Transfer Learning in Body Sensor Networks Using Ensembles of Randomized Trees

Pierluigi Casale, *Member, IEEE*, Marco Altini, and Oliver Amft, *Member, IEEE*

***Abstract*—We investigate the process of transferring the activity recognition models within the nodes of a body sensor network (BSN). In particular, we propose a methodology that supports and makes the transferring possible. Based on a collaborative training strategy, classifier ensembles of randomized trees are used to create activity recognition models that can successfully be transferred within the nodes of the network. The methodology has been applied in scenarios where a node present in the network is replaced by a new node located in the same position (replacement scenario) and relocated to a previously unknown position (relocation scenario). Experimental results show that the transferred recognition models achieve high-recognition performance in the replacement scenario and good-recognition performance are achieved in the relocation scenario. Results have been validated with multiple K-folds cross-validations in order to test the performance of the methodology when different amount of data are shared between nodes.**

***Index Terms*—Activity recognition, body area networks, transfer learning.**

I. Introduction

THE recognition of physical activity using wearable sensors and body sensor networks (BSNs) is a fundamental task in smart-health applications. In particular, physical activity recognition is at the base of more advanced services such as energy expenditure estimation and healthy lifestyle recommenders. In the recognition process of physical activities, data-driven techniques based on machine learning algorithms are nowadays the most prominent choice due to their capability to deal with noisy data and uncertainty. Nevertheless, these algorithms represent a limitation when the position of the sensor changes from the position where the recognition models have been trained. In order to avoid this situation, several models can be generated, each one trained with data from different positions but still not all the body positions may be completely covered. This situation will be more prominent when the use of commercial wearable devices will become widespread. For example, let us consider a scenario where the user wants to upgrade to a new fitness tracker from a different manufacturer. The possibility to keep the same capability on the new device may represent an appealing feature. On the other side, it is becoming every day more common to wear multiple wearable devices. We have our smart-phone, we wear our smart-watch, and our shoes are able to track our steps. If those devices might openly communicate and exchange information, we could exploit their recognition models to create a new model able to deal with a new sensor placed in new positions. The capability of exploiting the knowledge of a classification model in a different domain is known as transfer learning [1]. In this work, we investigate the process of transferring the activity recognition models of the nodes of a BSN to a new node previously untrained. The transferring process is accomplished through the combination of a collaborative training strategy and the use of classifiers ensembles based on randomized trees. Using the collaborative training strategy, a limited amount of data shared between all the nodes is used in the combination to the data of the node for training an ensemble of classifiers. This ensemble, while still able to provide high-recognition performance, contains a degree of redundancy helpful during the transferring process. When algorithms based on randomized trees such as bagging (Bag) [2], random forest (RF) [3], and rotation forest (Rot) [4] are considered, the proposed strategy allows us to learn ensembles that can be transferred within the BSN and are able to recognize the activities sensed by nodes positioned at different locations. Since the performance of the recognition models depends on the training data, the amount of data shared between the nodes at training time is a quantity that needs to be taken into account to find a good tradeoff between the performance at the node and the performance of the transferred classifiers. We applied the methodology in situations where a node is replaced by a new node located in the same position (replacement scenario) and a node already present in the network is relocated to a previously unknown position (relocation scenario). Experimental results show that the recognition model of a node can successfully be transferred achieving high-recognition performance in the replacement scenario and good recognition performance in the relocation scenario. Multiple K-folds cross-validations have been used to test the performance of the methodology when different amount of data are shared between nodes. This work extends and complements our previous paper [5] and the following contributions are presented here.

1) We evaluate the transfer learning methodology on two additional publicly available datasets with multiple sensors nodes.
2) In addition to nearest centroid classifier (NCC) and k-nearest neighbors (kNNs), support vector machine (SVM) has been used as comparison method.

Manuscript received June 20, 2014; revised December 05, 2014; accepted December 29, 2014. Date of publication January 07, 2015; date of current version February 04, 2015. This work was supported by the EU FP7 Marie Curie Network iCareNet under Grant 264738.

P. Casale is with the Holst Centre-IMEC, Eindhoven 5656 AE, The Netherlands, and also with the Signal Processing Group, Technical University (TU) Eindhoven, Eindhoven 5656 AE, The Netherlands (e-mail: pierluigi.casale@imec-nl.nl).

M. Altini is with Bloom Technologies, Diepenbeek, Belgium, and also with the Signal Processing Group, Technical University (TU) Eindhoven, Eindhoven 5656 AE, The Netherlands.

O. Amft is with the Department of Electrical Engineering, Signal Processing Systems, University of Passau, Passau 94032, Germany.

Digital Object Identifier 10.1109/JIOT.2015.2389335

3) We consider a further step in the cross-validation for testing the proposed methodology when very small amounts of data are shared between nodes.

II. Transfer Learning and Its Application in BSN

Classifier ensembles have been already considered as a mechanism for transfer learning [6]–[9]. Of particular interest, the work of Kamishima *et al.* [10] applied a bagging approach to transfer the learning capabilities of a model through different domains. In their work, a high number of trees was learned on data from both source and target domains and a pruned version of the final ensemble was used to predict examples of the target domain. The pruning step was used to avoid the decreasing in performance due to negative transfer. Although they used training examples from both source and target domains, no experiments were conducted to quantify the amount of data needed to achieve a reasonable accuracy of the transferred classifier. Moreover, natural extensions of bagging such as RF or Rot were not considered. In activity recognition, Calatroni *et al.* [11] showed the feasibility of transfer learning in BSN and used a transferring approach to evaluate the classification performance of transferring to a new deployed sensor in the BSN. The transfer approach was based on a teacher–learner paradigm where all the pairs of nodes were considered as possible teachers and learners. The direct transferring of the classifiers in the nodes was compared to a *system-supervised* approach where the teacher provides labels to the learner and a new classifier was thus trained in the node. Good results were obtained using the system-supervised approach specially when the transferring is done between nodes located in similar body parts. Distance-based classifiers such as NCC, kNN, and SVM were used. Although we take into account the direct transfer of classifiers between the nodes, in our work, we do not consider the possibility to train a classifier in the node. In order to avoid this training, we use the collaborative training strategy and ensembles of randomized trees that provide transformations of the feature space that are potentially beneficial in the classification of activities using nodes placed at different positions. Blanke and Schiele [12] applied transfer learning methodologies for the composition of complex activities from a time ordered sequence of events. In their work, they transferred the activity events shared between similar composite activities minimizing the training effort for new events. Results show good performance in the recognition of composite activities, also when different application domains are considered. Although the methodology does not use a direct transferring approach, the work shows that transfer learning is feasible when a level of abstractions is considered in the learning mechanism. This consideration is also reported in van Kasteren *et al.* [13] where a level of abstraction is used to map features with the aim of transferring recognition models between different scenarios.

III. Transfer Learning Approach

Given a BSN constituted by the set of sensor nodes

$$\mathbf{S} = \{S_i\}, \qquad i = 1, \ldots, N \tag{1}$$

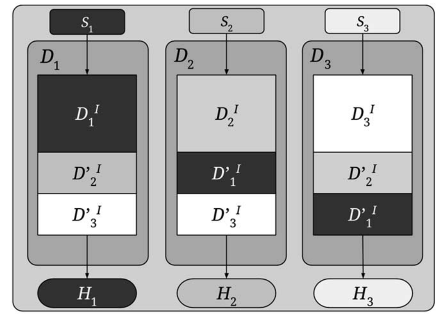


Fig. 1. Example of the collaborative training strategy for three sensors nodes. An amount of *junk data* from other nodes is added to the training set of the sensor to provide redundancy useful during the transferring of the recognition models.

the correspondent set of data is considered where each D_i^I represents the data of node S_i

$$\boldsymbol{D}_{\text{in}} = \left\{D_i^I\right\}, \qquad i = 1, \ldots, N. \tag{2}$$

For each S_i, an additional set of data $\boldsymbol{D}_i^O$ is defined as the composition of subsets of data coming from the other nodes of the network

$$\boldsymbol{D}_i^O = \left\{ {D'}_j^I \subset D_j^I \right\} \qquad \forall\, j = 1, \ldots, N; \quad j \neq i. \tag{3}$$

For each node S_i, an ensemble H_i is trained using the dataset D_i as defined in (4), constituted by the composition of D_i^I and $\boldsymbol{D}_i^O$. An example of this procedure of three nodes is shown in Fig. 1,

$$D_i = D_i^I \bigcup \boldsymbol{D}_i^O. \tag{4}$$

The classifier ensemble obtained in this step is used to classify activities on S_i during the normal operation of the node. An important consideration needs to be outlined from the previous definition. The datasets ${D'}_j^I$ defined in (3) are proper subsets of D_j^I, i.e., only a subset of data from the other nodes should be considered for training the classifier ensemble for the node S_i. Injecting this subset of *junk data* in the training set ensures that the recognition performance of the classifier ensemble H_i will be kept high since it will mostly learn from S_i. Nevertheless, the ensemble learning algorithm will learn some base classifiers that are barely able to classify data from other nodes. The performance of these classifiers will be boosted when they will be transferred to a new node and will create a new ensemble. These *junk data* injected in the training set provide a degree of redundancy useful during the transferring. The amount of *junk data* represents a quantity that will be evaluated in cross-validation (see Section IV). When a node S_i needs to be replaced, e.g., after a failure (replacement scenario), all the nodes transfer their classifiers H_j to the new node S_i^{rep}. The recognition model in S_i^{rep} will be created by the aggregation of all trees constituting the classifier ensembles H_j from the other nodes as shown,

$$H_i^{\text{rep}} = \bigcup_{j \neq i} H_j, \qquad j = 1, \ldots, N. \tag{5}$$

When a network node S_i is relocated to a different position (relocation scenario), a model assembled as in (5) can be used.

Nevertheless, as noticed in [10], the use of a smaller ensemble may generally avoid a decaying in performance due to negative transfer. In this case, the ensemble H_i^{rel} is generated by selecting from the full set of transferred classifiers, only the base classifiers h_j^t that perform better than a predefined accuracy α as defined as

$$H_i^{\mathrm{rel}} = \bigcup_{j=1,\ldots,N,t=1,\ldots,T} h_j^t \text{ s.t. accuracy} \left(h_j^t\right) > \alpha. \quad (6)$$

In this case, the corresponding labels should also be transferred in order to test the performance of each single base classifier.

IV. Evaluation Methodology

A. Evaluation Datasets

1) IMEC Dataset: The IMEC dataset contains data from a BSN of five wearable sensors with ECG and accelerometer from 17 participants. These data have previously been used for research on energy expenditure using data from a single node [14]. A wide range of sedentary, lifestyle, and sport activities have been considered and manually annotated on two levels of granularity. For the purpose of this work five macro-activities are considered, i.e., *household activities*, *resting*, *sitting*, *sport,* and *walking*. Nodes were positioned on the chest, the dominant ankle, the dominant thigh, the dominant wrist, and the waist at the right hip. Nodes were attached to the body using elastic bands and synchronized over a wireless network. Accelerometers were configured to acquire data at 64 Hz. A large set of features have been computed including statistics in the temporal domain and features from the spectral domain obtaining a 54-dimensional features vector for each node. The set of features includes mean, standard deviation, skewness and kurtosis for each acceleration axis and magnitude, correlations between each pairwise axis combination, entropy, signal spectral amplitude, power, and frequency.

2) Oulu Dataset: The Oulu dataset [15] contains data collected from 13 subjects from four sensor nodes placed at chest, right knee, left, and right wrist. Nodes were three axial accelerometers sampled at 10 Hz. A total of 17 activities are present in the dataset that have been aggregated into six macro activities, i.e., *walking*, *standing*, *running*, *cycling*, *manual activities,* and *low-level activity*. Precalculated mean and standard deviation on nonoverlapping windows of 0.7 s are used.

3) Daliac Dataset: The Daliac dataset [16] contains data collected from 19 subjects from four sensor nodes placed on right hip, chest, right wrist, and ankle. Each sensor node is constituted of a three axial accelerometer and gyroscope sampled at 204.8 Hz. For the aim on this work, only accelerometer data have been considered. A total of 13 activities are present in the dataset that have been aggregated into seven macro activities, i.e., *walking*, *standing*, *running*, *cycling*, *household activities*, *sitting,* and *lying*. Features on the temporal domain have been computed on the raw accelerometer data obtaining a 12-D features vector for each node. No frequency features have been computed for this dataset.

B. Analysis Methodology

1) Validation Strategy: Cross-validation [17] has been used for evaluation purposes. Datasets from each node have been split in K folds with approximately the same number of examples. In order to test the methodology with different amount of data shared, five K-folds cross-validation schemes have been considered, with $K = \{20, 10, 5, 3, 2\}$ corresponding to 5%, 10%, 20%, 33%, and 50% of data shared. For the replacement scenario, the training-set for sensor S_i has been constructed using K-1 folds from sensor S_i and all the Kth folds from the other nodes. Testing is executed on the Kth fold from sensor S_i and the $K-1$ folds for the remaining nodes. For the relocation scenario, a leave-one-sensor-out protocol has been used where at each iteration the training sets for the considered nodes have been generated as previously described and the testing is performed on the sensor left out. In order to provide statistical robustness, all the validation procedures have been performed twice using different splits of the datasets.

Classifiers Implementation and Parameters Selection: The methodology was applied with Bag, RF, and Rot. The number of trees was arbitrarily set to 31. Bagging subsamples were composed of 60% of the original training set. In RF and Rot, the number of features sampled was set as the root square of the total number of features. The methodology was compared with NCC, kNN, and SVM. The number of nearest neighbors in kNN was set to 11, as suggested in [11]. SVM uses a radial basis function kernel with parameters automatically selected using three-folds cross validation on the training set. The SVM multiclass extension was provided by ECOC-1vsAll [18].

Performance Metrics: Performance was measured using classification accuracy. In order to avoid interpersonal variability between the activities performed, experiments were conducted separately for each subject. Final results have been averaged over cross-validation executions and subjects.

V. Experimental Results

A. Replacement Scenario

Results reported in Figs. 2–4 show the classification accuracy obtained in the replacement scenario for the considered datasets using different subsets of data shared. Results show the nodes with the best and worst accuracy and the mean accuracy averaged over all nodes. White bars in the rightmost side of the pictures report the accuracy of the classifier H trained on D as defined in (4). Gray-scale bars report the results of the transferred ensemble H^{rep} when the collaborative training set is obtained sharing 5%, 10%, 20%, 33%, and 50% of data between nodes. As expected, accuracy of the transferred ensembles increases with the percentage of data shared. This fact is highlighted in Figs. 2(d)–4(d) where the accuracy is averaged over all nodes. SVM performs generally better than NCC and kNN but its performance is still lower than in the ensemble methods. It is worth to note how the performance of NCC varies between the Oulu and Daliac dataset, where different sets of features are considered. In particular, since no highly informative features but mean value and standard deviation are considered in the Oulu dataset, the performance of the NCC

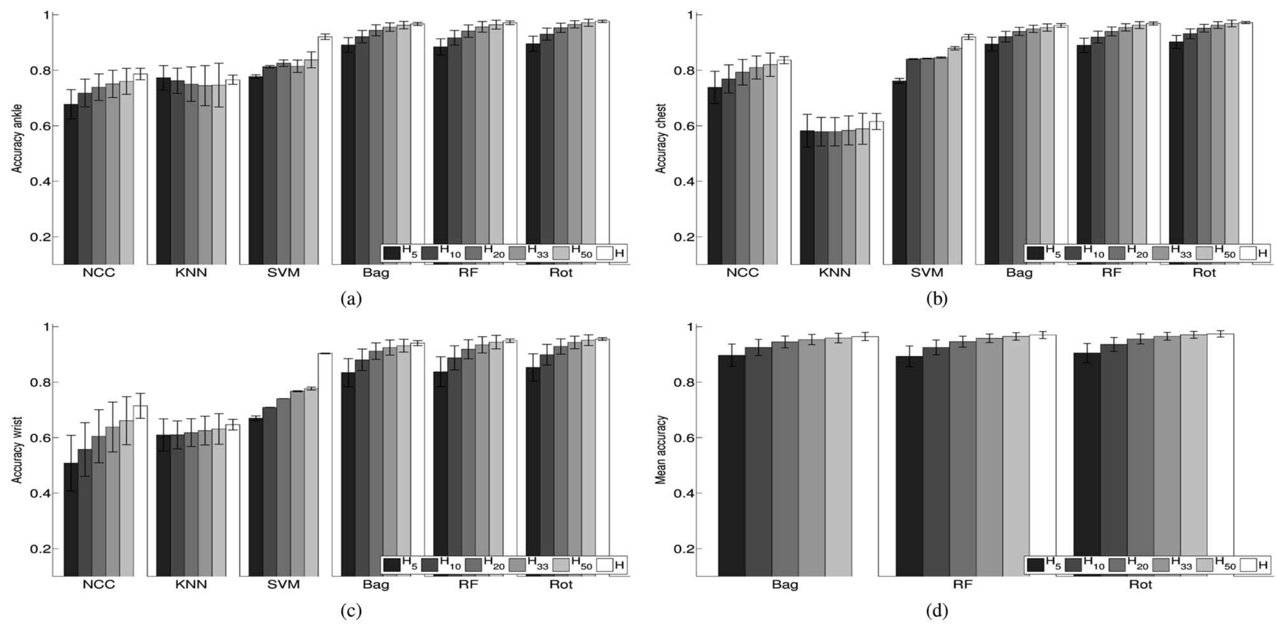


Fig. 2. Results on IMEC dataset in replacement scenario for nodes at (a) ankle; (b) chest; and (c) wrist. (d) Accuracy of Bag, RF, and Rot averaged over subjects and nodes. The performance of the classifier H trained on the node is reported in white bars in the rightmost side of the picture. Performance of the transferred ensembles is indicated with H_x, where x is the percentage of data shared at training time.

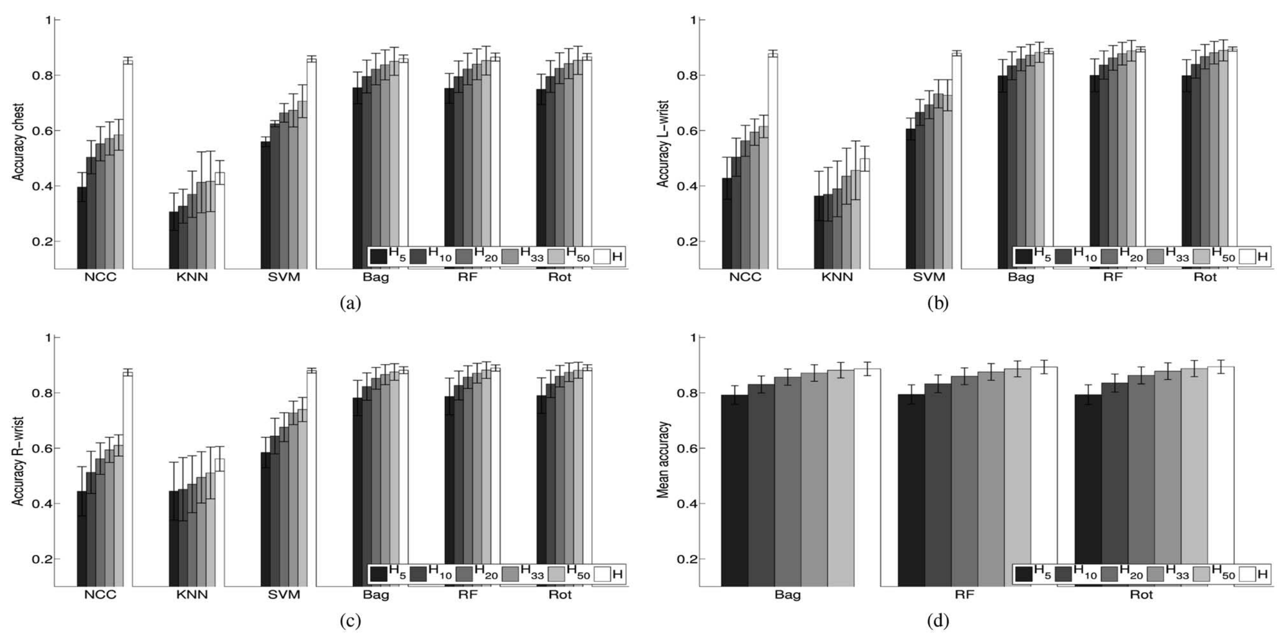


Fig. 3. Results on Oulu dataset in replacement scenario for nodes at (a) chest; (b) left wrist; and (c) right wrist. (d) Accuracy of Bag, RF, and Rot averaged over subjects and nodes. The performance of the classifier H trained on the node is reported in white bars in the rightmost side of the picture. Performance of the transferred ensembles is indicated with H_x, where x is the percentage of data shared at training time.

classifier results really low when data coming from different positions are considered. On the other hand, in the Daliac dataset the accuracy increases with the amount of data shared, in accordance with results obtained in the IMEC dataset. SVM shows an atypical behavior in the Daliac dataset where the classification performance obtained using data sharing is higher than the original performance. This behavior is due to the combination of the features and the short-term acquisitions present in the dataset that generates more support vectors and hence a more accurate recognition rate when more examples are added to the training set.

B. Relocation Scenario

Classification results for the relocation scenario are reported in Figs. 5–7 for the three datasets using different subsets of data shared. Results show the nodes with the best and worst accuracy and the mean accuracy averaged over all the nodes. Gray-scale bars report the results of the transferred ensemble H^{rep} when the collaborative training is obtained sharing 5%, 10%, 20%, 33%, and 50% of data between sensors. Although higher than NCC and kNN, the accuracy obtained by the transferred ensembles in this scenario does not provide a satisfactory

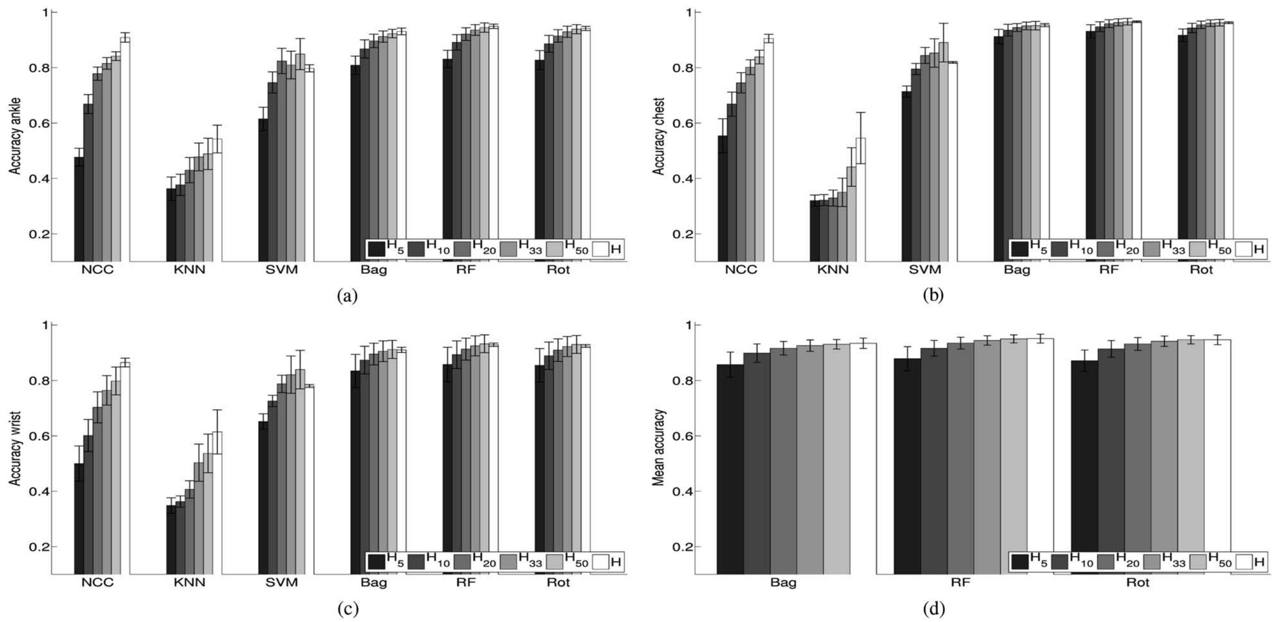


Fig. 4. Results on Daliac dataset in replacement scenario for nodes at (a) ankle; (b) chest; and (c) wrist. (d) Accuracy of Bag, RF, and Rot averaged over subjects and nodes. The performance of the classifier H trained on the node is reported in white bars in the rightmost side of the picture. Performance of the transferred ensembles is indicated with H_x, where x is the percentage of data shared at training time.

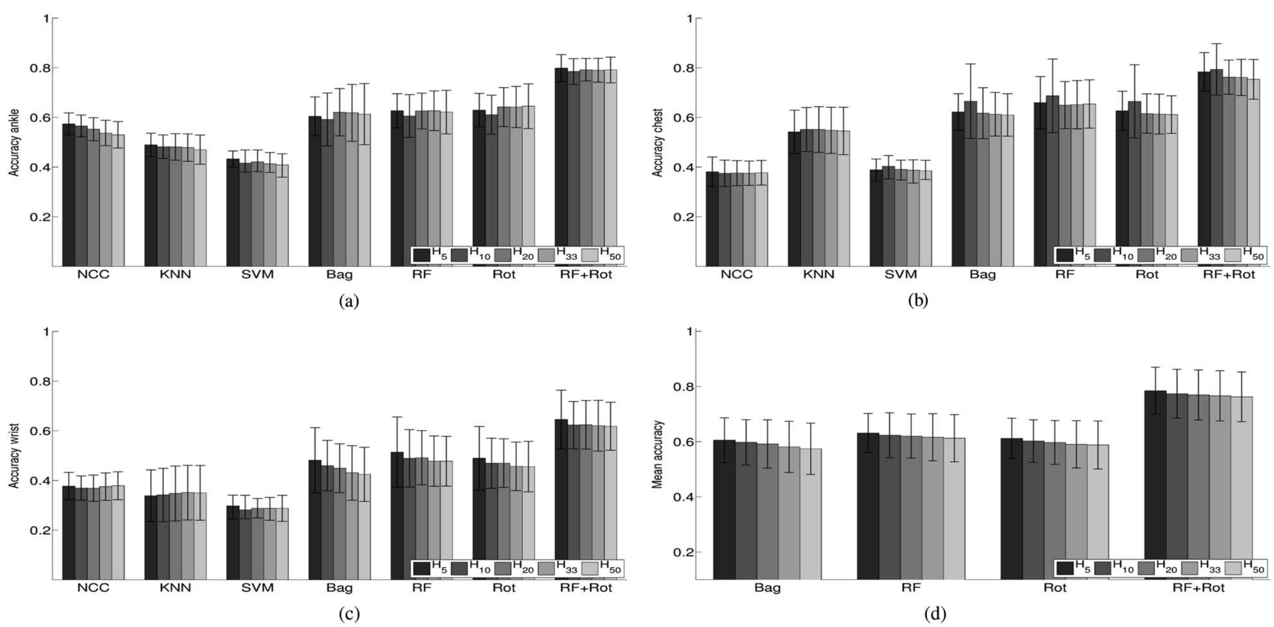


Fig. 5. Results on IMEC dataset in relocation scenario for nodes at (a) ankle; (b) chest; and (c) wrist. In (d), accuracy is averaged over all nodes. Performance of the transferred classifiers is indicated with H_x, where x is the percentage of data shared at training time.

result. In this scenario, a considerable improvement is provided when the final ensemble is obtained combining base classifiers from both RF and Rot. This heterogeneous ensemble achieves in many cases a recognition accuracy equal or higher than 80% in the IMEC and Daliac dataset and an improvement of 15% in the Oulu dataset. This fact is shown in Fig. 5(d)–7(d) where the accuracy of the ensembles averaged on all the nodes is reported. An unexpected effect is generally shown in these figures. Higher accuracy is achieved when a lower amount of data are shared between nodes. This effect is a consequence of negative transferring. The use of more data from different sources provides classifiers that are more accurate in the recognition process but they perform with lower generalization performance when transferred to a new position. Although more prevalent in the combined ensemble of RF and Rot, this situation is general for all the classifiers considered.

C. Ensemble Composition

The composition of the transferred ensembles can provide useful insights about which base classifiers help in the activity classification process after the transferring. Moreover, the

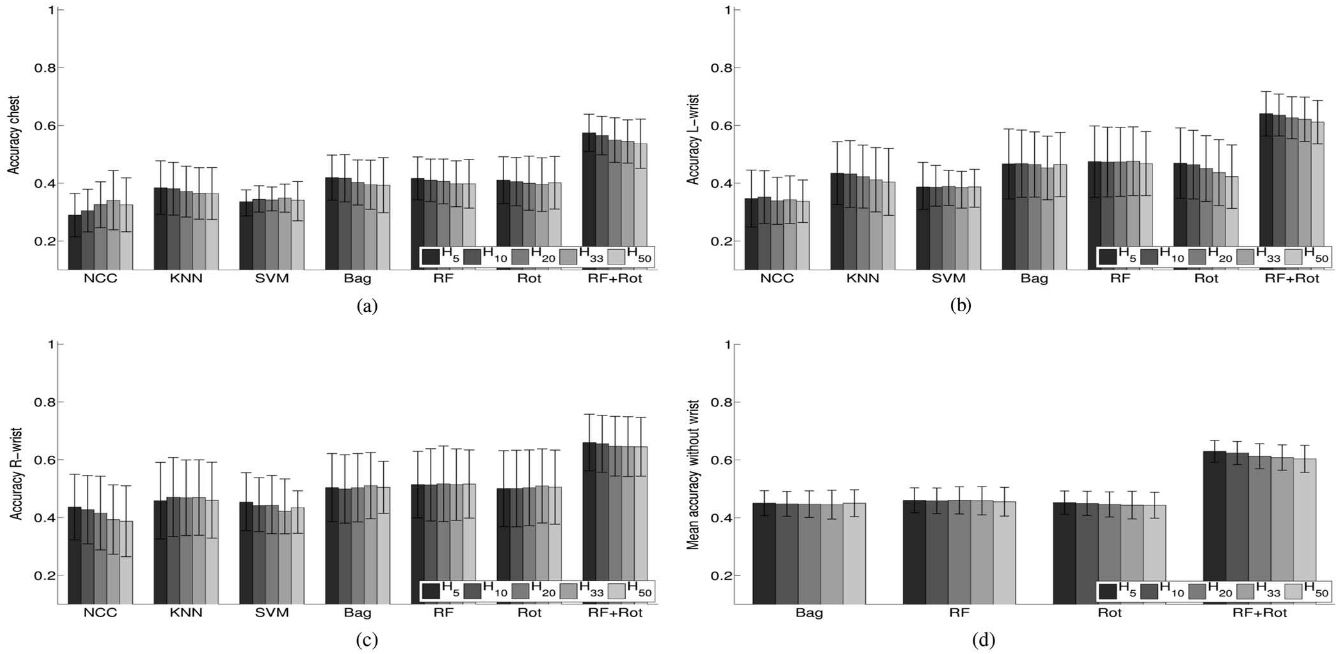


Fig. 6. Results on Oulu dataset in relocation scenario for nodes at (a) chest; (b) left wrist; and (c) right wrist. In (d), accuracy is averaged over all nodes. Performance of the transferred classifiers is indicated with H_x, where x is the percentage of data shared at training time.

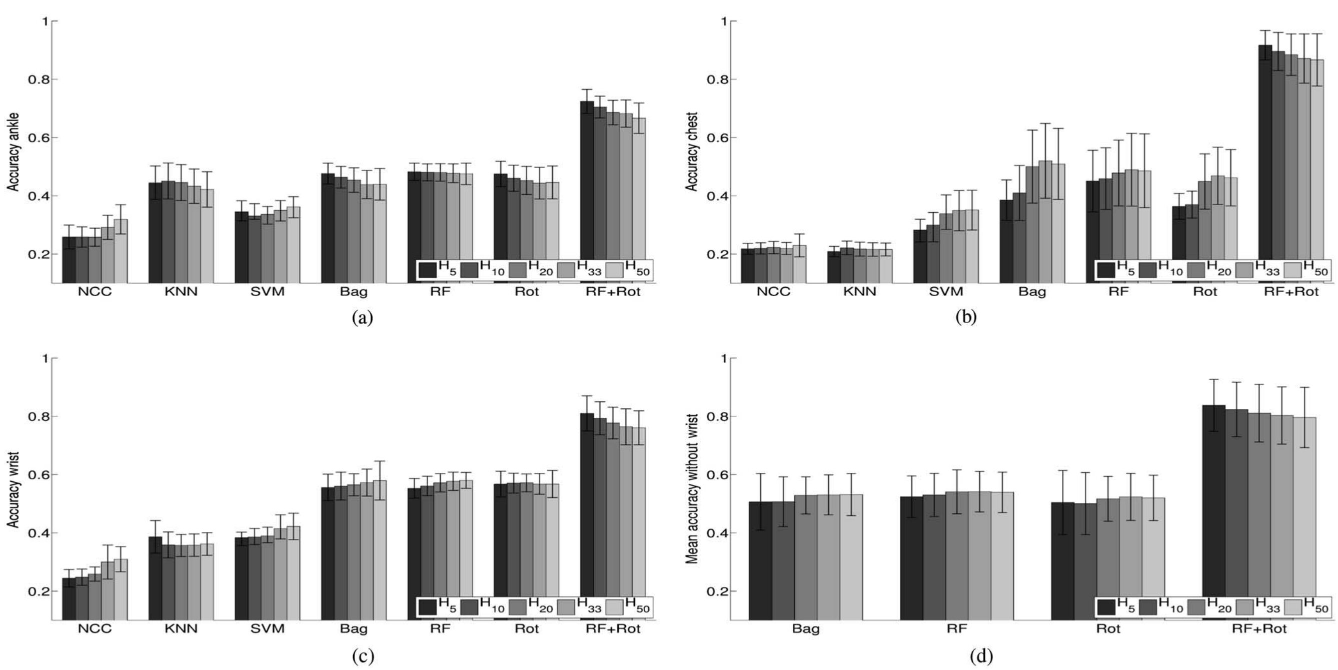


Fig. 7. Results on Daliac dataset in relocation scenario for nodes at (a) ankle; (b) chest; and (c) wrist. In (d), accuracy is averaged over all nodes. Performance of the transferred classifiers is indicated with H_x, where x is the percentage of data shared at training time.

composition of these ensembles can show if there exists a node that share some similarity with the node that receives the transferred ensemble. In Fig. 8, statistics about the composition of the transferred ensembles in the tight node for the IMEC dataset are shown for (a) replacement scenario and (b) relocation scenario. The base classifiers have been separated by nodes. Boxplots are computed over all the participants. For both scenarios, the classifier ensemble at thigh is constituted in average by base classifiers equally selected from all nodes. No node seems to mainly contribute with a predominant number of classifiers. This result is expected as a consequence of the collaborative training strategy. In each ensemble, there exist some base classifiers that learn from the combination of data coming from different nodes. These base classifiers are used to create the final ensemble and they do not depend on the particular position of the node. From the boxplot, an interesting situation also emerges. The average number of base classifiers in replacement is higher than the average

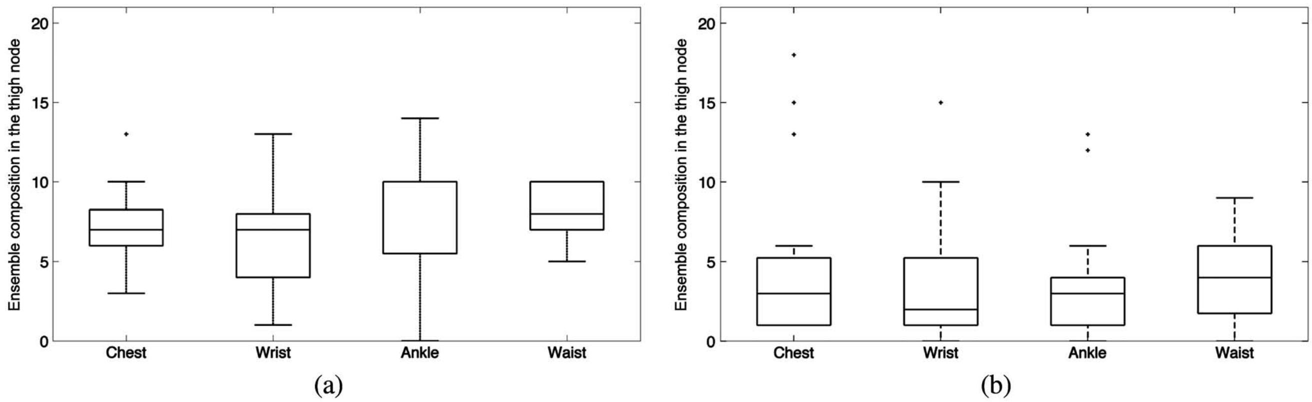


Fig. 8. Example composition of the ensemble at Thigh node in IMEC dataset for (a) replacement and (b) relocation scenario, computed over data from all participants. In average, the transferred ensemble is constituted by base classifiers selected from all the nodes of the network.

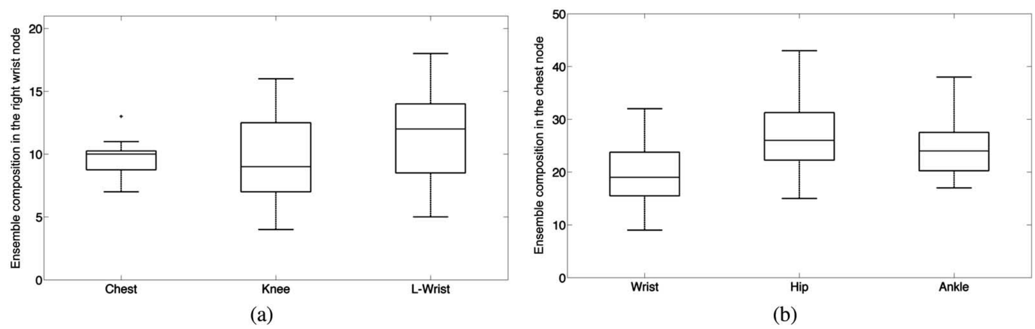


Fig. 9. Example composition of the ensemble at right wrist in Oulu Dataset (a) and chest in Daliac dataset (b) computed over data from all participants. In average, the transferred ensemble is constituted by base classifiers selected from all the nodes of the network.

number of base classifiers used in relocation. This behavior is a consequence of the selection process defined in (6) since only the best performing classifiers are retained for creating the final ensemble in the node. The same statistics are shown for the ensembles at Right Wrist in the Oulu dataset and chest in the Daliac dataset, reported in Fig. 9(a) and (b), respectively. Also in this case, the final ensemble is in average equally generated from base classifiers from all nodes. A slightly higher influence is given by base classifiers from the left wrist (in Oulu) and from the Hip (in Daliac) that obviously have similarity with the considered node.

VI. Discussions

Although the collaborative training has been used in NCC, kNN, and SVM, these classifiers do not provide the same level of performance achieved by the transferred ensemble methods. Experimental results have shown that a successful transfer is provided by the combination of sharing training data and using ensembles of randomized trees. The transformations performed on the training set by randomization and rotations provide an abstraction level that might model a set of activities independently from the particular location. This effect is predominant in the relocation scenario where results have shown that the combination of using subsamples of data and features (RF) and the use of rotated versions of these subsamples (Rot) can generate a new ensemble that provides recognition capabilities for nodes placed in a completely new position. As outlined in Section III, in the relocation scenario, labels should be transferred in order to check the performance of the transferred classifiers on the new node. Although this process could look similar to the *system supervised* approach used in [11], there exists a substantial difference. In our approach, labels are transferred only for testing the base classifiers. No training is performed in the node. Since the base classifiers are constituted by decision trees, just a sequence of *if-then* statements need to be executed in the node. Although not required, the transferring of labels can also be beneficial in the replacement scenario. In this scenario, the final classifier is aggregated using all the base classifiers transferred from all nodes generating a very large ensemble. Hence, the ensemble can be pruned in both scenarios retaining only the best performing base classifiers. Results obtained using pruned versions of the transferred ensemble in the replacement scenario show that there is no loss in the recognition performance in the node. Results in replacement scenario show that with 33% and 50% of data shared the performance of the transferred classifiers are equals or slightly lower than the performance of the classifier trained on the node. In this cases, the total amount of *junk data* injected in the training set is 67% and 80%, respectively. In particular, when 50% of data are shared between nodes, the final training set is equally composed from data of all the nodes. This fact brings an interesting conclusion in terms of using a single node of the BSN. The collaborative learning strategy can be used to train a single classifier able to provide good classification performance over multiple positions with no significant differences in the recognition of the activity performed. With a lesser extent, this conclusion is also valid for the comparison methods where, in some cases, the classification results of the transferred classifier is better of than the original classifier.

VII. Conclusion

In this work, the process of transferring the activity recognition models within the nodes of a BSN has been investigated and a methodology that supports and makes the transferring possible has been proposed. The methodology is based on a collaborative training strategy and the use of ensembles of randomized trees. During collaborative training, nodes share a subset of their data with other nodes of the network and activity classification models based on ensembles of randomized trees are trained on these datasets. The amount of *junk data* injected in the training set for each nodes is a quantity that influences the recognition performance of the activity classifier. The transformations on the training set that the randomized trees provide are beneficial to the creation of activity recognition models able to classify activity from different positions. The methodology has been evaluated on three datasets with multiple body sensor nodes. In two scenarios, we analyzed the situation: 1) when a node is replaced by a new node in the same position and 2) a node is relocated to a new unknown position. Several K-folds cross-validation protocols have been used to evaluate the methodology with different amounts of *junk data* for each node. The amount of *junk data* has been varied between a minimum of 17% up to a maximum of 80%. Experimental results have shown that the transferred recognition models can achieve a classification accuracy up to 98% in the replacement scenario and 80% in the relocation scenario. Results have also shown that, even with a low amount of junk data (17%), the performance of the transferred classifiers is not significantly influenced. Results also opens interesting possibilities for the use of a single node into several body position maintaining intact the level of recognition performance of the classifier.

References

[1] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[2] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[3] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[4] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.

[5] P. Casale, M. Altini, and O. Amft, "Transfer learning in body sensor networks using ensemble of randomised trees," in *Proc. Int. Conf. Wearable Implantable Body Sensor Netw. (BSN'14)*, 2014, pp. 1–8.

[6] S. Mei, "Probability weighted ensemble transfer learning for predicting interactions between HIV-1 and human proteins." *PLoS One*, vol. 8, no. 11, p. e79606, 2013.

[7] A. Acharya, E. R. Hruschka, J. Ghosh, and S. Acharyya, "Transfer learning with cluster ensembles," in *ICML Unsupervis. Transfer Learn.*, vol. 27, 2012, pp. 123–132.

[8] L. Jiang, K. Zhang, and G. Allen, "Transferred correlation learning: An incremental scheme for neural network ensembles." in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2010, pp. 1–8.

[9] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML'07)*, 2007, pp. 193–200.

[10] T. Kamishima, M. Hamasaki, and S. Akaho, "Trbagg: A simple transfer learning method and its application to personalization in collaborative tagging," in *Proc. 9th Int. Conf. Data Mining (ICDM)*, 2009, pp. 219–228.

[11] A. Calatroni, D. Roggen, and G. Tröster, "Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion," in *Proc. 8th Int. Conf. Netw. Sens. Syst. (INSS'11)*, Jun. 2011.

[12] U. Blanke and B. Schiele, "Remember and transfer what you have learned—Recognizing composite activities based on activity spotting," in *Proc. Int. Symp. Wearable Comput. (ISWC'10)*, 2010, pp. 1–8.

[13] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Transferring knowledge of activity recognition across sensor networks," in *Proc. 8th Int. Conf. Pervasive Comput. (Pervasive'10)*, 2010, pp. 283–300.

[14] M. Altini, J. Penders, R. Vullers, and O. Amft, "Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 219–226, Jan. 2015.

[15] S. Pirttikangas, K. Fujinami, and T. Nakajima, "Feature selection and activity recognition from wearable sensors," in *Proc. Ubiquit. Comput. Syst.*, 2006, vol. 4239, pp. 516–527.

[16] H. Leutheuser, D. Schuldhaus, and B. Eskofier, "Hierarchical, multi-sensor based classification of daily life activities: Comparison with state-of-the-art algorithms using a benchmark dataset," *PLoS One*, vol. 8, no. 10, p. e75196, 2013.

[17] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surveys*, vol. 4, no. 11, pp. 40–79, 2010.

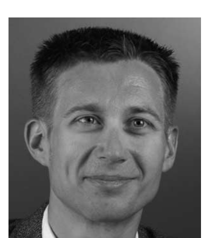
[18] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Sep. 2001.

Pierluigi Casale (M'11) received the B.Sc. and M.Sc. degrees in electronic engineering from the University of Rome "La Sapienza," Rome, Italy, in 2007, and the Ph.D. degree in applied mathematics from the University of Barcelona, Barcelona, Spain, in 2011.

He is a Researcher with the Holst Centre-IMEC, Eindhoven, The Netherlands, and former Postdoctoral Researcher with the ACTLab Research Group, Technical University (TU) Eindhoven, Eindhoven, The Netherlands. His research interests include machine learning and pattern recognition methods for wearable sensors and body sensor networks with applications, among others, in digital and mobile health, distributed learning methodologies, and big-data analysis for sensor networks with applications in human behavior analysis.

Marco Altini received the M.Sc. degree in engineering and computer science from the University of Bologna, Bologna, Italy, in 2010, and is currently working toward the Ph.D. degree at the Technical University (TU) Eindhoven, Eindhoven, The Netherlands.

His research interests include machine learning techniques for the development and implementation of activity recognition, energy expenditure and cardiorespiratory fitness estimation algorithms based on acceleration, and physiological signals using wearable sensors.



Oliver Amft (GSM'07–M'08) received the Dipl.-Ing. (M.Sc.) degree from Chemnitz Technical University, Chemnitz, Germany, in 1999, and the Dr. sc. ETH (Ph.D.) degree from ETH Zurich, Zurich, Switzerland, in 2008, both in electrical engineering and information technology.

He is a Full Professor heading the Chair of Sensor Technology and the ACTLab research group, University of Passau, Passau, Germany. In 2009, he initiated the ACTLab research group, TU Eindhoven. He has coauthored more than 100 publications. His research is focused on multi-modal activity pattern recognition, human behavior inference algorithms, and ubiquitous sensing systems, with applications (among others) in healthcare, sports, and building automation.