STAR: A Scalable Self-taught Learning Framework for Older Adults' Activity Recognition

Sreenivasan Ramasamy Ramamurthy†, Indrajeet Ghosh†, Aryya Gangopadhyay†,
Elizabeth Galik‡, Nirmalya Roy†
†*Department of Information Systems, University of Maryland, Baltimore County, Baltimore, United States*
‡*Department of Nursing, University of Maryland, Baltimore, Baltimore, United States*
rsreeni1@umbc.edu, indrajeetghosh@umbc.edu, gangopad@umbc.edu, galik@umaryland.edu, nroy@umbc.edu

***Abstract*—Activity Recognition (AR) in older adults living with Neurocognitive disorders caused by diseases such as Alzheimer's is still a challenging research problem. The inherent natural variation in performing an activity increases while repeating the same activity for an older adult, let alone the variation introduced when another older adult performs the same activity. Moreover, the challenges in acquiring the labeled data while preserving the privacy, availability of annotators with domain knowledge, aversion towards cameras even for a minimal amount of time for ground truth data collection, and psychological and mental health status make AR for older adults challenging. In this paper, we postulate a self-taught learning-based approach that helps recognize activities with variations that are not being directly seen during the training phase. We hypothesize that the features extracted using deep architectures from unlabeled data instances can learn general underlying representations of activities efficiently and help improve activity classification in a supervised setting, although the data instances in labeled data do not follow the generative distribution of that of unlabeled data. We posit real data from a retirement community center using our in-house SenseBox infrastructure and survey-based assessments concurrently done by a clinical evaluator to study the relationship between activities and functional/behavioral health of older adults. We evaluate our proposed self-taught learning-based approach, *STAR*, using the presented in-house Alzheimer's Activity Recognition (AAR) dataset acquired in a real-world deployment in 25 homes which outperforms the state-of-the-art algorithm by about 20%.**

***Index Terms*—Unsupervised Learning, Self-taught Learning, Pre-training, Human Activity Recognition**

I. Introduction

Recent innovative designs and developments of Internet of Things (IoT) devices have caused an increase in the research and commercialization of technologies that use wearable and ambient sensor modalities. Such consumer-grade IoT devices have helped develop various state-of-the-art methods and frameworks in the domain of human health monitoring relying on activities of daily living (ADLs) and instrumental activities of daily living (IADLs) for early detection of physical mobility and cognitive health impairments. Monitoring health conditions and diagnosing the precursor to a specific functional and/or behavioral health symptom are usually performed by conducting various clinical and survey-based assessments, which involve multiple trips to the hospital and always turn out to be a hassle, especially for older adults. On the other hand, an automated system that can help with the early detection of anomalies or abnormalities in an older adult's habitual activities or behavior could alert the caregiver to take an appropriate action just in time. Such automated systems could potentially improve the quality of life for older adults while they live independently in their preferred living environment. A variety of commercial wearable devices[1,2] recently has flooded the market for healthcare applications. Ambient sensors such as passive infrared sensors [1]–[3], magnetic sensors, radars [4], acoustic sensors [5], [6], everyday smart objects, and device-free sensing (WiFi [7]) have helped researchers to build novel human activity recognition models.

Although existing smart-home sensor systems [8], [9] for activity recognition can help detect early symptoms of diseases, they are often compromised by significant practical challenges. Our overarching objective in this work is to identify such challenges through our real-world deployment in 25 community-dwelling apartments in a retirement community center and postulate appropriate solutions to mitigate underlying system and data-related problems. Below, we articulate the practical challenges that exist with smart home sensor systems to scale and adapt the older adults' activity recognition (AR) models. *(i)* The reliability of AR models depends on the availability of large-scale annotated ground truth or labeled data. In this work where the older adults living with neurodegenerative disorders, it is challenging to monitor and recruit a vast number of human subjects to collect labeled data and train our proposed *STAR* model, *(ii)* Obtaining labels for the data is a humungous task where the dependence on the caretakers to provide the labels becomes a necessity. However, the labels provided by the caretaker can be irrelevant because of the lack of domain knowledge which eventually reduces the robustness of the machine learning algorithm. *(iii)* Camera systems could be leveraged to obtain the ground truths. However, cameras pose privacy concerns, especially when individuals live with neurodegenerative disorders. On the other hand, older adults generally have an aversion towards having a camera around them for a prolonged period and feel that as an invasion of their privacy, which makes acquiring ground truth a herculean

We acknowledge NSF CAREER Award #1750936 and Alzheimer's Association, Grant/Award #AARG-17-533039.

[1]https://www.fitbit.com/home
[2]https://www.empatica.com/en-eu/research/e4/

2693-8340/21/$31.00 ©2021 IEEE
DOI 10.1109/SMARTCOMP52413.2021.00037

task. Motivated by this, we propose *STAR*, an unsupervised self-taught activity recognition model that could leverage the benefits of both minimal labeled data and abundant unlabeled data and learn the generic feature representations from limited labeled data to work effectively in the presence of unlabeled older adults' data.

Several machine learning techniques have been leveraged in order to overcome the problem of limited labeled data instances. Most algorithms fall into the broad classification of *Active Learning*, *Transfer Learning*, *Semi-Supervised* techniques. Active learning helps acquire the labels from an oracle by sampling the most informative data instances [10]. The major challenge associated with active learning techniques is high computation time, as the algorithms require online training. Another challenge with active learning involves the quality of the annotation acquired through crowd-sourcing platforms. The second category of algorithms is transfer learning, which involves transferring meaningful information from a source domain to a target domain. Due to the difficulty of acquiring data with older adults, transfer learning is one of the most appropriate techniques. Finally, some Semi-Supervised techniques have been found useful where partial label information from the ground truth is leveraged in fine-tuning the algorithm's parameters. Most studies involving transfer learning and semi-supervised techniques have been tested on a smaller set of activities, usually confining to ADLs.

Another major challenge in the field of activity recognition for older adults has been scaling AR that can help adapt a developed sensing system and developed model to a new scenario. Unfortunately, AR data is generally susceptible to high inter-class and intra-class variability [10]. Older adults, in particular, exhibit such variations more frequently when compared to young adults. Such increased variations negatively impact the performance of AR models for older adults. Therefore, designing an AR model for a general population, and adapting it for older adults may cause the AR model to fail. Due to the additional variations introduced by older adults due to the evolving physical and cognitive health changes due to ageing, developing a generalized and scalable AR model for older adults is non-trivial. The major contributions of the paper are enumerated below.

- **Self-Taught Learning Framework**: We identified the most challenging problem of the availability of abundant unlabeled data and minimal labeled data. To tackle this problem, we propose a self-taught learning-based methodology to address the inherent activity variations across the same and different older adults with Alzheimer's. This methodology leverages a pre-training phase to learn the representations from unlabeled data and represent the labeled data in the new representation space to utilize the information from the activities not listed in the labeled space.
- **Data Collection from Multiple Smart Homes**: We presented a data collection system for activity recognition in a smart home setting deployed and used in this study. We described the data collection procedure and the dataset (namely *Alzheimer's Activity Recognition (AAR) dataset*, which comprises sensor data and survey-based data that can help assess an older adult's functional and behavioral health.
- **Evaluation in practical setting**: We demonstrated that the proposed framework, *STAR* helps reduce the misclassifications related to the system-level faults and data collection errors. Besides, we showcased that a simpler CNN network is sufficient to boost the performance of the downstream task using *STAR*. Finally, we evaluated *STAR* using the in-house Alzheimer's Activity Recognition (AAR) dataset collected from 25 apartments in a retirement community center with IRB approval and contrasted the performance to a publicly available dataset that comprises a different age-group population.

II. Related Work

This section will discuss the related work regarding techniques that improves classification performance with a minimal amount labeled dataset and unlabeled dataset. Active learning, a popular method that falls under the former category, determines the most uncertain data instances and queries the label from an annotator. Such methods drastically reduce the labeling efforts, time and help improve the classification performance in a supervised setting. Another popular method, Transfer learning, lets us transfer knowledge from a source domain to a target domain and thus minimize the training effort. The target domain could be a different person, scenario, surroundings, and so on. Transfer learning can be broadly classified as Inductive, unsupervised, and transductive based on the similarity between the source and target domains. *STAR* falls under the category of the inductive transfer. Several algorithms have been proposed for activity recognition using transfer learning to boost the classification performance in the target domain [1], [11], [12]. In a study, the authors have proposed a framework that leverages autoencoder features from unlabeled data and transfer the first two layers to the target domain to recognize unseen activities in the target domain [13]. In such works, the assumption remains that the source and target domain label space follows the same distribution in contrast to our approach, *STAR*, which assumes otherwise.

Our proposed framework, *STAR*, falls under the category of Self-taught learning, which is a special case of inductive transfer learning, where the model learns from the unlabeled data. Besides, the primary assumption is that the labeled data and the unlabeled data follow different generative conditional distributions (a different subset of activities are present in both labeled and unlabeled data domains). Such an assumption has proven to learn good representations of the input data from the unlabeled data and help aid the classification task in a supervised setting [14]. Self taught learning has shown to be effective in various fields such as audio classification [15], [16], E-Nose in Wound Infection Detection [17], facial beauty prediction [18], image classification [14]. In a study, the authors created codebooks of features called basis vectors

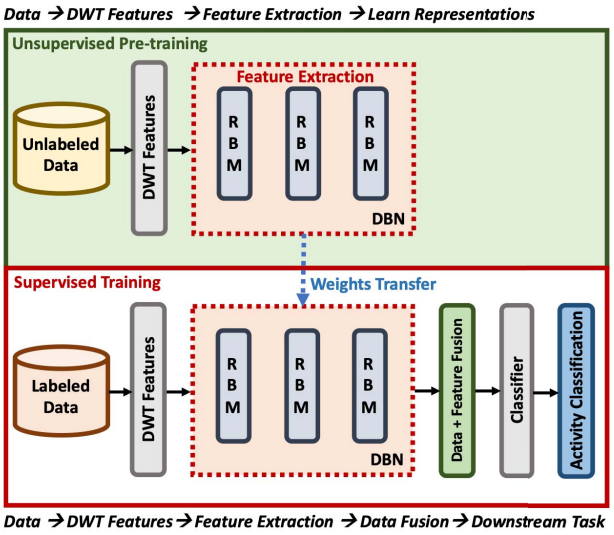

Fig. 1. Overall Architecture to illustrate the pre-training phase, the weight transfer and how it aids the classification in supervised setting

corresponding to each activity using sparse encoding [19] from the unlabeled data. A distance-based methodology was employed to determine the most appropriate basis vector, and the basis vector-label pair were classified in a supervised setting. The computational complexity involved in determining the activity-specific basis vectors and removing redundant basis vectors were the limitations of this methodology.

Another paradigm of machine learning that leverages unlabeled data to aid supervised classification is self-supervised learning. Self-supervised learning uses representations learned from unlabeled data through unsupervised transformations and further uses transfer learning to assist the downstream task in a supervised setting similar to our approach. However, *STAR* does not leverage any label information during the representation learning phase. Additionally, we believe that the augmentations such as jittering, cropping, rotation [20] of the data do not represent the true variation in the activity recognition dataset. We believe that the feature engineering step of using wavelet coefficients would expose the variations in the time-frequency domain to the pre-training and classification phases. In AR, authors of [20] proposed a multi-task learning framework to learn auxiliary tasks (augmentation of the unlabeled data) during the pre-training phase and used the primary task to perform classification, which makes it a semi-supervised approach.

III. Self Taught Learning

This section describes the proposed methodology that comprises an unsupervised feature learning phase, a feature fusion component, and the supervised classification phase. A detailed pictorial representation of the pipeline is described in Figure 1.

A. Unsupervised Pre-training

We propose to leverage the unlabeled data to learn the parameters using generative stochastic neural networks in the pre-training phase. The notion for using the unlabeled data is because the unlabeled data comprises activities outside of the labels included in the labeled dataset. In this study, we propose to use Deep Belief Networks (DBN) to learn high-level representations (from unlabeled data) and further use the learned parameters (from unlabeled data) to generate high-level representations for labeled data for the classification task. Let us now describe the mathematical formulation of the pre-training phase using DBN. A DBN is a probabilistic generative network that is formed by stacking Restricted Boltzmann Machines (RBM). RBMs are probabilistic neural networks that comprise of neurons in two groups (visible and hidden) and form a bipartite graph between them, with a restriction that the neurons are not connected to each other in the same group. Let us denote the visible layers as v_i and hidden layers as h_j, where i and j represent the neurons in the visible and hidden layers, respectively. The weight update required during the training phase of the RBM is performed using gradient descent and is given by equation 1.

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial P(v)}{\partial w_{ij}} \tag{1}$$

In eq 1, η represents the learning rate, w_{ij} is the weight vector connecting the visible layer and the hidden layer, and $P(v)$ is the probability distribution over the visible vectors defined using an energy function is shown in equation 2:

$$P(v) = \frac{1}{Z} \sum_{h} exp(-E(\mathbf{v},\mathbf{h})) \tag{2}$$

Here, $E(\mathbf{v},\mathbf{h})$ is the energy function. A lower value of energy function is desirable and is thus minimized during training by adjusting the weights and biases. The derivative of the log probability in equation 1 with respect to the weights are defined as

$$\frac{\partial P(v)}{\partial w_{ij}} = < v_i h_j >_{data} - < v_i h_j >_{model} \tag{3}$$

where $<>$ denotes the expectations under the distribution by the subscript. Since there is no connection among the hidden layers and among visible layers, it is possible to obtain $< v_i h_j >_{data}$ for the visible layer given the hidden layer (using eq. 5) and hidden layer given visible layer (using eq. 4).

$$p(h_j = 1|v) = \sigma(b_j + \sum_i v_i w_{ij}) \tag{4}$$

$$p(v_i = 1|h) = \sigma(a_i + \sum_j h_j w_{ij}) \tag{5}$$

However, obtaining $< v_i h_j >_{model}$ is cumbersome as it requires alternating Gibbs Sampling. Hence, we use [21] to train RBM using Contrastive Divergence (as explained in algorithm 1). Besides, training DBNs is a greedy process; each

Algorithm 1 : Pseudocode of Contrastive Divergence, $CD()$

Input: unlabeled data (X_u)
Output: RBM trained weights (W_{RBM})
Initialization : visible units, $v \leftarrow X_u$
for epoch = 1 to E, total epochs **do**
  for k = 1 to N, total number of instances in X_u **do**
    update hidden units, $p(h_j = 1|v)$
    reconstruction step: $p(v_i = 1|h)$
    update hidden units, $p(h_j = 1|v)$
    update weight, $w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial P(v)}{\partial w_{ij}}$
  end for
end for
return W_{RBM}

Algorithm 2 : Pseudocode of extracting self-taught features

Input: unlabeled data (X_u), labeled data (X_l), $CD(X_u)$
Output: self-taught data, X_{self}
for each RBM in DBN **do**
  $W_{RBM} \leftarrow$ training_cd(X_u)
  $W_{DBN} \leftarrow$ stack (W_{RBM})
end for
for m = 1 to M, total number of instances in X_l **do**
  $R_t \leftarrow$ Compute forward pass, $f_{DBN}(X_l^m)$
  $temp \leftarrow$ append($X_l^m, R(t)$)
end for
$X_{self} \leftarrow temp$
return X_{self}

RBMs are individually trained before moving on to the next RBM. The activations of the first RBM (after training) are fed as an input to the second RBM, and so on (as explained in algorithm 2).

B. Supervised Classification

Before performing the classification task, it is essential to obtain the representations of labeled data through the unlabeled data. We use the same feature extraction methodology (using DBN) proposed in the unsupervised pre-training step initialized with the parameters learned using the unlabeled data to achieve these representations. This step allows us to represent the labeled data in the new representation feature space (from unlabeled data). Let us assume to have m training data instances denoted by X_l with class labels y_l. The unlabeled data be denoted as X_u. Besides, let us denote the DBN feature extraction function as $f_{DBN}(.)$, so, the representation of the labeled data in the new representation space, R_l is defined as $R_l = f_{DBN}(X_l)$. Now, the new data instance for the supervised classification task becomes $([X_l, R_l], y_l)$.

IV. Alzheimer's Activity Recognition (AAR) Dataset

This section presents the Alzheimer's Activity Recognition (AAR) dataset created to study the relationship between activities and behavioral health, especially Dementia. AAR dataset was acquired from a population of 25 older adults living in a retirement living facility and possessing symptoms consistent with Dementia from a mix of individuals who are healthy, mild cognitive impairment, and cognitive impairment. The dataset has two components, sensor-based and survey-based data. The aim of capturing the sensor-based dataset was to record the movement patterns (activities performed on a daily basis: ADLs and IADLs; Table I). In contrast, the survey-based questionnaire aimed to acquire the current state (clinical evaluation) of the individual's functional and behavioral health. Below, we describe the data collection procedure, system architecture, and details of both the sensor and survey-based datasets.

A. Sensor System Architecture

To collect the AAR dataset (especially the sensor-based data), we developed a raspberry-pi based system to integrate various sensors (extended from SenseBox [8]). Raspberry-pi is a Linux-based miniaturized computer that consists of Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit processor, which clocks at 1.4GHz with 1GB LPDDR2 SDRAM onboard. The role of the raspberry-pi in a smart-home is to act as a hub, connect to an existing network, provide internet connectivity to various sensors in the smart-home, and finally log the data. The notion of such a system was to develop a system that can be readily deployable in a new home. Further, the system is comprised of two categories of sensors: wearable and ambient. Empatica-E4 was used as the wearable sensor to record the movement (through accelerometry; 32HZ sampling frequency), Electro-Dermal Activity (EDA), skin temperature, and heart-rate variability. Besides, ambient sensors such as passive infrared sensors (PIR; in each room), reed switches (on each door), and object tags (on objects used on a daily basis) were connected to the hub. The hub was in continuous connection with a server (present in the author's laboratory) via a reverse-ssh tunnel for constant monitoring of the state of the hub and the connected sensors. Finally, we integrated cameras with the hub (for limited time usage) to acquire the ground truth and the cameras were place in each room. We further elaborate the dataset collection in the following sections.

B. Sensor-based Data Collection

As soon as the hub and the ambient sensors are placed in strategic positions, we request the participating individuals to wear the wearable device on their dominant hand (preferably, although not compulsory). Now, we divide our data collection duration into two phases: scripted activities and unscripted activities. For the first two hours, the participants are requested to perform scripted activities that involve both ADLs and IADLs with a camera online for ground truth collection. For the following 2 hours, the participants are monitored with a camera for any activity (unscripted); however, none of the team members of data collection will be present in the house. Further, we leave the ambient sensors and the wearable device to operate in the house (no cameras) to collect unlabeled data for the next 20 hours. This procedure was repeated for 25 different participants in their own homes, where the home's layout was different from each other.

TABLE I
LIST OF ACTIVITIES IN AAR DATASET

Categories	Activities
Activities of Daily Living(ADLs)	Sitting **(Sit)**, Standing **(St)**, Walking **(Wa)**
Instrumental Activities of Daily Living (IADLs)	Brushing Hair **(BH)**, Folding Laundry **(FL)**, Phone **(Ph)**, Preparing Sandwich **(PS)**, Sweeping **(Sw)**, Taking Trash Out **(TT)**, Using Toothbrush **(UT)**, Wash Hands **(WH)**, Wear Jacket **(WJ)**, Wear Shoes **(WS)**, Writing **(Wr)**

C. Survey-based Data Collection

During the scripted activities collection (described in the previous section), we requested the participants to perform certain activities based on the scales used for clinical evaluation of functional and behavioral health assessment (described in this section). First, the survey-based questionnaire collects the basic demographics of the participants. The inclusion criteria to participate in the study includes i) must be above 65 years of age, ii) must be a candidate for the symptoms of dementia due to old age or any underlying neurodegenerative disorders. Hence, the first survey we conducted uses the Saint Louis University Mental Status (SLUMS) Examination to help screen for Alzheimer's or any other type of dementia. The advantage of SLUMS is that it can identify people with milder cognitive problems even if it has not risen to the level of dementia. The second survey we collected was related to mental health. We used the Geriatric Depression Rating Scale to measure the participant's mental depression state. Besides, we used the Geriatric Anxiety Scale to measure older adults' anxiety levels. Next, we used the Barthel Scale and the Lawton Instrumental Activities of Daily Living Scale to measure the performance of ADLs and IADLs, respectively. These measures were ideal for this study as they were performance-based and would provide us with the current state of functional health (can be compared to sensor-based data). We would ask the participants to perform the scripted activities, and based on their performance, we would score them, which we believe can help us measure their ability to carry out everyday activities required for independent living.

V. EXPERIMENTATION PIPELINE

The experiments were conducted on a Linux server consisting of i7-6850K with 4x NVIDIA GeForce GTX 1080Ti GPUs and 64GB RAM. Data preprocessing, feature extraction, classification, deep learning techniques were implemented and visualized using python. The deep learning tasks were implemented using PyTorch libraries.

A. Preprocessing

This study only considers the body-worn sensor data for analysis (only accelerometer for AAR dataset). Body-worn sensors are affected by motion artifacts. For instance, if the Empatica E4 is not tight enough, then the devices' slight movements on the wrist can cause motion artifacts. Motion artifacts are nothing but high-frequency noise. Hence, we employ a median filter of window size four, which acts as a low pass filter. We compared the effect of 2^{nd} order Butterworth filter and Kalman filter and found the results comparable. Hence, we chose the median filter as it was computationally faster. Further, a windowing approach was employed to prepare the data of the AAR dataset. The body-worn sensors provide continuous time-series, and a patch of the series represents an activity takes, thus requiring a windowing approach. A window size of 1 second with an overlap of 0.5 seconds was used for each axis and then concatenated. Additionally, we performed a feature extraction step after the windowing phase. We extracted discrete wavelet transform coefficients (using Daubechies 5(db5) scaling function) for each window and used that as the input data for the unsupervised feature learning phase.

B. Pipeline and Model Parameters

In the pre-training phase, we feed the unlabeled data (widowed (1sec) with overlap (0.5sec)) after wavelet transform to the feature extraction function that uses DBN to learn the representations. The raw input data instance vector's size was $(3 * 32)$, where 3 is the number of accelerometer channels and 32 is the window size. Besides, the wavelet coefficients remains with the same dimension. We stacked three RBMs with $90, 80, 70$ neurons in each RBM to form DBN. We opted to use a decreasing number of neurons so that the representations are both sparse and with a reduced dimension. Following the pre-training phase using the unlabeled dataset, we preserved the network structure and froze the weights. The labeled dataset was then passed into the preserved pre-trained network to acquire the representations, making the data instance to a size of $(3 * 32 + 70)$ for the supervised classification task.

C. Evaluation Strategy

To evaluate the effectiveness of the proposed methodology, we first visualize the representations learned from the DBNs for the labeled data. Further, we evaluate the performance of the classifiers using classification accuracy, precision, recall, f1-score, markedness and Informedness [22]. We noticed that the AAR dataset was imbalanced which suggests that accuracy is not the best choice of metric for model comparison. Instead, we use Informedness, and markedness for comparison because Informedness tells us about how well the model has learned the positives and negatives of classification while markedness tells us the trustworthiness of those positive and negative predictions [22], [23]. We also employ a 70-20-10 split for training, testing, and validation for shallow learning and deep learning approaches. The validation set for deep learning algorithms was used to tune the hyper-parameters and select the best model. The validation set and test set were truly held out data and were neither a part of the training phase nor the feature extraction phase. We have a massive set of unlabeled

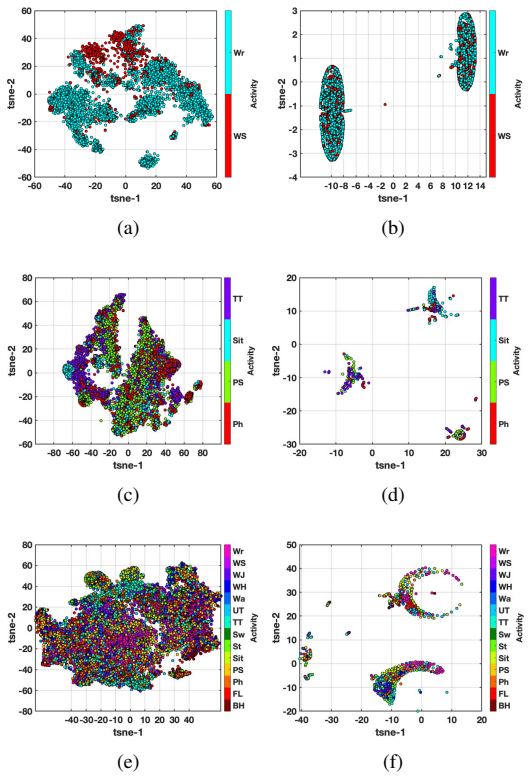

Fig. 2. Visualization of representations of labeled data in the new representation space learned using self-taught-learning for a) 2 activities , c) 4 activities and e) 14 activities without pre-training; b) 2 activities , d) 4 activities and f) 14 activities after pre-training.

instances (50936470 instances) compared to 651844 labeled instances for the AAR dataset before windowing.

VI. Results

This section discusses the various analysis performed using the AAR dataset. Before analyzing the classification task, it is imperative to assess the quality of the representation of the labeled data in the new self-taught space (obtained due to pre-training). Thus, we leverage visualization techniques to assess the quality of the representations. We perform principal component analysis (PCA) of the multi-dimensional representations to reduce the dimension to [1*20] for each instance, and further perform t-SNE and use the first two t-SNE components for visualization in 2D. Besides, the visualizations are color-coded with respect to the class labels. Figure 2 shows such a visualization for the AAR dataset comparing the representations of the raw data with the self-taught features. Figure 2 a, c, e corresponds to the visualization of the data after passing through the wavelet transform stage and Figure 2 b, d, f corresponds to that of the self-taught features of the labeled data in the new representation space. In Figures 2 a, c and e, we infer that the data belonging to different classes overlap with each other, which makes classifying such data a challenging problem. On the contrary, Figure 2 b, d, and f, shows clustering behavior in the new representation phase. Such a visualization asserts our assumption that the pre-training phase is able to learn representations corresponding to different classes (underlying conditional probability distribution of the data) in a way that similar classes are grouped together. On the other hand, there is also an overlap of data instances for some classes such as *Wear Shoe and Writing*. We believe there are two explanations for such findings. First, the visualization of the t-SNE is performed in 2-D; however, the underlying data is multi-dimensional, which may cause information loss. Secondly, the visualization is performed at the feature level, that is, before the classification. The deep learning algorithm especially contains a feature learning phase that hierarchically learns higher-level representations from input data and uses it for classification. Most learning algorithm performance declines as the number of classes; in our case, the number of activities decreases. To demonstrate the scalability of *STAR*, we show the separability of clusters of data instances belonging to the same class in the context of the increasing number of classes in Figure 2.

Now that we have obtained a good representation of the labeled data in a new representation space, we now aim to perform the classification task. First, we used some popular shallow learning algorithms such as k-NN, SVM, Random forest, and Multi-layer perceptron to perform the classification task. The notion behind this step was to check the efficacy of our pre-training and self-taught learning step for algorithms that highly depend on feature engineering. In contrast, we also investigated if we could extract more meaningful features from the self-taught features using deep learning approaches such as Convolutional Neural Networks (CNN). We chose CNNs for the deep learning approach because the convolution operation makes the algorithm shift-invariant. In the AAR dataset, some activities may have been performed in 3 seconds, and other instances of the same activity may have been performed in 2 seconds. The shift-invariant property of convolution operation in CNN ensures (at various depths of convolutional layers) to gather this information and represent them as a higher-level representation for the classification task. Table II shows the hyper-parameters used to train the convolutional neural network. Besides, the classification metrics have been listed in Table III. Comparing the shallow learning algorithms for the AAR dataset with/without self-taught features, we see an increase of 2% in both informedness and markedness metrics with a maximum of 44.56% for k-NN using the proposed pre-training methodology. Such a result ascertains the need to improve the feature representation of labeled data and supports our decision for a deep learning-based approach. Performing the same comparison with CNN-based classification, we noticed a 36% improvement in informedness and markedness metrics. Figure 3 shows the learning curves of the CNN model comparing both with/without self-taught learning. For AAR dataset, the model converges relatively quickly at 10 epochs with a higher accuracy when compared to the baseline CNN. Besides, Figure 4 shows the confusion matrix for *STAR*. Interestingly, we notice that some of the misclassifications are related to *Standing* activity. Activities such as *Wash Hands, Walking, Using Tooth Brush, Sweeping,*

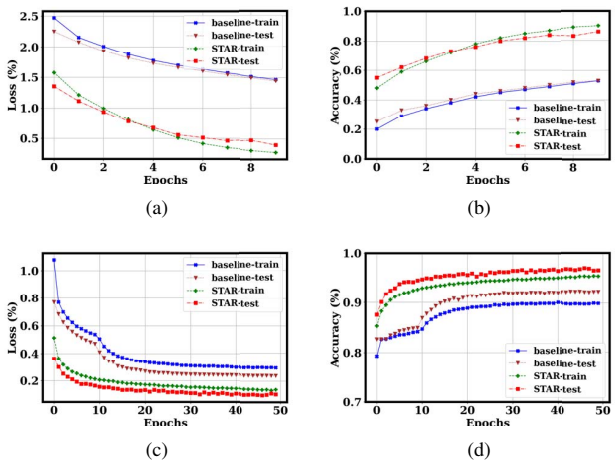

Fig. 3. AAR Dataset: a) loss vs epoch, b) accuracy vs epoch; OPPORTUNITY Dataset: c) loss vs epoch, d) accuracy vs epoch

Preparing Sandwich have been misclassified as *Standing* (2 - 15 %). We believe the reason for this is because each of these activities was performed while standing. A part of these activities constitutes *standing*, which makes it plausible for some windows labeled as *Wash Hands, Walking, Using Tooth Brush, Sweeping, Preparing Sandwich* activities could contain the pattern of *Standing*. In addition, *STAR* can successfully mitigate the practical challenges that arise during our data collection drive. For instance, let us consider some activities/sub-activities such as *brushing hair, picking up phone call, writing, wearing shoes*. Most of the time, these activities may not be detected by the wearable device based on its dominant/non-dominant hand position. Alternatively, some of these activities may be performed while sitting, as previously discussed. However, based on the confusion matrix (Figure 4), we notice a tremendous improvement in the detection of such activities using the proposed methodology. We believe our approach can also help mitigate the shortcomings of such system-related and practical challenges as well.

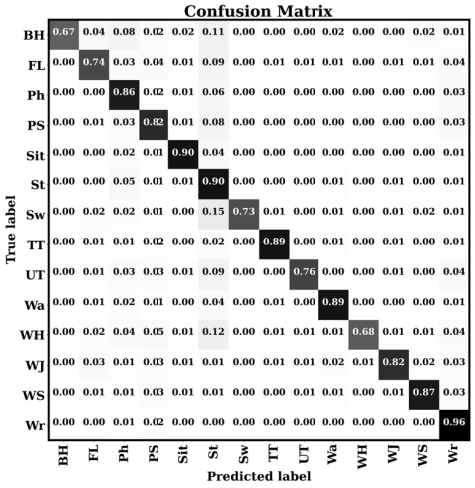

Fig. 4. AAR Dataset: Confusion matrix generated for *STAR*

TABLE II
HYPER-PARAMETERS OF CNN MODEL

Hyper-parameters	Values
No. of convolution layers	3
No. of filters in convolution layers	256, 256, 512
Convolution filter dimensions	5, 5, 5
No. of maximum fully connected layers	6
No. of neurons in fully connected layers	512, 256, 256, 128, 64, 14
Batch size	32

A. Comparison with Baseline

To better highlight the efficacy of our approach, we perform two types of comparisons. First, we compare how the state-of-the-art algorithm performs on the AAR dataset. Secondly, we use a publicly available dataset to show how well the algorithm scales to different scenarios.

1) State-of-the-art algorithms: The state-of-the-art algorithm for a similar category of data includes pre-training phase combined with a multi-task learning framework falling under the category of self-supervised learning [20](discussed in related work section). We evaluated the AAR dataset on this algorithm and found the informedness and markedness to be 68.09% and 68.11%. Comparing [20] with *STAR*, we see that *STAR* outperforms the state-of-the-art algorithm. The key difference between [20] and *STAR* lies in the pre-training phase where we use a generative energy based model while compared to a learning a focused variation or augmentation version of the data. Additionally, we also compare our work with other deep architectures known to perform well on activity recognition datasets such as LSTM [24]. With LSTM, we found the informedness and markedness to be 49.02% each with the AAR dataset, which suggests that our framework outperforms LSTM based networks.

2) Publicly available dataset: We chose OPPORTUNITY dataset [25], [26] as the publicly available dataset for the analysis. The OPPORTUNITY dataset comprises of both wearable sensors and ambient sensor data totaling to a sum of 145 attributes. In this study, we limit the usage of 77 attributes corresponding to the worn body sensors placed at different locations on the body. The 77 attributes correspond to the 3D accelerometer and 3D inertial measurement data (3D acceleration, 3D rate of turn, 3D magnetic field, and orientation of the sensor), and the sensors were placed at various positions on the body for 4 participants. In this dataset, some high level and low-level activities such as Relaxing, coffee time, early morning, cleanup, sandwich time, unlock the door, stir, lock the door, close door, reach the door, open door, sip, clean, bite, cut, spread, release the door and move are considered. Data corresponding to 2 participants constituted the labeled data, and the remaining unlabeled data of the same participants and all the data of another 2 participants were considered unlabeled data. We also considered the null class data to be a part of the unlabeled data pool from which the features were derived during the pre-training phase. We performed the same experiments as that of the AAR dataset and report the analysis below. For shallow learning algorithms, we found a substantial increase in the accuracies, approx. 8%

TABLE III
EVALUATION METRICS: COMPARISON OF SHALLOW LEARNING AND DEEP LEARNING FOR DOWNSTREAM CLASSIFICATION TASK

Classifiers	Without Self-taught learning					Using Self-taught learning				
	Precision	Recall	F1-score	Informedness	Markedness	Precision	Recall	F1-score	Informedness	Markedness
k-NN	46.38 %	46.38 %	46.38 %	44.56 %	44.56 %	48.30 %	48.30 %	48.30 %	44.32 %	44.32 %
SVM	21.56 %	21.56 %	21.56 %	15.53 %	15.53 %	40.16 %	40.16 %	40.16 %	35.55 %	35.55 %
Random Forest	21.91 %	21.91 %	21.91 %	15.91 %	15.91 %	23.82 %	23.82 %	23.82 %	17.96 %	17.96 %
MLP	20.62 %	20.62 %	20.62 %	14.51 %	14.51 %	36.66 %	36.66 %	36.66 %	31.78 %	31.78 %
CNN	53.23 %	53.23 %	53.23 %	49.63 %	49.63 %					
STAR						**86.18 %**	**86.18 %**	**86.18 %**	**85.12 %**	**85.12 %**

for SVM and approx. 7% using random forest. Besides, *STAR* shows a 6% boost in accuracy for deep learning. We believe that the both shallow and deep learning algorithms showed improvement because of the following reasons: 1) the dataset was captured from 4 users only, 2) the dataset comprises of numerous features (77 attributes as discussed earlier) 3) the dataset was captured in a lab environment due to which there is less intra-class variability.

VII. CONCLUSION

In this paper, we have demonstrated that the challenging problem of AR for older adults can be boosted using self-taught learning, especially leveraging abundant unlabeled data to aid the downstream classification task. We have presented self-taught learning-based activity recognition that has achieved approx. 36% increase in informedness and markedness; and outperforms the state-of-the-art research. We have also demonstrated that the proposed method, *STAR* is scalable to other population as well by evaluating on a publicly available dataset obtained for a different population. Besides, we have deployed our data collection system, *SenseBox*, in the real-world (in a retirement community) and collected 25 participants' data envisioned to study the relationship between activities and underlying functional and behavioral health. We believe *STAR* could be scaled to other data sources, such as sports analytics, fitness applications where capturing labeled sensory data is challenging and acquiring abundant unlabeled data is plausible.

REFERENCES

[1] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge and information systems*, vol. 36, no. 3, pp. 537–556, 2013.

[2] M. Alam, N. Roy, A. Misra, and J. Taylor, "Cace: Exploiting behavioral interactions for improved activity recognition in multi-inhabitant smart homes," in *ICDCS*. IEEE, 2016, pp. 539–548.

[3] M. Alam, N. Roy, S. Holmes, A. Gangopadhyay, and E. Galik, "Automated functional and behavioral health assessment of older adults with dementia," in *CHASE*. IEEE, 2016, pp. 140–149.

[4] M. Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "Ram: Radar-based activity monitor," in *INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, IEEE*. IEEE, 2016, pp. 1–9.

[5] M. Khan, H. Hossain, and N. Roy, "Infrastructure-less occupancy detection and semantic localization in smart environments," in *12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2015, pp. 51–60.

[6] N. Pathak, M. Khan, and N. Roy, "Acoustic based appliance state identifications for fine-grained energy analytics," in *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*. IEEE, 2015, pp. 63–70.

[7] J. Ma, H. Wang, D. Zhang, Y. Wang, and Y. Wang, "A survey on wi-fi based contactless activity recognition," in *2016 Intl IEEE Conferences (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*. IEEE, 2016.

[8] J. Taylor, H. S. Hossain, M. Alam, M. Khan, N. Roy, E. Galik, and A. Gangopadhyay, "Sensebox: A low-cost smart home system," in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 2017.

[9] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "Casas: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, 2012.

[10] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition — a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1254, 2018.

[11] K. D. Feuz and D. J. Cook, "Collegial activity learning between heterogeneous sensors," *Knowledge and information systems*, vol. 53, no. 2, pp. 337–364, 2017.

[12] R. Fallahzadeh and H. Ghasemzadeh, "Personalization without user interruption: boosting activity recognition in new subjects using unlabeled data," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*. ACM, 2017, pp. 293–302.

[13] M. Khan and N. Roy, "Untran: Recognizing unseen activities with unlabeled data using transfer learning," in *Internet-of-Things Design and Implementation (IoTDI), 2018 IEEE/ACM Third International Conference on*. IEEE, 2018, pp. 37–47.

[14] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07. New York, NY, USA: ACM, 2007.

[15] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariance sparse coding for audio classification," *arXiv preprint arXiv:1206.5241*, 2012.

[16] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in neural information processing systems*, 2009.

[17] P. He, P. Jia, S. Qiao, and S. Duan, "Self-taught learning based on sparse autoencoder for e-nose in wound infection detection," *Sensors*, vol. 17, no. 10, p. 2279, 2017.

[18] J. Gan, L. Li, Y. Zhai, and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, vol. 144, pp. 295–303, 2014.

[19] S. Bhattacharya, P. Nurmi, N. Hammerla, and T. Plötz, "Using unlabeled data in a sparse-coding framework for human activity recognition," *Pervasive and Mobile Computing*, vol. 15, pp. 242–262, 2014.

[20] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *CoRR*, vol. abs/1907.11879, 2019. [Online]. Available: http://arxiv.org/abs/1907.11879

[21] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, pp. 1527–1554, 2006.

[22] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[23] H. Hossain, M. Al Haiz Khan, and N. Roy, "Deactive: Scaling activity recognition with active deep learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018.

[24] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-lstm for human activity recognition using wearable sensors," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[25] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha *et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*. IEEE, 2010, pp. 233–240.

[26] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 2033–2042, 2013.