

WIJEKOON, A. and WIRATUNGA, N. 2020. Personalised meta-learning for human activity recognition with few-data. In Bramer, M. and Ellis, R. (eds.) *Artificial intelligence XXXVII: proceedings of 40th British Computer Society's Specialist Group on Artificial Intelligence (SGAI) Artificial intelligence international conference 2020 (AI-2020), 8-9 December 2020, [virtual conference]*. Lecture notes in artificial intelligence, 12498. Cham: Springer [online], pages 79-93. Available from: [https://doi.org/10.1007/978-3-030-63799-6\\_6](https://doi.org/10.1007/978-3-030-63799-6_6)

# Personalised meta-learning for human activity recognition with few-data.

WIJEKOON, A. and WIRATUNGA, N.

2020

The final authenticated version is available online at: [https://doi.org/10.1007/978-3-030-63799-6\\_6](https://doi.org/10.1007/978-3-030-63799-6_6). This pre-copyedited version is made available under the Springer terms of reuse for AAMs: <https://www.springer.com/gp/open-access/publication-policies/aam-terms-of-use>

 OpenAIR  
@RGU

This document was downloaded from  
<https://openair.rgu.ac.uk>

SEE TERMS OF USE IN BOX ABOVE

DISTRIBUTED UNDER LICENCE

# Personalised Meta-Learning for Human Activity Recognition with Few-data

Anjana Wijekoon<sup>[0000-0003-3848-3100]1</sup> and Nirmalie Wiratunga<sup>[0000-0003-4040-2496]1</sup> \*

School of Computing, Robert Gordon University,  
Aberdeen AB10 7GJ, Scotland, UK  
{a.wijekoon, n.wiratunga}@rgu.ac.uk

**Abstract.** State-of-the-art methods of Human Activity Recognition (HAR) rely on having access to a considerable amount of labelled data to train deep architectures with many train-able parameters. This becomes prohibitive when tasked with creating models that are sensitive to personal nuances in human movement, explicitly present when performing exercises. Also, it is not possible to collect training data to cover all persons in the target population. Accordingly, learning personalised models with few data remains an interesting challenge in HAR research. We present a meta-learning methodology for learning to learn personalised HAR models for HAR; with the expectation that the end-user need only provides a few labelled data. These personalised HAR models benefit from the rapid adaptation of a generic meta-model using only a few end-user data. We implement the personalised meta-learning methodology with two algorithms, Personalised MAML and Personalised Relation Networks. A comparative study shows significant performance improvements against state-of-the-art deep learning algorithms and personalised algorithms in multiple HAR domains. In addition, we show how personalisation improved meta-model training, to learn a generic meta-model suited for a wider population while using a shallow parametric model.

**Keywords:** Personalisation · Human Activity Recognition · Meta-Learning · Few-shot Learning

## 1 Introduction

Machine Learning research in Human Activity Recognition (HAR) has a wide range of high impact applications in gait recognition, fall detection, orthopaedic rehabilitation and general fitness monitoring. A HAR dataset consists of sensor data streams collected from multiple persons. Unavoidably, sensors capture personal traits and nuances in some activity domains more than others. Typically with activities that involve greater degrees of freedom. Thus, learning a single

---

\* This work was part funded by SELFBACK, a project funded by the European Union's H2020 research and innovation programme under grant agreement No. 689043. More details available at <http://www.selfback.eu>

reasoning model to recognise the set of activity classes using a HAR dataset can be challenging, which calls for the personalisation.

We propose it is more intuitive to treat a “person-activity” pair as the class label. Accordingly, each person’s data can be viewed as a dataset in its own right, and the HAR task involves learning a reasoning model for the person. Learning from only specific persons’ data has shown significant performance improvements in early research with both supervised learning and active learning methods [13, 4]. But these methods require considerable amounts of data obtained from the end-user, periodical end-user involvement and model re-training. Also, current state-of-the-art Deep Learning algorithms require a large number of labelled data instances to avoid under-fitting.

Here we explore the “person-activity” classes concept but attempt to learn with a limited number of data instances per class. This can be viewed as a few-shot classification task [14, 10] where the aim is to learn a classifier with one or few labelled data instances for each class. Meta-learning methods are the state-of-the-art in few-shot classification for image recognition [2, 6]. In a nutshell, meta-learning is described as learning-to-learn, where a wide range of tasks abstract their learning to a meta-model, such that, it is transferable to any unseen task. Meta-learning algorithms such as MAML [2] and Relation Networks (RN) [12] implement this methodology for few-shot classification, by learning a generic models, and rapidly adapting to new tasks with only a few instances of data.

The concept of learning-to-learn aligns well with personalisation where modelling a person can be viewed as a single task; whereby the meta-model must help learn a model that is rapidly adaptable to a new person. We propose “personalised meta-learning” to create personalised models, by leveraging a small amount of sensing data (i.e. calibration data) extracted from a person. Accordingly, in this paper, we make the following contributions,

1. formalise Personalised Meta-Learning and implement with two Algorithms, Personalised MAML and Personalised RN;
2. perform a comparative evaluation in 9 experiments with 3 HAR datasets representing a wide range of activity domains; and
3. visualise how personalisation methodology enhanced the training and testing of meta-learners.

Importantly, we show that personalised meta-learning achieve significant performance improvement with simple shallow parametric models that only require a limited amount of labelled data compared to conventional DL models.

## 2 Related Work

Human Activity Recognition (HAR) is an active research challenge, where Deep Learning (DL) methods claim the state-of-the-art in many application domains [7, 18, 16, 15]. Learning a generalised reasoning model adaptable to many user groups is a unique transfer learning challenge in the HAR domain. Given access to large

quantities of end-user data, early research has achieved improved performance by learning personal models [1]. Follow on work attempts to reduce the burden on end-user, by adopting semi-supervised [4], active learning [4] and multi-task [11] methods that rely on periodical model re-training and continuous user involvement post-deployment. Recent advancements in few-shot learning are adopted as an approach to personalisation in Personalised Matching Networks ( $MN^p$ ) [8, 14].  $MN^p$  learns a parametric model, that is learning to match, leveraging a few data instances from the same user. This few-shot method avoids post-deployment re-training but is restricted by the similarity metric used for matching.

Meta-Learning or “learning-to-learn” is the learning of a generalised classification model that is transferable to new learning tasks with only a few instances of labelled data. In recent research it is interpreted mainly in three approaches; firstly, “learning-to-match” implemented by Relation Networks (RN) [12]; secondly, model-specific approach like SNAIL [5]; and finally, optimisation based algorithms like MAML [2]. MAML including its variants (FOMAML [2], Reptile [6]) and RN [12] are “model-agnostic”, where parametric feature learners are interchangeable. In contrast to model-specific meta-learners, such as SNAIL [5] and MANN [9], where meta-learning is achieved using specific neural network constructs such as LSTM and Neural Turing Machine [3]. Model-agnostic methods are preferred in a HAR setting, where heterogeneous sensor modalities may require different feature learners. While MN is an early interpretation of “learning-to-match”, RN is not limited by a similarity metric which makes RN more generalisable to many new tasks. In contrast to MAML, RN has the potential to perform Open-ended HAR, by modelling the classification task as a matching task, similar to Open-ended MN [17]. In this paper, we implement personalised meta-learning for HAR with the two model-agnostic meta-learners; MAML and RN.

### 3 Methods

Given a dataset,  $\mathcal{D}$ , Human Activity Recognition (HAR), like any other supervised learning problem, is the learning of the feature mapping  $\theta$  between data instances,  $x$ , and activity classes,  $y$ , where  $y$  is in the set of activity classes,  $\mathcal{C}$ . In HAR, each data instance in  $\mathcal{D}$  belongs to a person,  $p$ . Given the set of data instances obtained from person  $p$  is  $\mathcal{D}^p$ ,  $\mathcal{D}$  is the collection of data instances from the population  $\mathcal{P}$  (Equation 1). As before, all data instances in  $\mathcal{D}^p$  will belong to a class in  $\mathcal{C}$ .

$$\mathcal{D} = \{\mathcal{D}^p \mid p \in \mathcal{P}\} \text{ where } \mathcal{D}^p = \{(x, y) \mid y \in \mathcal{C}\} \quad (1)$$

Importantly, looking at individual  $\mathcal{D}^p$  performing the same set of exercises we see how sensor data capture personal nuances, and display different data distributions as in Figure 1.

#### 3.1 Personalised Meta-Learning for HAR

Meta-learning for few-shot classification can be seen as optimisation of a generic parametric model over many few-shot tasks (i.e. meta-train), and rapid adap-

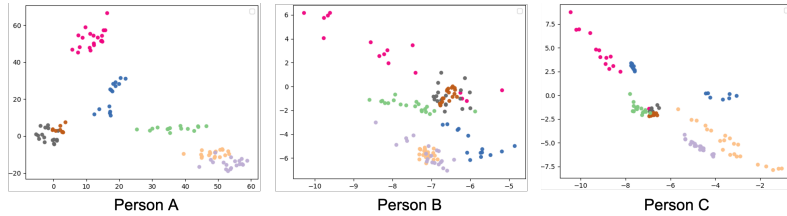


Fig. 1: Data distribution visualisation with PCA for 3 persons from MEx dataset

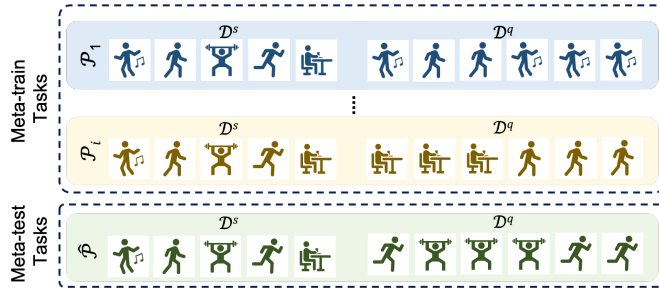


Fig. 2: Personalised Meta-Learning Tasks design for HAR

tation of the generic model for an unseen few-shot task (i.e. meta-test). More formally, a few-shot classification task has a “support set”,  $\mathcal{D}^s$ , and a “query set”,  $\mathcal{D}^q$ . The support set act as training data, with one or few representatives data instances for each class, and query set act as test data.

Personalised meta-learning for HAR is the learning of a meta-model  $\theta$  from a population  $\mathcal{P}$  while treating activity recognition for a person as an independent few-shot classification task. We propose the task design in Figure 2 for Personalised Meta-Learning. Given a dataset  $\mathcal{D}$ , of population  $\mathcal{P}$ , we create tasks such that, each “person-task”,  $\mathcal{P}_i$ , only contain data from a specific person,  $p$ . We randomly select a  $K^s \times |\mathcal{C}|$  number of labelled data instances from person  $p$  stratified across activity classes,  $\mathcal{C}$ , such that there are  $K^s$  amount of representatives for each class. We follow a similar approach when selecting a query set,  $\mathcal{D}^q$ , for  $\mathcal{P}_i$ . Typically  $\mathcal{D}^q$  has no overlap with  $\mathcal{D}^s$  similar to a train/test split in supervised learning. Given that existing HAR dataset are not strictly few-shot learning datasets, there can be a few to many data instances available to be sampled for the query set,  $\mathcal{D}^q$ . Each resulting “person-task” is learning to classify the set of “person-activity” class labels.

At test time, the test person,  $\hat{p}$ , provides a few seconds of data for each activity class while being recorded by recommended sensor modalities, which forms the support set,  $\mathcal{D}^s$ , of the person-task,  $\hat{\mathcal{P}}$ . Thereafter, the meta-model, in conjunction with the support set, predicts the class label for each query data instance,  $x_i^q$ , in  $\mathcal{D}^q$ . It is noteworthy that, contrary to conventional Meta-Learning, all personal models and the meta-model are learning to classify the same set of

activity classes  $\mathcal{C}$ , but of different persons (i.e. “person-activity”). Therefore, it is seen as a few-shot classification problem with a  $|\mathcal{C}| \times |\mathcal{P}|$  number of classes. Personalised meta-learning is a methodology adaptable with any meta-learner to perform personalised HAR, and next we show how with two meta-learners, MAML and RN.

### 3.2 Personalised MAML

---

#### Algorithm 1 Personalised MAML Training

---

**Require:**  $p(\mathcal{P})$ : HAR dataset; distribution over persons

**Require:**  $\alpha, \beta, n, gs, meta\_gs$ : step sizes, batch size and gradient-steps hyper-parameters

- 1: randomly initialise  $\theta$
- 2: **while** not done **do**
- 3:   Sample  $n$  person-tasks  $\mathcal{P}_i \sim p(\mathcal{P})$
- 4:   **for all**  $\mathcal{P}_i$  **do**
- 5:      $\mathcal{D}^s = \{(x, y) \in \mathcal{P}_i : |\mathcal{D}^s| = K^s \times |\mathcal{C}|\}$
- 6:     **for**  $i = 0$  **to**  $gs$  **do**
- 7:       Compute  $\nabla_{\theta} \mathcal{L}_{\mathcal{P}_i}(\theta)$  w.r.t.  $\mathcal{D}^s$
- 8:       Compute updated parameters:  $\theta_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{P}_i}(\theta)$
- 9:     **end for**
- 10:     $\mathcal{D}^q = \{(x, y) \in \mathcal{P}_i : \mathcal{D}^s \cap \mathcal{D}^q = \emptyset, |\mathcal{D}^q| = K^q \times |\mathcal{C}|\}$
- 11:    Compute  $\mathcal{L}_{\mathcal{P}_i}(\theta_i)$  w.r.t.  $\mathcal{D}^q$
- 12:    **end for**
- 13:    Meta-update:  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{P}_i \sim p(\mathcal{P})} \mathcal{L}_{\mathcal{P}_i}(\theta_i)$
- 14: **end while**

---

MAML [2] is a versatile meta-learner applicable to any parametric model optimised with Gradient Descent (GD). Personalised MAML ( $MAML^p$ ) for HAR is optimised to learn the generic parametric model (i.e. meta-model),  $\theta$ , such that it is adaptable to any new person encountered at test time. Task design for  $MAML^p$  is as follows. For a person-task  $\mathcal{P}_i$ , a support set,  $\mathcal{D}^s$ , and a query set  $\mathcal{D}^q$  is selected. The number of instance in the support set,  $|\mathcal{D}^s|$ , determines the number of instances that need to be requested from a new person,  $\hat{p}$ , during testing. Thus, we keep  $K^s$  small, similar to a few-shot learning scenario. We use all remaining data instances from each class, in  $\mathcal{D}^q$ . More formally, given there are  $K$  instances per “person-activity”,  $K^q = (K - K^s)$  and  $|\mathcal{D}^q| = K^q \times |\mathcal{C}|$ .

We present the training of Personalised MAML in Algorithm 1. At each training epoch, a set of person-tasks are sampled where each optimises its person-task-model,  $\theta_i$ .  $\theta_i$  is trained with  $\mathcal{D}^s$  using one or few steps of GD ( $gs$ ). The meta-modal  $\theta$  is then trained using GD with the losses computed by the trained person-task-models,  $\theta_i$ , against their respective  $\mathcal{D}^q$ s. This is referred to as the

---

#### Algorithm 2 Personalised MAML Testing

---

**Require:**  $\mathcal{D}^s$  for test person  $\hat{P}$  obtained via micro-interactions,

**Require:**  $\theta$ ; Meta-model

- 1: Initialise  $\hat{\theta} = \theta$
- 2: **for**  $i = 0$  **to**  $meta\_gs$  **do**
- 3:   Compute  $\nabla_{\hat{\theta}} \mathcal{L}_{\mathcal{P}_i}(\hat{\theta})$  w.r.t.  $\mathcal{D}^s$
- 4:   Compute updated parameters:  $\hat{\theta}' = \hat{\theta} - \alpha \nabla_{\hat{\theta}} \mathcal{L}_{\mathcal{P}_i}(\hat{\theta})$
- 5: **end for**
- 6: **for all**  $\mathcal{D}_i^q$  **do**
- 7:   predict  $y_i^q = \hat{\theta}'(\mathcal{D}_i^q)$
- 8: **end for**

---

meta-update. This process is repeated for many epochs with many person-tasks, to learn a generic model,  $\theta$  that can be rapidly adapted to a new person. The meta-model learning, is influenced by the categorical cross-entropy loss generated by the  $\mathcal{D}^q$  as in Equation 2.

$$\mathcal{L}_{\mathcal{P}_i}(\theta_i) = \sum_{x^q, y^q \sim \mathcal{D}^q} y^q \log \theta_i(x^q) + (1 - y^q) \log(1 - \theta_i(x^q)) \quad (2)$$

A meta-test person  $\hat{p}$ , not seen during training, uses its support set,  $\mathcal{D}^s$  to train a parametric model  $\hat{\theta}$ , initialised by the meta-model  $\theta$ , for few gradient steps (*meta\_gs*). Thereafter, the personalised,  $\hat{\theta}$  is used to classify instances in its query set,  $\mathcal{D}^q$  as in Algorithm 2. We note that we prefer First-Order MAML [2] when implementing Personalised MAML, which is computationally less intensive, yet achieves comparable performances in comparison to MAML [2].

### 3.3 Personalised RN

---

#### Algorithm 3 Personalised RN Training

---

**Require:**  $p(\mathcal{P})$ : HAR dataset; distribution over persons

**Require:**  $\alpha$ : step size hyper-parameter

- 1: randomly initialise  $\theta_f$  and  $\theta_r$
- 2: **while** not done **do**
- 3:   Sample  $n$  person-tasks  $\mathcal{P}_i \sim p(\mathcal{P})$
- 4:   **for all**  $\mathcal{P}_i$  **do**
- 5:      $\mathcal{D}^s = \{(x, y) \in \mathcal{P}_i : |\mathcal{D}^s| = K^s \times |\mathcal{C}|\}$
- 6:      $\mathcal{D}^q = \{(x, y) \in \mathcal{P}_i : |\mathcal{D}^q| = K^q \times |\mathcal{C}|, \mathcal{D}^s \cap \mathcal{D}^q = \emptyset\}$
- 7:     **for all**  $x_i^q$  **do**
- 8:       Create train data instance  $(x_i^q, \mathcal{D}^s)$
- 9:     **end for**
- 10:   **end for**
- 11:   Compute  $\nabla \mathcal{L}_{\mathcal{P}_i}(\theta_f, \theta_r)$  w.r.t. train data instances of size  $n \times K^q \times |\mathcal{C}|$
- 12:   Update  $(\theta_f, \theta_r) \leftarrow (\theta_f, \theta_r) - \alpha \nabla_{\theta_f} \mathcal{L}_{\mathcal{P}_i}(\theta_f, \theta_r)$
- 13: **end while**

---

#### Algorithm 4 Personalised RN Testing

---

**Require:** Support set  $\mathcal{D}^s$  for test person  $\hat{p}$ ,

**Require:**  $\theta_r, \theta_f$

- 1: **for all**  $x_i^q$  **do**
  - 2:   predict  $y_i^q = \theta_f, \theta_r(\mathcal{D}_i^q, \mathcal{D}^s)$
  - 3: **end for**
- 

Relation Network (RN) [12] is a Few-shot Meta-Learning algorithm that “learns-to-match”. RN has a similar goal to other Meta-Learners, of generalising over many tasks. Personalised Relation Networks  $RN^p$  learns a matching, generalisable to new persons encountered at test time. The meta-task design for  $RN^p$  is similar to  $MAML^p$ , where the the support set,  $\mathcal{D}^s$ , and the query set  $\mathcal{D}^q$ , is selected from the same person. Meta-training instance for person-task,  $\mathcal{P}_i$ , is created by combining each data instance  $x_i^q$ , in  $\mathcal{D}^q$ , with the support set,  $\mathcal{D}^s$ .

During training (Algorithm 3),  $RN^p$  learns to match  $x_i^q$  to a matching instance in  $\mathcal{D}^s$ . A parametric model,  $\theta_f$  learns feature representations for every

instance; next each instance, in  $\mathcal{D}^s$  is paired with the  $x_i^q$  to create  $|\mathcal{D}^s|$  number of pairs. The parametric model,  $\theta_r$ , then learns to estimate the similarity of the paired instances. With the personalised approach, the similarity is always estimated against ones own data in the support set. The network is trained end to end using mean squared error loss as in Equation 3. Here the output of  $\theta_r$  is of size 1 which is expected to be 1 if a matching pair or 0 if not matching pair. A meta-test person  $\hat{p}$ , not seen during training, can use trained  $RN^p$  to match a query instance to an instance in it’s own support set  $\mathcal{D}^s$  provided during calibration, and therein use the class of the matched support instance as the predicted class label (Algorithm 4).

$$\mathcal{L}_{\mathcal{P}_i}(\theta_i) = \sum_{x^q, y^q \sim \mathcal{D}_q} \|\theta_f, \theta_r(x^q, \mathcal{D}^s) - y^q\|_2^2 \quad (3)$$

## 4 Evaluation

We compare the performance of **personalised MAML** ( $MAML^p$ ) and **personalised RN** ( $RN^p$ ) against a number of baselines as listed below;

**DL**: Best performing DL algorithm from benchmarks published in [16]

**MN**: Few-shot Learning classifier Matching Networks from [14]

**MN<sup>p</sup>**: Personalised Matching Networks from [8]

**MAML**: Model-Agnostic Meta-Learner [2]

**MAML<sup>p</sup>**: Personalised MAML from

**RN**: Relation Networks [12]

**RN<sup>p</sup>**: Personalised RN from Section 3.3 Section 3.2

### 4.1 Datasets and Pre-processing

We use three data sets to create 9 single modality sensing experiments. Feature learners in both  $MAML^p$  and  $RN^p$  are model agnostic, such that the feature representation learning models are interchangeable to suit any modality combination. MEx <sup>1</sup> is a Physiotherapy Exercises dataset compiled with 30 participants performing 7 exercises. A depth camera (DC), a pressure mat (PM) and two accelerometers on the wrist (ACW) and the thigh (ACT) provide four sensor data streams creating four experiments. PAMAP2 <sup>2</sup> dataset contains 8 Activities of Daily Living recorded with 8 participants. Three accelerometers on the hand (H), the chest (C) and the ankle (A) provide three sensor data streams creating 3 experiments. SELFBACK <sup>3</sup> is a HAR dataset with 9 activities. These activities are recorded with 33 participants using two accelerometers on the wrist (W) and the thigh (T), creating 2 experiments.

<sup>1</sup> <https://archive.ics.uci.edu/ml/datasets/MEx>

<sup>2</sup> <http://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring>

<sup>3</sup> <https://archive.ics.uci.edu/ml/datasets/selfBACK>



A sliding window method is applied on each sensor data stream to obtain data instances. Window size of 5 seconds is applied for all 9 datasets and an overlap of 3, 1 and 2.5 for data sources MEx, PAMAP2 and SELFBACK, resulted in 30, 76 and 88 data instance per person-activity on average. A few pre-processing steps are applied on data instances, adapted from previous work [16]. Resulting input sizes for  $\theta_f$  of RN and  $\theta$  of MAML are  $(5 \times 12 \times 16)$ ,  $(5 \times 16 \times 16)$  and  $(5 \times 3 \times 60)$  for DC, PM and AC modalities respectively.

## 4.2 Experiment Design

We use the DL results for MEx from previous work [16] and for comparability we implement the same network architectures for PAMAP2 and SELFBACK datasets. We use a 1 layer dense network with 1200 units as the feature learners in  $MN$ ,  $MN^p$ ,  $MAML$ ,  $MAML^p$  algorithms.  $RN$  and  $RN^p$  use a 1 layer 2D convolutional network as the feature learner and a 3 layer network with 1 2D convolutional layer, and 2 dense layers as the relation learner. All networks use batch normalisation for regularisation. Importantly, these architectures are significantly shallower compared to feature learners used in original  $MAML$  and  $RN$  architectures. We present results in a 5-shot setting where all algorithms are trained with early stopping.

We follow Leave-One-Person-Out (LOPO) evaluation where the data from one person is used to create meta-test tasks the rest to create meta-train tasks. We note that during testing, even  $MN$ ,  $MAML$  and  $RN$  preserve the personalisation aspect because only one user is present in the meta-test task when using LOPO. The meta-train and meta-test tasks are created while maintaining class balance; accordingly we report the accuracy of each experiment averaged over the number of person folds. We use the Wilcoxon signed-rank test for paired samples to evaluate the statistical significance at 95% confidence and highlight the significantly improved performances in bold text.

## 4.3 Results

Table 1: Performance comparison with MEx for exercise recognition

Algorithm	$MEx_{ACT}$	$MEx_{ACW}$	$MEx_{DC}$	$MEx_{PM}$
DL	0.9015	0.6335	0.8720	0.7408
$MN$	0.9073	0.4620	0.5065	0.6187
$MN^p$	0.9155	0.6663	0.9342	0.8205
$MAML$	0.8673	0.6525	0.9629	0.9283
$MAML^p$	0.9106	0.6834	<b>0.9795</b>	<b>0.9408</b>
$RN$	0.9327	0.7279	0.8189	0.8145
$RN^p$	<b>0.9436</b>	<b>0.7719</b>	0.9205	0.8520

Table 1 presents the comparative performances for 4 experiments on the exercise recognition task using the MEx dataset. We remind that MEx experiments

create few-shot classification settings, where one “person-activity” class has only 30 data instances and there are  $30 \times 7 = 210$  classes. Overall, personalised meta-learning models significantly outperformed conventional DL and meta-learning models. Notably with visual data;  $\text{MEx}_{DC}$  and  $\text{MEx}_{PM}$ , best performance is achieved with the optimisation based personalised meta-learner  $\text{MAML}^p$ , in contrast, accelerometer data prefer learning to compare method  $RN^p$ . It is noteworthy that the personalised few-shot learning algorithm  $MN^p$ , achieves comparable performance against  $\text{MAML}^p$  with  $\text{MEx}_{ACT}$  data and outperform  $RN^p$  with  $\text{MEx}_{DC}$  data, yet fails to outperform best personalised meta-learner. Overall, with ExRec we confirm the importance of personalisation and demonstrate that personalised meta-learners successful adapt to new unseen persons with few-data.

Table 2: Performance comparison with SELFBACK and PAMAP2 for general human activity recognition

Algorithm	$\text{SB}_T$	$\text{SB}_W$	$\text{PMP}_H$	$\text{PMP}_C$	$\text{PMP}_A$
DL	0.7880	0.6997	0.7505	0.7878	0.8075
$MN$	0.8392	0.7669	0.6625	0.7536	0.7361
$MN^p$	0.9124	<b>0.8653</b>	0.7484	<b>0.8548</b>	<b>0.8330</b>
$\text{MAML}$	0.8398	0.7532	0.7593	0.7626	0.6830
$\text{MAML}^p$	0.8625	0.8075	<b>0.8037</b>	0.7822	0.7256
$RN$	0.9334	0.8276	0.7818	0.8170	0.7527
$RN^p$	<b>0.9487</b>	0.8528	0.7868	0.8294	0.7761

In comparison to MEx experiments, PAMAP2 and SELFBACK experiments do not create strictly few-shot classification settings with 76 and 88 data instances per “person-activity” class. We compare their performance on our personalised methodology against conventional DL and few-shot learning methods. These experiments helps to understand if improvements we observe in MEx experiments can be re-produced in a not-strictly few-shot classification settings.

Results show that personalised meta-learners have outperformed conventional meta-learners in all 5 experiments and outperformed conventional DL models in 4 out of 5 experiments. It is noteworthy that in experiment with  $\text{PMP}_H$  data,  $RN^p$  performance is comparable with  $RN$ . While two of the five experiments significantly outperform personalised  $MN^p$ , three experiments fail to outperform  $MN^p$ . But all experiments achieve their best performance with a personalised algorithms further confirming the significance of personalisation in different domains of HAR. The failure to outperform DL methods in one occasion is as expected given the larger amount of data available for training. In addition, all 5 experiment use accelerometer data, where  $MN^p$ 's simpler similarity metric is proven to be sufficient to discriminate significant similarity relations between different classes.

Considering all 9 experiments, we find that visual data prefer the optimisation based meta-learning algorithm (i.e.  $MAML^p$ ) and experiments with time-series data prefer learning to compare methods (i.e.  $MN^p$  and  $RN^p$ ). It is noteworthy that  $MAML^p$  and  $MN^p$  use a 1-dense layer network and  $RN^p$  uses a 1-convolution layer network for feature learning while achieving significant performance improvements. Overall, personalisation strategy introduced in this paper has elevated the conventional meta-learners significantly when using few-data and shallow network architectures. Personalisation has positively contributed towards eliminating the need for parametric models with many deep layers that require a large labelled data collection for training. This is highly significant outcome in the domain of HAR, where even a comprehensive data collection fails to cover all possible personal nuances that a reasoning model may encounter during deployment.

## 5 Conventional vs. Personalised Meta-Learners

Here we look closer at training of meta-learners to understand how personalisation methodology improved performance using an experiment with the  $MEXP_M$  data.

### 5.1 MAML vs. $MAML^p$

We first investigate the performance improvements achieved by  $MAML^p$  against  $MAML$ . Here we compare three variants,  $MAML$  where meta-train and test tasks are created disregarding any person identifiers;  $MAML^p$ , as described in Section 3.2; and person-aware  $MAML$ . Here person-aware  $MAML$  can be seen as a lazy personalisation of MAML where a meta-train task is sampled from a set of persons, where one person contributes data for only one exercise class in the support set. The query set will have data from a single person who may not have been selected to form the support set. This method still preserves the concept of “person-activity” only at the class label level, but not over the entire support set level. We visualise the impact of model adaptation at test time using the three different algorithms in both,  $K^s = 1$ , and  $K^s = 5$ , settings on the  $MEXP_M$  dataset in Figures 3 and 4.

Here we plot test-person accuracy (y-axis) evaluated at every 10 meta-train epochs ( $2^{nd}$  row of the x-axis); at each of these evaluation points, the meta-test support set is used to adapt the current meta-model for a further 10 steps ( $1^{st}$  row of the x-axis). During each adaptation steps we record accuracy using the meta-test query. Through this process we can observe the impact on the partially optimised general meta-model when being adapted at the test time at increasing adaptation steps.  $MAML^p$  and person-aware  $MAML$  significantly outperformed  $MAML$  in both settings. When comparing  $MAML^p$  and person-aware  $MAML$ ,  $MAML^p$  algorithm achieves a more generic meta-model even without performing meta-gradient steps for meta-model adaptation (every 0 on  $1^{st}$  row of the x-axis); this is most significant in the  $K^s = 1$  setting. These

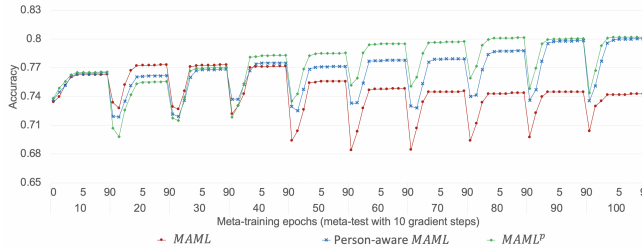


Fig. 3: MAML vs. person-aware MAML vs.  $MAML^P$  with  $MEX_{PM}$  when  $K^s = 1$

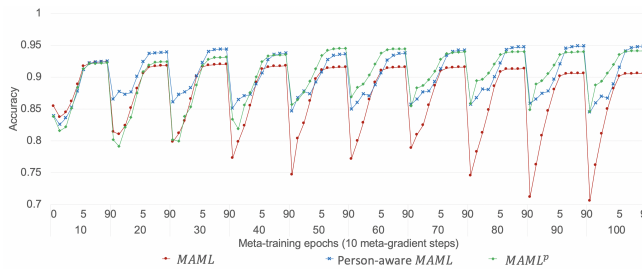


Fig. 4: MAML vs. person-aware MAML vs.  $MAML^P$  with  $MEX_{PM}$  when  $K^s = 5$

observations verify the advantage of creating personalised tasks; that even with the person-aware  $MAML$  algorithm where a task contains data from multiple people. But a “person-activity” class only contains data from one person has allowed learning a more generic meta-model as seen with  $MAML$ ,  $MAML^P$ .

$MAML^P$  where all “person-activity” data belongs to the same person, provides further generalisation with faster adaptation. Another indication of the significance of personalisation is found when investigating  $MAML$  performance over the training epochs. While  $MAML$  overall performance is indifferent as the meta-model train,  $MAML$  meta-test accuracy before adaptation (every 0 on 1<sup>st</sup> row of the x-axis), declines consistently. This is most significant when  $K^s = 5$ , which indicates that the meta-model learned with  $MAML$  is not generalised when an activity class in a meta-train task support set contains data from multiple people. In comparison, meta-model learned with  $MAML^P$ , performs well on meta-test tasks, even before adaptation.

### 5.2 RN vs. $RN^P$

Similarly we compare the performance between the two algorithms Relation Networks (RN) and personalised RN ( $RN^P$ ) to understand the effect of personalisation on training and testing. For this purpose we create experiments with the  $MEX_{PM}$  dataset in two settings  $K^s = 1$  and  $K^s = 5$  and evaluate the model at every 10 epochs using meta-test tasks, which we plot in Figures 5a and 5b.

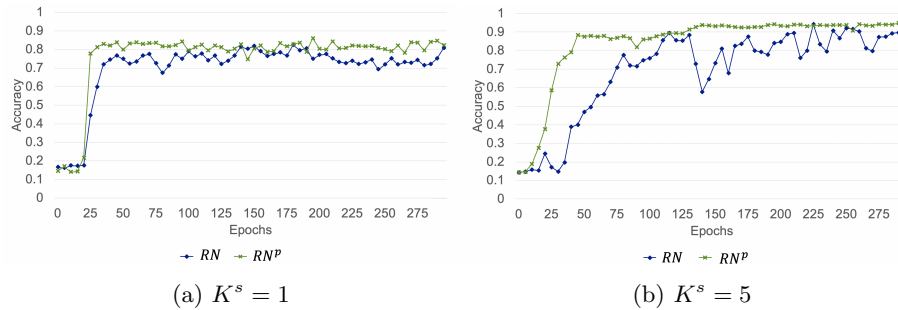


Fig. 5: RN vs.  $RN^P$  with  $MEx_{PM}$  meta-model tested at every 10 meta-train epochs

It is evident that personalisation has stabilised the meta-training process, where meta-testing performance consistently improve with  $RN^P$  models. In contrast meta-test evaluation on the  $RN$  is erratic, especially evident in the  $K^s = 5$  setting. When training  $RN$  in the  $K^s = 5$  setting, a task is created by disregarding the person parameter, as a result, an activity class contains data instances from more than one person and learning similarities to many people has adversely affected the learning of the  $RN$  meta-model. Similarly, in the  $K^s = 1$  setting, when a task contains only one data instance per class, learning from ones own data with  $RN^P$  is advantages in comparison to  $RN$  where the data instance for a class is from another person. Overall these results confirm the strong presence of personal nuances in sensor data, that need to be considered when creating classification models for exercise recognition.

## 6 Discussion

While  $RN^P$  does not require model-retraining, obtaining the activity class label for a given query involves a more complex inference process; each data instance in the end-user provided support set and the query instance is converted to feature vectors and later concatenated to obtain the relation scores and the predicted class. We calculate the average time elapsed for obtaining a prediction on the  $MEx_{ACT}$  query data instance, using both algorithms in a computer with 8GB RAM and 3.1 GHz Dual-Core processor. While  $MAML^P$  takes 0.0156 milliseconds for a single prediction,  $RN^P$  takes 2.4982 milliseconds when  $K^s = 1$  and 3.7218 milliseconds when  $K^s = 5$ . A HAR algorithm should be able to recognise activities as they are performed in real-time for the best user experience, and the processor and memory requirements along with the response time are crucial considerations for edge device deployment. In comparison,  $MAML^P$  inference is a simple classification task but requires post-deployment model re-training which calls for deployment in a development-friendly environment.

A limitation of  $MAML^P$  is the inability to perform open-ended HAR. Originally both  $MAML$  and  $RN$  perform zero-shot Learning for image classifica-

tion [2, 12] with a fixed class length. Specifically, *MAML* is restricted to performing multi-class classification with a conventional soft-max layer; for instance 5 outputs for a 5-class classification task. Open-ended HAR requires dynamic expansion of the decision layer as the person adds new activities in addition to the activities that are already included. Few-shot classifiers such as Matching Networks (MN) [14] does not have a strict decision layer which inspired Open-ended MN [17] for Open-ended HAR. Similarities of Relation Networks (RN) to MN presents the opportunity to improve Open-ended HAR using RN, which we will explore in future.

When a Personalised Meta-Learning model is trained and embedded in the fitness application, there is an initial configuration step that is required for collecting the calibration data (i.e. support set) of the end-user. The end-user will be instructed to record a few seconds of data for each activity using the sensor modalities synchronised with the fitness application. This is similar to demographic configurations users perform when installing new fitness applications (on-boarding). Thereafter this support set will be used by the algorithm either to re-train the model (*MAML<sup>p</sup>*) or for comparison (*RN<sup>p</sup>*). Importantly, both *MAML<sup>p</sup>* and *RN<sup>p</sup>* provide the opportunity to provide new calibration data if the physiology or preferences of the user change over time.

## 7 Conclusion

In this paper, we presented “personalised meta-learning”, a methodology optimised for personalisation of Human Activity Recognition (HAR) using only a few labelled data. This is achieved by treating the “person-activity” pair in a HAR dataset as an activity class, where each class now has only a few instances of data for training. We implement personalised meta-learning with two meta-learners for few-shot classification personalised MAML (*MAML<sup>p</sup>*) and personalised Relation Networks (*RN<sup>p</sup>*) where a meta-model is learned, such that it can be rapidly adapted to any person not seen during training. Both algorithms require only a few instances of calibration data from the end-user to personalise the meta-model. At deployment, *MAML<sup>p</sup>* uses calibration data for model re-training and *RN<sup>p</sup>* uses calibration data directly for matching (without re-training). Our evaluation with 9 experiments shows that both algorithms achieve significant performance improvements in a range of HAR domains while outperforming state-of-the-art deep learning and conventional meta-learning algorithms. We highlight that personalisation achieves higher meta-model generalisation, compared to conventional methods, allowing rapid adaptation. Importantly we find, real-time inference with *MAML<sup>p</sup>* is significantly faster with fewer memory requirements compared to *RN<sup>p</sup>* where calibration data need to be retained in memory.

## References

1. Berchtold, M., Budde, M., Gordon, D., Schmidtke, H.R., Beigl, M.: Actiserv: Activity recognition service for mobile phones. In: International Symposium on Wearable

- Computers (ISWC) 2010. pp. 1–8. IEEE (2010)
2. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 1126–1135. JMLR. org (2017)
  3. Graves, A., Wayne, G., Danihelka, I.: Neural turing machines. arXiv preprint arXiv:1410.5401 (2014)
  4. Longstaff, B., Reddy, S., Estrin, D.: Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In: 2010 4th International Conference on Pervasive Computing Technologies for Healthcare. pp. 1–7. IEEE (2010)
  5. Mishra, N., Rohaninejad, M., Chen, X., Abbeel, P.: A simple neural attentive meta-learner. arXiv preprint arXiv:1707.03141 (2017)
  6. Nichol, A., Achiam, J., Schulman, J.: On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999 (2018)
  7. Ordóñez, F.J., Roggen, D.: Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **16**(1), 115 (2016)
  8. Sani, S., Wiratunga, N., Massie, S., Cooper, K.: Personalised human activity recognition using matching networks. In: International Conference on Case-Based Reasoning. pp. 339–353. Springer (2018)
  9. Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., Lillicrap, T.: Meta-learning with memory-augmented neural networks. In: International conference on machine learning. pp. 1842–1850 (2016)
  10. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in neural information processing systems. pp. 4077–4087 (2017)
  11. Sun, X., Kashima, H., Ueda, N.: Large-scale personalized human activity recognition using online multitask learning. *IEEE Transactions on Knowledge and Data Engineering* **25**(11), 2551–2563 (2012)
  12. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1199–1208 (2018)
  13. Tapia, E.M., Intille, S.S., Haskell, W., Larson, K., Wright, J., King, A., Friedman, R.: Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: 2007 11th IEEE international symposium on wearable computers. pp. 37–40. IEEE (2007)
  14. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in neural information processing systems. pp. 3630–3638 (2016)
  15. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* **119**, 3–11 (2019)
  16. Wijekoon, A., Wiratunga, N., Cooper, K.: Mex: Multi-modal exercises dataset for human activity recognition. arXiv preprint arXiv:1908.08992 (2019)
  17. Wijekoon, A., Wiratunga, N., Sani, S., Cooper, K.: A knowledge-light approach to personalised and open-ended human activity recognition. *Knowledge-Based Systems* **192**, 105651 (2020)
  18. 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. pp. 351–360 (2017)