



**AUTHOR(S):**

**TITLE:**

**YEAR:**

**Publisher citation:**

**OpenAIR citation:**

**Publisher copyright statement:**

This is the \_\_\_\_\_ version of proceedings originally published by \_\_\_\_\_  
and presented at \_\_\_\_\_  
(ISBN \_\_\_\_\_; eISBN \_\_\_\_\_; ISSN \_\_\_\_\_).

**OpenAIR takedown statement:**

Section 6 of the "Repository policy for OpenAIR @ RGU" (available from <http://www.rgu.ac.uk/staff-and-current-students/library/library-policies/repository-policies>) provides guidance on the criteria under which RGU will consider withdrawing material from OpenAIR. If you believe that this item is subject to any of these criteria, or for any other reason should not be held on OpenAIR, then please contact [openair-help@rgu.ac.uk](mailto:openair-help@rgu.ac.uk) with the details of the item and the nature of your complaint.

This publication is distributed under a CC \_\_\_\_\_ license.

\_\_\_\_\_

# kNN Sampling for Personalised Human Activity Recognition

Sadiq Sani<sup>1</sup>, Nirmalie Wiratunga<sup>1</sup>, Stewart Massie<sup>1</sup>, and Kay Cooper<sup>2</sup>

<sup>1</sup> School of Computing Science and Digital Media,

<sup>2</sup> School of Health Sciences,

Robert Gordon University,

Aberdeen AB10 7GJ, Scotland, UK

{s.a.sani,n.wiratunga,s.massie,k.cooper}@rgu.ac.uk

**Abstract.** The need to adhere to recommended physical activity guidelines for a variety of chronic disorders calls for high precision Human Activity Recognition (HAR) systems. In the SELFBACK system, HAR is used to monitor activity types and intensities to enable self-management of low back pain (LBP). HAR is typically modelled as a classification task where sensor data associated with activity labels are used to train a classifier to predict future occurrences of those activities. An important consideration in HAR is whether to use training data from a general population (subject-independent), or personalised training data from the target user (subject-dependent). Previous evaluations have shown that using personalised data results in more accurate predictions. However, from a practical perspective, collecting sufficient training data from the end user may not be feasible. This has made using subject-independent data by far the more common approach in commercial HAR systems. In this paper, we introduce a novel approach which uses nearest neighbour similarity to identify examples from a subject-independent training set that are most similar to sample data obtained from the target user and uses these examples to generate a personalised model for the user. This nearest neighbour sampling approach enables us to avoid much of the practical limitations associated with training a classifier exclusively with user data, while still achieving the benefit of personalisation. Evaluations show our approach to significantly out perform a general subject-independent model by up to 5%.

## 1 Introduction

Human Activity Recognition (HAR) is the computational discovery of human activity from sensor data and is increasingly being adopted in health, security, entertainment and defense applications [9]. An example of the application of HAR in healthcare is SELFBACK [2], a system designed to assist users with low back pain (LBP) by monitoring their level of physical activity in order to provide advice and guidance on how best to adhere to recommended physical activity guidelines. Guidelines for LBP recommend that patients should not be sedentary for long periods of time and should maintain moderate physical

activity. SELFBACK continuously reads sensor data from a wearable device worn on the user’s wrist, and recognises user activities in real time. This allows SELFBACK to compare the user’s activity profile to the recommended guidelines for physical activity and produce feedback to inform the user on how well they are adhering to these guidelines. Other information in the user’s activity profile include the durations of activities and, for walking, the counts of steps taken, as well as intensity e.g. slow, normal or fast. The categorisation of walking into slow, normal and fast allows us to better match the activity intensity (i.e. low, moderate or high) recommended in the guidelines. HAR is typically modelled as a classification task where sensor data associated with activity labels are used to train a classifier to predict future occurrences of those activities. This introduces two important considerations, representation and personalisation.

Many different representation approaches have been proposed for HAR. In this paper, we broadly classify these approaches into three: hand-crafted, transformational and deep representations. Previous works have not provided a definitive answer as to which feature extraction approach is best due to the often mixed or contradictory results reported in different works [14]. This may be attributed to the differences in the configurations (e.g. sensor types, sensor locations, types of activities etc.) used in different works. For this reason, we conduct a comparative study of five different representation approaches from the three representation classes in order to determine which representation works best for our particular configuration (single wrist-mounted accelerometer) and our choice of activity classes.

The second consideration for HAR is personalisation, where training examples can either be acquired from a general population (subject-independent), or from the target user of the system (subject-dependent). Previous works have shown using subject-dependent data to result in superior performance [3, 8, 16]. The relatively poorer performance of subject-independent models can be attributed to variations in activity patterns, gait or posture between different individuals [11]. However, training a classifier exclusively with user provided data is not practical in a real-world configuration as this places significant cognitive burden on the user to provide sufficient amounts of training data required to build a personalised model. In this paper, we introduce a nearest neighbour sampling approach for subject-independent training example selection. In doing so, we achieve personalisation by ensuring only those examples that best match a user’s activity pattern influence the generation of the HAR model. Our approach uses nearest neighbour to identify subject-independent examples that are most similar to a small number of labelled examples provided by the user. In this way, our approach avoids the practical limitations of subject-dependent training. Our work draws inspiration from selective sampling in CBR where useful cases are sampled from the set of available cases for building effective case-bases [7, 17].

The rest of this paper is organised as follows: in Section 2, we discuss important related work on personalised HAR and selective sampling of examples. Section 3 discusses the different feature representation approaches considered in this work, while our kNN sampling approach is described in Section 4. A de-

scription of our dataset is presented in Section 5, evaluations are presented in Section 6 and conclusions follow in Section 7.

## 2 Related Work

The common approach to classifier training in HAR is to use subject-independent examples to create a general classification model. However, comparative evaluation with personalised models, trained using subject-dependent examples, show this to produce more accurate predictions [3, 8, 16]. In [16], a general model was compared with a personalised model using a c4.5 decision tree classifier. The general model produced an accuracy of 56.3% while the personalised model produced an accuracy of 94.6% using the same classification algorithm, which is an increase of 39.3%. Similarly, [3] and [8] reported increases of 19.0% and 9.7% between personalised and general models respectively. However, all rely on access to subject-dependent training dataset. Such an approach has limited practical use for real-world applications because of the burden it places on users to provide sufficient training data.

Different types of semi-supervised learning approaches have been explored for personalised HAR e.g. Self-learning, Co-learning and Active learning, which bootstrap a general model with examples acquired from the user [11]. Both Self-learning and Co-learning attempt to infer accurate activity labels for unlabelled examples without querying the user. This way, both approaches manage to avoid placing any labelling burden on the user. In contrast, Active learning selectively chooses the most useful examples to present to the user for labelling. Hence, while Active learning does not avoid user labelling, it attempts to reduce it to a minimum using techniques such as uncertainty sampling which consistently outperform random sampling [12]. Our work does not focus on uncertainty, but instead uses similarity as the focus.

While semi-supervised learning approaches address the data acquisition bottleneck of subject-dependent training, they do not address the presence of noisy or inconsistent examples in the general model. It is our view that part of the reason why general models do not perform very well is that some examples are sufficiently distinct from the activity pattern of the current user that they contribute more to noise in the training set. Therefore, an attempt at selecting only the most useful examples from the training set for classifier training is likely to improve classification performance.

In CBR, sampling methods have been employed for casebase maintenance. Here the aim is to delete cases that fail to contribute to competence such that edited case-bases consistently lead to retrieval gains [15]. Case selection heuristics commonly exploit neighbourhood properties as a cue to identify areas of uncertainty and in doing so, active sampling approaches are adopted to inform case selection [4, 17]. For our intended application, the criterion for example selection is very well defined. We seek to select examples from the available training set that are similar to examples supplied by the user, in order to personalise our classifier to the user’s activity pattern. Accordingly, we use a  $k$  Nearest Neigh-

hour sampling approach where the  $k$  most similar examples to the user’s data are selected.

### 3 Feature Representation

Feature representation approaches for accelerometer data for the purpose of HAR can be divided into three categories: handcrafted features, frequency transform features and deep features.

#### 3.1 Hand-crafted Features

This is the most common representation approach for HAR and involves the computation of a number of defined measures on either the raw accelerometer data (time-domain) or the frequency transformation of the data (frequency domain) obtained using Fast Fourier Transforms (FFTs). These measures are designed to capture the characteristics of the signal e.g. average acceleration, variation in acceleration, dominant frequency etc. that are useful for distinguishing different classes of activities. For both time and frequency domain hand-crafted features, the input is a vector of real values  $\vec{v} = v_1, v_2, \dots, v_n$  for each axis  $x, y$  and  $z$ . A function  $\theta_i$  (e.g. mean) is then applied to each vector  $\vec{v}$  to compute a single feature value  $f_i$ . The final representation is a vector of length  $l$  comprised of these computed features  $\vec{f} = f_1, f_2, \dots, f_l$ . The time-domain and frequency domain features used in this work are presented in Table 1. Further information on these features can be found in [20] and [5] respectively.

| <i>Time-Domain Features</i>  | <i>Frequency Domain Features</i> |
|------------------------------|----------------------------------|
| Mean                         | Dominant frequency               |
| Standard deviation           | Spectral centroid                |
| Inter-quartile range         | Maximum                          |
| Lag-one-autocorrelation      | Mean                             |
| Percentiles (10,25,50,75,90) | Median                           |
| Peak-to-peak amplitude       | Standard deviation               |
| Power                        |                                  |
| Skewness                     |                                  |
| Kurtosis                     |                                  |
| Log-energy                   |                                  |
| Zero crossings               |                                  |
| Root squared mean            |                                  |

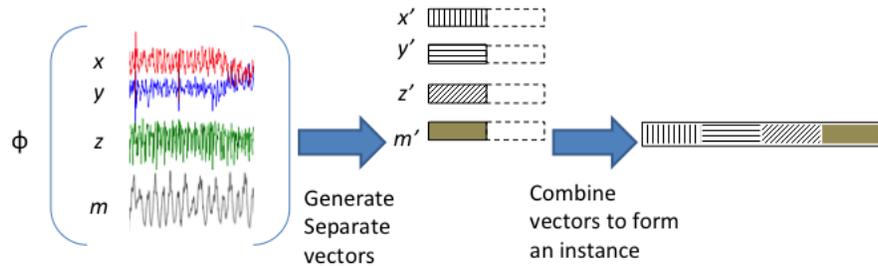
**Table 1.** Hand-crafted features for both time and frequency domains.

While hand-crafted features have worked well for HAR [9], a significant disadvantage is that they are domain specific. A different set of features need to be defined for each different type of input data i.e. accelerometer, gyroscope, time-domain or frequency domain values. Hence, some understanding of the

characteristics of the data is required. Also, it is not always clear which features are likely to work best. Choice of features is usually made through empirical evaluation of different combinations of features or with the aid of feature selection algorithms [19].

### 3.2 Frequency Transform Features

Frequency transform feature extraction involves applying a single function  $\phi$  on the vectors of raw accelerometer data to transform these into the frequency domain where it is expected that distinctions between different activities are better emphasised. Common transformations that have been applied include FFTs and Discrete Cosine Transforms (DCTs) [6, 13]. FFT is an efficient algorithm optimised for computing the discrete Fourier transform of a digital input by decomposing the input into its constituent sine waves. DCT is a similar algorithm to FFT which decomposes an input into its constituent cosine waves. Also, DCT returns an ordered sequence of coefficients such that the most significant information is concentrated at the lower indices of the sequence. This means that higher DCT coefficients can be discarded without losing information, making DCT better for compression. The main difference between frequency transform and frequency-domain hand-crafted features is that here, the coefficients of the transformation are directly used for feature representation without further feature computations. An overview of transform feature representation is presented in Figure 1.



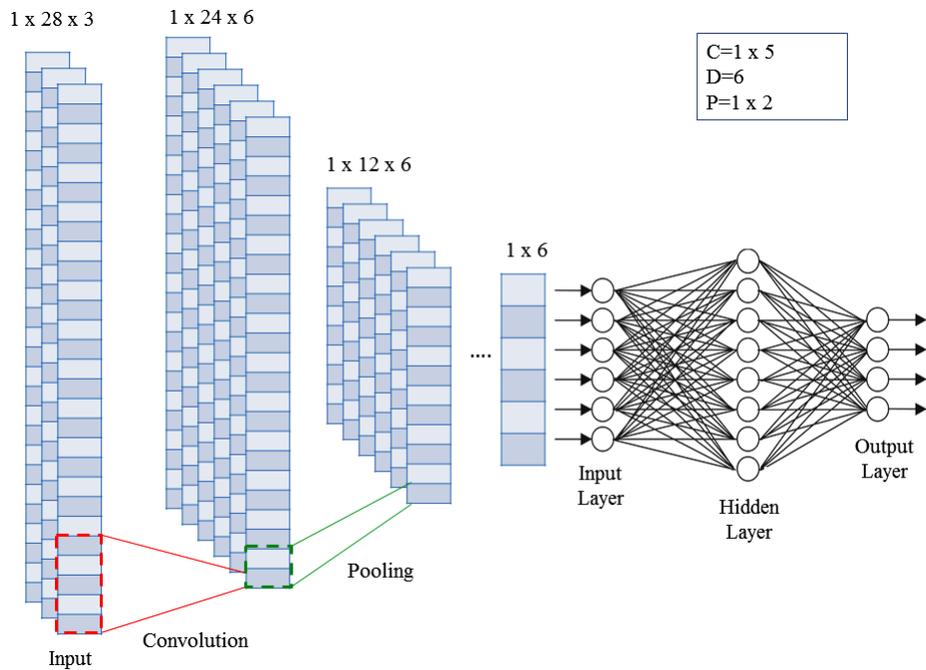
**Fig. 1.** Feature extraction and vector generation using frequency transforms.

A transformation function (DCT or FFT)  $\phi$  is applied to the time-series accelerometer vector  $\vec{v}$  of each axis  $\mathbf{x}' = \phi(\mathbf{x})$ ,  $\mathbf{y}' = \phi(\mathbf{y})$  and  $\mathbf{z}' = \phi(\mathbf{z})$ , as well as for the magnitude vector  $\mathbf{m} = \{m_{i1}, \dots, m_{il}\}$ . The output of  $\phi$  is a vector of coefficients which describe the sinusoidal wave forms that constitute the original signal. The final feature representation is obtained by concatenating the absolute values of the first  $l$  coefficients of  $\mathbf{x}'$ ,  $\mathbf{y}'$ ,  $\mathbf{z}'$  and  $\mathbf{m}'$  to produce a single feature

vector of length  $4 \times l$ . The value  $l = 48$  is used in this work, which is determined empirically.

### 3.3 Deep Features

Recently, deep learning approaches have been applied to the task of HAR due to their ability to extract features in an unsupervised manner. Deep approaches are able to stack multiple layers of operations to create a hierarchy of increasingly more abstract features [10]. Early work using Restricted Boltzmann Machines for HAR have only shown comparative performance to FFT and Principal Component Analysis [13]. More recent applications have used more of Convolutional Neural Networks (CNNs) due to their ability to model local dependencies that may exist between adjacent data points in the accelerometer data [18]. CNNs are a type of Deep Neural Network that have the ability for feature extraction by stacking multiple convolutional operators [10]. An example of a CNN is shown in Figure 2.



**Fig. 2.** Illustration of CNN

The input into the CNN in Figure 2 is a 3-dimensional matrix representation with dimensions  $1 \times 28 \times 3$  representing the width, length and depth respectively.

Tri-axial accelerometer data typically have a width of 1, a length  $l$  and a depth of 3 representing the  $x$ ,  $y$  and  $z$  axes. A convolution operation is then applied by passing a convolution filter over the input which exploits local relationships between adjacent data points. This operation is defined by two parameters,  $D$  representing the number of convolution filters to apply and  $C$ , the dimensions of each filter. For this example,  $D = 6$  and  $C = 1 \times 5$ . The output of the convolution operation is a matrix with dimensions  $1 \times 24 \times 6$ , these dimensions being determined by the dimension of the input and the parameters of the convolution operation applied. This output is then passed through a Pooling operation which basically performs dimensionality reduction. The parameter  $P$  determines the dimensions of the pooling operator which in this example is  $1 \times 2$ , which results in a reduction of the width of its input by half. The output of the pooling layer can be passed through additional Convolution and Pooling layers. The output of the final Pooling layer is then flattened into a 1-dimensional representation and then fed into a fully connected neural network. The entire network (including convolution layers) is trained through back propagation over a number of generations until some convergence criteria is reached. Detailed description of CNNs can be obtained in [10].

## 4 kNN Sampling

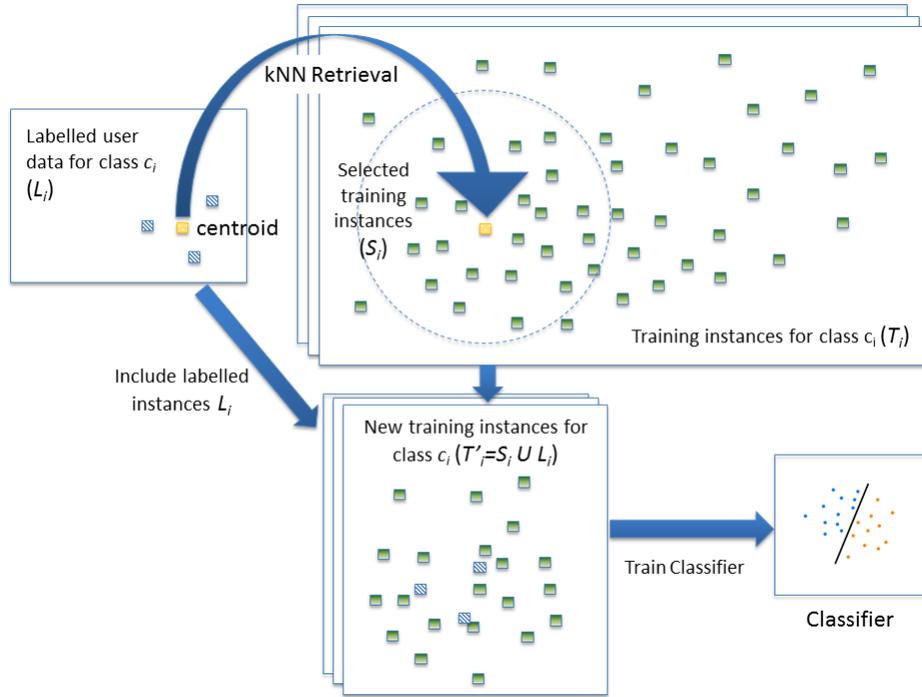
The main limitation of a general activity recognition model is that it fails to account for the slight variations and nuances in movement patterns of individuals. However, we hypothesise that similarities do exist in activity patterns between users. Hence, by identifying data that is most similar to the current user’s movement pattern, we can build a more effective HAR model that is personalised to the current user.

In order to identify similar data to the current user’s activity pattern, we need sample data from the user. In our current approach, we assume that the user provides a small sample of annotated data for each type of activity. This is similar to the calibration approach which is commonly employed in gesture control devices and is also used by the Nike + iPod fitness device [11]. Our selective sampling approach is illustrated in Figure 3.

The user provides a sample of  $n_i$  annotated examples for each class  $c_i \in C$ . These annotated examples are passed through feature extraction (e.g. DCT) to obtain a set of labelled examples  $L_i$ . The centroid of these examples ( $m_i$ ) is then obtained as the average of all examples in  $L_i$  using Equation 1.

$$m_i = \frac{\sum_j^{n_i} l_{ij}}{n_i} \quad (1)$$

Where  $l_{ij} \in L_i$ . The centroid  $m_i$  is used along with kNN to obtain the  $k$  most similar training examples  $S_i$  from the set of training examples  $T_i$  that belong to class  $c_i$ . The selected examples  $S_i$  are then combined with the user labelled examples  $L_i$  to form a new training set  $T'_i$  which is used for training a personalised classifier.



**Fig. 3.** Nearest neighbour sampling approach.

## 5 Dataset

A group of 50 volunteer participants was used for data collection. The age range of participants is 18 - 54 years and the gender distribution is 52% Female and 48% Male. Data collection concentrated on the activities provided in Table 2.

| Activity       | Description                                |
|----------------|--|
| Lying          | Lying down relatively still on a plinth    |
| Sitting        | Sitting still with hands on desk or thighs |
| Standing       | Standing relatively still                  |
| Walking Slow   | Walking at slow pace                       |
| Walking Normal | Walking at normal pace                     |
| Walking Fast   | Walking at fast pace                       |
| Up Stairs      | Walking up 4 - 6 flights of stairs         |
| Down Stairs    | Walking down 4 - 6 a flights of stairs     |
| Jogging        | Jogging on a treadmill at moderate speed   |

**Table 2.** Details of activities classes.

The set of activities in Table 2 was chosen because it represents the range of normal daily activities typically performed by most people. Three different walking speeds (slow, normal and fast) were included in order to have an accurate estimate of the intensity of the activities performed by the user. Identifying intensity of activity is important because guidelines for health and well-being include recommendations for encouraging both moderate and vigorous physical activity [1].

Data was collected using the Axivity Ax3 tri-axial accelerometer <sup>1</sup> at a sampling rate of 100Hz. Accelerometers were mounted on the right-hand wrists of the participants using specially designed wristbands provided by Axivity. Activities are roughly evenly distributed between classes as participants were asked to do each activity for the same period of time (3 minutes). The exceptions are Up stairs and Down stairs, where the amount of time needed to reach the top (or bottom) of the stairs was just over 2 minutes on average. This data is publicly available on Github <sup>2</sup>.

## 6 Evaluation

Evaluations are conducted using a leave-one-person out methodology where all data for one user is held out for testing and the remaining users' data are used for training the model. A time window of 5 seconds is used for signal segmentation and performance is reported using macro-averaged F1 score, a measure of accuracy that considers both precision (the fraction of examples predicted as class  $c_i$  that correctly belong to  $c_i$ ) and recall (the fraction of examples truly belonging to class  $c_i$  that are predicted as  $c_i$ ) for each class.

Our evaluation is composed of two parts. Firstly, we compare the different representations discussed in Section 3 using 2 classifiers: kNN and SVM. In the second section, we use the best representation/classifier combination to compare different selection approaches for generating personalised HAR models.

### 6.1 Feature Representations

In this section, we compare the feature representation approaches presented in Section 3 as follows:

- Time: Time domain features
- Freq: Frequency domain features
- FFT: Frequency transform features using FFT coefficients
- DCT: Frequency transform features using DCT coefficients

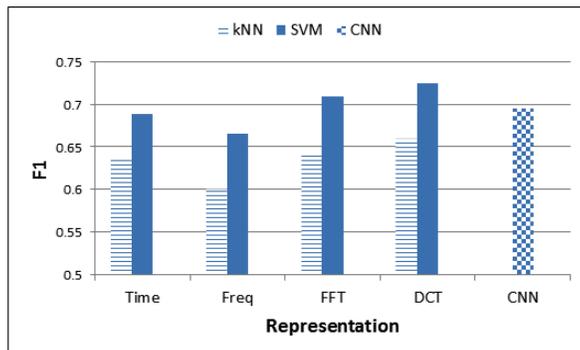
Each representation is evaluated with both a kNN and SVM classifier. In addition, we include a CNN classifier. The architecture of our CNN uses 3 convolution layers with convolution filter numbers  $D$ , set to 40, 20 and 10 respectively.

---

<sup>1</sup><http://axivity.com/product/ax3>

<sup>2</sup>[https://github.com/selfback/activity-recognition/tree/master/activity\\_data](https://github.com/selfback/activity-recognition/tree/master/activity_data)

Dimensions of each convolution filter  $C$ , are set to  $1 \times 10 \times 3$ . Each convolution layer is followed by a pooling layer with dimension  $P$ , set to  $1 \times 2$ . The output of the convolution is fed into a fully connected network with 2 hidden layers with 900 and 300 units respectively and an output layer with soft-max regression. Training of the CNN is performed for a maximum of 300 generations as longer training generations did not improve performance. The inclusion of CNN allows us to compare the performance of a state-of-the-art approach against conventional HAR approaches.



**Fig. 4.** Evaluation of different representations and classifiers.

Note from Figure 4 that the best result is achieved using *DCT* representation with SVM classifier, while second best is *FFT* with SVM. In general, SVM out performed kNN on all representation types. The poor performance of kNN might be because the dataset does not provide clearly separable neighbourhoods. Indeed it is intuitive to think many examples from similar classes e.g. sitting and lying, as well as slow, normal and fast walking would be within close proximity in the feature space and might not be easily distinguishable using nearest neighbour similarity. CNN came in third best in the comparison. This indicates the potential of CNN for HAR, however, in our evaluation, it did not beat the much simpler frequency transform approaches. Our results are consistent with the findings of [14] where CNNs did not out perform conventional approaches. Also, the high cost of retraining a CNN makes this approach impractical for personalisation using our approach.

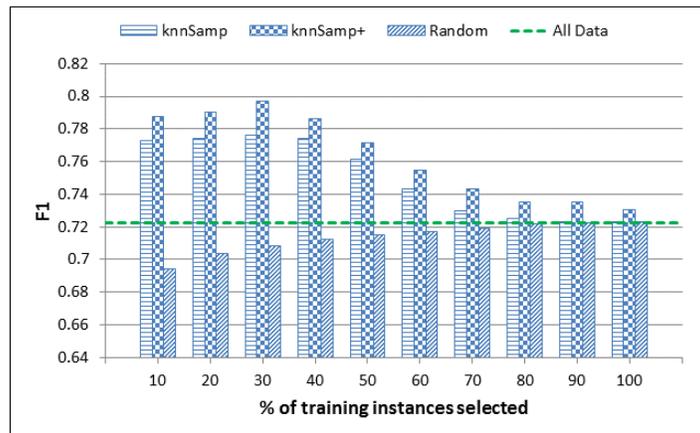
## 6.2 Selective Sampling

The second part of the evaluation uses the best representation/classifier combination, i.e. DCT+SVM to compare different sampling approaches for generating a personalised HAR model. The sampling approaches included in the comparison are as follows:

- All-Data: uses entire training set  $T$  without sampling;

- knnSamp: uses the kNN sampling approach presented in Section 4, but uses only the selected training examples  $S$  for classifier training;
- knnSamp<sup>+</sup>: uses the kNN sampling approach presented in Section 4 and uses the combined set  $T' = S \cup L$  for model generation; and
- Random: selects training examples at random for classifier training.

For any given user, 30% of test data is held-out to simulate user provided annotated data for personalisation. The remaining 70% forms the test data.



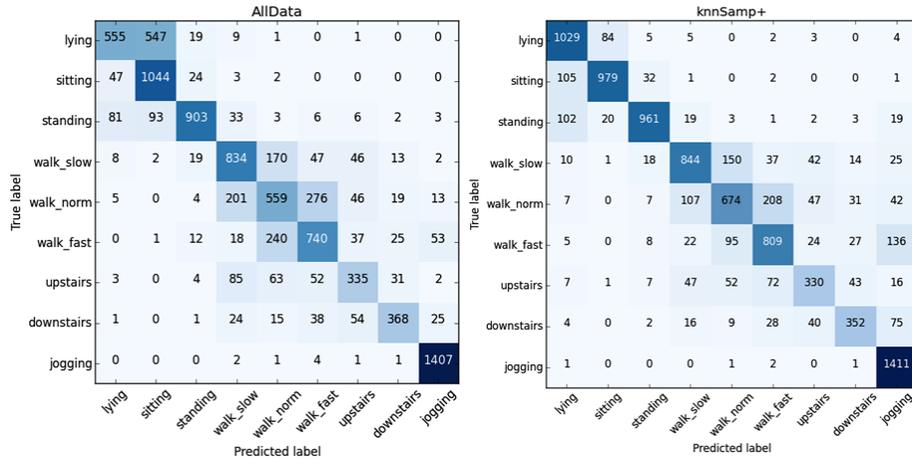
**Fig. 5.** Results for personalised model generation strategies.

Figure 5 shows the results of the different sampling approaches where the x-axis shows the percentage of training examples selected while the y-axis shows the F1 score. The horizontal line shows the result for All-Data. Observe that both knnSamp and knnSamp<sup>+</sup> significantly outperform all other approaches when no more than 50% of the training set is used, with the best result achieved using only 30% of the training set. F1 score declines after 50% as more of the noise from dissimilar examples in the training set are introduced into the model. The high accuracy of knnSamp compared to the other approaches indicates that the nearest neighbour selection strategy effectively selects useful similar examples for activity recognition. The best improvements of both knnSamp and knnSamp<sup>+</sup> compared to the other approaches are statistically significant at 99% using a paired T-test. Unsurprisingly, no improvement is achieved through random selection of training examples. Note that adding the user data to the entire training set without sampling produces only marginal improvement (+0.008 F1 Score).

### 6.3 Discussion

To further understand the performance gain of our personalisation approach, we present the break down of the performance (precision, recall and F1 score)

by class for the best performing sampling method  $\text{knnSamp}^+$  (with 30% sampling of training data) in Table 3. Here, we can see that personalisation had produced considerable improvement in the F1 scores of lying, sitting, walk\_slow, walk\_normal and walk\_fast. From the confusion matrix in Table 6, we can observe that without personalisation, about 50% (547) of lying examples are predicted as sitting. However, personalisation produces better separation between lying and sitting which is evidenced by the higher recall score of lying (0.91) and higher precision of sitting (0.90) after personalisation.



**Fig. 6.** Confusion matrices for All-Data (left) and  $\text{knnSamp}^+$  (right).

|             | All-Data  |        |          | $\text{knnSamp}^+$ |        |          |
|-------------|-----------|--------|----------|--------------------|--------|----------|
|             | Precision | Recall | F1 Score | Precision          | Recall | F1 Score |
| lying       | 0.80      | 0.49   | 0.61     | 0.81               | 0.91   | 0.86     |
| sitting     | 0.62      | 0.92   | 0.74     | 0.90               | 0.87   | 0.89     |
| standing    | 0.91      | 0.82   | 0.86     | 0.92               | 0.85   | 0.89     |
| up stairs   | 0.63      | 0.57   | 0.60     | 0.68               | 0.57   | 0.62     |
| down stairs | 0.79      | 0.70   | 0.74     | 0.75               | 0.67   | 0.71     |
| walk_fast   | 0.63      | 0.66   | 0.65     | 0.70               | 0.72   | 0.71     |
| walk_normal | 0.53      | 0.51   | 0.52     | 0.68               | 0.60   | 0.64     |
| walk_slow   | 0.69      | 0.72   | 0.70     | 0.80               | 0.74   | 0.77     |
| jogging     | 0.93      | 0.98   | 0.95     | 0.82               | 1.00   | 0.90     |

**Table 3.** Precision, Recall and F1 scores by class for All-Data and  $\text{knnSamp}^+$ .

A similar pattern can also be observed with the three different walking speeds. From Figure 6, it can be observed that without personalisation, only about half

(559 examples) of walking normal are predicted correctly, with most of the other half split between walking slow (201 examples) and walking fast (276 examples) giving a low recall score of 0.51. In addition, 240 walking fast examples are misclassified as walking normal which results in a low precision score for walking normal of 0.53. However, with personalisation, the number of walking normal examples predicted correctly increases to 674 while the number of misclassified walking fast examples reduces to 95 which improves the recall and precision scores to 0.68 and 0.60 respectively.

In contrast, the activities down stairs and jogging suffer a slight decline in F1 score, from 0.74 and 0.95 without personalisation to 0.71 and 0.90 with personalisation respectively. With personalisation, more examples from other classes are being misclassified as jogging. This requires further investigation to identify the root cause. However, jogging still benefits from higher recall from 0.98 to 1.0 with personalisation.

## 7 Conclusion

In this paper, we have presented a novel nearest neighbour sampling approach for personalised HAR that selects examples from a subject-independent training set that are most similar to a small number of user provided examples. In this way, much of the irrelevant examples in the general model are eliminated, the model is personalised to the user, and accuracy is improved. Evaluation shows our approach to outperform a general model by up to 5% of F1 score. Another advantage of our approach is that it avoids the practical limitation of subject-dependent training by reducing the data collection burden on the user.

Many different representation approaches have been proposed for HAR without a definitive best approach, partly due to the differences in configurations (e.g. sensor types, sensor locations etc.) and partly due to different mix of activity classes. Therefore, it is important to determine which representation approach is best suited for the configuration used in SELFBACK i.e., a single wrist-mounted accelerometer, as well as the types of activities. Accordingly, another contribution of this paper is a comparative study of five representation approaches including state-of-the-art CNNs on our dataset. Results show a frequency transform approach using DCT coefficients to outperform the rest.

A number of considerations have been identified for future work. Firstly, a method that further reduces if not eliminates the need for user annotated data will further improve the user experience of our system. Secondly, evaluations in this paper have only been applied on short time durations that immediately follow the user examples. Test data covering longer durations are needed in order to evaluate the performance of the personalised model over longer periods of time. If the accuracy of the model drops due to long term changes in context, it would be interesting to be able to identify these context changes in order to initiate further rounds of personalisation. Note that success in automatic acquisition of labelled examples should significantly aid this process of continuous personalisation with minimal impact on user experience.

## 8 Acknowledgment

This work was fully sponsored by the collaborative project SELFBACK under contract with the European Commission (# 689043) in the Horizon2020 framework. The authors would also like to thank all students and colleagues who volunteered as subjects for data collection.

## References

1. Abel, M., Hannon, J., Mullineaux, D., Beighle, A.: Determination of step rate thresholds corresponding to physical activity intensity classifications in adults. *Journal of Physical Activity and Health* 8(1), 45–51 (2011)
2. Bach, K., Szczepanski, T., Aamodt, A., Gundersen, O.E., Mork, P.J.: Case representation and similarity assessment in the selfback decision support system. In: *Proceedings of 24th International Conference on Case-Based Reasoning, ICCBR 2016*. pp. 32–46. Springer International Publishing (2016)
3. Berchtold, M., Budde, M., Gordon, D., Schmidtke, H.R., Beigl, M.: Actiserv: Activity recognition service for mobile phones. In: *Proceedings of International Symposium on Wearable Computers, ISWC 2010*. pp. 1–8 (2010)
4. Craw, S., Massie, S., Wiratunga, N.: Informed case base maintenance: A complexity profiling approach. In: *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*. p. 1618. AAAI Press (2007)
5. Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.: Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing* 14(7), 645–662 (2010)
6. He, Z., Jin, L.: Activity recognition from acceleration data based on discrete cosine transform and svm. In: *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, SMC 2009*. pp. 5041–5044. IEEE (2009)
7. Hu, R., Delany, S., MacNamee, B.: Sampling with confidence: Using k-nn confidence measures in active learning. In: *Proceedings of the UKDS Workshop at 8th International Conference on Case-based Reasoning, ICCBR 2009*. p. 50 (2009)
8. Jatoba, L.C., Grossmann, U., Kunze, C., Ottenbacher, J., Stork, W.: Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. In: *Proceedings of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 5250–5253 (Aug 2008)
9. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE* 15(3), 1192–1209 (2013)
10. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series. In: *Arbib, M.A. (ed.) The Handbook of Brain Theory and Neural Networks*, pp. 255–258. MIT Press (1998)
11. Longstaff, B., Reddy, S., Estrin, D.: Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In: *Proceedings of 4th International Conference on Pervasive Computing Technologies for Healthcare*. pp. 1–7 (March 2010)
12. Miu, T., Missier, P., Plötz, T.: Bootstrapping personalised human activity recognition models using online active learning. In: *Proceedings of IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and*

- Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM) 2015. pp. 1138–1147. IEEE (2015)
13. Plötz, T., Hammerla, N.Y., Olivier, P.: Feature learning for activity recognition in ubiquitous computing. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. pp. 1729–1734. IJCAI'11, AAAI Press (2011)
  14. Ronao, C.A., Cho, S.B.: Deep convolutional neural networks for human activity recognition with smartphone sensors. In: Arik, S., Huang, T., Lai, W.K., Liu, Q. (eds.) Neural Information Processing: 22nd International Conference, ICONIP 2015, November 9–12, 2015, Proceedings, Part IV. pp. 46–53. Springer International Publishing (2015)
  15. Smyth, B., Keane, M.T.: Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1. pp. 377–382. IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995)
  16. 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: Proceedings of 11th IEEE International Symposium on Wearable Computers. pp. 37–40. IEEE (2007)
  17. Wiratunga, N., Craw, S., Massie, S.: Index driven selective sampling for cbr. In: Proceedings of 5th International Conference on Case-Based Reasoning, ICCBR 2003. pp. 637–651. Springer (2003)
  18. Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P., Zhang, J.: Convolutional neural networks for human activity recognition using mobile sensors. In: Proceedings of 6th International Conference on Mobile Computing, Applications and Services. pp. 197–205 (2014)
  19. Zhang, S., Mccullagh, P., Callaghan, V.: An efficient feature selection method for activity classification. In: Proceedings of IEEE International Conference on Intelligent Environments. pp. 16–22 (2014)
  20. Zheng, Y., Wong, W.K., Guan, X., Trost, S.: Physical activity recognition from accelerometer data using a multi-scale ensemble method. In: IAAI (2013)