

CogAx: Early Assessment of Cognitive and Functional Impairment from Accelerometry

Sreenivasan Ramasamy Ramamurthy^{†*}, Soumyajit Chatterjee[§], Elizabeth Galik[‡], Aryya Gangopadhyay^{†*},
Nirmalya Roy^{†*}, Bivas Mitra[§], Sandip Chakraborty[§]

[†]Department of Information Systems, University of Maryland, Baltimore County, United States

^{*}Center for Real-time Distributed Sensing and Autonomy, University of Maryland, Baltimore County, United States

[§]Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India

[‡]School of Nursing, University of Maryland, Baltimore, United States

rsreeni1@umbc.edu, soumyachat@iitkgp.ac.in, galik@umaryland.edu, {gangopad, nroy}@umbc.edu,
{bivas, sandipc}@cse.iitkgp.ac.in

Abstract—An individual’s cognitive and functional abilities are commonly assessed through physical and mental status examination, observational performance measures, surveys and proxy reports of symptoms. These strategies are not ideal for early impairment detection as the individual needs to be present physically at the clinic to avail the assessments, especially for older adults who require assistance from a caregiver, and experience mobility, cognitive and functional disabilities from neurodegenerative disorders. Moreover, these strategies rely on self-reporting and proxy reports for evaluation which often leads to under-reporting of symptoms and decrease the validity of these measures. We argue that an early assessment of functional, and cognitive health impairment can be obtained from the individual’s daily activities captured through accelerometry. In this work, we postulate to learn high-level motion related representations from accelerometer data to better correlate with underlying functional and cognitive health parameters of older adults using a contrastive and multi-task learning framework. In particular, we posit a novel indicator, *Impairment Indicator* using the proposed multi-task learning framework that can indicate functional or cognitive decline as neurodegenerative disease progresses. An extensive 24-hour data collection from 25 older adults with the clinician in-the-loop was carried out in a retirement community center with IRB approval. We collected the activity patterns using wearables in their homes in addition to survey-based assessments and observational performance measures recorded by a clinical evaluator to infer their current cognitive and functional impairment status. Our evaluation on the acquired dataset reveals that the representations learned using contrastive learning aids in improving the detection of activities, activity performance score, and stage of dementia to 92%, 97%, and 98%, respectively.

Index Terms—Multi-task learning, Contrastive Learning, Activity Recognition, Functional, Cognitive Impairment Assessment

I. INTRODUCTION

Underdiagnosis of cognitive impairment is a prevalent problem, which can lead to severe health hazards like Dementia. People living with cognitive impairment often lack insight into their symptoms and may conceal their symptoms from

family caregivers and health care providers [1], [2]. These hindrances make the early detection of cognitive impairment a challenging problem, and a mild cognitive impairment may lead to more severe issues of Dementia. For a person living with neurodegenerative disorders, their cognitive decline is manifested as subtle to severe changes in their daily activities. Such collective symptoms of impairment in memory, language, motor skills, and judgment are hallmark symptoms of Dementia. Interestingly, these symptoms can induce anomalies in the activity patterns of individuals. Nevertheless, activity patterns may also change due to underlying functional health conditions, like dependability due to aging or other medical issues. Therefore, it is crucial to study the relationship between the activities and underlying cognitive (Dementia) and functional (ability to live independently) health to detect such impairments well before they become hazardous. The general approach to detect cognitive and functional impairments is through a set of medical surveys to analyze the memory, language, and motor skills of the subject [3], [4]. In contrast, this paper explores whether and how ubiquitous sensing can be used to understand such impairments at their early stages.

Existing studies in this context primarily focused on developing ubiquitous systems for supporting people living with cognitive impairments [5]–[7]. Although few works in the literature have focused on early detection of cognitive impairments, they primarily analyze specific activity patterns monitored through specialized sensors [8]–[10]. Although such platforms can work well under medical supervision, deploying them in the wild is generally challenging. However, various studies [3], [11] have shown that subtle differences exist in the daily activity patterns of an average individual and a person living with even a mild cognitive impairment.

Daily activities of individuals can broadly be classified as *Activities of Daily Living* (ADL) and *Instrumental Activities of Daily Living* (IADL). ADLs refer to simple activities that are essential to living independently such as ambulatory tasks, dressing, and eating. On the other hand, IADLs refer to complex activities, which are not crucial to survival but can help lead an independent life and take care of their needs and health. Notably, various works in the literature have focused on

This work has been partially supported by NSF CAREER Award #1750936, Alzheimer’s Association Grant/Award #AARG-17-533039, U.S. Army Grant #W911NF2120076, and MHRD funded SPARC collaborative project #SFE_SKI-1220.

understanding ADL and IADL patterns from pervasive sensing, particularly from motion sensors like accelerometers [12], [13]. Therefore, this paper explores whether we can develop a sophisticated model that can predict the stages/labels of the cognitive and functional impairments directly from the accelerometry sensing. One straightforward solution to this problem would be to identify the activity patterns from the accelerometry. The identified activity patterns can then be used to detect the stages of Dementia, particularly, *no cognitive impairment*, *mild cognitive impairment* (MCI), and *cognitive impairment* or Dementia. Similarly, one can also use the same activity patterns to infer the functional health represented by a *Functional performance score* that indicates the ability of the individual to live independently.

However, developing such a model has multiple research challenges. **First**, the major challenge is the *sparsity of data points*. As the target user group majorly contains the old and aged people and people living with Dementia, continuous monitoring with explicit supports, possibly with medical interventions, are required for collecting the labeled data properly. **Second**, for many ADLs and IADLs, the data in the input space has *marginal separability*; as there can be multiple patterns for the same activity. For example, a person may use different limb motions while combing her hair. With such different limb motions, it is difficult to identify whether a change in the pattern is intentional or due to the impact of Dementia. **Third**, different ADLs/IADLs have *different distributions by nature*; a person is likely to ‘brush’ maximum once/twice a day, but may perform ‘writing’ multiple times. Therefore, an irregular class distribution is inherent in the problem, which we cannot avoid. Further, as we are concerned with identifying anomalous activity patterns, using standard class balancing techniques such as SMOTE [14] may affect the inherent patterns in different activities. Interestingly, we do not know a priori which activities will be impacted due to cognitive impairment or functional dependency due to aging.

Owing to these challenges, we present *CogAx*, an end-to-end generalized pipeline to study the relationship between activity patterns and the underlying health conditions for two cases – (i) *Functional Impairment Assessment*, where we explore the Functional performance score of individuals, indicating how independently a person can live, and (ii) *Cognitive Impairment Assessment*, where we analyze activity patterns to infer the stages of Dementia. The input of the *CogAx* pipeline is accelerometer data, and the expected outcomes are two labels: (i) *activity labels* and (ii) *functional performance scores*, or *stages of dementia*, depending on the target output. In contrast to the existing works, our contributions in this paper are as follows.

(1) Studying the relationship between activities and underlying health conditions: We develop a novel model based on *Contrastive Multi-task Learning* (MTL) framework, where MTL takes the accelerometry data as input and captures the differences in the activity patterns of individuals. The cumulative contrastive and cross entropy losses measure

the orthogonality between the two output labels of the model – the activities and the underlying health conditions. One of the characteristics of the *CogAx* pipeline is its generalizability that the Contrastive Multi-Task Learning introduces. The proposed *CogAx* pipeline helps us learn both – (i) the activity dependency of a person while performing the ADL/IADL classification, and (ii) the Dementia stages while classifying ADL/IADL labels.

(2) Development of a framework to indicate impairments from accelerometry: Interestingly, we observe that the *CogAx* pipeline starts misclassifying the activities when the underlying health condition deteriorates. By leveraging this observation, we propose a novel parameter, *Impairment Indicator*, that is capable of longitudinally tracking an underlying health impairment for both Functional and Cognitive assessments.

(3) Model Development and evaluation through multi-source data: We conducted an extensive data collection drive to obtain sensor- and survey-based data from 25 older adults living with the symptoms of Dementia. Using this multi-source data (sensor data as input and survey data for ground truth), we develop *CogAx* and evaluate it for activity-functional impairment and activity-cognitive impairment assessments. We observe that *CogAx* achieves an F1-score of 97% for functional impairment, 98% for cognitive impairment, and 92% for activity classification.

II. RELATED WORK

A plethora of research has been conducted in the domain of *Sensor-based Activity Recognition* [12], [13], especially with wearable devices that use IMUs, PPG, GSR, and ambient sensors. In most recent studies based on sensor-based activity recognition, variants of deep learning algorithms such as autoencoders [15], RBM [16], [17] DBN [18], [19], CNN [20], [21], and RNN [20], [21] have been proven to capture the temporal variations of different ADLs and IADLs, thus improving the classification of activities. Most of these works leverage categorical cross-entropy loss to optimize the deep architectures to minimize the error between the true activity labels and the model estimates. In this paper, we propose to jointly optimize categorical cross-entropy loss with a similarity-based loss to better group the representations of the activity data.

Contrastive Learning, a type of deep metric learning paradigm, learns representations of the input data by pulling the instances in the same class closer while pushing away the instances otherwise. These learning paradigms leverage the inter-and/or intra-class similarities to learn the representations. Some examples of such losses are max-margin contrastive loss [22], triplet loss [23], quadruplet [24], multi-class N-pair loss [25], and NT-Xent loss [26]. These losses depend on the number of data instances and how they optimize inter-and intra-class distances. For instance, [27], [28] have demonstrated enhanced classification performance for the wearable sensor data by optimizing the inter-and intra- class distances

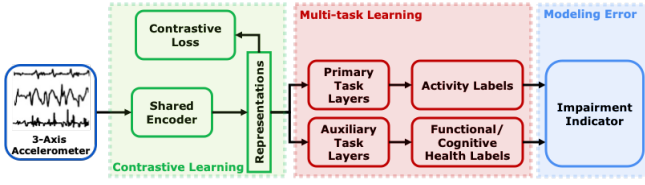


Fig. 1. Overview of *CogAx* pipeline

of raw and augmented data in the representation space using NT-Xent loss (similar to SimCLR [26] for Computer Vision applications). [27] leveraged NT-Xent loss to predict the future timesteps of wearable data by enhancing the quality of representations for better classification performance. In another work, the authors used hierarchical triplet losses that leverage subject similarities to mine informative triplets for an LSTM-based deep architecture with an attention mechanism to detect activities from sensor data [29]. Triplet loss has also been used with CNN [30] and Siamese loss with CNN+LSTM [31] based architectures. It is critical to note that the computational complexity increases when loss considers more data during optimization. *CogAx* pipeline uses the max-margin contrastive loss to minimize computational complexity.

Multi-task Learning (MTL) is a well-established deep learning paradigm that classifies multiple tasks simultaneously. Recent work in activity recognition leverages the MTL framework, where primary tasks are classified using representations learned from eight auxiliary tasks from the unlabeled data [32]. This paper aims to estimate functional and cognitive health impairment as the auxiliary task, which aids activity classification performance while providing meaningful clinical inferences.

The relationship between *activities and cognitive/functional impairment* using a data-driven approach is a less-investigated topic due to the limitations of collecting real-world datasets. Researchers have managed to collect multi-modal datasets [33] to predict the clinical health scores by fusing ambient and mobile sensor features into a behaviorome. However, this scheme involves multiple stages of training a module-specific shallow learning model to combine the multi-modal data and predict the clinical health score. Instead of a hierarchical approach, we aim to develop an end-to-end framework that simultaneously predicts the clinical score and activities. In addition, *CogAx* relies on a comprehensive data-driven approach that are robust while scaling and generalizing the model compared to handcrafted feature selection as used in previous studies including [34], [35].

III. PROBLEM STATEMENT & CHALLENGES

Problem. Fig. 1 depicts the broad overview of the *CogAx* pipeline that inputs the accelerometer data and learns the representations using a deep neural network architecture. The learned representations are used to learn – (i) activity labels (known as *Primary Task* (PT) labels), and (iii) underlying health condition labels (known as *Auxiliary Task* (AT) labels). Mathematically, *CogAx* helps us learn the mapping function, $F : X \rightarrow (Y_{PT}, Y_{AT})$, where $X = \{X_1, X_2, \dots, X_N\} \in \mathbb{R}$

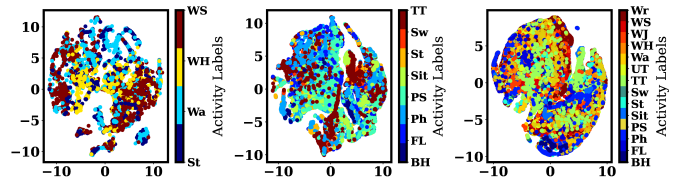


Fig. 2. Plot shows increasing overlap in data points as the number of activities increases (4, 8, 14; from left to right). This introduces marginal separability between the clusters of different classes.

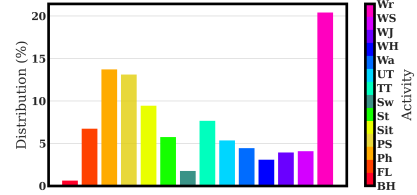


Fig. 3. Class distribution of the sensor-based accelerometer dataset that indicates the problem of imbalanced dataset.

denotes the input accelerometer data. The labels are defined as $y^T = \{y_1^T, y_2^T, \dots, y_N^T\} \subset Y_T$, where Y_T is the label space corresponding to either PT (Y_{PT}) or AT (Y_{AT}).

The highlight of the *CogAx* pipeline is that the same pipeline can address different objectives (characterized by a difference in the number of PT and AT labels, labeling granularity of AT labels; discussed in § V-B2) by executing the same pipeline with the corresponding PT-AT label pairs and minimal adjustments to deep architectures (discussed in § V-B1).

Challenges. The first challenge that the sensor-based data (explained in § V-A) poses is the marginal separability of the data in input space, as shown in Fig. 2. For any learning algorithm, marginal separability of the data makes the learning process difficult. For activity data captured from wearable sensing modalities, various reasons contribute to making the data marginally separable, such as higher complexity in activities, more number of activities, overlapping of data corresponding to different activities within a window while switching from one activity to another, overlapping motions of limbs for different activities (inter-class variations), multiple motions to perform the same activity (intra-class variations), etc. In Fig. 2, we show that as the number of activities increases (4, 8, and 14 activities), the complexity in the data increases, causing marginal separability in the data. The effect of marginal separability in the input data hampers the performance of any learning algorithm, especially when there is a smaller amount of data. The second challenge that the data poses is highlighted in Fig. 3 where the class distribution ranges from 2% (for *Brush Hair*) to 20% (*Writing*) of the total number of data instances. Such an imbalanced dataset causes learning algorithms to be biased towards estimating classes with the majority of data instances while failing to estimate the classes with minority classes. In an attempt to mitigate the effect of marginal separability in the data and

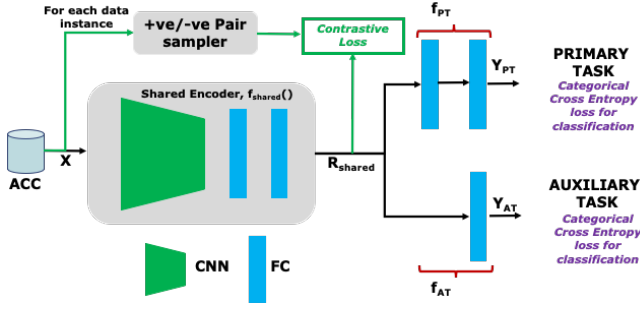


Fig. 4. *CogAx* : Network Architecture depicting the flow of the data, and highlighting how the contrastive and multi-task learning blocks are combined.

ease the learning process, we propose to leverage the inter- and intra- class distances in the optimization process using contrastive learning. Besides, contrastive learning constitutes a sampling process that mitigates the imbalanced class distribution problem.

***CogAx* Overview:** Fig. 1 outlines the crucial building blocks of *CogAx* pipeline, which are **contrastive learning**, **multi-task learning**, and **modelling error** blocks. The goal of the contrastive learning block is to learn the representations that are easily separable in the representation space. The second (multi-task learning) block is responsible for simultaneously learning the tasks (PT and AT). By combining MTL with contrastive learning using a joint optimization of the losses (as shown in Fig. 4), we believe that the model training would require fewer epochs (quicker convergence), because contrastive learning makes the representations easily separable in the representation space. The final block – the modeling error block, is responsible for computing a parameter representing the conditional missed detection of the PT labels based on the AT labels. We call this parameter *Impairment Indicator* (ζ) that can indicate a change in a person’s underlying health condition based on their activity patterns by analyzing the *CogAx* model error.

IV. DEVELOPMENT OF *CogAx*

Fig. 4 shows the network architecture of the contrastive MTL part of *CogAx* pipeline. The detail follows.

A. Design of Shared Encoder/Layers

The accelerometer data is first encoded using the shared layers, F_{shared} (comprises of Convolutional and fully connected layers). The role of the shared encoder/layers is to learn representations (R_{shared}) from the input data that are easily separable in the representation space and are attained using the contrastive loss discussed below.

Let $x^+ = \{x_i, x_j\}$ and $x^- = \{x_k, x_l\}$ denote the positive and the negative pairs, respectively, $x_i, x_j, x_k,$ and x_l are data points having the labels from Y_{PT} . Here, x_i and x_j have the same conditional probability distribution and belong to the same class (say, *Using Toothbrush*). On the other hand, x_k (e.g. *Using Toothbrush*) and x_l (e.g. *Brush Hair*) are data instances

that belong to different class. Let us assume the data instance x_j belongs to the class *Using Toothbrush*, but the user used a new electric toothbrush instead of a normal one. In this unique case, the underlying class labels are same, however, x_i and x_j will have slightly different patterns. The principle of contrastive loss is to project the representations of such data instances from x^+ together in the representation space. Conversely, data instances belonging to x^- are projected farther away from each other in the representation space. Minimizing the distance between $x_i, x_j \in x^+$ is achieved by maximizing the mutual information between $x_i, x_j \in x^+$ in the representation space [22]. On the other hand, to push the negative pairs apart, the mutual information is minimized. The representations that are projected in the representation space based on their class labels are obtained from the shared encoders. The positive pair, x_i and x_j , is fed to the shared network F_{shared} that provides us the representations v_i and v_j ($v_i = F_{shared}(x_i)$). The mutual information between the representations v_i and v_j can be defined as the KL divergence (D_{KL}) of the ratio of the joint distribution and the marginal distribution, $D_{KL}[\frac{P(v_i, v_j)}{P(v_i)P(v_j)}]$. To approximate mutual information, we use a similarity metric to reduce the computational complexity. We compute the contrastive loss using a logistic regression like loss function [22] as follows.

$$L_C = y_{cont} * \psi(v_i, v_j) + (1 - y_{cont}) * (\eta - \psi(v_i, v_j)) \quad (1)$$

In Eq. (1), y_{cont} is assigned either 1 or 0, if v_i and v_j belong to the same class (positive pair) or different class (negative pair), respectively. This phase of learning representation is semi-supervised because the actual label of either task is not used to learn the representation. However, the only label information that is used in this loss function is if the data pairs belongs to the same class or otherwise. Besides, the label information used to determine y_{cont} is based on the primary task labels Y_{PT} . In Eq. (1), we also define a hyperparameter η that determines how much farther away the representations must be if they are negative pairs. Besides, we use the similarity measure, $\psi()$, to approximate the mutual information between the two representations of the pair. We experimented with two similarity measures, cosine similarity and max-margin (Eq. (2)) for the contrastive loss.

$$Max - margin(v_i, v_j) = \begin{cases} \|v_i - v_j\| & , \text{ same class} \\ \max(0, \eta - \|v_i - v_j\|) & , \text{ different class} \end{cases} \quad (2)$$

Using this contrastive loss function, we design a discriminative network that maps the data into a new representation space where the data are clustered based on their classes, while being easily separable, by maximizing their agreement between the representations if they belong to the same class.

B. Design of Task Specific Layers

Given the representations from the shared layers (R_{shared}), the task-specific layers input R_{shared} to classify both the primary and auxiliary tasks. These task specific layers comprise of fully connected layers with the final layer equipped with a softmax activation function to obtain the probability of the data instance belonging to a certain class.

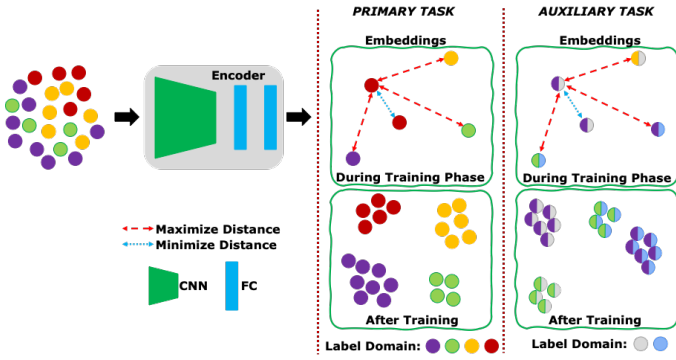


Fig. 5. Contrastive setup in a multi-task learning setting

1) *Primary Task Classification*: To learn the probability of a data instance belonging to a class, we leverage the task-specific layers corresponding to the primary task, F_{PT} . A visualization of the learned representations in the representation space is shown in Fig. 5 that shows the intuition of how the representations are grouped in the representation space, color-coded with Y_{PT} and Y_{AT} . For any given representation, there are two labels associated with it, which forms a $\{y_p \in Y_{PT}, y_a \in Y_{AT}\}$ pair. Since the contrastive loss is directly related to the primary task, it is straightforward to group together the representations based on Y_{PT} . However, when auxiliary task labels Y_{AT} are considered in tandem, the grouped representations are further split with respect to Y_{AT} , as demonstrated in Fig. 5.

To learn the mapping of the representations to the class labels, we use categorical cross entropy loss to learn the network parameters. Now, the total loss is defined as in Eq. (3).

$$L_{PT} = L_C - \sum_i^{|Y_{PT}|} y_{(PT|true)} * \log(f_{PT}(v)) \quad (3)$$

The combined L_{PT} loss leverages the semi-supervised contrast based on the primary task labels and help predict the primary task labels in a supervised fashion.

2) *Auxiliary Task Classification*: To convert the previously discussed supervised contrastive learning setup to an MTL setup, we use the representations learned from the shared layers and feed them to the Auxiliary task specific layers, F_{AT} . Since we are modeling the auxiliary task as a classification problem, we use a categorical cross entropy loss to obtain the logits for the auxiliary task (L_{AT}). Accordingly, the overall loss function is defined in Eq. (4).

$$L_{total} = L_{PT} - \sum_i^{|Y_{AT}|} y_{(AT|true)} * \log(f_{AT}(v)) \quad (4)$$

The total loss is used to back-propagate the gradients to learn all the network parameters.

C. Modeling Errors: Impairment Indicator

Understandably, like any other learning algorithm, this model also misclassifies input instances depending on the

representations learned by the model during the training time. In this context, the usage of contrastive loss provides us with a richer scope to analyze and model this error to provide better insights. This is due to the inherent nature of the contrastive setup, which reduces the overall distances between the representations corresponding to the same activity labels. However, with the complicated nature of the problem and a diverse dataset containing input data from people involving different underlying health conditions, there can be certain conditions where the model may fall prey to the problem of data variability. For example, consider two subjects P_1 and P_2 performing the same IADL, *Sweeping*, with P_1 living with *MCI* whereas P_2 has no underlying health issues. The contrastive setup forces the model to learn representations that are close enough in the latent space. However, with constant variations in the input signatures from the subjects living with *MCI*, the model may eventually get confused and misclassify the activities (the primary task for this model). Focusing on this behavior of the contrastive setup in misclassifying when the representations change beyond a certain threshold, we capture and model this error through a parameter called *Impairment Indicator* (ζ) defined as follows.

Formally, we define ζ as the dependent missed detection rate of the primary tasks conditioned on the correct detection of auxiliary tasks. Mathematically, ζ_c for a primary task class c is represented as in Eq. (5).

$$\zeta_c = \frac{1}{M} \left[\sum_{m=0}^M \{ \delta[y_m^{PT} \neq \hat{y}_m^{PT}] * \delta[y_m^{AT} = \hat{y}_m^{AT}] \} \right] * 100 \quad (5)$$

Here, $c \in Y_{PT}$ refers to a specific activity (listed in Table I based on the experimental objectives), $m = \{0 : M\}$ denotes the index of the data instances belonging to the class c , PT and AT refers to primary task and auxiliary task respectively, and δ refers to a binary indicator function that returns 1 when the condition in the parenthesis is satisfied and 0 otherwise. In simpler words, ζ provides an insight into the variations caused in the sensing pattern in the presence of an underlying health condition and thus have a potential to provide a *clinical assessment* of both functional and cognitive impairment assessments using simpler accelerometry data.

V. DATASET DETAILS AND EXPERIMENT

A. Data Collection

For this work, we collect two types of data: Sensor-and survey-based. The sensor-based data was collected to record the patterns of activities. Whereas the survey-based data aimed at clinically assessing functional, behavioral, and cognitive impairment. The intention was to obtain a dataset that would facilitate studying the nuances in the activity patterns and their relationship with various aspects of a person's health.

Sensor-based Data Collection. Empatica E4, a wearable device, was used to capture a person's motion through accelerometry (sampling rate 32Hz), electrodermal activity through galvanic skin response sensor, and cardiac activity

TABLE I
LIST OF THE LABELS CONSIDERED FOR BOTH PRIMARY TASK (PT) AND
AUXILIARY TASK (AT) CLASSIFICATION IN BOTH SCENARIOS

Objectives	Labels
Objective 1 (PT)	Brush Hair (BH), Folding Laundry (FL), Phone (Ph), Preparing Sandwich (PS), Sitting (Sit)
Objective 1 (AT)	0, 1, 2, 3, 4, 5, 6, 7, 8
Objective 2 (PT)	Brush Hair (BH), Folding Laundry (FL), Phone (Ph), Preparing Sandwich (PS), Sitting (Sit), Standing (St), Sweeping (Sw), Taking Trash Out (TT), Using Toothbrush (UT), Walking (Wa), Wash Hands (WH), Wear Jacket (WJ), Wear Shoes (WS), Writing (Wr)
Objective 2 (AT)	Normal (NOR), Mild Cognitive Impairment (MCI), Dementia (DEM)

using photoplethysmograph (PPG). In addition, ambient sensors (e.g., passive infrared sensors, object tags, reed switches) were used to obtain the context. The data were collected in a real-world scenario with 25 older adults living in a retirement home facility out of which 10 were living with MCI (50.42%), 8 with DEM (28.57%), and 7 were NOR (20.99%). The sensor data was collected for 24 hours for each of the older adults. For the first 2 hours, the participants were requested to perform scripted activities (based on a survey-based assessment) comprising both ADLs and IADLs. Thereafter, for the next 2 hours, the participants were requested to continue their daily routine, albeit no explicit restrictions were imposed regarding the repetition of activities (ADLs and IADLs). Notably, subjects with underlying cognitive/functional impairment tend to have inconsistent behavior even for the same set of activities; Nonetheless, these repetitions of activities from varying subjects with different underlying impairments made our collected dataset more naturalistic and diverse. In compliance with the IRB approval (#Y18NR12035), we collected a video feed for the first four hours for the ground truth data annotation while keeping the rest of the 20 hours of a day uninterrupted without having any cameras.

Survey-based Data Collection. The survey-based questionnaire was collected in tandem during sensor data collection. Besides demographic information, we conducted – a) *SLUMS* Examination [36] to help screen for dementia, b) Geriatric Depression Rating [37] to measure the participant’s mental depression state, c) Barthel and the Lawton Instrumental Activities of Daily Living Examinations [38] to measure the functional performance of ADLs and IADLs respectively.

The inclusion criteria for participation in this study include a) must be above the age of 65 b) be a candidate for the symptoms of dementia either due to old age or any underlying neurodegenerative disorders. Although multi-modal data were acquired (both sensor- and survey-based), we only used the accelerometer data as the sensor data source in this paper. The notion was to investigate if it could extract any other information from accelerometer data other than activities in an end-to-end manner. Besides, this study considered only *SLUMS* and Lawton’s scales for cognitive and functional impairment assessment scores, respectively.

TABLE II
HYPERPARAMETER OF *CogAx* MODEL

Hyperparameters	Shared	Objective 1		Objective 2	
		PT	AT	PT	AT
Convolutional layers	4	-	-	-	-
Convolutional filters	96, 128, 256, 512	-	-	-	-
Convolutional filter shapes	1x5, 1x5, 1x5, 1x5	-	-	-	-
Fully Connected (FC) layers	4	2	3	2	3
FC Neurons	512, 256, 265, 128	128, 5	64, 32, 6	128, 14	64, 32, 3
Batch Size	16				

Annotations. The first step in labeling the sensor-based time-series dataset was aligning the timestamps of the wearable devices with the cameras. We noted each activity’s start and end time and extracted the corresponding wearable data for each activity. We appointed two annotators (with a background in ML), and the labels that matched both the annotators were chosen as the final label for this study. In the case of survey-based labels, both the Cognitive Impairment labels (that differs from person to person) and the activity performance score (that differs for each activity performed by every individual) were labeled by a clinical evaluator. Finally, we matched the sensor- and survey-based labels based on the {participant-id, activity} pair tuple.

B. Experiment procedure

The proposed framework was implemented using Pytorch, and we provide the details of all the parameters required to replicate our analysis below.

1) *Model Implementation and Training:* We used the raw data directly for training the model without any explicit feature extraction [39], after normalizing the accelerometry data. Next, separate held-out data was used as the test and validation set to assess the model performance. In addition, we employed a sliding window of 1 second with 50% overlap to capture the temporal patterns of micro-activities [40]–[42]. Notably, these different short-spanned involuntary movements often contribute significantly to discerning all the activities, especially for older adults with underlying cognitive/functional impairments. Based on this intuition, in the *CogAx* pipeline, we rely on a shorter window size to extract more meaningful feature representations instead of longer windows which are known to work well with handcrafted features [34], [35].

The training is performed for 50 epochs with a scheduled learning rate that decreases as the number of epochs increases. We chose Adam [43] optimizer to optimize the weights. The hyperparameters of the shared and task-specific layers are listed in Table II. A crucial point to note is that for both functional and cognitive impairment assessments, the same *CogAx* pipelines were used. Besides, the number of convolutional and fully connected layers were same as well. However, the only change in the pipeline were the number of neurons in the final layer to adjust to the number of labels in PT and AT, giving us an objective specific models.

2) *Experiment Scenarios*: Table I summarizes the primary task (PT) and the auxiliary task (AT) labels for the two objectives – (1) functional impairment assessment, and (2) cognitive impairment assessment. For functional impairment, the scores range from 0 (low function) to 8 (high function) for women and from 0 (low function) to 5 (high function) for men. Although the scores range from 0 - 8, our dataset only consisted of older adults with a maximum score of 5. Hence, we only used six neurons in the final fully connected layer for this objective. Both the tasks were annotated at the same granularity. For cognitive impairment assessment, the scores for stages of dementia were computed using the SLUMS examination, and the labels were decided based on the categorization provided in the SLUMS examination (Normal (NOR), Mild Cognitive Impairment (MCI), Dementia (DEM)).

C. Performance Metrics and Evaluation Strategy

Since the sensor data exhibits imbalanced class distribution, the F1-score would be the appropriate performance measure. Besides, the data used for *CogAx* was captured in a real-world setting. As a result, we were unable to get all the fourteen activities (Table I) from all the participants due to their underlying conditions. This aspect weighed in on our decision to choose completely held out data with random stratified sampling for evaluation over Leave One Out Cross Validation (LOOCV). Moreover, LOOCV typically minimizes the bias in the evaluation and increases the variance. However, since we do not have the same set of activities across all participants, the categorical cross-entropy loss would suffer a higher variance due to the absence of certain activities during training. In addition, referring to the discussion in §IV-A, the selection of +ve and -ve pairs for the contrastive setup involves random sampling of +ve/-ve pairs irrespective of the user, thereby reducing the user-related bias. Collectively, we believe a better balance of the bias-variance was obtained in our evaluation by choosing a completely held out data. Besides we employed a 70-20-10 ratio in a stratified fashion to for train, validate, and test the MTL model and reported all the evaluation scores by averaging the scores obtained with five different random seeds. Note that though there is a likelihood of having data from the same subject in both test and train sets, a variety of involuntary movements (discussed in section V-B1) in conjunction with a 1-second window ensures increased intra-class variations (as seen in figure 2), and thereby mitigate the chance of overfitting.

D. Baseline

As a baseline, we implemented CNN and fully connected (FC) layers-based network for each of the individual tasks listed in Table I. The same procedure were followed for the input data, windowing with overlap, and used a completely held out data for test and validation set for the baselines.

VI. EVALUATION

In this section, we discuss the outcomes of our experiments and show that the proposed architecture is efficient in addressing the research problems.

TABLE III
BASELINE AND ABLATION STUDY EVALUATION; CNN + FC LAYERS WERE USED FOR CLASSIFICATION.

Impairment Assessment	Loss Functions Considered			F1-score
	L_C	L_{PT}	L_{AT}	
Functional	✓	✓	X	0.92
Functional	X	X	✓	0.76
Functional	✓	X	✓	0.98
Cognitive	X	✓	X	0.46
Cognitive	✓	✓	X	0.92
Cognitive	X	X	✓	0.75
Cognitive	✓	X	✓	0.97

L_C : Contrastive loss, L_{PT} : Categorical Cross Entropy loss for PT, L_{AT} : Categorical Cross Entropy loss for AT

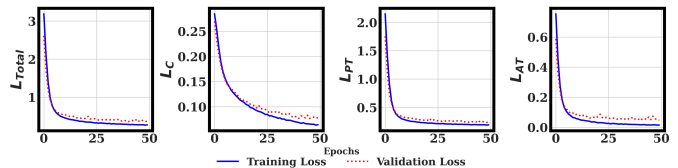


Fig. 6. Total and Individual Loss Curves of *CogAx* pipeline for Cognitive Impairment Assessment Objective.

A. Model Performance and Key Observations

1) *Model Performance*: Here we compare the performance metrics for the classification (both PT and AT) of both objectives and compare them with the baseline. Table III shows the classification performance for the baselines for activity, functional, and cognitive impairment assessment. For the 14 class activity classification baseline, the best model was achieved using more CNN and FC layers than *CogAx*. In addition, the model converged at 50 in comparison to 20 epochs for *CogAx*. The *CogAx* model for 14 class activity classification outperformed the baseline by an F1-score of 45%. For either objectives, *CogAx* models outperformed the baseline for AT by 22%. One of the reasons for *CogAx* to perform better than their baseline counterparts can be attributed to the +ve/-ve pair sampling strategy employed as explained in §IV-A. When the pairs are sampled, we allow an equal number of data instances from each class to form pairs with the same or different classes. Hence, the variable class distribution is mitigated in the *CogAx* pipeline, whereas the irregular class distribution persists for the baseline. We make the two scenarios comparable by using the weighted F1-score that leverages the support for each class in computing the average F1-score. The weighted F1-score is equal to the averaged F1-score for the proposed contrastive MTL model.

2) *Model Convergence*: The primary concern with the *CogAx* pipeline was the convergence of the network, given that the model is optimized using a cumulative loss of three orthogonal losses. The gradual decrease in loss values towards 0 as the number of epochs increases is an indicator of model convergence. We observed a similar pattern for each of the losses. We have visualized the total loss, PT categorical cross-entropy loss (L_{PT}), and AT categorical cross-entropy loss (L_{AT}) for both the objectives in Fig. 6. Interestingly, at epoch=0, the L_{PT} loss was very high when

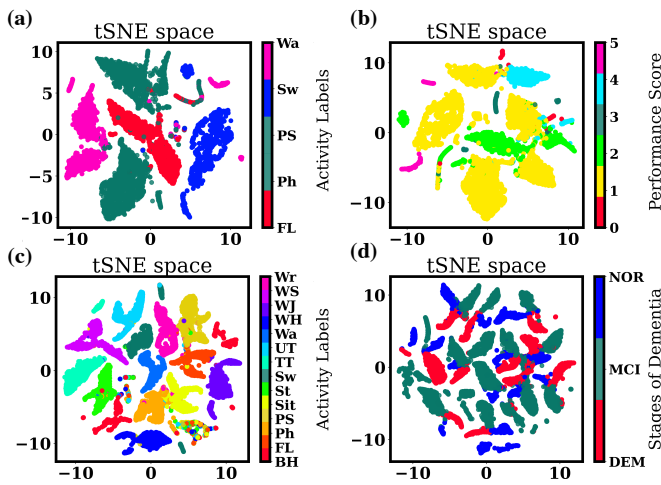


Fig. 7. Learned representations by the shared layers color coded for the PT and the AT for Functional Impairment Assessment (a, b) and Cognitive Impairment Assessment (c, d).

compared to other losses suggesting that the primary task classification was very poor. As the epochs progressed, the lower and gradual reduction of L_C and L_{AT} helped reduced L_{PT} . Such a loss pattern profile confirms our claim that although the losses were orthogonal to each other, the combined loss optimization allowed them to regularize each other and learn the model parameters quicker. Although Fig. 6 shows the loss curves for cognitive impairment assessment, we obtained similar curves for functional impairment assessment as well. In either objective, the total loss converged at epoch 20.

B. Dissecting the CogAx pipeline

1) *Representations learned using Contrastive Setup:* Fig. 7 shows the tSNE plot of representations obtained from the shared encoders. Figs. 7(a) and (c) are color-coded with respect to the PT labels, and Figs. 7(b) and (d) are color-coded with respect to the AT. For Figs. 7(a) and (c), we observe a single group or cluster for each activity label, indicating that the contrastive learning succeeded in pulling the data instances from the same class together. For Figs. 7(b) and (d), we observe multiple clusters for each AT label, suggesting that the L_{AT} has induced a contrastive effect. Now, the induced contrastive behavior is observed between AT-PT label pairs instead of only PT labels used implicitly in the contrastive loss (L_C). We believe that through this mechanism, the categorical cross-entropy loss of auxiliary task (L_{AT}) counters the orthogonality between the tasks.

Although tSNE plots show separability w.r.t. various PT and AT labels, tSNE pulls data instances from the same class together and form clusters by minimizing the KL Divergence. This is an iterative and parameterized method that requires tuning to obtain a good visualization. Over-tuning parameters may cause instances farther away from the cluster center to be grouped along with the data instances closer to the cluster center. Thus, tSNE visualization would inhibit us from obtaining deeper insight into which data instance

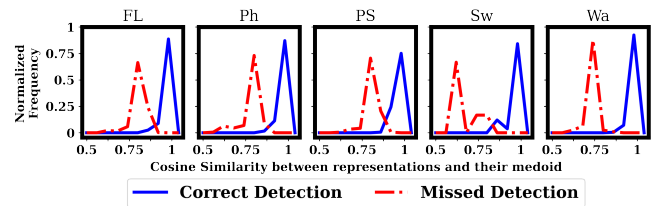


Fig. 8. Functional Impairment Assessment: Probability density function of pairwise cosine similarity between each representation with their medoid for each activity for correct and missed detected PT; higher value of cosine similarity indicates that the correct detected data instances and the medoid are closer to each other in the representation space.

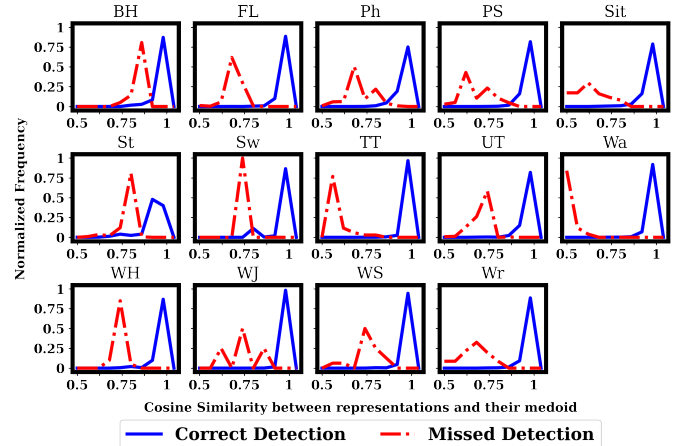


Fig. 9. Cognitive Impairment Assessment: Probability density function of pairwise cosine similarity between each representation with their medoid for each activity for correct and missed detected PT; higher value of cosine similarity indicates that the correct detected data instances and the medoid are closer to each other in the representation space.

causes the *CogAx* pipeline to incorrectly detect the PT and AT labels. Given that the primary notion of this work is based on contrastive learning, the naive assumption for incorrectly detecting labels by *CogAx* would be that the model experiences difficulty in learning representations that are farther away from the cluster center (representative embedding of a particular label) or closer to the cluster center of another class. To corroborate this assumption, we determine the representative embedding (representation) of each PT label, and we refer to such a representation as medoid. The medoid is chosen when the pairwise cosine similarity between a representation and all the other representations (belonging to the same PT label) is maximum. Further, we compute the probability density function (*pdf*) of the cosine similarity between the medoid and the representations that are correctly detected (based on PT labels). A similar *pdf* of the cosine similarity between the medoids and the representations that are missed detected (based on PT labels) was computed. We plot the computed *pdf* of correct and missed detections for functional and cognitive impairment assessments in Fig. 8 and Fig. 9 respectively. We observe that the cosine similarity of the missed detected representations is lower than that of

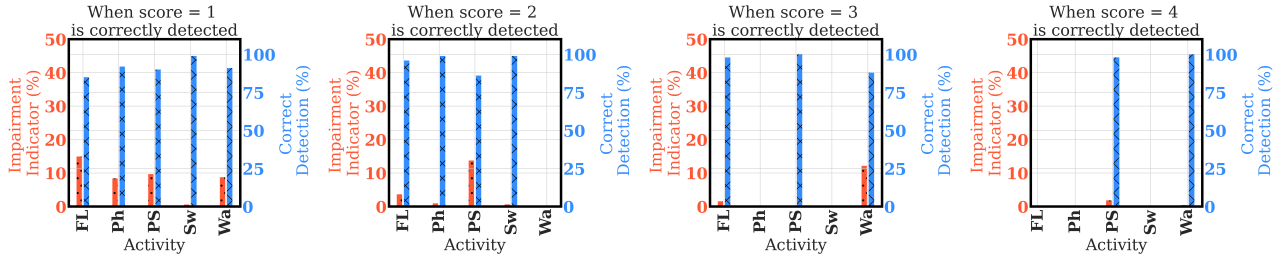


Fig. 10. ζ values for Functional Impairment Assessment. Lower the functional performance score, the person is more dependant.

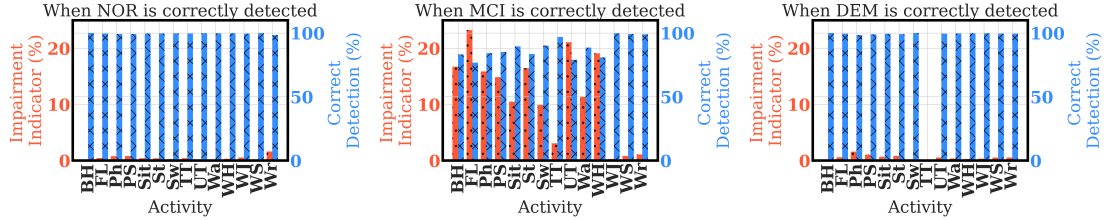


Fig. 11. ζ values for Cognitive Impairment Assessment for Normal Control, MCI and Dementia.

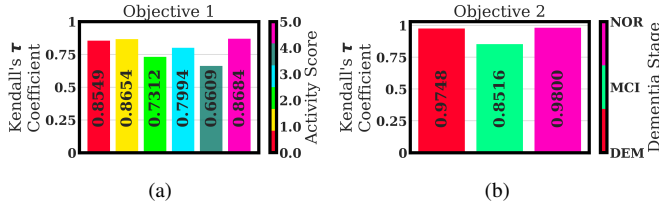


Fig. 12. Kendall Tau Hypothesis testing to show the association between correctly detected PT labels conditioned on correct detection of their corresponding AT label for Objective 1 (a) and Objective 2 (b).

the correctly detected for either objective. Thus, we can infer that the representations farther from the medoid contribute to the *CogAx* pipeline to detect PT labels incorrectly.

2) *Clinical Significance of Impairment Indicator, ζ* :

Functional Impairment Assessment. Fig. 10 shows the plot of ζ that enables the study between activities and Functional performance score. We observe that the ζ values increase as the scores reduce. The intuition is that as a person becomes more dependant, *CogAx* immediately misclassifies activities (due to a underlying health condition), and the proposed ζ value captures these misclassifications.

Cognitive Impairment Assessment. Fig. 11 shows the plots of ζ for various stages of dementia. In comparison to Scenario 1, the ζ parameter is a much prominent indicator of impairment. We can infer that the ζ is very close to zero when the impairment stage is *Dementia* and *Normal*. On the contrary, the ζ value increases when the dementia stage is *MCI*. During the data collection drive, interestingly, we observed some behaviors for each category of stages of Dementia. One such observation is that a *Normal* person could follow the instructions clearly during the scripted activity data collection

phase. However, a person with *Dementia* faced interruptions while performing activities. However, when they performed the activity well, they were as good as that of a *Normal* person. Interestingly, when the person lives with *MCI*, the level of confusion was more, and the order in which the activities were performed were shuffled. Despite a contrastive loss trying to pull data closer, when data with *MCI* was presented, the *CogAx* model could not classify them correctly, which is exactly what we highlight through the Impairment Indicator (ζ).

Consequently, ζ can potentially be used to track a person's health decline longitudinally for either objective. As noted earlier, there is a well-defined trend of ζ (decreasing as the health deteriorates and then increasing when the health further deteriorates). To further validate this trend, we statistically show the significance of ζ . Since ζ is defined based on conditional missed detection, we show the significance of ζ based on the conditional correct detections. So we define the null hypothesis H_0 as there is no association between the y_m^{PT} and \hat{y}_m^{PT} . Since both y_m^{PT} and \hat{y}_m^{PT} are categorical in nature, it is not possible to obtain their distributions. Besides, the true labels and the predicted labels have one to one relationship. Thus a non-parametric statistical approach, such as Kendall's Tau test, would be ideal. *Kendall's Tau Coefficient* is a statistics that is used to measure ranked or ordinal association between two measured quantities. The value of tau coefficient closer to 1 indicates a strong association between the measured quantities, suggesting that the null hypothesis can be rejected. Fig. 12 shows the tau statistics for both objectives. We observe that the tau statistics decrease as the Functional performance score reduces and increases when the score reaches near 5 for objective 1. Similarly, the tau coefficient decreases when the stage of dementia moves from NOR to MCI, and the same increases when the stages of dementia change from MCI to DEM. In other words, the association between y_m^{PT} and \hat{y}_m^{PT} varies similar to ζ , however in an opposite way (tau decreases

and then increases; ζ increases and then decreases). For the tau statistics, we obtained a p-value < 0.05 indicating that the null hypothesis can be rejected.

3) *Ablation Study*: In this section, we discuss the ablation study to demonstrate that we have chosen the best network for *CogAx*. First, we removed the auxiliary task-specific layers and retained L_C and L_{PT} from *CogAx* and obtained the same F1-score for classifying PT labels (Table III). However, the number of epochs needed for the model convergence was 50 epochs compared to 20 epochs for the *CogAx* pipeline. We further removed the L_C with the same network configuration and observed an F1-score of 55% obtained at 120 epochs, suggesting that the network failed to learn PT labels. Next, we removed the L_C from the *CogAx* pipeline and only used the categorical cross-entropy loss for both PT and AT. We obtained an F1-Score of 56% in this scenario and observed that the model overfit the data. Finally, we observed the *CogAx* network’s performance after reducing the network size in terms of one CNN layer and one FC layer reduction. The network performance drastically reduced, and we report an F1-score of 58% and 73% respectively for the PT, suggesting that the best network was chosen for the *CogAx* framework.

VII. DISCUSSIONS

Here we summarize some interesting observations encountered while developing *CogAx* framework.

Orthogonality in PT and AT: We define orthogonality in terms of which task the model tries to learn naturally. Referring to Eq. (1), the Contrastive loss has been set up with the PT labels. Thus, the model is biased towards learning PT labels, which makes learning AT labels difficult. Another crucial aspect that influences orthogonality is the mapping between the PT and AT labels. The PT and AT for the Functional Impairment Assessment objective are activity labels and Functional performance scores. A one-to-one mapping between PT and AT exists for a given user as each activity was evaluated once per user during sensor and survey-based data collection (§V-A). However, such a one-to-one mapping does not exist for the objective of cognitive impairment assessment. Thus, the effect of orthogonality is limited in the prior scenario compared to the latter. Evaluating *CogAx* for two scenarios with varying levels of orthogonality demonstrates its generalizability.

Deviations in Impairment Indicator (ζ): The MTL aspect of the *CogAx* pipeline provides both PT (L_{PT} and L_C leading to biased supervision) and AT (using L_{AT} with less supervision) labels concurrently by optimizing three different loss functions. Due to the biased supervision on PT labels, *CogAx* is more inclined towards estimating the PT labels correctly leading to our assumption that, any missed detection of the PT labels (albeit biased supervision) are caused due to the prevalent underlying cognitive/functional impairment. On the other hand, there are data instances whose AT labels were incorrectly detected; these are the data

instances where *CogAx* is not sure if the missed detection of PT labels is contributed by cognitive/functional impairment. As the parameter ζ is designed based on the missed detection of PT labels conditional to the correct detection of AT labels, ζ does not consider these data instances, making ζ resilient to other causes contributing to the missed detection of PT. In conclusion, *CogAx* pipeline estimates activity labels in a biased way so that any missed detection of activity labels can be attributed to cognitive/functional impairment (especially when cognitive/function impairment labels are correctly detected). For instance, if a person is physically hurt temporarily without any cognitive impairment, *CogAx* will automatically classify the person as *Normal* (due to a very high performance of AT label classification), while the ζ value will still be deviated compared to others under the *Normal* criteria. On the contrary, if *CogAx* misclassifies the person as either *MCI* or *DEM*, the conditional missed detection criteria would prevent the ζ parameter from deviating. Comparing these deviated ζ values against an anxiety scale or a depression scale would further allow us to relate the cause of these deviations from the usual routine to either anxiety, depression, or any other health impairment, which we plan to investigate in the future.

Selection of +ve/-ve pairs: We have shown that the proposed sampling of +ve/-ve pairs can drastically boost the model performance for both PT and AT (although they are orthogonal to each other). In addition, this pipeline can solve classical issues related to Human Activity Recognition (HAR). If the +ve pairs belong to the same user, it solves the intra-class variation problem caused due to behavioral randomness exhibited by any person. On the other hand, when the +ve pair belongs to different users, the subject-level heterogeneity is mitigated. In either case, the aforementioned problems are solved by grouping the positive pairs closer to each other in the representation space.

VIII. CONCLUSION

In this paper, we presented a pipeline, *CogAx*, to study the relationship between activities and underlying cognitive/function health impairment. *CogAx* feeds on accelerometer data and simultaneously estimates both activity and cognitive/functional impairment assessment scores using a novel contrastive multi-task learning approach. Our key insight is that when *CogAx* pipeline detects the activity labels with superior supervision (using multiple loss functions), any missed detections of activity labels are caused by an underlying cognitive/functional health impairment. Relying on this insight, we designed an Impairment Indicator (ζ) to quantify the error in the activity prediction model induced by underlying health conditions, which can help posit a decline in an older adults’ impairment. Finally, we demonstrated the effectiveness of ζ for the two objectives of Functional and Cognitive Impairment Assessment, annotated at two different levels (subject- and activity- level of granularity).

REFERENCES

- [1] N. Falk, A. Cole, and T. J. Meredith, "Evaluation of suspected dementia," *American family physician*, vol. 97, no. 6, pp. 398–405, 2018.
- [2] P. S. Farsai, "Cognitive impairment in older adults and oral health considerations," *Geriatric Dental Medicine, An Issue of Dental Clinics of North America, eBook*, vol. 65, no. 2, p. 345, 2021.
- [3] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, 2005.
- [4] S. Borson, L. Frank, P. J. Bayley, M. Boustani, M. Dean, P.-J. Lin, J. R. McCarten, J. C. Morris, D. P. Salmon, F. A. Schmitt *et al.*, "Improving dementia care: the role of screening and detection of cognitive impairment," *Alzheimer's & Dementia*, 2013.
- [5] N. T. Ly, R. Tscharn, J. Preßler, S. Huber, S. Akinie, A. Serna, and J. Hurtienne, "Smart lighting in dementia care facility," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1636–1639.
- [6] D. C. Sandoval and J. Favela, "Co-designing ambient-assisted interventions using digital interlocutors for people with dementia," in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 813–821.
- [7] N. Sakhnini, J. E. Yu, and D. Chattopadhyay, "myCityMeter: Helping older adults manage the environmental risk factors for cognitive impairment," in *2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 235–238.
- [8] L. A. M. Cano, J. Beltrán, R. Navarro, M. S. García-Vázquez, and L. A. Castro, "Towards early dementia detection by oculomotor performance analysis on leisure web content," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 800–804.
- [9] M. Ianculescu, A. Alexandru, E. Tudora, and G. Neagu, "Mapping the behavioral change evaluation of elderly with mild cognitive impairment in a smart environment," in *2020 International Conference on e-Health and Bioengineering (EHB)*. IEEE, 2020, pp. 1–4.
- [10] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, "Fine-grained recognition of abnormal behaviors for early detection of mild cognitive impairment," in *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2015.
- [11] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: clinical characterization and outcome," *Archives of neurology*, vol. 56, no. 3, pp. 303–308, 1999.
- [12] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—a survey," *Wiley Interdisciplinary Reviews: DMKD*, vol. 8, no. 4, p. e1254, 2018.
- [13] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.
- [14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [15] 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.
- [16] T. Plötz, N. Y. Hammerla, and P. L. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.
- [17] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, "Pd disease state assessment in naturalistic environments using deep learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.
- [18] S. Ramasamy Ramamurthy, I. Ghosh, A. Gangopadhyay, E. Galik, and N. Roy, "Star: A scalable self-taught learning framework for older adults' activity recognition," in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2021, pp. 121–128.
- [19] L. Zhang, X. Wu, and D. Luo, "Recognizing human activities from raw accelerometer data using deep neural networks," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2015, pp. 865–870.
- [20] F. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.
- [21] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.
- [22] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.
- [23] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in neural information processing systems*, 2006, pp. 1473–1480.
- [24] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: a deep quadruplet network for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [25] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Advances in neural information processing systems*, 2016.
- [26] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *arXiv preprint arXiv:2002.05709*, 2020.
- [27] C. I. Tang, I. Perez-Pozuelo, D. Spathis, and C. Mascolo, "Exploring contrastive learning in human activity recognition for healthcare," *arXiv preprint arXiv:2011.11542*, 2020.
- [28] B. Khaertdinov, E. Ghaleb, and S. Asteriadis, "Contrastive self-supervised learning for sensor-based human activity recognition," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2021, pp. 1–8.
- [29] —, "Deep triplet networks with attention for sensor-based human activity recognition," in *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2021, pp. 1–10.
- [30] M. G. Abdu-Aguye and W. Goma, "Robust human activity recognition based on deep metric learning," in *ICINCO (1)*, 2019, pp. 656–663.
- [31] T. Sheng and M. Huber, "Siamese networks for weakly supervised human activity recognition," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 4069–4075.
- [32] A. Saeed, T. Ozcebe, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019.
- [33] D. J. Cook and M. Schmitter-Edgecombe, "Fusing ambient and mobile sensor features into a behaviorome for predicting clinical health scores," *IEEE Access*, 2021.
- [34] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment using smart home monitoring of complex tasks," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 43, no. 6, pp. 1302–1313, 2013.
- [35] M. A. U. Alam, N. Roy, S. Holmes, A. Gangopadhyay, and E. Galik, "Automated functional and behavioral health assessment of older adults with dementia," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2016, pp. 140–149.
- [36] S. H. Tariq, N. Tumosa, J. T. Chibnall, M. H. Perry III, and J. E. Morley, "Comparison of the saint louis university mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—a pilot study," *The American journal of geriatric psychiatry*, vol. 14, no. 11, pp. 900–910, 2006.
- [37] D. Weintraub, K. A. Oehlberg, I. R. Katz, and M. B. Stern, "Test characteristics of the 15-item geriatric depression scale and hamilton depression rating scale in parkinson disease," *The American journal of geriatric psychiatry*, vol. 14, no. 2, pp. 169–175, 2006.
- [38] C. Graf, "The Lawton instrumental activities of daily living scale," *AJN The American Journal of Nursing*, vol. 108, no. 4, pp. 52–62, 2008.
- [39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [40] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *IMWUT*, vol. 1, no. 2, pp. 1–28, 2017.
- [41] N. Y. Hammerla, R. Kirkham, P. Andras, and T. Plötz, "On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution," in *Proceedings of the 2013 international symposium on wearable computers*, 2013, pp. 65–68.
- [42] H. Haresamudram, A. Beedu, V. Agrawal, P. L. Grady, I. Essa, J. Hoffman, and T. Plötz, "Masked reconstruction based self-supervision for human activity recognition," in *Proceedings of the 2020 International Symposium on Wearable Computers*, 2020, pp. 45–49.
- [43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.