_G} \mathbb{J}_D(D, G, X^t, X^s) + \lambda \nabla_{\theta_G} \mathbb{J}_C(C, G, X^s)\end{aligned} \tag{4}$$

During the training phase, these fake generated samples getting as similar as possible so that the discriminator will not be able to discriminate them from the original target data.

The training samples are commonly feeding into the adversarial frameworks in mini-batch form, to avoid the mode collapse problem in which, the generator learns only to generate fake samples from a few classes (modes) of the data distribution, albeit the samples from the ignored modes appears in the training set [41]. Therefore, the generator collapses into the few modes that discriminator assumes them highly realistic. In contrary, feeding the samples in mini-batch to the discriminator rather than in isolation, gives a broader horizon to the discriminator and possibly avoids the mode collapse. However, it would not be practical enough in all cases, especially for the highly imbalanced small-size datasets such as LISSI. Mini-batch selection of the dataset would even aggravate the problem in practice since it scales down the sample size. It is highly probable that the less populated classes be left without a representative in some batches, as the sample size shrinks and the proportions of the classes in the sample space cannot be taken for granted. Consequently, the discrimination of that certain classes would be unconcerned in the gradient computation of the batch due to the disappearance of their samples.

3.3. Proposed method

To tackle this issue, we applied a *micro-mini batch* strategy of learning. Each mini-batch consists of $\mathbb{C}$ micro-batches whereas $\mathbb{C}$ is the number of classes of activities in the dataset. Micro-batches contains m samples randomly drawn without replacement from each one of $\mathbb{C}$ classes while m can be set by the value of the least populated class population in the dataset. In this way, the proposed adversarial approach do not only prevent the Mode Collapse and its related issues but also enables a kind of concurrency of instance transfer and re-training of the target domain's classifier. Every iteration, the generator and the classifier modules concurrently aim to improve through our adversarial framework; meaning that the generator is supposed to generate more realistic samples (instance transfer) and the classifier is supposed to show better classification performance (classifier update).

The optimization problem for discriminator can be solved by ascending the gradient of a Mean Squared Error (MSE) loss function $\mathbb{J}_D$:

$$\begin{aligned}\mathbb{J}_D(D, G, X^t, X^s) = \frac{1}{m\mathbb{C}} \sum_{j=1}^{\mathbb{C}} \sum_{i=1}^{m} \Bigg[\Big(1 - D\big(x_t^{(i)}\big)\Big)^2_{x_t^{(i)} \in b_t^j} \\ + \Big(1 + D\big(G(x_s^{(i)}, z)\big)\Big)^2_{x_s^{(i)} \in b_s^j} \Bigg]\end{aligned} \tag{5}$$

where b_s^j and b_t^j refer to the j-th micro part of the current mini-batch from the source X^s and target X^t domain samples, respectively. The ground truth for real/fake indication is considered as ± 1.

Correspondingly, the classifier attempts to assign a right label to its inputs including source domain data and the synthetic data generated by optimizing the cross-entropy loss function $\mathbb{J}_C$:

$$\begin{aligned}\mathbb{J}_C(C, G, X^s) = \frac{1}{m\mathbb{C}} \sum_{j=1}^{\mathbb{C}} \sum_{i=1}^{m} \\ \Big[- y_s^{(i)} \log C\big(x_s^{(i)}\big) - y_s^{(i)} \log \Big(C\big(G(x_s^{(i)}, z)\big)\Big) \Big]_{x_s^{(i)}, y_s^{(i)} \in b_s^j}\end{aligned} \tag{6}$$

Both the discriminator and classifier components get updated based on their gradient $\frac{\partial}{\partial \theta_D}\mathbb{J}_D$ and $\frac{\partial}{\partial \theta_C}\mathbb{J}_C$, yet any gradient-based optimizer can be exploited for each component independently. Finally, when the training loss values converge, the training phase can be terminated and the classifier component will be functional independently. It is interesting to note that in the absence of the micro-mini batch training strategy, training batches b_s and b_t are random subsets of data of size m. The average loss value of the batch obtained by equations (5), (6), and (3) will be used to update the model. However, dealing with class imbalanced data, the classes with lower number of samples would have a much higher chance to be under-represented (or even not represented at all) in a small batch of data, by random selection approach and hence the update would be blind to the information of the missing class/classes.

Algorithm 1: Micro mini-batch training of the proposed model

Input: X^s, X^t, Y^s

Output: C

for *number of training iterations or until convergence* **do**

  for *number of mini-batches of data* **do**

1. Sample a micro-batch of size m from X^s, X^t, Y^s of class $j = 1, ..., \mathbb{C}$ of data:
   $b_s^j = \{(x_s^{(1)}, y_s^{(1)}), (x_s^{(2)}, y_s^{(2)}), ..., (x_s^{(m)}, y_s^{(m)})\}$,
   $b_t^j = \{x_t^{(1)}, x_t^{(2)}, ..., x_t^{(m)}\}$
2. Make a mini-batch of source $\{b_s^1, b_s^2, ..., b_s^{\mathbb{C}}\}$ and target domain $\{b_t^1, b_t^2, ..., b_t^{\mathbb{C}}\}$ samples
3. Update discriminator D by ascending its stochastic gradient:
   $$\nabla_{\theta_D} \frac{1}{m\mathbb{C}} \sum_{j=1}^{\mathbb{C}} \sum_{i=1}^{m} \Big[\Big(1 - D\big(x_t^{(i)}\big)\Big)^2_{x_t^{(i)} \in b_t^j} + \Big(1 + D\big(G(x_s^{(i)}, z)\big)\Big)^2_{x_s^{(i)} \in b_s^j} \Big]$$
4. Update classifier C by ascending its stochastic gradient:
   $$\nabla_{\theta_C} \frac{1}{m\mathbb{C}} \sum_{j=1}^{\mathbb{C}} \sum_{i=1}^{m} \Big[- y_s^{(i)} \log C\big(x_s^{(i)}\big) - y_s^{(i)} \log \Big(C\big(G(x_s^{(i)}, z)\big)\Big) \Big]_{x_s^{(i)}, y_s^{(i)} \in b_s^j}$$
5. Update generator G by ascending its stochastic gradient:
   $$\nabla_{\theta_G} \frac{1}{m\mathbb{C}} \sum_{j=1}^{\mathbb{C}} \sum_{i=1}^{m} \Bigg[\lambda \Bigg(-y_s^{(i)} \log C\big(x_s^{(i)}\big) - y_s^{(i)} \log \Big(C\big(G(x_s^{(i)}, z)\big)\Big) \Bigg) + \mu \Bigg(\Big(1 - D\big(x_t^{(i)}\big)\Big)^2 + \Big(1 + D\big(G(x_s^{(i)}, z)\big)\Big)^2 \Bigg) \Bigg]$$
   $$x_s^{(i)}, y_s^{(i)} \in b_s^j, x_t^{(i)} \in b_t^j$$

Return classifier C.

Algorithm 1 outlines the training steps. In summary, each iteration of the training procedure consists of 3 steps for the mini-batch update of discriminator D, classifier C, and generator G, respectively. Reordering the steps may affect convergence flow. In theory, training the discriminator, classifier, and generator reduces the error, so the order of training should not matter. However, from a practical standpoint, the degree of error reduction for each of them depends on the training order. The components of the model together struggle to close in the distribution of target domain samples on those of source domain where the labels are available. Having samples with approximately the same distribution, source domain labels are compatible to be exploited in supervised training of the classifier. As the distribution of generated samples $P(G(X^s, z))$ getting closer to $P(X^t)$, the performance of the classifier C improves since the source labels more deeply cohere with target inputs.

3.4. Adversarial Training Techniques

Training GANs includes obtaining a Nash equilibrium to a two-player non-cooperative game while each player wishes to minimize its own cost function [42]. A Nash equilibrium for our problem is a triple $(\theta_D, \theta_G, \theta_C)$ such that $\mathbb{J}_D(G, D, X^s, X^t)$ is at the maximum with respect to θ_D, and $\mathbb{J}(G, C, D, X^s, X^t)$ and $\mathbb{J}_C(C, G, X^s)$ are at the minimum with respect to θ_G and θ_C, respectively. Thought, finding Nash equilibria is quite problematic since maximizing $\mathbb{J}_D$ contradicts the minimization of two remaining cost functions. In addition, a very confident discriminator pose several challenges to adversarial training such as vanishing gradients phenomena on the generator, as well as instability of the generator gradient updates [43]. Different techniques have been utilized in order to overcome these challenges [42, 43]. The mini-batch discrimination strategy applied to avoid the mode collapse failure by allowing the discriminator to look at multiple samples in combination, rather than in isolation.

Label smoothing is another technique that replaces the 0 and 1 targets for a classifier with smoothed values, such as 0.9 or 0.1. Replacing positive class labels with α and negative class labels with β, the optimal discriminator is formulated as the following:

$$D^*(x) = \frac{\alpha P(X_t = x) + \beta P(X_g = x)}{P(X_t = x) + P(X_g = x)} \tag{7}$$

Adding continuous noise to the inputs of the discriminator can fix the instability and vanishing gradients issues and smoothen the distribution of the probability mass. Therefore, the optimal discriminator is re-written as follows:

$$D^*(x) = \frac{\alpha P(X_t = x + \epsilon) + \beta P(X_g = x + \epsilon)}{P(X_t = x + \epsilon) + P(X_g = x + \epsilon)} \tag{8}$$

Whereas ϵ denotes a continuous random variable with density P_ϵ. In our case, we chose the noise inputs from uniform distribution within the interval [0, 1).

4. Experimental Setup

4.1. Dataset

To evaluate the performance and robustness of the proposed model, we selected three representative datasets of large, medium, and small sizes. Several HAR-related datasets are publicly available, though they mostly have their samples distributed into many subjects. The goal is to take each subject as a domain and transfer the knowledge among them. Hence, we consider the *average sample size per subject* in the datasets to measure the size and select appropriate datasets. The experiments have been conducted on Opportunity and PAMAP2 benchmark datasets [7, 8] and the LISSI dataset [9]. The latter is collected for the study at the LISSI laboratory. Those datasets are briefly introduced in the following lines.

- PAMAP2: Totally 9 subjects were participating in the data collection of PAMAP2 following a protocol of 12 activities and 6 optional ones while wearing 3 IMUs and a Heart Rate-monitor [8]. Among them, 4 subjects with the more well-distributed classes of samples have been selected for evaluation.

- Opportunity: It consists of wearable sensors' output worn by 4 human subjects who were performing predefined unscripted daily living activities [7]. Four groups of activity are defined in this dataset based on the abstraction level. Generally, the more abstract activities may have more complicated and diverse patterns. To exemplify, let us compare the high-level *Coffee*

Time to the low level *Sit* locomotion activity. Since the activities are not scripted, there are different attitudes in making and drinking coffee. Therefore, the related patterns would be more challenging to be recognized by machine learning techniques. The Opportunity dataset contains 5 classes of high-level activities that we aim to recognize in this research.

- LISSI: This dataset has been collected in the context of the Medolution European project [1] in order to develop HAR methods that can be applied for the remote monitoring of the rehabilitation self exercises performed at home. The dataset consists of sensors' data of human subjects, which repeated a complete sequence of five rehabilitation self exercises. The subjects are equipped with wearable inertial sensors and follow a video tutorial in order to perform the exercises in the sequencing indicated by the tutorial. The data acquisition platform comprises 5 wearable sensors (Xsens Inertial Measurement Units) and 3 Kinect RGB-D cameras, which can capture the body movements from three perspectives: front, side, and top. The LISSI dataset consists of 8 high-level classes corresponding to the self exercises; each one encapsulates a series of low-level classes corresponding to the self exercise components, also called low-level exercises. The latter are composed of repetitive or non-repetitive movements, which may occur multiple times during the high-level exercise. The dataset additionally includes contextual labels that indicate how the self exercise has been performed. The dataset was collected during two different periods involving 7 and 11 subjects. Only the data belonging to 6 human subjects involved in first period have been annotated and used for the present study.

4.2. Data Preparation

Data preparation for all 3 datasets includes replacing missing values, data segmentation, and normalization. A min-max normalization has been applied

[1] https://itea3.org/project/medolution.html

based on the sensors range of each dataset. Dimension reduction was applied for Opportunity and LISSI datasets.

Approximately 3, 1.5, and 5 seconds length sliding window was employed to segment data for Opportunity, LISSI, and PAMAP2 dataset respectively, taking into account 70% of overlap between consecutive windows of data. Although segmenting the samples into windows provides more information about ongoing activity for the system, it enlarges the dimensionality of data to be processed.

To exemplify, let us consider 3 and 1.5 seconds windows of data collected with 60 and 30 Hz frequency for Opportunity and LISSI datasets, respectively. Having feature sizes of 113 and 95, segmentation leads into input vectors with thousands dimensions. Handling high-dimensional data samples is way more laborious and time-consuming for the artificial neural network. Hence, a dimension reduction technique may come as a boon to facilitate the training process. Using Principal Component Analysis (PCA), we scaled dimensions down to 1% and 2% for Opportunity and LISSI by investigating the precision-time trade-off. However, dimension reduction imposed a crucial decrease in the performance of the classification task of the PAMAP2 dataset. Therefore, this step is skipped for this dataset.

Statistics of the LISSI dataset show it has a lower *Sample per Subject* rate while it has more subjects in comparison to the Opportunity dataset [7]. Furthermore, There exist very short-length classes of activity such as *Kneeling* or *Warm up*, which have fewer samples compared to the rest and make the dataset imbalanced. Since the research is focused on subject-level adaptation, and a deep architecture is used in the proposed approach, we demand a high amount of data for training. We figured out this problem through the use of *micro-mini-batch* learning strategy which is discussed in section 3.

The datasets are split into training, validation, and testing sets as follows. As for the Opportunity dataset, the samples of each subject were held in 5 Activity of Daily Living (ADL) les. The first 3 ADL files were dedicated to the training set, and the fourth and fifth ones were considered for validation and testing sets, respectively. PAMAP2 and LISSI dataset, the samples of each subject divided into the training, validation and testing set proportional to 0.6, 0.1, 0.3, respectively. In summary, each experiment has a source and target subject whose samples are categorized into 3 disjoint subsets. The reported results were obtained over the testing set for the target subject.

5. Experiments

As earlier discussed in section 4, Opportunity, PAMAP2, and LISSI datasets are prepossessed for evaluating our proposed method. A set of subject translation experiments has been held to practically demonstrate the necessity of domain adaptation and its effectiveness as well.

For each dataset, samples of each subject are considered independently as the source or target domain. A classifier is adapted for each target domain (subject), examined and compared with two other adaptation methods, Stratified Transfer Learning (STL) [10] and Geodesic Flow Kernel (GFK) [13], as well as adaptation performance upper bound. The implementations of all the experiments for STL and the proposed model have been done in Python using Keras and Tensorflow libraries. The codes of GFK and STL are provided in Matlab and can be obtained online[2].

Table 2 details the architecture of the proposed method components. The complexity of the problem is different for each source-target pair of subjects. Therefore, the complexity of the proposed method should be adjusted for the problem by its controlling parameters, including cb, gf, cf, and df. The high learning capacity of deep one-dimensional Convolutional Neural Networks makes them more suitable for this application. Additionally, employing CNN models facilitates the recognition task by extracting more abstract feature values than time and frequency-based handcrafted feature extraction approaches.

[2]https://github.com/jindongwang/transferlearning/tree/master/code

Table 2: Hyper-parameters of the proposed adversarial model's components

Element	**Parameter**	**Values**
Generator	Input layer dimension	(100,), (100,), (4000,)†
	Number of convolutional blocks	cb
	Kernel size of 1^{st}, 2^{end} Conv. layers in convolutional blocks	3, 3
	Sliding stride of 1^{st}, 2^{end} Conv. layers in the blocks	1, 1
	Number of filters in 1^{st}, 2^{end} Conv. layers in the blocks	gf, gf *
	Number of filters of output Conv. layer	(100,) , (100,), (4000,)†
	Kernel size of output Conv. layer	3
Classifier	Input Layer dimension	(100,) , (100,), (4000,) †
	Number of filters for 1, 2, 3^{th} Conv. layers	$cf, cf/2, cf/4$ *
	Number of neurons in Dense output layer	6, 8, 10 †
Discriminator	Input Layer dimension	(100,) , (100,), (4000,)†
	Number of filters for Conv. layers	$2 \times df, 4 \times df, 8 \times df, 4 \times df, 2 \times df$ *
	Number of the units in the output Dense layer	1

† *for Opportunity, LISSI, and PAMAP2 datasets, respectively.*

* *the variablea gf (generator filters), cf (classifier filters), and df (discriminator filters) should be tuned in training.*

5.1. Results and Analysis

We evaluate the proposed model by conducting extensive experiments on three datasets of different sizes to assess its functionality and robustness. We opted for W-F1 measure as the evaluation metric since it gives better insight compared to accuracy, precision, and recall deliberating imbalance distribution of classes in the dataset:

$$\begin{gathered} W\text{-}F1 = \sum_{i=1}^{\mathbb{C}} 2\omega_i \frac{precision_i \times recall_i}{precision_i + recall_i} \\ \omega_i = ||class_i||, \;\; \sum_{i=1}^{\mathbb{C}} \omega_i = 1 \end{gathered} \tag{9}$$

Fig. 2 depicts the mean sample rate for each class on these datasets. Opportunity dataset is the most populated one among the selected datasets; nonetheless, it is supposed to be challenging due to its highly imbalanced distribution. PAMAP2 dataset includes the smoothest distribution of classes; however, it contains more classes of activities with less sample/class rate. Finally, the LISSI dataset is expected to be the most challenging one owing to its imbalanced, less populated distribution and remarkably low amount of samples. It is worth noting that the learning task could be challenging in each of the selected datasets due to various reasons such as having a high number of classes and complication level, a small number of samples, imbalance distribution of samples, and multiple underrepresented classes.

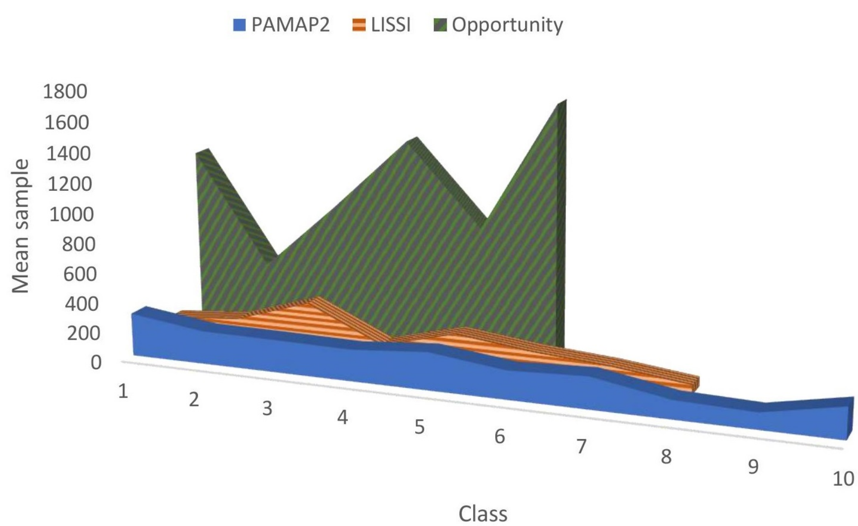


Figure 2: Distribution of the average number of samples of each class for subjects in Opportunity, PAMAP2, and LISSI datasets.

Analysis of the model during the training: Fig. 3 portrays the loss value trends for three main components of the proposed model during a translation experiment on the PAMAP2 dataset. The plot highlights the adversary of generator and discriminator as presumed. The loss values for D and G fluctuate oppositely; in the periods that G is improving by decreasing its loss values, D is deteriorating and vice versa. The whole framework converges after about 150 epochs in this experiment; when the discriminator achieves desired uncertainty(D^*), the generator and classifier both converge and no further improvement can be reached by optimizing the loss functions.

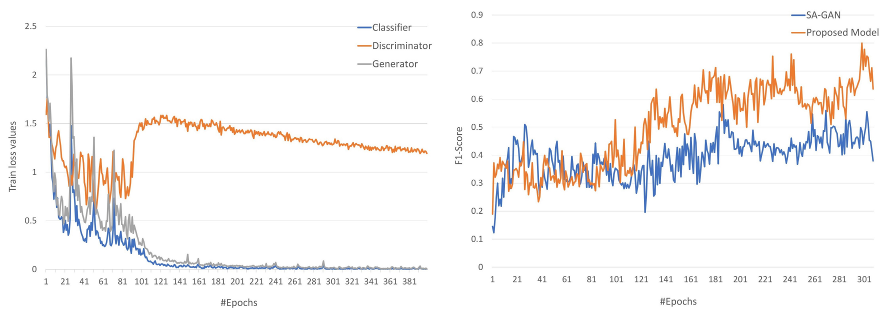


Figure 3: Left:Average loss values for the discriminator, generator and the classifier elements of the proposed model adversarial framework for translation of subject 5 to subject 6 in PAMAP2 dataset. Right: Accuracy of the classifier over the validation set of the target domain during the training of the framework to translate subject 5 to subject 1 of PAMAP2 dataset.

The trend of accuracy improvement of the proposed model and SA-GAN in Fig. 3 approves the effectiveness of our proposed model, which improves the performance of adversarial adaptation. The plot compares the accuracy of the models over the validation set of the samples from the target subject in the PAMAP2 dataset. Although both models set out similar performance in the beginning, the proposed model shows more superiority over SA-GAN after 100 epochs as it is reaching closer to convergence.

Evaluation of the model over different datasets and comparison with state-of-the-art (test phase): Table 3 outlines the results of subject-level transfer learning on the Opportunity dataset compared with other states of the art models. This dataset is considered a large-scale one since it holds 6400 windows of samples per subject on average. The *No Transfer* column refers to the experiments that have been done to justify the necessity of domain adaptation by training a Convolutional Neural Network over the source domain and testing it against the target domain samples, directly without any adapta-

Table 3: Comparison of the generic proposed approach performance and GFK [13], STL [10], SA-GAN [6], Supervised and Not transferred model, in terms of Weighted F1 measure on **Opportunity** dataset. *No Transfer* and *Fully Supervised* columns refer to models which have been exclusively trained with the source and target subject data, respectively. The most dominant performance in each transformation experiment marked in bold.

Source Subject	Target Subject	Distance (Source,Target)	No Transfer	STL	GFK	SA-GAN	Proposed Model	Fully Supervised
	2	46.69	0.45	0.65	0.59	0.73	**0.74**	0.75
1	3	45.10	0.27	0.37	0.43	0.45	**0.58**	0.71
	4	77.15	0.40	0.47	0.55	0.49	**0.57**	0.59
	1	40.47	0.48	0.52	**0.62**	0.56	0.56	0.65
2	3	34.38	0.44	0.46	0.51	**0.52**	0.40	0.71
	4	72.80	0.29	**0.46**	0.40	0.39	0.42	0.59
	1	38.38	0.23	0.40	0.45	0.42	**0.52**	0.65
3	2	37.54	0.21	0.54	0.53	**0.61**	0.52	0.75
	4	73.69	0.31	0.37	0.44	0.44	**0.50**	0.59
	1	73.53	0.26	0.38	0.51	0.51	**0.52**	0.65
4	2	70.80	0.29	0.54	0.45	0.55	**0.68**	0.75
	3	69.44	0.24	0.48	0.37	0.49	**0.53**	0.71

tion technique. The *Distance* column represents an estimation of Wasserstein distance between source and target subjects [6]. The best knowledge transfer source could be selected either based on the distance or validation metrics. Since the proposed method is a semi-supervised one, we considered the outcomes of the CNN trained under supervision as the upper bound for domain adaptation. Training a model with the annotated data from the same domain is considered as the highest margin that can be accomplished. Hence, the adaptation performance is regarded as convincing as it is close to the upper bound. We find that our proposed method improved the performance of the classification task in comparison with other adversarial domain adaptation methods SA-GAN by 7.3% on average, which admits the effectiveness of the micro-mini-batch training approach. Entirely, the proposed model improves the classification performance in 8 out of 12 experiments accomplished best and second-best W-F1 measure in

Table 4: Comparison of the generic proposed approach performance and DANN[34], VADA[35], IIMT[37], and Not Transferred model, in terms of Weighted F1 measure on Opportunity dataset. No Transfer column refers to the model which have been exclusively trained with the source subject data. The most dominant performance in each transformation experiment marked in bold. [37]

Source Subject	**Target Subject**	**No Transfer**	**DANN**	**VADA**	**IIMT**	**Proposed Model**
S1	S2	0.652	0.768	0.776	0.809	**0.866**
S3	S2	0.631	0.725	0.72	0.787	**0.807**
S1	S3	0.64	0.731	0.747	0.78	**0.79**
S2	S3	0.637	0.694	0.734	**0.745**	0.703
S3	S1	0.659	0.746	0.726	**0.83**	0.785
S2	S1	0.696	0.785	0.797	0.813	**0.831**
Average		0.652	0.741	0.75	0.794	**0.797**

more than 83% of the experiments. Moreover, it can be observed from Table 4 that our proposed model improves the W-F1 score up to 6% for gesture classes of activity on the Opportunity dataset and obtained the best performance for each target domain.

The results on the Opportunity dataset concur with the competence of the proposed approach in large-scale datasets. Let us consider all three available sources of knowledge transfer for each source domain in this dataset. The classifier adapted by the proposed model presented the best classification results for 3 out of 4 target domains and competitive results for the remaining one.

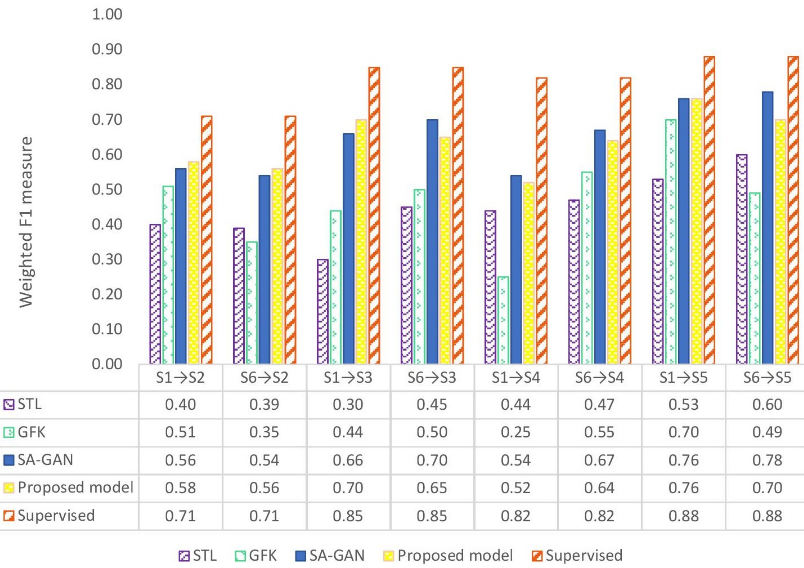
Table 5 summarizes the classification results on a medium-size dataset, PAMAP2, which has 2049 segmented samples per subject on average. However, this dataset comes to be challenging as it contains 10 classes of activities that is higher than the Opportunity dataset with 6 classes. It might adversely affect the performance due to the more probable imbalanced distribution of the classes. Though, the results in Table 5 demonstrate that the proposed micro-mini batch learning technique overcame this challenge. The proposed model dominated other states of the art methods in more than 90% of the experiments and improved W-F1 measure up to 13%.

Table 5: Comparison of the generic proposed approach performance and GFK [13], STL [10], SA-GAN [6], Supervised, and Not Transferred model, in terms of Weighted F1 measure on **PAMAP2** dataset. *No Transfer* and *Fully Supervised* columns refer to models which have been exclusively trained with the source and target subject data, respectively. The most dominant performance in each transformation experiment marked in bold.

Source Subject	Target Subject	Distance (Source,Target)	No Transfer	STL	GFK	SA-GAN	Proposed Model	Fully Supervised
	5	91.82	0.37	0.62	0.72	0.69	**0.77**	0.98
1	6	91.58	0.32	0.56	0.64	0.66	**0.70**	0.97
	8	107.07	0.04	0.57	0.49	0.65	**0.72**	0.92
	1	91.82	0.32	0.76	0.66	0.71	**0.76**	0.99
5	6	42.13	0.64	0.83	0.75	0.83	**0.83**	0.97
	8	56.01	0.26	0.52	0.69	0.66	**0.73**	0.92
	1	91.58	0.16	0.67	0.56	0.61	**0.78**	0.99
6	5	42.13	0.41	0.74	0.75	0.79	**0.83**	0.98
	8	56.76	0.17	**0.86**	0.58	0.63	0.82	0.92
	1	107.07	0.10	0.54	0.58	0.68	**0.76**	0.99
8	5	56.01	0.25	0.55	0.41	0.60	**0.73**	0.98
	6	56.76	0.27	0.58	0.61	**0.73**	0.71	0.97

The results in Fig. 4 presents the proposed model's competitive performance over the LISSI dataset. It dominated the STL and GFK model in all the experiments and came close to the supervised learning performance. It can be inferred from the results that both adversarial models overall performed well on this imbalanced small size dataset in terms of W-F1 measure.

Performance of the model over a given sample: As can be inferred from Table 6 and Fig. 5, the proposed model performed well in classifying samples from Subject 2 by leveraging knowledge from subject 4 in the Opportunity dataset. The precision values are notably higher than recall and W-F1 measure in *null* and *Sandwich time* classes. Reminding the precision and recall formula



	S1→S2	S6→S2	S1→S3	S6→S3	S1→S4	S6→S4	S1→S5	S6→S5
STL	0.40	0.39	0.30	0.45	0.44	0.47	0.53	0.60
GFK	0.51	0.35	0.44	0.50	0.25	0.55	0.70	0.49
SA-GAN	0.56	0.54	0.66	0.70	0.54	0.67	0.76	0.78
Proposed model	0.58	0.56	0.70	0.65	0.52	0.64	0.76	0.70
Supervised	0.71	0.71	0.85	0.85	0.82	0.82	0.88	0.88

Figure 4: Results of applying Subject to Subject Transfer Learning on **LISSI** dataset, comparing the generic proposed approach performance and GFK [13], STL [10], SA-GAN [6], Supervised model, in terms of Weighted F1 measure. *Supervised* column refers to models which have been exclusively trained with the target subject data.

as bellow, it can be concluded:

$$\begin{aligned} &Recall = \frac{TP}{TP + FN}, Precision = \frac{TP}{TP + FP} \\ &Precision > Recall \rightarrow FN > FP \end{aligned} \tag{10}$$

which means the proposed model falsely rejected samples more frequently than it falsely predicted that the samples belong to these two classes. The most likely explanation of this phenomenon is the tension of the adversarial models to collapse, which may push the model to falsely tag samples as if they are drawn from the class that the generator collapsed on. Furthermore, the classifier did not present significant results in recognizing samples from *Coffee time* and *Clean up* classes. It is interesting to note that these classes of activities have been among the most problematic classes, even for a model that had supervised training by the samples of the target subject. This discrepancy could be attributed to their more complicated patterns than the rest; the pattern

Table 6: The main classification metrics of a sample classification task on Opportunity, LISSI, and PAMAP2 datasets.

Opportunity Dataset				LISSI Dataset				PAMAP2 Dataset			
Subject 4 → Subject 2				Subject 1 → Subject 5				Subject 5 → Subject 6			
class	precision	recall	support	class	precision	recall	support	class	precision	recall	support
Null	**0.98**	0.56	292	Kneeling	0.63	0.83	53	Ironing	0.87	0.99	137
Relaxing	0.81	**0.84**	190	Lying	0.78	0.78	79	Lying	0.89	**0.99**	85
Coffee time	0.34	0.33	225	Relaxing	**0.99**	0.87	142	Sitting	0.86	0.99	84
Early morning	0.61	0.84	372	Sitting	0.74	0.47	49	Standing	0.88	0.65	89
Clean up	0.34	0.57	224	Sit to Stand	0.80	**0.88**	100	Walking	0.87	0.86	93
Sandwich time	0.94	0.71	679	Standing	0.87	0.71	76	Running	0.82	0.94	82
				Dance walk	0.69	0.51	57	Cycling	**0.93**	0.88	74
				Warm up	0.38	0.73	41	Ascending	0.59	0.82	49
								Descending	0.80	0.49	41
								Cleaning	0.70	0.43	76
Accuracy	0.67		1982	Accuracy	0.76		597	Accuracy	0.83		810
W-Avg	0.74	0.67	1982	W-Avg	0.79	0.76	597	W-Avg	0.84	0.83	810

of making coffee may vary highly from one human to another due to the unscripted data collection approach [44]. Therefore, models require more samples to generalize, yet the datasets contain a limited number of samples.

The number of samples, overlap between classes, duration of the activities, and level of abstraction are some of the factors associated with the complexity of the recognition task. Supposing slight similarity overlap between classes, the activities that contain sharper movements could be straightforward in recognition since the peaks in acceleration data are more discernible. According to Table 6, the proposed model delivers low precision in recognizing the samples from the *Warm up* class of the LISSI dataset. The *Warm up* activity is short-length and contains an extremely rapid sharp movement, making the recognition more challenging for the model.

The cleaning activity in both PAMAP2 and Opportunity datasets is amongst the most perplexing activities owing to its high abstraction level. An extensive set of sub-activities combinations exist that form the cleaning activity, and no dataset affords full coverage of it. Samples of one activity of the same human subject (domain) may considerably vary based on human beings' innate non-deterministic behavior. Thus, the high generalization capability is required to

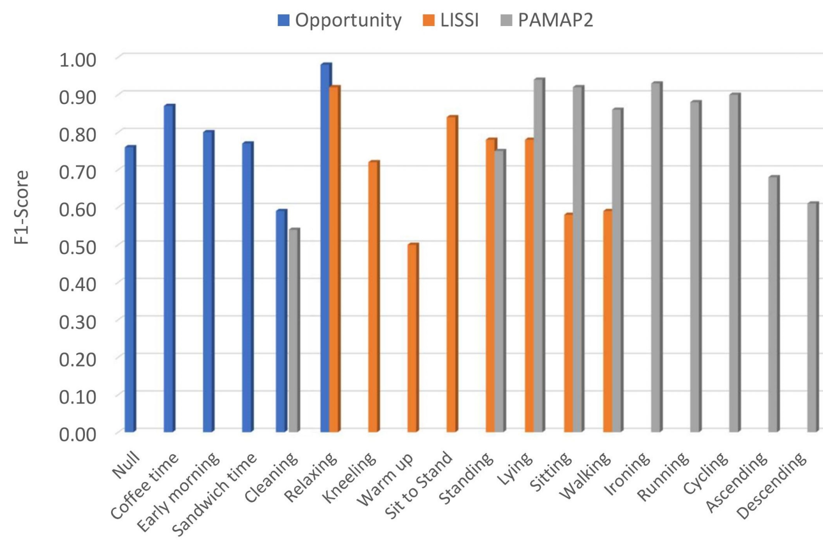


Figure 5: Comparison of per class F1-score in Opportunity, PAMAP2, and LISSI datasets.

overcome this intra-domain shift between the train and test set of the same domain, as well as the inter-domain shift between source and target domains. On the other hand, activities such as *Cycling*, *Relaxation*, or *Lying* contain a common inter and intra-domain acceleration pattern. Consequently, these activities are more tangible for the model to classify as it can be observed in Table 6.

Analysis of Fig. 5 reveals prevailing recognition performance in terms of F1 measure over PAMAP2 dataset. Nevertheless, slight phenomena of imbalance learning still can be found over the samples of *Standing*, *Descending* and *Cleaning* class which can be addressed in future work.

To further demonstrate the efficiency of our proposed model, we provided a detailed per class comparison in terms of F1-score between our proposed approach and SA-GAN [6] for the experiment reported in Table 6 (Subject 4 $\rightarrow$ Subject 2). It can be concluded from Fig. 6 that the proposed model mostly improved the per-class performance, even for classes with a low number of samples (such as Relaxing and Coffee Time). As the plot shows, the proposed model

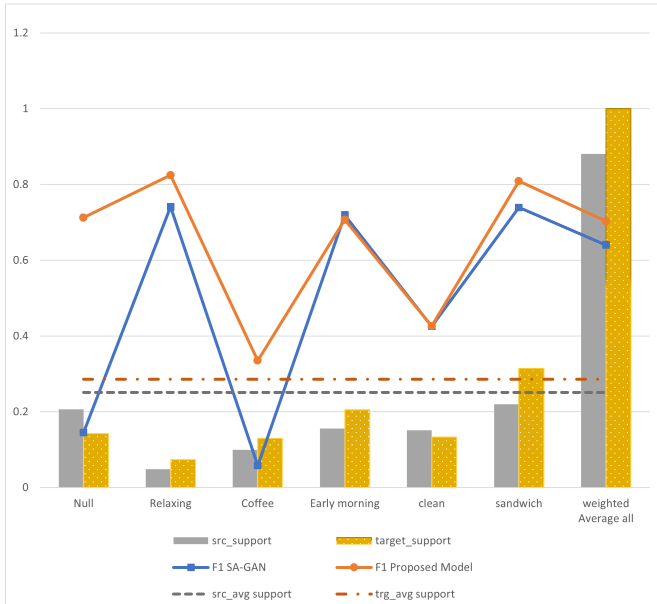


Figure 6: Per-Class W-F1 score comparison between SA-GAN and the proposed model for S4 $\rightarrow$ S1 experiment on Opportunity dataset

hugely improved the performance on the Null and Coffee time, activities that were a minority class in the source and/or target training set. It is worth noting that the samples of the Relaxing class are well classified despite the low number of samples that could be due to its more uncomplicated nature.

6. Conclusions and Future Work

Recent trends in Generative Adversarial Networks have led to a proliferation of studies that offer adversarial solutions for a variety of applications in Artificial Intelligence, mostly Image Processing and Vision. This study set out to propose a generic adversarial framework for knowledge transfer in the domain of Human Activity Recognition. The proposed semi-supervised model has been evaluated against three datasets with different challenges to assess its robustness to the scale and imbalance of the data.Our research findings are quite convincing, and thus the following conclusions can be drawn:

The proposed model provides striking results on the PAMAP2 benchmark medium-size multi-class datasets. It improved the adversarial domain adaptation performance by applying a micro-mini-batch learning technique on the Opportunity large-scale yet highly imbalanced dataset. Interestingly, the proposed model revealed competitive results compared to other states of the art models on the LISSI dataset, which is very challenging in terms of both the number of samples and the balance of classes. Our comprehensive assessment was carried out over high-dimensional data of highly abstract activities in all three datasets. Though the framework is not HAR-exclusive and it can be potentially utilized to solve domain adaptation problems.

The proposed framework deals with mode collapse and imbalanced data problems in GAN. Thus, its limitations are the same as a classic GAN, such as vanishing gradients, failure to converge, the imbalance between the generator and discriminator causing overfitting, and high sensitivity to the hyperparameter selections. Besides, regardless of the balance in the distribution, Neural Networks are known to be prone to small training sets. Therefore, providing more samples could provide improvement to the final classification results.

Further studies, will need to be undertaken with more focus on the lack of samples. In future investigations, it might be possible to use multiple sources of knowledge or a combination of transferred models from different source domains and the source/model selection policies. Besides, integrating an ontology-based inference module into the framework could be a mean of improvement of the classification results obtained by Machine Learning.

References

[1] Y. Bengio, Deep learning of representations: Looking forward, in: International Conference on Statistical Language and Speech Processing, Springer, 2013, pp. 1–37 (2013).

[2] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (10) (2010) 1345–1359 (2010).

[3] W. M. Kouw, M. Loog, A review of domain adaptation without target labels, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019) 1–1 (2019). `doi:10.1109/TPAMI.2019.2945942`.

[4] M. A. A. H. Khan, N. Roy, A. Misra, Scaling human activity recognition via deep learning-based domain adaptation, in: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2018, pp. 1–9 (2018).

[5] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in neural information processing systems, 2014, pp. 3320–3328 (2014).

[6] E. Soleimani, E. Nazerfard, Cross-subject transfer learning in human activity recognition systems using generative adversarial networks, Neurocomputing 426 (2021) 26–34 (2021).

[7] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, D. Roggen, The opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognition Letters 34 (15) (2013) 2033–2042 (2013).

[8] A. Reiss, D. Stricker, Introducing a new benchmarked dataset for activity monitoring, in: 2012 16th International Symposium on Wearable Computers, IEEE, 2012, pp. 108–109 (2012).

[9] LISSI Dataset rehabilitation self exercises dataset, `http://www.lissi.fr/parkinson-rehabilitation-dataset-2/`, accessed: 2019-09-30.

[10] J. Wang, Y. Chen, L. Hu, X. Peng, S. Y. Philip, Stratified transfer learning for cross-domain activity recognition, in: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2018, pp. 1–10 (2018).

[11] Y. Chen, J. Wang, M. Huang, H. Yu, Cross-position activity recognition with stratified transfer learning, Pervasive and Mobile Computing 57 (2019) 1–13 (2019).

[12] J. Wang, V. W. Zheng, Y. Chen, M. Huang, Deep transfer learning for cross-domain activity recognition, in: Proceedings of the 3rd International Conference on Crowd Science and Engineering, ACM, 2018, p. 16 (2018).

[13] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2066–2073 (2012).

[14] H. S. Hossain, M. A. A. H. Khan, N. Roy, Active learning enabled activity recognition, Pervasive and Mobile Computing 38 (2017) 312–330 (2017).

[15] S. J. Pan, I. W. Tsang, J. T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Transactions on Neural Networks 22 (2) (2011) 199–210 (2011).

[16] R. Yasarla, V. A. Sindagi, V. M. Patel, Syn2real transfer learning for image deraining using gaussian processes, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2726–2736 (2020).

[17] J. Wang, Y. Chen, S. Hao, W. Feng, Z. Shen, Balanced distribution adaptation for transfer learning, in: 2017 IEEE International Conference on Data Mining (ICDM), IEEE, 2017, pp. 1129–1134 (2017).

[18] J. Wang, Y. Chen, W. Feng, H. Yu, M. Huang, Q. Yang, Transfer learning with dynamic distribution adaptation, ACM Transactions on Intelligent Systems and Technology (TIST) 11 (1) (2020) 1–25 (2020).

[19] M. A. A. H. Khan, N. Roy, Transact: Transfer learning enabled activity recognition, in: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2017, pp. 545–550 (2017).

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680 (2014).

[21] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 417–425 (2017).

[22] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 146–157 (2017).

[23] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, Generative adversarial text to image synthesis, arXiv preprint arXiv:1605.05396 (2016).

[24] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. N. Metaxas, Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5907–5915 (2017).

[25] A. Bidgoli, P. Veloso, Deepcloud. the application of a data-driven, generative model in design, arXiv preprint arXiv:1904.01083 (2019).

[26] T. Ueda, M. Seo, I. Nishikawa, Data correction by a generative model with an encoder and its application to structure design, in: International Conference on Artificial Neural Networks, Springer, 2018, pp. 403–413 (2018).

[27] M. Alzantot, S. Chakraborty, M. B. Srivastava, Sensegen: A deep learning architecture for synthetic sensor data generation, 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (2017) 188–193 (2017).

[28] R. Saeedi, K. Sasani, S. Norgaard, A. H. Gebremedhin, Personalized human activity recognition using wearables: A manifold learning-based knowledge transfer, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 1193–1196 (2018). doi:10.1109/EMBC.2018.8512533.

[29] J. Wang, Y. Chen, Y. Gu, Y. Xiao, H. Pan, Sensorygans: An effective generative adversarial framework for sensor-based human activity recognition, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–8 (2018).

[30] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176 (2017). doi:10.1109/CVPR.2017.316.

[31] Z. Luo, Y. Zou, J. Hoffman, L. F. Fei-Fei, Label efficient learning of transferable representations acrosss domains and tasks, in: Advances in Neural Information Processing Systems, 2017, pp. 165–177 (2017).

[32] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, D. Krishnan, Unsupervised pixel-level domain adaptation with generative adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3722–3731 (2017).

[33] G. He, X. Liu, F. Fan, J. You, Classification-aware semi-supervised domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 964–965 (2020).

[34] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The journal of machine learning research 17 (1) (2016) 2096–2030 (2016).

[35] R. Shu, H. H. Bui, H. Narui, S. Ermon, A dirt-t approach to unsupervised domain adaptation, arXiv preprint arXiv:1802.08735 (2018).

[36] O. Chapelle, A. Zien, Semi-supervised classification by low density separation, in: International workshop on artificial intelligence and statistics, PMLR, 2005, pp. 57–64 (2005).

[37] S. Yan, H. Song, N. Li, L. Zou, L. Ren, Improve unsupervised domain adaptation with mixup training, arXiv preprint arXiv:2001.00677 (2020).

[38] Z. Yu, L. Chen, Z. Cheng, J. Luo, Transmatch: A transfer-learning scheme for semi-supervised few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12856–12864 (2020).

[39] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognition Letters 119 (2019) 3–11 (2019).

[40] J. Jiang, A literature survey on domain adaptation of statistical classifiers, URL: http://sifaka. cs. uiuc. edu/jiang4/domainadaptation/survey 3 (2008) 1–12 (2008).

[41] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, C. Sutton, Veegan: Reducing mode collapse in gans using implicit variational learning, in: Advances in Neural Information Processing Systems, 2017, pp. 3308–3318 (2017).

[42] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Advances in neural information processing systems, 2016, pp. 2234–2242 (2016).

[43] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, arXiv preprint arXiv:1701.04862 (2017).

[44] P. Asghari, E. Soleimani, E. Nazerfard, Online human activity recognition employing hierarchical hidden markov models, Journal of Ambient Intelligence and Humanized Computing (2019). doi:10.1007/s12652-019-01380-5.