

MUSIC-CAR: A Multi-label Complex Health-Related Activity Recognition Dataset from Smartphone Sensors

Kavin Chandrasekaran Luke Buquicchio
Data Science Program
Worcester Polytechnic Institute
Worcester, MA, USA
{kchandrasekaran, ljuquicchio}@wpi.edu

Emmanuel Agu
Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA, USA
emmanuel@wpi.edu

Elke Rundensteiner
Data Science Program
Worcester Polytechnic Institute
Worcester, MA, USA
rundenst@wpi.edu

Abstract—Automated recognition of complex activities (CA), such as making a sandwich while talking on the phone, offers crucial insights into daily routines for context-aware applications like patient monitoring. While machine learning models effectively recognize simple activities (SAs), identifying CAs composed of sequential, interleaved, or concurrent SAs remains challenging, hindered by a dearth of suitable public datasets. To address this gap, we introduce a comprehensive, publicly available dataset called MUSIC-CAR for Multilabel Sequential, Interleaved, and Concurrent Complex Activity Recognition. Collected from 47 participants using two smartphones, one wrist mounted to mimic a smartwatch and the other in a pants pocket, MUSIC-CAR provides IMU sensor data for 9 CA types (e.g., cooking). It uniquely features multi-level labels, containing coarse CA labels and fine-grained SA labels for 30 constituent SAs, capturing their complex relationships (sequential, concurrent, interleaved) within CAs. Our evaluation establishes baseline performance for Complex Activity Recognition (CAR), with XGBoost achieving the highest F1-score of 58.8%. This performance, significantly lower than typical SA recognition scores, demonstrates the inherent difficulty of multi-label CAR from IMU data. MUSIC-CAR represents a valuable resource for the ubiquitous computing and health communities, enabling researchers to develop and evaluate more sophisticated approaches for recognizing complex activities.

Index Terms—Human activity recognition, Complex activity recognition, Dataset, Machine Learning

I. INTRODUCTION

Motivation: The wide availability of sensor-equipped smartphones and Internet of Things (IoT) devices has sparked increased research interest in Human Activity Recognition (HAR) from sensor data [1]. HAR applications span diverse domains including health monitoring, elder care, security, patient care, and context-aware computing [2] [3] [4] [5] [6]. Human ambulatory activities can be classified into three types:

Simple activity (SA) Atomic unit-level, repetitive actions that cannot be further decomposed. Examples include walking and standing.

Activity transitions occur when a person changes from one SA to another. Examples include sitting to standing and sitting to lying down.

Complex activity (CA) consist of multiple SAs that may be performed sequentially, concurrently, or interleaved. Ex-

amples include talking on the phone while making a sandwich or packing a backpack.

Complex Activity Recognition (CAR) plays a vital role in passively assessing the health of a person. Changes in a patient’s daily activity patterns can indicate shifts in their health status, providing valuable insights into their physical and mental well-being. Ranasinghe et al. [7] describe the three distinct types of CAs consisting of SAs that can be sequential, concurrent, or interleaved. They are illustrated in Figure 1.

Sequential CAs consist of multiple SAs that follow one another in some order. For example, the CA of cooking involves multiple sequential steps such as going to the fridge, retrieving ingredients, going to the stove, getting utensils, cutting up the ingredients.

Interleaved CAs involves the interruption of SA a by another SA b . Once activity b is completed, the person resumes activity a . For example, while performing the CA of cooking, a person may stop cooking to talk on the phone, and then may resume cooking, once the phone call is completed. In fact, a CA may be interleaved with other CAs. For instance, a person may stop cooking, then go and pack a backpack (another CA) before returning to cooking.

Concurrent CAs involves multiple SAs that occur at the same time. For example, while performing the CA of cooking, a person may simultaneously talk on the phone.

Automated Complex Activity Recognition (CAR) using data from sensors such as the smartphone’s built-in accelerometer and gyroscope can help in assessing the daily activities of a person without the need for specialized hardware [8]–[10]. But prior work has demonstrated that machine learning HAR models that are designed for SAs perform relatively poorly on CAs [11], for example, they achieved over 90% accuracy for recognizing SAs versus 52% for CAs. Complex Activities are inherently harder to recognize from sensor data compared to simple activities. Challenges in CAR stem from the fact that there is a variability in the exact constituent SAs and their order in a given CA. For example, different people might

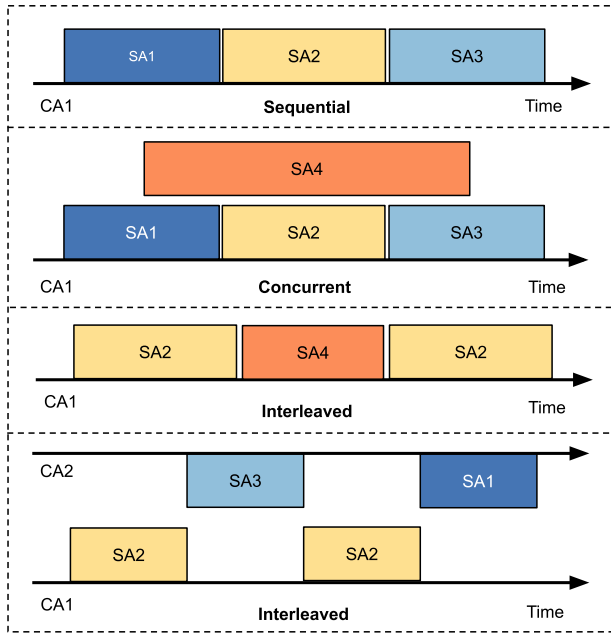


Fig. 1: Types of complex activities: sequential, interleaved and concurrent. The image shows the three types of Complex Activities (CAs), namely sequential, concurrent and interleaved. Each rectangular box with labels represents a simple activity in time along the x-axis. The CA is labeled as CA1 along the x-axis. The sequential CA shows 3 boxes labeled SA1, SA2, and SA3 that occur one after the other in a sequence. The concurrent CA has 4 boxes labeled SA1, SA2, SA3, and SA4, where the first three boxes labeled SA1, SA2, and SA3 are placed one after the other and the fourth box labeled SA4 is placed beside the boxes labeled SA1, SA2, SA3. The SA4 is partially over SA1, completely over SA2, and partially over SA3, indicating that the SA4 is performed simultaneously with part of SA1, all of SA2, and part of SA3. Two examples of interleaved complex activities are shown. The first interleaved CA has three boxes labeled SA2, SA4, and SA2, which occur sequentially, but the activity SA2 was interrupted by an activity SA4, which is not part of the CA. The second interleaved CA has two CAs CA1 and CA2, which are represented by two parallel lines along the x-axis. The line with label CA1 contains two boxes both labeled SA2 but separated by a gap. The line with label CA2 contains two boxes labeled SA3 and SA1 that are separated by a gap. The gap in each complex activities are aligned so that SA2 is followed by SA3, which is followed by SA2 and then by SA1. This pattern indicates that the two complex activities are performed in an interleaved manner altering between CA1 and CA2

make a sandwich in different ways. Some people might toast the bread before making the sandwich, while others might not use certain ingredients such as cheese and mayonnaise. In fact, the same person may make a sandwich differently at different times. The dish being cooked also affects the order of SAs performed for the CA of cooking. A person cooking a stir-fry

TABLE I: Rationale for selecting each Complex Activity in our MUSIC-CAR dataset

Complex Activity	Reason for Choosing the Activity
Sequential Complex Activities	
Bathroom	CA representative of health status
Bathroom with blowing nose	Simulates challenges of changing order in sequential CA
Bathroom with blowing nose first	Simulates challenges of changing order in sequential CA
Making a sandwich	Unique CA representative of the cooking category
Packing a backpack	Unique CA
Interleaved Complex Activities	
Cleanup interleaved with packing sandwich	Simulates challenges of interleaved CA where the interleaving activities are not in training data
Making a sandwich interleaved with packing a backpack	Simulates challenges of interleaved CA where the two interleaving CAs were in training data separately and during testing, the recognition model should recognize the data as belonging to two separate CAs
Making a sandwich interrupted by phone call	Simulates challenges of interleaved CA. Also provides the data for the activity of 'simulate phone call' with the help of simple activity label
Concurrent Complex Activities	
Bathroom with phone call	Simulates challenges of concurrent CA where the activities were observed separately in the training data but the concurrency is unobserved
Making a sandwich concurrent with phone call	Simulates scenario of concurrent CA where the activities were observed separately in the training data but the concurrency is unobserved

will also have a different pattern of activities than a person making a sandwich. Due to these temporal- as well as inter- and intra-subject variabilities, machine learning HAR models designed to recognize SAs perform poorly on CAR. Additional issues exist with recognizing concurrent and interleaved CAs. Specifically, the SAs that make up the complex activity can not only appear sequentially in order, but they can also be performed concurrently or interleaved with other activities. For instance, a person may choose to make a phone call while cooking a stir fry or stop cooking for some minutes to go and use the bathroom. Consequently, research into new models, which capture the inherent complexities of CAs and account for the afore-mentioned variability, is needed.

Prior work. While extensive research has focused on recognizing and monitoring SAs, much less work has addressed the inherent challenges in CAR from smartphone sensor data. This is due to the dearth of publicly available CA datasets. There are 17 SA HAR datasets, which contain data from inertial sensors or smartphones, while only 9 CA datasets exist. These 9 datasets also have important limitations, lacking fine-grained multi-level labels for both the CA and constituent SAs or lacking data for one or more of the three CA types (sequential, interleaved, and concurrent). Several existing datasets have advanced research on CAR from alternate modalities including video-based approaches [12], [13], [14], [15] and smart-home environments [16], [17] [18]. However, while valuable, these datasets fall outside our comparison scope as they monitor only locations where cameras are installed, or are recorded in spaces with specialized camera- or sensor-augmented en-

vironments. Moreover video monitoring can compromise user privacy. Instead, our work focuses on CAR from data captured from the built-in sensors of commodity smartphones that over 91% of the US population now own [19].

To facilitate and spur CAR research, we have collected a rich comprehensive CA dataset called MUSIC-CAR for Multilabel Sequential, Interleaved, and Concurrent Complex Activity Recognition that includes all three types of CAs. We have carefully designed specific realistic scenarios that highlight the challenges unique to each CA type. A key point to emphasize is that due to variability in the performance of CAs, the SAs performed in a given CA and their order may differ between the training and test sets. Since our dataset provides detailed labels for both CAs and SAs, it is highly customizable, allowing the simulation of new real-world scenarios beyond those we collected data on, by repeating SA data and/or masking data and/or labels. Table I summarizes the CAs in our dataset and provides rationale for selecting them. The CAs are a representative set including CAs that are common in daily life (cooking-related) and also some that are indicative of health status (bathroom-related activities). The cooking-related activities can also be important for assessing mood, affect, and quality of life [20].

The contributions of this paper include:

- 1) **MUSIC-CAR Dataset:** A rich publicly available dataset for Complex Activity Recognition (CAR) from smartphone sensors. It contains accelerometer and gyroscope data from 47 subjects performing realistic examples of sequential, interleaved, and concurrent CAs. The dataset features 9 unique Complex Activities (CAs) consisting of 30 unique Simple Activities (SAs), with approximately ≈ 54 minutes of sensor data per participant. The dataset consists of smartphone sensor data from two phones, one placed in the participant's wrist and the other in their pant pocket. The dataset contains fine-grained hierarchical labels capturing both CAs and their constituent SAs with key timing information.
- 2) **Open-source Tools and Software:** including a data collection application, a custom video annotation tool for labeling SAs within CAs, and scripts for extracting comprehensive statistical features from raw sensor data. These resources are designed to encourage further data collection and extension of this initial corpus, supporting reproducibility and advancement of the field.
- 3) **Statistical Analysis Framework:** for comprehensive statistical characterization and visualization of the dataset, including descriptive statistics across activity types, distributional analyses after systematic outlier removal, and feature importance quantification. All analysis code is made publicly available to ensure reproducibility and facilitate further exploration of the dataset's properties at https://wp.wpi.edu/carelab/music_car/.
- 4) **Performance Evaluation of Baseline Machine Learning Models:** on our CAR dataset, including Support Vector Machines (SVM), Random Forest, XGBoost, Convolutional LSTM, and Multilayer Perceptron. XGBoost

TABLE II: Simple Activity Datasets

Dataset	# of Participants	# of Activities	Sensors
HARTH [31]	22	12	2 Activity Accelerometers
CZU-MHAD [22]	5	22	Depth videos, Skeletal positions, and 10 Inertial sensors
RealWorld HAR [28]	15	8	Accelerometer, GPS, Gyroscope, Light, Magnetometer, Audio
UniMiB-SHAR [30]	30	17	Accelerometer in a Smartphone
HAR70+ [33]	18	8	2 Accelerometers
Wild-SHARD [34]	40	6	1 Smartphone
MHEALTH [27]	10	12	3 Shimmer2 Inertial sensors, and 2-lead ECG sensor
REALDISP [26]	17	33	9 Inertial sensors
HHAR [29]	9	6	8 Smartphones, 4 Smartwatches
UCI HAR [32]	30	7	Accelerometer and Gyroscope from a Smartphone
UTD-MHAD [23]	8	27	RGB, Depth, Skeletal, Inertial sensor
WEAR [24]	22	18	Video, Inertial sensor
MEx [25]	30	7	2 Accelerometers, Pressure Mat, and Depth camera

TABLE III: Datasets with Activity Transitions

Dataset	# of Participants	# of Activities	Sensors
C-MHAD [35]	12	5 Smart TV gestures, 7 Transitions	Video, Inertial sensor
UCI HAPT [36]	30	6 Simple activities and 6 Postural Transitions	Accelerometer and Gyroscope from a Smartphone
FORTH-TRACE [37]	15	7 Simple activities and 9 Postural Transitions	5 Shimmer Inertial sensors
CSL-SHARE [38]	20	20 Actions and 2 Postural Transitions	2 Accelerometers, 2 Gyroscopes, 4 surface electromyography (sEMG) sensors, 1 biaxial electrogoniometer, and 1 microphone integrated into a knee bandage

achieved the highest performance with 58.8% F1-Score that is significantly lower than the 98.9% reported for SA recognition in prior work [21] and demonstrating the challenging nature of CAR.

II. RELATED WORK

We now review publicly available datasets in the simple activities, transitions, and complex activities categories. There are several simple activity datasets, which include [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [31]. Table II summarizes these datasets. Only 4 datasets exist with labels for activity transitions [35] [36] [37] [38] Table III summarizes these datasets.

Novelty of our dataset: To the best of our knowledge, ours is the first publicly available complex human activity dataset collected from smartphone sensors that contains both simple and complex labels and includes scenarios for all three types of CAs. Fine-grained SA labels are essential for understanding

TABLE IV: Datasets with complex activity data but missing simple or complex labels

Dataset	# of Participants	# of Activities	Sensors
WISDM [8]	51	18	1 Smartphone and 1 Smartwatch
PAMAP2 [39]	9	18	3 Inertial measurement units, 1 Heart rate monitor
Ruzzon et al. [40]	10	9	6 Inertial measurement units
KU-HAR [9]	90	18	Accelerometer and Gyroscope from a Smartphone

TABLE V: Datasets with simple and complex activity labels, but lack one or more of the 3 types of complex activities

Dataset	# of Participants	# of Activities	Sensors
CMU-MMAC [41]	36	5	Video, Audio, Motion capture, IMUs, Wearable devices
OPPORTUNITY [43]	4	5	72 sensors of 10 modalities
Cooking Activity [42]	7	5	5 IMUs
Kitchen Dataset [10]	10	76 simple activities, 2 Complex activities	2 Smartwatches
Cooking Scenario [44]	4	10 simple activities, 3 complex activities	Motion capture, 2 Smartphones, 2 Smartwatches

concurrent and interleaved CAs, and for developing robust machine learning models as they precisely mark the beginning and end of concurrency or interleaving. Existing CA datasets fall into one of two categories, which differentiate them from our dataset.

Existing datasets containing CAs and Activities of Daily Living(ADL) but do not have fine-grained simple activity labels or are missing some sub-types of complex activities: including WISDM [8], PAMAP2 [39], Ruzzon et al’s dataset [40] and KU-HAR [9]. Table IV summarizes these datasets.

Complex activity datasets with both simple and complex labels, but only contain sequential CAs: including the CMU-MMAC Database [41], cooking dataset [42], Opportunity dataset [43] and kitchen dataset [10]. Table V summarizes these datasets.

While the above datasets contain labels for both SAs and CAs, they only focus on sequential CAs and do not contain concurrent or interleaved CAs. To facilitate new research in CAR, our dataset captures data for all three types of CAs along with fine-grained labels of the constituent SAs.

III. METHOD

Figure 2 shows the overview of the steps involved in processing the data and the final classification from that data using machine learning models. The diagram illustrates a complete Human Activity Recognition pipeline that progresses through seven key phases: data collection from pocket and wrist devices, temporal alignment of sensor streams with video footage, video annotation to establish ground truth labels, feature engineering across various window sizes, preparation of aligned datasets, and model development through algorithm selection and parameter optimization. This systematic

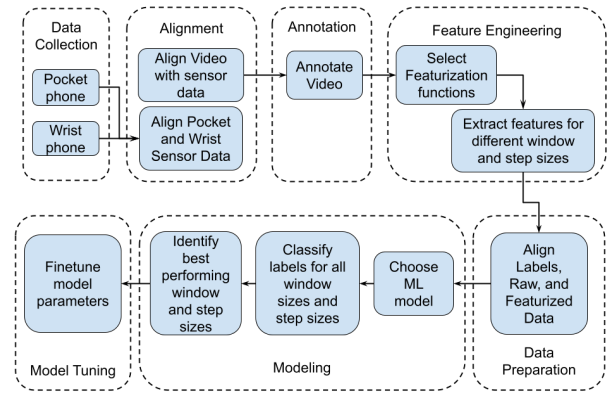


Fig. 2: Overview of data processing pipeline and machine learning pipeline. The diagram illustrates a comprehensive Human Activity Recognition (HAR) pipeline organized in seven sequential phases, depicted as dashed-outline boxes containing each containing one or more process blocks. Starting with the “Data Collection” phase in the top-left, the workflow captures data from two sources: a pocket phone and a wrist phone. This data flows into the “Alignment” phase where video and sensor data are synchronized, with specific processes for aligning pocket and wrist sensor data. Next, in the “Annotation” phase, the synchronized video is annotated to label human activities. The pipeline continues to the “Feature Engineering” phase where featurization functions are selected and features are extracted using different window and step sizes. The extracted features move downward to the “Data Preparation” phase which aligns labels with raw and featurized data. This prepared data then flows into the “Modeling” phase where machine learning models are chosen, labels are classified across various window and step sizes, and the best performing window and step size combination are identified. Finally, in the “Model Tuning” phase at the bottom-left, the model parameters are fine-tuned for optimal performance. Arrows throughout the diagram indicate the directional flow of information processing, showing how raw sensor data is progressively transformed into an optimized activity recognition model.

workflow transforms raw sensor data into an effective activity classifier by carefully synchronizing multi-modal inputs, extracting relevant features, and tuning the recognition model to identify human activities accurately.

A. Study Protocol

The study was conducted such that data was collected for one participant at a time. The entire study was video recorded to facilitate labeling. For privacy reasons the bathroom activities were not recorded by video. The phones were disinfected before each participant participated in the study. The study started with the protocol being explained to the participants including the protocol, risks and benefits of participation, followed by answering of participants’ questions. Thereafter the informed consent was signed. Participants were then given two android smartphones: one to keep in their pant pocket

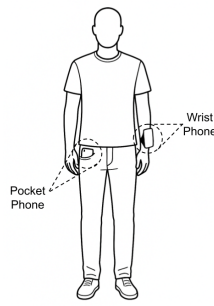


Fig. 3: Location of phones on the participants: Outline of a person with a phone in their right pant pocket and another phone strapped to their left wrist.

and the other to be strapped on their wrist (mimicking a smartwatch). The Wrist phone was attached to a phone mount with an armband that was used to strap the phone to the participant's wrist as shown in Figure 3. The participant was asked to wear the phone on the wrist on which they normally wore their watch. Phones were utilized for data collection from both the wrist and pant pocket phone placements so that the same app could be used for data collection.

During the study, the participants were instructed on what CA they should perform but were given autonomy to perform the activity as they normally would. For example, in the CA of 'Packing a backpack', five items were spread out throughout the room and the participant was asked to pack the backpack with those items. Participants chose different ways to close the backpack after packing it. Some closed just the drawstring, some closed the drawstring and the top flap, and some closed the flap and the clamps, but did not pull the drawstring. Similarly, in the CAs involving 'making a sandwich', the participants were instructed to make a sandwich, but the order in which they performed activities such as opening the fridge to get the ingredients, taking the bread, and getting the plate and knives were left to the participants' discretion. This was done to simulate the dynamic nature of the CAs. The activities performed by the participants are listed below:

- 1) **Packing a backpack** Five items were spread out throughout the room and the participants were asked to pack all the items into the backpack.
- 2) **Simulate bathroom usage** The participants were instructed to simulate urinating in the bathroom. No further instructions were given on the usage of urinal or stall to accommodate the participants' preference.
- 3) **Simulate bathroom usage and then blow nose twice** The participants were instructed to enter the bathroom and repeat the same process in the previous step and then blow their nose twice before washing their hands.
- 4) **Blow the nose twice and then simulate bathroom usage** The participants were instructed to enter the bathroom and blow their nose twice first before repeating the process in step 2.
- 5) **Simulate bathroom usage with a concurrent phone call**

The participants were instructed to pick up the phone in their pocket and simulate a phone call while repeating the process in step 2.

- 6) **Making a sandwich** The plate and knives were placed on a dish rack, the bread on the counter, and the rest of the ingredients inside the fridge. The participants were asked to make a turkey and mayo sandwich if the ingredients were available. If not, or if the participants did not want to handle meat, they were asked to make a peanut butter and jelly sandwich instead. If they were allergic to peanut butter, they were instructed to make a jelly sandwich. 7 participants prepared a turkey and mayo sandwich, 38 participants made a peanut butter and jelly sandwich, and 2 participants made a jelly sandwich.
- 7) **Clean up interleaved with packing the sandwich** The participants were asked to wash the knives first, using soap and a sponge. A separate sponge was given to participants who were allergic to peanut butter. Then the participant was asked to pack sandwich in a plastic wrap. Finally, they were asked to wash the plate they used and return all the items to their original locations to maintain consistency in item placement throughout the study.
- 8) **Making a sandwich while interrupted by a phone call** Participants were instructed to make the same sandwich as step 6. In the middle of making the sandwich, they were asked to stop making the sandwich, simulate talking on the phone and then go back to making the sandwich.
- 9) **Clean up interleaved with packing the sandwich** Participants were asked to perform the same cleanup and sandwich packing procedure as in step 7, ensuring all items were returned to their original locations in preparation for the next activity.
- 10) **Making a sandwich while concurrently simulating a phone call** Participants were instructed to start simulating a phone call and then continue making the same sandwich as in step 6.
- 11) **Clean up interleaved with packing the sandwich** Participants were asked to perform the same cleanup and sandwich packing procedure as in step 7, ensuring all items were returned to their original locations in preparation for the next activity.
- 12) **Making a sandwich interleaved with Packing a backpack** The participants were instructed to pack the backpack and make a sandwich, alternating between each activity. Each time participants switched activities, they would perform one step in the activity before switching activities. For example, They were instructed to take the plate and knives, then pick up one item to be put in the backpack and put it in the backpack. Take the bread out and then pick up one item to be put into the backpack and put it in the backpack. They alternated between activities until both the activities were completed.
- 13) **Cleanup interleaved with packing the sandwich** Participants were asked to perform the same cleanup and sandwich packing procedure as in step 7, ensuring all items were returned to their original locations to conclude

the study protocol.

To simplify the dataset description, when two CAs are interleaved (e.g., making a sandwich and packing a backpack), it is represented as a separate label resulting in a label ‘Making a sandwich interleaved with packing a backpack’. During evaluation, this is considered as a multi-class multi-label classification problem. When two activities are interleaved, it is appropriately represented by a positive label for both the interleaving CAs. Figure 4 shows the layout out of the room where the study was conducted and shows the location of the various objects used in the study.

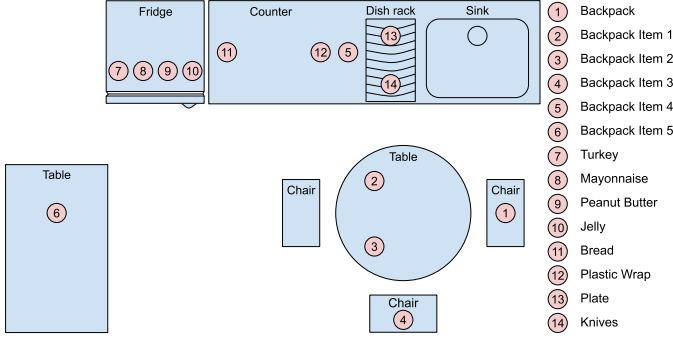


Fig. 4: Study Area Layout

B. Participants

There were a total of 47 participants which included undergraduate and graduate students between ages 18-37. The study risk was minimal and participants were offered to choose between \$30 in cash or a course credit for a class in the social sciences program. The demographic information was collected after the end of the study through an online form. Out of the 47 participants, 44 filled out the demographic information.

Figure 5 showing a participant performing some of the study activities. Figures 6a, 6d, 6c and 6b show the distribution of participants ages, genders, race and handedness respectively.

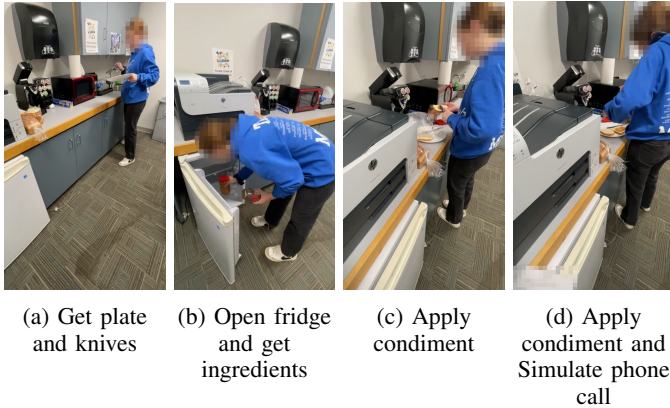


Fig. 5: Examples of a participant performing various activities in our study.

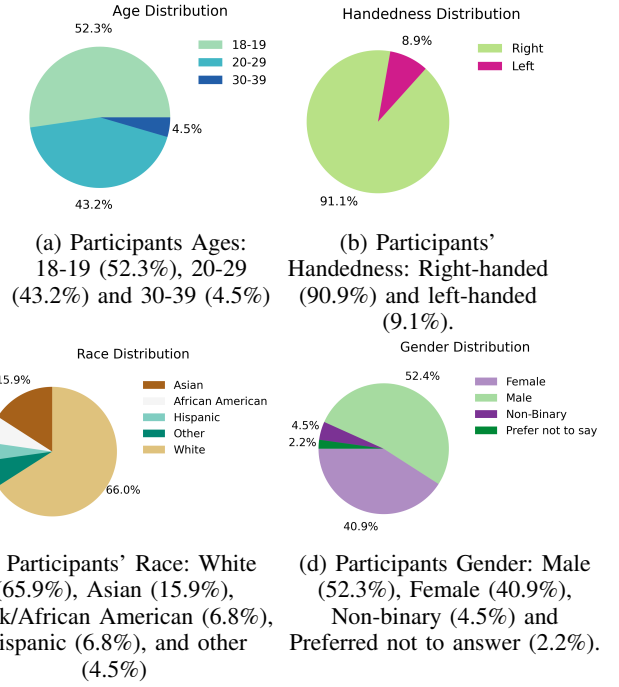


Fig. 6: Participants' age, handedness distribution, race, and gender distributions

C. Phone Sensors

All phones used for data collection were of the same model and specifications, Motorola G5 plus. The phones were equipped with accelerometer, gyroscope, magnetometer, light and proximity sensors.

Accelerometer A triaxial sensor that measures the linear acceleration along the X, Y and Z axes, measured in meters per second squared (m/s^2). The data collection sampling rate was set to 100Hz for this sensor.

Gyroscope A triaxial sensor that measures the angular velocity or rotation movement around the X,Y and Z axes, measured in units of radians per second. The sampling rate was set to 200Hz.

Magnetometer A sensor that measures the magnetic field around the phone that can be used to determine the orientation and position relative to the earth's magnetic field and is measured in units of Micro-Tesla (μT). The sampling rate was set at 50Hz.

Light An event-driven sensor that measures the amount of ambient light, which is recorded whenever there is a change in light values. It is measured in units of Lux.

Proximity An event-driven sensor that measures the proximity of an object to the phone, which is recorded whenever its value changes. It is measured in units of centimeters (cm).

D. Data annotation

A custom video annotation tool, shown in Figure 7, was developed to label the sensor data using the video that was recorded during the study. The tool was programmed in

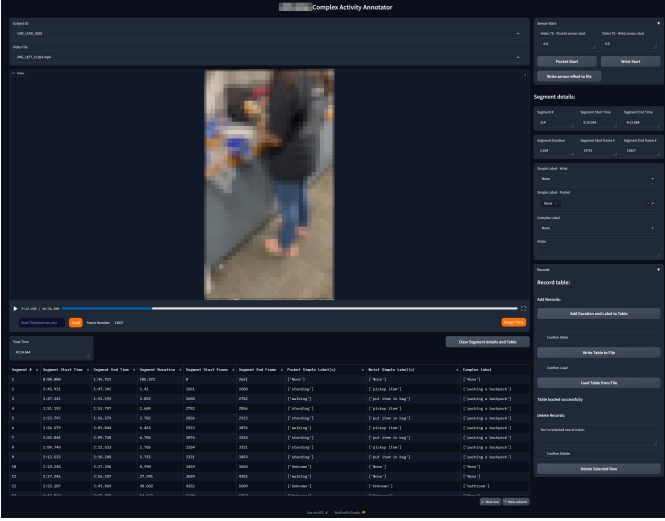


Fig. 7: Video Annotation Tool: Screenshot of the video annotation tool.

Python using the Gradio framework. We share the code for this annotation tool along with code for the data processing pipeline and evaluation models.



Fig. 8: Sample video timeline: timeline of the sequence of actions that occurred at the start of the study.

To begin the annotation process, the sensor data and the video were aligned by resampling the sensor data to 50Hz and converting the video to 25fps before the annotation. The video captured the moment the data collection on each phone was switched on. The figure 8 shows a sample sequence of events that occurred during data collection. The video was started at time vs and the phone data collection was started one at a time. The order in which data collection was started on the phones was arbitrary. The pocket phone data collection was started at time ps and the wrist phone data collection was started at time ws . The first timestamp from the polling sensors in the pocket phone is $pfts$, and the first timestamp from the polling sensors in the wrist phone is $wfts$. The time delta between the start of the data collection and first recorded sensor data are represented as pd and wd for pocket and wrist phones respectively. They are given by $pd = pfts - ps$ and $wd = wfts - ws$. In this dataset we assume that pd and wd are negligible and $ps \approx pfts$ and $ws \approx wfts$. The labeling process involved splitting the timeline into segments and recording the segment details and the corresponding simple and complex activity labels. The time duration between vs and ps , ws were also recorded as wof and pof .

TABLE VI: Simple and Complex Activities by primary sensor suite location.

Pocket Simple Activities	Wrist Simple Activities	
1) Putting phone back in pocket	1) Apply condiment	15) Open fridge and get ingredients
2) Simulate phone call	2) Close bag	16) Open fridge and put ingredients back
3) Standing	3) Close bread	17) Pickup item
4) Taking phone from pocket	4) Close condiment	18) Put item in bag
5) Walking	5) Close sandwich	19) Take bread
	6) Cut sandwich	20) Take plastic wrap
	7) Dry hands	21) Take turkey slices and place on bread
	8) Dry knife	22) Wash hands
	9) Dry plate with paper towel	23) Wash knife
	10) Get knife	24) Wash plate
	11) Get plate	25) Wrap sandwich
	12) Get plate and knife	
	13) Open bread	
	14) Open condiment	

E. Dataset Summary

For labeling purposes, the dataset is split into two parts: wrist dataset and pocket dataset. When considering the SA label classification, each dataset has different classification problems. The wrist dataset is a multi-class uni-label dataset and the pocket dataset is a multi-class multi-label dataset. Table VI contains the lists of SA labels available in the dataset. The SAs are split into those available in the pocket dataset and those in the wrist dataset. Bathroom-related CAs possess only the complex label, whereas all other CAs include both the complex label and the constituent SA labels.

IV. VALIDATION

The dataset was validated by classifying the data using several traditional machine learning classification models including SVM with a Radial Basis Function (RBF) kernel function, Random Forest, XGBoost, and deep learning models such as Multilayer Perceptron (MLP) and DeepConvLSTM [45]. MLPs have performed well on HAR tasks [46] and are effective baselines for modeling non-linear relationships within static feature vectors. DeepConvLSTM models, integrating convolutional and recurrent architectures, are adept at automatically learning hierarchical spatial features and concurrently modeling temporal dependencies from sequential sensor data, which is crucial for representing the inherently dynamic characteristics of human activities.

We utilized an MLP architecture with 4 hidden layers each containing units of size 256, 128, 64, 32 respectively. The final layer is a sigmoid layer with 9 outputs. A dropout layer was used for regularization. The model used Adam as the optimizer with weight decay. The optimizer used a binary cross-entropy loss function with custom weights for each class. The class weights were calculated using normalized inverse class frequency. $f_i = \frac{n_i}{N}$ where f_i is the frequency of class i , n_i is the number of samples in class i , and N is the total number of samples. $w_i^{\text{norm}} = \frac{w_i}{\sum_{j=1}^C w_j} \times C$ where w_i^{norm} is the normalized weight for class i , and C is the total no. of classes.

We utilized a DeepConvLSTM [45] with 1 CNN layer followed by 1 LSTM layer and a fully connected sigmoid layer. Adam was used as the optimizer with weight decay. A binary cross-entropy loss function was used with custom weights for each class.

Before performing classification, several features were extracted from the raw sensor data. Two sets of features were utilized. The first set, referred to as (ExtraSensory or ES Features) were generated based on the features from [46]. The second feature set, referred to as (Transition Aware Human Activity Recognition or TAHAR Features) were those extracted in prior work by Anguita et al. [47]. For simplicity, the validation results shown here are only based on the ES Features.

A. Processing Pipeline

Raw sensor data were collected at different frequencies: Accelerometer at 100Hz, Gyroscope at 200Hz, and Magnetometer at 50Hz. All sensor data were resampled to 50Hz, the lowest sampling rate among all the sensors. Based on prior work [48] [49], a sampling rate ≥ 20 Hz is sufficient for HAR from accelerometer sensor. This value is based on the Nyquist-Shannon sampling theorem [50], which states that for lossless reconstruction, the sampling rate has to be higher than twice the highest frequency. $T < \frac{1}{2f_{\max}}$, which is also represented as $f_s > 2f_{\max}$ where T = Sampling period, f_s = Sampling frequency and $f_{\max}$ = Maximum frequency component. As human activities are performed at a frequency less than 10Hz, a sampling rate of 50Hz is adequate for recognizing human activities. From the resampled sensor data, various features are extracted based on the selected feature set. The feature extractor takes window size and step size as input parameters. To determine optimal values, we explored the performance achieved using various combinations of window size and step size.

1) *ES Features*: include statistical features, spectral features and auto-correlation features. Statistical features include, mean, standard deviation, momentum, quartiles, and entropy. Spectral features include the log energies of various sub-bands and their entropy. Table VII shows the equations for calculating the features.

2) *TAHAR Features*: Accelerometer signals are a combination of the acceleration by the body and the gravitational force. As shown in previous work [47], gravitational force affects the lower frequencies of the signals and can be removed. The windowed raw data samples are filtered to separate the body signals from the influence of gravity. The jerk and magnitude of the filtered signals are calculated. The jerk denotes the rate of change in acceleration. The sudden changes in acceleration can indicate a beginning and end of transitions between different activities. The provided time-domain signals are transformed into frequency domain signals using Fast Fourier Transform. Various transformations are processed using the functions in Table VIII to create features. The TAHAR features uses the same featurization functions as ES features, but they differ in the pre-processing steps which includes filtering,

TABLE VII: Feature formulas used by both ES Features and TAHAR Features generated from our featurization code

Feature	Formulation
Mean	$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$
Standard deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2}$
Median Absolute Deviation	$\text{median}_i (s_i - \text{median}_j (s_j))$
Largest values in array	$\max_i (s_i)$
Smallest value in array	$\min_i (s_i)$
Signal magnitude area	$\frac{1}{3} \sum_{i=1}^3 \sum_{j=1}^N s_{i,j} $
Signal Entropy	$\sum_{i=1}^N (c_i \log(c_i)), c_i = s_i / \sum_{i=1}^N s_j$
Interquartile range	$Q3(s) - Q1(s)$
4th order Burg Autoregression coefficients	$a = \text{arburg}(s, 4), a \in \mathbb{R}^4$
Pearson Correlation coefficient	$C_{1,2} / \sqrt{C_{1,1} C_{2,2}}, C = \text{cov}(s_1, s_2)$
Frequency signal weighted average	$\sum_{i=1}^N (i s_i) / \sum_{j=1}^N s_j$
Spectral energy of a frequency band [a, b]	$\frac{1}{a-b+1} \sum_{i=a}^b s_i^2$
Frequency signal Skewness	$E \left[\frac{(s-\bar{s})^3}{\sigma^3} \right]$
Frequency signal Kurtosis	$E \left[\frac{(s-\bar{s})^4}{\sigma^4} \right] / E \left[\frac{(s-\bar{s})^2}{\sigma^2} \right]^2$
Largest frequency component	$\arg \max_i (s_i)$

s: signal vector, N: signal vector length Q: quartile

separating body and gravitational signals, and calculating jerk and magnitude of signals.

TABLE VIII: TAHAR Feature generation functions

Functions	Features											
	Time Domain						Frequency Domain					
	BodyAcc-XYZ	GravityAcc-XYZ	BodyAccJerk-XYZ	BodyGyro-XYZ	BodyGyroJerk-XYZ	BodyAccMag	GravityAccMag	BodyAccJerkMag	BodyGyroMag	BodyGyroJerkMag	BodyAcc-XYZ	BodyAccJerk-XYZ
											BodyGyro-XYZ	BodyAccMag
											BodyGyroJerkMag	BodyAccJerkMag
											BodyGyroMag	BodyGyroJerkMag
Mean	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard Deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Median Absolute Deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Max	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Energy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Signal Magnitude Area	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Entropy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IQR	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Autoregression Coefficients	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pearson Correlation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Frequency	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Energy Band	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Skewness	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kurtosis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MaxFreqInd	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

B. Results

Optimal values for the combination of window and step size were based on the values that yielded the best performance for the baseline machine learning models. Since this is a multi-class multi-label classification problem, we use the samples average of f1-score calculated as shown in equation 4.

$$\text{Exact Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Contain Accuracy} = \frac{1}{N} \sum_{i=1}^N 1(\hat{y}_i \supseteq y_i). \quad (2)$$

where N = The number of samples, $\hat{y}_i$ = The set of predicted labels for sample i , y_i = The set of true labels for sample i and $\supseteq$ = The “contains” operator.

$$\text{Hamming Accuracy} = 1 - \left(\frac{1}{N \cdot L} \sum_{i=1}^N \sum_{j=1}^L \mathbb{I}(y_{ij} \neq \hat{y}_{ij}) \right) \quad (3)$$

where N = The number of samples, L = The number of labels, y_{ij} = The true label value for sample i and label j and $\hat{y}_{ij}$ = The predicted label value for sample i and label j . Additionally, we utilize well known classification metrics including Precision, Recall and F1-score. We also utilize

$$\text{F1-Score Samples Average} = \frac{1}{N} \sum_{i=1}^N F1 - score_i \quad (4)$$

where N = The number of samples in the dataset.

Determining the optimal window and step size using XGBoost, it was observed that window size of 60 seconds and step size of 5 seconds yields the best performance. Figure 9 shows the box plots for f1-score across the 5-folds during cross-validation.

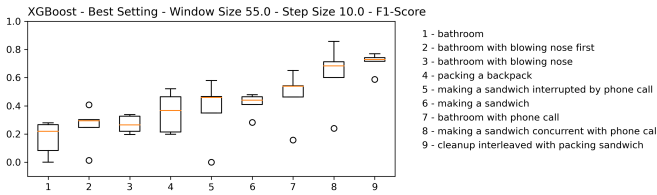
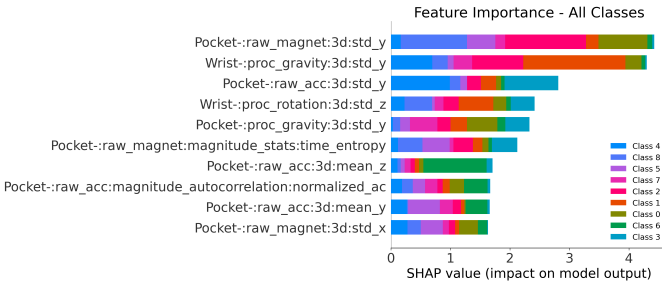


Fig. 9: F1 score for XGBoost using the optimal combination of window and step size, and settings

Table IX shows the results from all the baseline models for the setting window size of 60 seconds and step size of 5 seconds. The results show that the best performing model is XGBoost. While using raw data instead of the featurized data might affect the performance of deep learning models such as DeepConvLSTM, it is beyond the scope of this article.



(a) Feature importance across all classes

Fig. 10: SHAP-based Feature Importance from XGBoost Class 0-Bathroom, Class 1-Bathroom with blowing nose, Class 2-Bathroom with blowing nose first, Class 3-Bathroom with phone call, Class 4-Cleanup interleaved with packing sandwich, Class 5-Making a sandwich, Class 6-Making a sandwich concurrent with phone call, Class 7-Making a sandwich interrupted by phone call, Class 8-Packing a backpack

Figure 10 shows the top 10 important features based on the SHAP values, which represents the impact of each feature on the model's output. The plots are based on the best performing baseline XGBoost model. From the summary plot figure 10a, we can see that the top three important features

are standard deviation of y-axis from raw magnetometer in the pocket phone, gravity values of the wrist phone, and the raw accelerometer values of the pocket phone.

Recognition of sequential complex activities was particularly challenging for the baseline models evaluated, evidenced by low F1-scores across this category (Table X). Activities involving bathroom, especially those involving blowing nose, were difficult for most models to distinguish reliably, yielding F1-scores below 0.37. This scenario is one of the motivating reasons for creating this dataset as sequential CAs with only the order of the SAs switched leads to poor performance, thus elucidating the need for specialized CAR models.

Performance of recognizing interleaved activities generally surpassed that observed for sequential tasks. The 'Cleanup interleaved with packing sandwich' activity was classified with notably high accuracy, achieving a top F1-score of 0.826 with XGBoost, and strong performance also from Random Forest (F1 $\approx$ 0.760). It is worth noting this activity was also the most frequent CA label in the dataset, which may contribute to the higher recognition rates due to more training examples. Recognizing 'Making a sandwich interrupted by phone call' was comparatively more difficult, indicating that the nature and duration of the interruption still presents challenges to machine learning classifiers.

The models recognized concurrent activities relatively well, specifically tasks performed simultaneously with making a phone call. XGBoost again yielded the best results for both 'Bathroom with phone call' and 'Making a sandwich concurrent with phone call'. Random Forest was consistently the second-best performer for these concurrent tasks. This pattern suggests that tree-based ensemble methods (XGBoost and RF) are effective at handling the signal super-imposition generated by distinct simultaneous activities.

V. DISCUSSION

Dataset Existing publicly available datasets for complex human activity recognition often lack fine multi-level labels or comprehensive coverage of the three main types of CAs (sequential, interleaved, and concurrent), a limitation our dataset addresses. It includes sensor data for four CAs related to 'Bathroom' scenarios, which are labeled only at the CA level (for privacy reasons), and five activities with both complex and simple activity labels. This design creates a unique opportunity for approaches such as transfer learning; the available SA labels can be used to improve the recognition of the 'Bathroom' activities. For instance, SA labels exist for 'Simulate phone call' and 'Wash hands,' which are components of the 'Bathroom' activities but currently lack labels within those contexts. Further, these SA labels enhance the dataset's versatility, allowing for a flexible representation of CAs and opens up avenues for follow-on research.

Performance evaluation of baseline machine learning models revealed that XGBoost achieved the highest average F1-score (58.4%). Comparing the performance of XGBoost and Random Forest across individual CAs, we found that 'bathroom with phone call,' 'making a sandwich concurrent

TABLE IX: Baseline Results - F1-Score, Precision, Recall

Highest metric performance is in **Bold** and the second highest is underlined

Model	Exact Accuracy	Contain Accuracy	Hamming Accuracy	F1-Score	Precision	Recall
SVM	0.022±0.01	0.651±0.13	0.667±0.06	0.342±0.04	0.244±0.03	0.664±0.12
Random Forest	0.429±0.06	0.436±0.06	<u>0.922±0.01</u>	0.439±0.07	0.441±0.07	0.440±0.07
XGBoost	0.555±0.05	<u>0.577±0.06</u>	0.935±0.01	0.584±0.06	0.588±0.05	<u>0.588±0.06</u>
MLP	<u>0.451±0.11</u>	0.496±0.10	0.895±0.02	<u>0.494±0.10</u>	0.493±0.10	0.506±0.10
DeepConvLSTM	0.372±0.06	0.381±0.06	0.878±0.01	0.379±0.06	0.377±0.06	0.381±0.06

TABLE X: F1-score of baseline models per activity

Highest metric performance is in **Bold** and the second highest is underlined

Complex Activity	SVM	RF	XGBoost	MLP	DeepConvLSTM
Sequential Complex Activities					
Bathroom	0.178±0.01	0.149±0.08	0.368±0.07	<u>0.206±0.11</u>	0.176±0.15
Bathroom with blowing nose	0.215±0.01	0.011±0.01	<u>0.263±0.06</u>	0.266±0.09	0.245±0.18
Bathroom with blowing nose first	0.216±0.03	0.006±0.01	<u>0.229±0.03</u>	0.221±0.07	0.365±0.06
Making a sandwich	0.327±0.02	0.269±0.09	<u>0.503±0.12</u>	0.509±0.05	0.094±0.06
Packing a backpack	0.265±0.01	0.370±0.09	0.685±0.06	<u>0.433±0.08</u>	0.080±0.03
Interleaved Complex Activities					
Cleanup interleaved with packing sandwich	0.579±0.02	<u>0.760±0.04</u>	0.826±0.03	0.667±0.06	0.650±0.05
Making a sandwich interrupted by phone call	0.204±0.01	0.369±0.04	0.576±0.02	<u>0.481±0.09</u>	0.183±0.09
Concurrent Complex Activities					
Bathroom with phone call	0.303±0.05	<u>0.685±0.08</u>	0.748±0.06	0.578±0.06	0.259±0.08
Making a sandwich concurrent with phone call	0.427±0.05	<u>0.710±0.03</u>	0.760±0.01	0.481±0.09	0.303±0.20

with phone call,’ and ‘cleanup interleaved with packing sandwich’ were the easiest to classify. Further investigation showed that the inclusion of accelerometer data from a pocket phone improved classification accuracy for activities involving phone calls, likely contributing to their higher performance. The relatively high number of instances for ‘Cleanup interleaved with packing sandwich’ also likely facilitated its accurate classification. Conversely, ‘bathroom with blowing nose,’ ‘bathroom with blowing nose first,’ and ‘bathroom’ exhibited the lowest performance, potentially due to their inherent similarity and subtle variations in the sequence of constituent SAs. These results suggest that traditional machine learning models may struggle to capture the nuances of CAR, highlighting the need for custom models.

Baseline models struggled to recognize complex sequential activities, evidenced by low F1-scores, whereas interleaved and concurrent activities were generally classified more reliably. This may be partly attributed to the distinct signal pattern associated with ‘Simulate phone call’, making activities involving it potentially easier to identify. XGBoost consistently demonstrated the strongest performance across all activity types. Notably, tree-based ensembles (XGBoost and Random Forest) excelled at handling the signal superimposition present in concurrent and interleaved tasks, significantly outperforming MLP, SVM, and particularly the DeepConvLSTM model in most scenarios evaluated.

Limitations of our dataset The most significant is one inconsistent annotation detail. Namely, CAs that involve ‘sandwich’ such as ‘Making a sandwich’ include fine-grained SA labels, while ‘bathroom’ CAs do not. This lack of detail for bathroom CAs arose due to privacy concerns with recording bathroom use. The study was conducted in an uncontrolled bathroom environment where video recording was infeasible. Class imbalance is another key issue. Some CAs inherently take longer than others (e.g., ‘Making a Sandwich’ versus the shorter ‘Packing a Backpack’), leading to unequal durations.

Our study protocol, designed to minimize proctor interference, involved environmental resets after each sandwich task, causing the ‘Cleanup interleaved with Making a Sandwich’ CA to be recorded about four times more frequently than others. Finally, the time-consuming nature of video annotation restricted data collection to a small number of participants. To help overcome these data acquisition challenges for future work, we are sharing the tools used for collection, annotation, and analysis, hoping to foster community contributions.

VI. CONCLUSION

To address the lack of comprehensive complex activity smartphone datasets, this paper presents a new dataset for complex human activity recognition from smartphone sensor data. This dataset uniquely includes labels for both complex activities and their constituent simple activities, enabling a more granular analysis of human behavior. Our exploratory data analysis and detailed description of the data collection, labeling, and processing pipeline provide a solid foundation for future research. Preliminary validation results demonstrate the dataset’s potential for successful CAR, and we anticipate it will serve as a valuable resource for the research community.

ACKNOWLEDGMENT

This material is based on research sponsored by DARPA under agreement number FA8750-18-2-0077. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

DATA AVAILABILITY

More information on the dataset and the code can be found at https://carelab.wpi.edu/music_car/

REFERENCES

- [1] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, "Sensor-based datasets for human activity recognition—a systematic review of literature," *IEEE Access*, vol. 6, pp. 59 192–59 210, 2018.
- [2] G. Ogbuabor and R. La, "Human activity recognition for healthcare using smartphones," *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, 2018.
- [3] A. Chelli and M. Pätzold, "A machine learning approach for fall detection and daily living activity recognition," *IEEE Access*, vol. 7, pp. 38 670–38 687, 2019.
- [4] Z. Sun, S. Tang, H. Huang, Z. Zhu, H. Guo, Y.-E. Sun, and L. Huang, "Sos: Real-time and accurate physical assault detection using smartphone," *Peer-to-Peer Networking and Applications*, vol. 10, pp. 395 – 410, 2016.
- [5] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients," *PLoS ONE*, vol. 10, 2015.
- [6] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, "Gchar: An efficient group-based context—aware human activity recognition on smartphone," *Journal of Parallel and Distributed Computing*, vol. 118, pp. 67–80, 2018.
- [7] S. Ranasinghe, F. Al Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *International Journal of Distributed Sensor Networks*, vol. 12, no. 8, p. 1550147716665520, 2016.
- [8] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.
- [9] N. Sikder and A.-A. Nahid, "Ku-har: An open dataset for heterogeneous human activity recognition," *Pattern Recognition Letters*, vol. 146, pp. 46–54, 2021.
- [10] Y. F. O. Mohammad, K. Matsumoto, and K. Hoashi, "A dataset for activity recognition in an unmodified kitchen using smart-watch accelerometers," *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, 2017.
- [11] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in *2012 eighth international conference on intelligent environments*. IEEE, 2012, pp. 214–221.
- [12] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: a large video database for human motion recognition," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011.
- [13] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.
- [14] C. Schuld, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, 2004, pp. 32–36 Vol.3.
- [15] B. G. Fabian Caba Heilbron, Victor Escorcia and J. C. Nibbles, "Activity-net: A large-scale video benchmark for human activity understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 961–970.
- [16] D. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intelligent Systems*, vol. 27, no. 1, pp. 32–38, 2012.
- [17] R. G. Ramos, J. D. Domingo, E. Zalama, J. Gómez-García-Bermejo, and J. López, "Sdhar-home: A sensor dataset for human activity recognition at home," *Sensors*, vol. 22, no. 21, p. 8109, 2022.
- [18] F. Alves, M. Gairing, F. A. Oliehoek, and T.-T. Do, "Sensor data for human activity recognition: Feature representation and benchmarking," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8.
- [19] Pew Research Center, "Mobile fact sheet," Pew Research Center, Washington, D.C., Tech. Rep., Nov. 2024. [Online]. Available: <https://www.pewresearch.org/internet/fact-sheet/mobile/>
- [20] N. Farmer, K. Touchton-Leonard, and A. Ross, "Psychosocial benefits of cooking interventions: a systematic review," *Health Education & Behavior*, vol. 45, no. 2, pp. 167–180, 2018.
- [21] C. Alagoz, "Comparative analysis of xgboost and minirocket algorithms for human activity recognition," 2024. [Online]. Available: <https://arxiv.org/abs/2402.18296>
- [22] X. Chao, Z. Hou, and Y. Mo, "Czu-mhad: a multimodal dataset for human action recognition utilizing a depth camera and 10 wearable inertial sensors," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7034–7042, 2022.
- [23] C. Chen, R. Jafari, and N. Kehtarnavaz, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *2015 IEEE International conference on image processing (ICIP)*. IEEE, 2015, pp. 168–172.
- [24] M. Bock, H. Kuehne, K. Van Laerhoven, and M. Moeller, "Wear: An outdoor sports dataset for wearable and egocentric activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 4, pp. 1–21, 2024.
- [25] A. Wijekoon, N. Wiratunga, and K. Cooper, "Mex: Multi-modal exercises dataset for human activity recognition," *arXiv preprint arXiv:1908.08992*, 2019.
- [26] O. Baños, M. Damas, H. Pomares, I. Rojas, M. A. Tóth, and O. Amft, "A benchmark dataset to evaluate sensor displacement in activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 1026–1035.
- [27] O. Banos, C. Villalonga, R. Garcia, A. Saez, M. Damas, J. A. Holgado-Terriza, S. Lee, H. Pomares, and I. Rojas, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *Biomedical engineering online*, vol. 14, pp. 1–20, 2015.
- [28] T. Szttyler and H. Stuckenschmidt, "On-body localization of wearable devices: An investigation of position-aware activity recognition," in *2016 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2016, pp. 1–9.
- [29] H. Blunck, S. Bhattacharya, T. Prentow, M. Kjrgaard, and A. Dey, "Heterogeneity activity recognition," *UCI Machine Learning Repository*, vol. 10, p. C5689X, 2015.
- [30] D. Micucci, M. Mobilio, and P. Napolitano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones," *Applied Sciences*, vol. 7, no. 10, p. 1101, 2017.
- [31] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "Harth: A human activity recognition dataset for machine learning," *Sensors*, vol. 21, no. 23, 2021.
- [32] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *The European Symposium on Artificial Neural Networks*, 2013.
- [33] A. Ustad, A. Logacjov, S. Ø. Trollebø, P. Thingstad, B. Vereijken, K. Bach, and N. S. Maroni, "Validation of an activity type recognition model classifying daily physical behavior in older adults: the har70+ model," *Sensors*, vol. 23, no. 5, p. 2368, 2023.
- [34] N. A. Choudhury and B. Soni, "An adaptive batch size-based-cnn-lstm framework for human activity recognition in uncontrolled environment," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 10, pp. 10 379–10 387, 2023.
- [35] H. Wei, P. Chopada, and N. Kehtarnavaz, "C-mhad: Continuous multimodal human action dataset of simultaneous video and inertial sensing," *Sensors*, vol. 20, no. 10, p. 2905, 2020.
- [36] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.
- [37] K. Karagiannaki, A. Panousopoulou, and P. Tsakalides, "The forth-trace dataset for human activity recognition of simple activities and postural transitions using a body area network," 07 2016.
- [38] H. Liu, Y. Hartmann, and T. Schultz, "Csl-share: A multimodal wearable sensor-based human activity dataset," *Frontiers in Computer Science*, vol. 3, 2021.
- [39] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," *2012 16th International Symposium on Wearable Computers*, pp. 108–109, 2012.
- [40] M. Ruzzon, A. Carfi, T. Ishikawa, F. Mastrogiovanni, and T. Murakami, "A multi-sensory dataset for the activities of daily living," *Data in Brief*, vol. 32, p. 106122, 2020.
- [41] F. De la Torre, J. K. Hodgins, J. Montano, and S. Valcarcel, "Detailed human data acquisition of kitchen activities: the emu-multimodal activity database (emu-mmact)," in *Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, in conjunction with CHI*, vol. 2009, 2009.
- [42] F. Krüger, M. Nyolt, K. Yordanova, A. Hein, and T. Kirste, "Computational state space models for activity and intention recognition. a feasibility study," *PLoS ONE*, vol. 9, 2014.

- [43] D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. del R. Millán, "Collecting complex activity datasets in highly rich networked sensor environments," *2010 Seventh International Conference on Networked Sensing Systems (INSS)*, pp. 233–240, 2010.
- [44] P. Lago, S. Takeda, S. S. Alia, K. Adachi, B. Bennai, F. Charpillet, and S. Inoue, "A dataset for complex activity recognition with-micro and macro activities in a cooking scenario," *arXiv preprint arXiv:2006.10681*, 2020.
- [45] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, 2016. [Online]. Available: <https://www.mdpi.com/1424-8220/16/1/115>
- [46] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE pervasive computing*, vol. 16, no. 4, pp. 62–74, 2017.
- [47] D. Anguita, A. Ghio, L. Oneto, F. X. Llanas Parra, and J. L. Reyes Ortiz, "Energy efficient smartphone-based activity recognition using fixed-point arithmetic," *Journal of universal computer science*, vol. 19, no. 9, pp. 1295–1314, 2013.
- [48] H. Junker, P. Lukowicz, and G. Troster, "Sampling frequency, signal resolution and the accuracy of wearable context recognition systems," in *Eighth International Symposium on Wearable Computers*, vol. 1. IEEE, 2004, pp. 176–177.
- [49] X. Yang, A. Dinh, and L. Chen, "Implementation of a wearable real-time system for physical activity recognition based on naive bayes classifier," in *2010 International Conference on Bioinformatics and Biomedical Technology*, 2010, pp. 101–105.
- [50] A. Jerri, "The shannon sampling theorem—its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, 1977.