GeSmart: A Gestural Activity Recognition Model for Predicting Behavioral Health

Mohammad Arif Ul Alam
Department of Information Systems
University of Maryland Baltimore County
Baltimore, Maryland 21250
Email: alam4@umbc.edu

Nirmalya Roy
Department of Information Systems
University of Maryland Baltimore County
Baltimore, Maryland 21250
Email: nroy@umbc.edu

***Abstract*—To promote independent living for elderly population activity recognition based approaches have been investigated deeply to infer the activities of daily living (ADLs) and instrumental activities of daily living (I-ADLs). Deriving and integrating the gestural activities (such as talking, coughing, and deglutition etc.) along with activity recognition approaches can not only help identify the daily activities or social interaction of the older adults but also provide unique insights into their long-term health care, wellness management and ambulatory conditions. Gestural activities (GAs), in general, help identify fine-grained physiological symptoms and chronic psychological conditions which are not directly observable from traditional activities of daily living. In this paper, we propose GeSmart, an energy efficient wearable smart earring based GA recognition model for detecting a combination of speech and non-speech events. To capture the GAs we propose to use only the accelerometer sensor inside our smart earring due to its energy efficient operations and ubiquitous presence in everyday wearable devices. We present initial results and insights based on a C4.5 classification algorithm to infer the infrequent GAs. Subsequently, we propose a novel change-point detection based hybrid classification method exploiting the emerging patterns in a variety of GAs to detect and infer infrequent GAs. Experimental results based on real data traces collected from 10 users demonstrate that this approach improves the accuracy of GAs classification by over 23%, compared to previously proposed pure classification-based solutions. We also note that the accelerometer sensor based earrings are surprisingly informative and energy efficient (by 2.3 times) for identifying different types of GAs.**

Keywords*—*smart jewelry, behavioral health, change-point detection, energy efficiency, cognitive computing.

I. Introduction

Modeling and analyzing the physiological symptoms and psychological behaviors of older adults have a profound impact on future smart and connected elder care. The fine grained insights about the human health, wellness and independence obtained from the physiological and psychological data analysis if coupled with activities of daily living can help improve the mental health, stress disorders, ambulatory conditions, and social interactions of older adults. The wide availability of commodity smart healthcare appliances, stand-alone and integrated sensing devices make it increasingly easy to ubiquitously and continuously monitor an individual's health-related vital signals, activities, and behavior and to integrate such data into healthcare systems. We are witnessing early commercial activity, where a combination of *body-worn* medical and non-medical sensors (e.g., sensors to monitor blood oxygenation or accelerometers to monitor movement) and *in situ* sensors (e.g., thermal and motion detectors) continuously monitor and automatically determine an individual's context. Broadly speaking, *context* in smart health refers to a variety of dynamically changing states, related to either an individual's activities (e.g., ambulatory vs. sleeping), biomedical conditions (e.g., fatigue vs. anxiety), or behavioral conditions (e.g., shouting vs. agitation). In many health and wellness applications, such context enables critical capabilities, such as alerting a first responder if the individual is shouting for an abnormal period of time or flagging a health risk by analyzing wellness data related to continuous burping or hiccup after every day eating. In this paper, we particularly investigate the recognition and discovery of gestural activities (henceforth defined as GAs) which are observable and detected; provide significant insights about the long-term wellbeing of the elderly people. Our approach enables efficient abstraction and finer correlation of the activities of daily living with the acute physiological symptoms and chronic psychological conditions.

Providing both behavioral and physical health status in an unified setting is of utmost need for proactive healthcare. Mental disorders and cognitive impairments oftentimes evolve from chronic physiological symptoms and abnormal psychological behaviors. Suffering from different sort of mood disorders inhibit different patterns of infrequent gestures such as depression, sadness, crying, shouting etc. Likewise for different kinds of physiological health issues, the patient shows irregular gestures such as frequent coughing, burping, breathing problem etc. Therefore mental and physical health of elderly people are correlated and if harnessed appropriately may provide meaningful microscopic physiological and psychological contexts. For example, a person feeling a headache from anxiety or anger might shout loudly or show irregular interpersonal traits. Thus the mental hygiene or physical wellness of a particular person can be inferred by monitoring the GAs which reflect the emotional or behavioral state of the individual. On the other hand when a person shows infrequent gestures while being engaged in other activities, his or her body produces different kind of movements. The differences between these subtle movements, if captured and detected naturally could help infer the infrequent gestural activities.

A variety of activities of daily living (ADLs) recognition techniques have been studied extensively over the last few decades in different dimensions of smart healthcare [1], [5], [6], [7], [16], but very few of the work have addressed gestural activity recognition [16], [24]. Traditionally, activity recognition approaches can be classified into the following three categories based on the specific device usage and data source accessibility.

- Wearable sensor: Multiple body worn sensors or sensors embedded with everyday devices, such as, earbud, necklace, ring etc. have been used for recognizing ADLs [16], [19], [20].
- Smartphone sensor: Smartphone's microphone sensor has been used to capture acoustic signals of human surroundings to recognize non-speech human sounds in ambient living environment [5], [6], [7].
- Hybrid: Multiple or single body worn sensors along with smartphone's microphone sensor have been used to recognize ADLs [1]. To accommodate energy hungry microphone sensor intelligent on-chip and off-chip acoustic signal processing algorithms have been developed [3].

Previous works have focused on human speech processing extracting features from acoustic signal to detect human voice and non-speech human sounds. While acoustic sensor can certainly help determine the sound gesture of human but undermine significantly the operational cost and life longevity of wearable devices due to its energy hungry operations. Acoustic signal recording, pre-processing, ambient noise reduction, features extraction and classification process cause huge computational overhead which rapidly drains out the battery power of source devices. Sound signals generating from other individuals, surrounding the target user may cause severe misclassification problem creating unavoidable false positive results. Moreover, continuous sensing of sound signals may cause serious privacy violations. On the other hand embedding sensors on myriad objects of daily living, such as microwaves and kitchen cabinets [4] or mounting them on the ceiling has challenging operational costs and battery-life issues. Individuals, particularly, elderly patients appear reluctant to continually wear multiple sensors on the body [2]. Motivated by these shortcomings we propose to use an energy and computationally inexpensive accelerometer sensor in the form of a smart earring for detecting fine-grained gestural activities of the user.

Research Questions: Our investigations in this paper pursue the following research questions:

- Given the adaptation of activity recognition algorithms to help older adults in healthy independent living what sort of gestural activities may shed light on long-term physiological health and psychological behavior of older adults?
- What sort of models and algorithms are needed to detect such fine grained gestural activities along with traditional classification approach?
- How much quantitative improvement do we observe in our ability to recognize the correct GAs?

In this paper we first use real life data traces from 10 subjects with a variety of different gestural activities (max 5) and develop an adaptive C4.5 classification algorithm based on dynamic feature selection for recognizing potential gestural activities. While this approach helps to successfully identify different gestural activities but fails to identify when the gestural events are either instantaneous or continuous in nature. Realizing this we propose a novel change-point detection based hybrid classification model for gestural activity recognition that exploits the abrupt changes in gestural signals along with its inherent pattern to obtain divergence estimation between the time-series samples. We validate the proposed approach using real life data traces. Our work thus affirms how a microscopic gestural activity recognition model augmented with activities of daily living can provide practical insights that (a) helps capture the finer correlation of the activities of daily living with the acute physiological symptoms and psychological conditions, and (b) provide additional contexts which help devise novel interventions that can be effectively used in managing functional and cognitive health decline of older adults.

Key Contributions:

- Our key contribution lies in the proposed change point detection based gestural activity recognition approach (GeSmart) which represents the instantaneous perturbation of gestural signals as an abrupt change and continuous perturbation as a specific pattern and help detect the microscopic gestures. This provides a practical way to determine fine-grained discrimination of physiological and psychological health markers, without incurring the expensive and laborious in-situ laboratory testing.
- As a secondary contribution, we posit that low power, cheap accelerometer sensor is a potential option if integrated inside the smart jewelries (e.g., earrings, necklace etc) and provide better detection accuracy and substantial energy savings compared to the acoustic sensors.
- We evaluate the accuracy of GeSmart using real life activity traces from 10 domestic users, collected over several weeks. Our results show that, given normal everyday patterns of domestic living, GeSmart can provide very high accuracy in identifying microscopic gestural activities ($\approx 95\%$), and significantly decrease the energy consumption (by approx. 30%). These results demonstrate the viability of the GeSmart approach, for both finer-grained gestural activity recognition and long-term healthy independent living.

The rest of the paper is organized as follows. We first discuss the related work and then present the high-level overview of the proposed GeSmart framework. We highlight our initial findings on gestural activity recognition based on C4.5 classification algorithm. We then describe how change point detection based gestural activity detection method can be integrated with the regular classification approach. We develop an earring system using off-the-shelf commercially available accelerometer sensor and present our detailed experimental results. Finally, we identify future research directions and conclude our work.

II. Related Works

Most of the approaches in monitoring human gestures involve image or video feed analysis for tracking facial expressions or body postures. Early works for tracking gestures were unimodal which were based on only one criterion like vocal features, facial expressions, body postures or physiological changes [11], [12], [13], [14], [15], [17], [18]. After gathering data from different modalities, most of the work have focused on a supervised pattern classification algorithms to detect the gestures. But this approach fails to address the problem when a person has any overlapped gestures. The vocal features has also been used in gesture analysis based on the speech analysis techniques using signal processing. To differentiate between different variants of vocal sound *Mel-Frequency Cepstral Coefficient (MFCC)* has been applied and 66% average accuracy for detecting 6 emotions has been reported [21]. [22] added acceleration of pitch and MFCCs to form feature streams. It has applied different machine learning techniques for stressed/neutral style classification and Gaussian SVM for 4-class speaking style classification. Physiological signals like heart rate, skin conductivity, muscle activity etc have also been considered for inferring gestures achieving overall accuracy of 81% [15]. Recently researchers have proposed multimodal approaches where multiple sensor modalities have been considered simultaneously. [8] [9] proposed the bi-modal approach to capture human gestures by using both facial expression and body postures whereas [10] used multi-modal approach where facial expressions, vocal features, body movements and gestures have been fused altogether. [23] proposed a rule based approach by applying classification of audio-visual data. The multimodal approaches indicate that the performance of gesture or emotion recognition can be improved by multimodal sensor data fusion. A mobile sensing system leveraging the microphone sensor has been proposed to detect non-speech body sounds or gestural activities in [24].

In our work, we propose to use an accelerometer sensor based earring to detect the subtle movements users made during the course of a gestural activity occurrence. Our work is closest to [16] which used wearable accelerometer sensor to identify social actions. The main difference between [16] and *GeSmart* is that former uses HMM model to analyze face-to-face interactive conversing behaviors (e.g., speaking, laughing, gesturing, drinking, or stepping) in a densely crowded social gatherings to find out the correlation between movement acceleration and a person's social activity, ability of talking in a group with known/unknown persons or power of dominating in a group conversation. While [16] focused on building a model for only conversing behaviors using wearable accelerometer sensor incorporated with a badge (which is not ubiquitous), we focus in this paper to identify infrequent gestural activities which are independent of any specific posture (i.e., standing, sitting, running etc.) or predefined environment (i.e., alone or crowded) using ubiquitous device (i.e., earring). In general it is easier to detect fine-grained movements when the user posture and context are known a-priori such as conversing in a standing position [16]. But in this paper we focus on building a generic gestural activity recognition model independent of any specific location or postural position of a user. Indeed we particularly focus on specific gestural activities rather than social interactions which have long-term correlation with the physiological and psychological health of a person.

III. An Overview of GeSmart Framework

Fig. 1 shows a schematic representation of our proposed GeSmart model. It consists of the following logical steps. 1) Data Collection: gathering the accelerometer readings from Chronos through the bluetooth access point. 2) Device Position setup: Calculating average change-point score for each gestural activities from four different body positions. 3) Data Segmentation: Derive data segments corresponding to movement components (i.e., framing, windowing etc.); 4) Feature Extraction: Estimate accelerometer data features from each data segment incorporating change-point score; 5) Feature Selection: Rank features according to the contributions of achieving separation among classes associated with different change-point scores; and select feature sets that minimize the overlap among classes as associated with different change-point scores; 6) Robust Classification: Use 10-fold cross validation to estimate the quality of change-point score based classification for individual GAs; finally, 7) Gestural Activity Detection: Apply the hybrid classification model on real data traces to detect the fine-grained GAs.

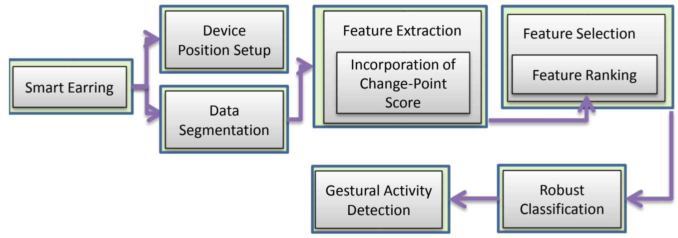

Fig. 1. An overview of our model

IV. Gestural Activity Detection: Initial study and Findings

Given our focus on detecting gestural activities, we first present the challenges of capturing and recognizing gestural activity's acceleration patterns in perspective of human body movement and motion. The goal here is to establish that with only an accelerometer based system we can leverage the non-speech body sound associated with a variety of GAs.

A. Anatomy of Gestural Patterns

The instantaneous or continuous periodicity of gestural events and their impact on human motion pose significant challenge on detecting them successfully. Different gestural activities has different intensity and motion characteristics, which may posit diagnostically valuable movement information to distinguish them. These underpinning characteristics are correlated with physical constraints of a person generating unique capturable human motion patterns at occurrences. Every spontaneous gestural activity (e.g., coughing, yawning etc.) occurs when a sequence of events is stimulated by the presence of sputum or foreign particles in the main, central airways of a person [28]. For example, in case of normal coughing the sequence of events are referred to as the sequence of irritation, inspiration, compression and expulsion [28]. Irritation is an abnormal stimulus (inflammatory, mechanical, chemical or thermal) which provokes sensory fibers to send impulses to the brain's medullary cough center. In the inspiration phase the glottis becomes wide open due to reflex muscle contraction. Movement of glottis, respiratory muscles and the branches of the bronchus are closely tied during the course of cough phase. Thus these four phases describe the major effects of the cough reflex. Each phases causes unique pattern of movement associated with the human body as shown in Fig. 2. In Fig. 2, we see a sudden downward acceleration change due to the movement of head from irritation to inspiration state. Then inspiration to compression state causes almost no

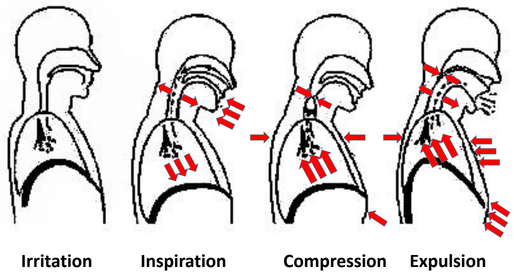

Fig. 2. Normal Coughing Consists of four events

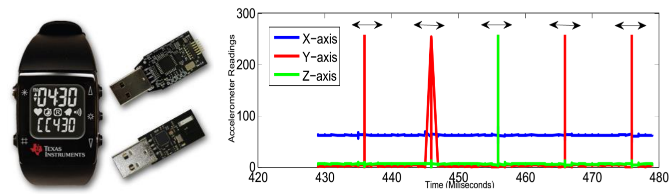

Fig. 3. (1) Chronos. Wrist Watch, CC1111 USB RF access point, eZ430 USB programming and debugging interface (b) Coughing data from Chronos used as earring

acceleration downward of head. Again, we see sudden upward acceleration change of head from compression to expulsion state. To capture these slightest movements and acceleration changes of the user, we place an accelerometer based system, Chronos (Fig. 3) corresponding to different body position as a smart jewelry such as earring and necklace.

B. Recognizing Gestural Events

Although we are able to define different micro events that construct human gestures, those micro events' duration, occurrence sequence and acceleration features (i.e., x, y and z axis data features) in terms of body movements vary from one gesture to another. For example, normal cough consists of four events: irritation, inspiration, compression and expulsion but normal yawning consists of only irritation, inspiration and expulsion with different duration ratio [28]. Even in case of dry coughing caused by tuberculosis, consecutive coughing may cause intense pain in the throat creating several extra compression events. Thus it is challenging to recognize gestural activities using an unified model. While using Chronos as an earring to capture these coughing events' acceleration due to its impact on head, we note that the x and z axis accelerometer sensor data are always steady, but y axis acceleration increases when the transition from irritation to inspiration occurs and decreases when the transition from compression to expulsion occurs (see Fig. 3). Fig. 4 depicts that different gestural activities has almost similar movement (i.e., acceleration) patterns making the classification problem more challenging. To distinguish these similar statistical features we propose to use change-point scoring method on each statistical feature which helps capture the fine-grained changes between the gestural activities. Fig. 4 shows the change-point scores of magnitudes applied on each statistical feature which enlightens the unique pattern for each of the gestural activity measures. Next we focus on developing the smart earring prototype, finding out its most informative position on the body, and designing robust classification and change point detection based hybrid classification model.

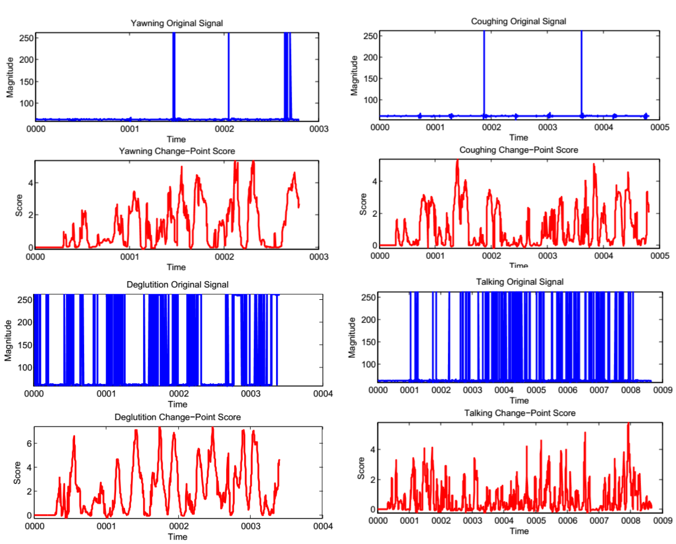

Fig. 4. Acceleration (magnitude $= \sqrt{x^2 + y^2 + z^2}$) and corresponding Change-Point Score of talking, coughing, yawning and deglutition

C. Device Setup and Customization

The goal for selecting a device for our earring prototype development was mainly cost, form factor, rapid customization and ease of deployment and data collection. The Texas Instruments Chronos development [37] was found to fit our needs and used for the development. The Chronos is a development platform built around an MSP430-compatible system-on-chip with an integrated wireless modem. Communications between a computer and the Chronos modem was done by a USB-interfaced "access poin" that comes with the kit. Data between the host system and the access point is communicated through a virtual COM port abstracted by the access point driver distributed by Texas Instruments. The accelerometer included with the Chronos platform is a Bosch BMA250 [39]. The BMA250 exposes an SPI and I2C interface for communication, and internally utilizes a 10-bit analog-to-digital converter. Serial communications are the limiting factor in sampling rates, offering a bandwidth of up to 1000 Hz. The BMA250 has a programmable range of 2g to 16g. The Chronos ships pre-programmed device with an evaluation firmware that demonstrates the features of the device, including the reading of raw accelerometer data, but the software from Texas Instruments does not provide an option to save the data received to disk. The simple binary serial protocol used is not described in Texas Instruments documents, and collecting information about the protocol by reading the publicly available source code for the firmware was found to be impractical. In an effort to develop interfacing software, communications between the Texas Instruments host-side software and the virtual COM port were monitored, and the protocol was elucidated by analyzing those communications. The protocol was re-implemented, and software was developed in C# to attach to the virtual COM port provided by the Chronos access point and poll for accelerometer data. Data was saved to disk in a CSV format.

D. Device Position Setup

We conducted various experiments to find out the most informative device position to capture the most significant acceleration of human body movements for different gestural

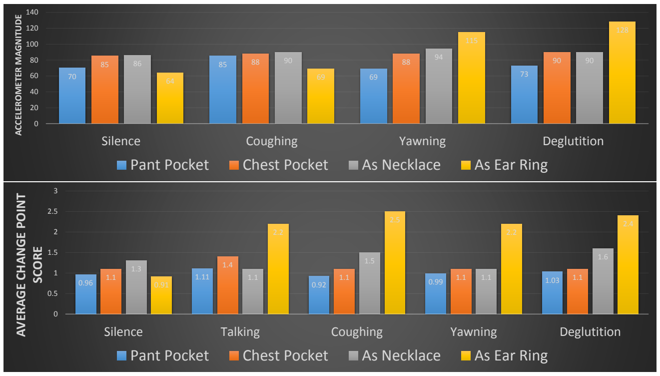

Fig. 5. Comparing different body positions (pant pocket, chest pocket, neck, ear) to capture different types of gestural motion

level activities. We performed extensive study on real data traces collected from 10 users using our prototype system to confirm the most informative position on the body. This test consisted of two parameters: the first being body position (pant pocket, chest pocket, neck and ear) and the second being gestural activities (silence, talking, coughing, yawning and deglutition). We recorded the five types of gestural activities' accelerometer data from the device and took the average accelerometer magnitudes and average change point scores of accelerometer magnitude (magnitude $= \sqrt{x^2+y^2+z^2}$) to compute the average acceleration changes in each gestural activity. Fig. 5 presents the change point score with respect to different body positions and gestural activities. The change-point score we used is an abrupt signal changes measure based on subsequence pattern matching [29]. We describe details and significant impact of this algorithm on gestural activity recognition in later part of our paper. We compared the average change-point scores and average magnitudes of the captured accelerometer data with respect to different body positions for different gestures. Fig. 5 shows that considering average magnitudes, it is impossible to detect the best position on the body for gestural activity recognition. But considering average change-point scores, it is noted that among the four locations, the ear gives us greater average change-point scores for all types of gestural activity, except silence. Intuitively, the breathing motion only affects chest and neck creating abrupt changes in acceleration than at pant pocket or ear. This continuous changes in acceleration due to the inherent breathing activity poses challenges to detect our finer GAs. In fact, through our device position experimentation as shown in Fig. 5 in presence of a variety of GAs along with the continuum regular ADLs, we establish that the position ear is always less affected by any external noise sources. To reduce this breathing noise, prior researchers [3], [5], [6], [7], [16], [19], [20] proposed to use multiple levels of noise reduction methods. In our case, intelligent determination of this position a-priori help reduce the unwanted noise created from the breathing gestural activity. Fig. 5 showed that the position ear is more informative and noise free than other three positions; pant pocket, chest pocket or necklace. Therefore, we postulate that given our goal of capturing a wide range of gestural activity events in presence of normal ADLs, the position ear is the most informative and noise free location for the GeSmart jewelry device design and real deployment.

Fig. 6. Chronos in different position

V. Change Point Detection Algorithm

Change point detection refers to identify the instances when the probability distribution of a stochastic process or time series changes. As we note previously that the standard statistical features fail to exploit the abrupt changes in gestural signals we propose to use change-point detection to capture the signal divergence. We design a hybrid classification model and use the relative Pearson divergence as a divergence measure estimated by a method of direct density-ratio estimation method [29]. We first mathematically describe the evolution of change point scoring for one dimensional time series sample of single valued sensor, and then consider multi-dimensional abrupt change-point detection estimation associated with the combination of three axis-accelerometer observational values and their standard statistical features.

To explain the method, let consider $y(t)$ as a 1-dimensional time series with single subsequence sample at time t where $y(t)$ =[x-axis reading, t]. Then, the subsequence of time series at time t with length k be,

$$Y(t) = [y(t)^T, y(t+1)^T, ..., y(t+k-1)^T]^T \in \mathbf{R^k} \quad (1)$$

where $y(t)^T$ represents the transpose of $y(t)$. Now, let consider $\mathbf{Y(t)}$ be a set of n retrospective subsequence samples starting at time t. Then,

$$\mathbf{Y(t)} = Y(t), Y(t+1), ..., Y(t+n-1) \in \mathbf{R}^{k \times n} \quad (2)$$

Now, let consider $y(t)$ be a d-dimensional time series with n subsequence sample where $y(t)$=$\begin{pmatrix} x-axis & t \\ y-axis & t \\ z-axis & t \end{pmatrix}$ Then, the subsequence of time series at time t be:

$$\mathbf{Y(t)} = Y(t), Y(t+1), ..., Y(t+n-1) \in \mathbf{R}^{dk \times n} \quad (3)$$

$\mathbf{Y(t)}$ forms a Hankel matrix and plays a key role in change-point detection based on subspace learning [33]. In our model, we considered $k = 10$, $n = 50$ and $d = 3$. We compute the dissimilarity measure between two consecutive segments $\mathbf{Y(t)}$ and $\mathbf{Y(t+n)}$, and use it as the plausibility of change points i.e., the higher the dissimilarity measure is, the more likely the point is a change point as depicted in Fig 7. Mathematically, we represent the dissimilarity measure as follows,

$$D(P_t|P_{t+n}) + D(P_{t+n}|P_t) \quad (4)$$

where P_t and P_{t+n} are probability distributions of samples in $\mathbf{Y(t)}$ and $\mathbf{Y(t+n)}$, respectively. $D(P|P')$ denotes the *f-divergence* [34]. We use *Pearson Divergence Estimation(PE)* [35] which is a modified version of *f-divergence* and represent PE divergence as follows,

$$PE(P|P') = \frac{1}{2} \int p'(Y) (\frac{p(Y)}{p'(Y)} - 1)^2 \times dY. \quad (5)$$

Where $p(Y)$ and $p'(Y)$ are probability densities. The formulation of *Pearson divergence* from *f-divergence* is omitted due to the space constraints. Since the probability densities $p(Y)$ and

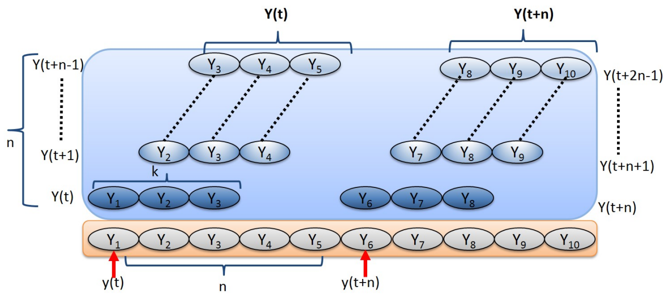

Fig. 7. One-dimensional time-series data.

$p'(Y)$ are unknown in practice, we cannot directly compute the $f-divergence$.

To estimate PE divergence, we use relative density-ratio estimator (RuLSIF) [36]. Considering *α-relative PE-divergence* for $0 < \alpha < 1$, we have,

$$PE_\alpha(P|P') = \int p'_\alpha(Y)(\frac{p(Y)}{p'(Y)} - 1)^2 \times dY. \quad (6)$$

where $p'_\alpha(Y) = \alpha p(Y) + (1-\alpha)p'(Y)$ is *α-mixture density*. So the final dissimilarity measure is

$$PE_\alpha(P|P') + PE_\alpha(P'|P) \quad (7)$$

Given $\alpha - relativedensity - ratio$ estimator $\hat{g}(\mathbf{Y})$, an approximation of the PE divergence is constructed as:

$$P\hat{E}_\alpha = -\frac{\alpha}{2n}\sum_{i=1}^{n}\hat{g}(\mathbf{Y_i})^\mathbf{2} - \frac{1-\alpha}{2n}\sum_{j=1}^{n}\hat{g}(\mathbf{Y'_j})^\mathbf{2} + \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}\hat{\mathbf{g}}(\mathbf{Y_i}) - \frac{\mathbf{1}}{\mathbf{2}} \quad (8)$$

We include the estimated change-point score (i.e., dissimilarity measure $PE_\alpha(P|P') + PE_\alpha(P'|P)$) of the most popular statistical features in our model. We use the change-point algorithm implementation [38] to estimate the change-point score of a variety of gestural activities as shown in Fig 4.

VI. Hybrid Classification Model

We propose a hybrid classification technique based on change point detection method combining traditional feature based technique with additional change point score based filtering. Detecting abrupt changes in time-series data, relying on change-point detection methods, can be classified into two categories:

- Real-time detection, targets applications that require immediate responses such as robot control, intrusion detection etc..
- Retrospective detection, useful for more robust and abrupt signal change detection although detection may require longer reaction periods.

In this work we propose to incorporate retrospective change point detection based method along with the traditional clarification technique to capture the finer movement changes in GAs.

A. Data Collection

We recruited 10 volunteers (including 1 female) with different heights and weights to collect five different gestural activities (i.e., silence, talking, coughing, yawning and deglutition) in two postural positions (i.e., standing and sitting).

TABLE I. List of Gestural Activities

Index	Gesture	Description
1	Silence	Without any gestural activities
2	Coughing	Natural two coughing
3	Yawning	Yawning as natural as possible
4	Deglutition	Natural water deglutition
5	Talking	Normal talking

The participants were asked to wear the "Chronos" on their ear and to adjust the position of the hook behind their ear such a way that it seems like he or she is wearing an earring. They were asked to perform 5 different gestural activities in two different postural states. The types of gestural activities and a short description of each task are listed in Table I. Most of the previous works considered talking and silence activities as noise in their classification methods [24] [16]. For example, [24] focuses on non-body sounds where talking and silence created some noises in their classification methods creating the need of filtering them out. In our system, the choice of device position (i.e., use of device as earring) and change-point detection algorithm conform fine grained classification for both silence and talking. In total, each of our participants contributed at least 15 minutes of continuous recordings consisting of a controlled sequence of five gestural activities. Table I shows the detail description of our captured gestural activities.

To examine the acceleration characteristics of the collected accelerometer data in different gestural activities, we plot their corresponding spectrograms in Fig. 4. Spectrogram illustrates a visual representation of the x, y and z axis spectrum of a gestural event as it varies with time. Silence is not shown separately because it is always present in between two consecutive gestural events. The distinct spectral pattern is not clearly visible in the original graph of x, y and z axis spectrum, while change-point scoring for all of the gestural activities generates a distinct spectral pattern.

B. Feature Extraction

The raw accelerometer data sampled from "Chronos" was first segmented into frames of uniform length. We considered the frame length of 48 ms and window size of 2 seconds. To characterize gestural events' body acceleration characteristics, we employed a two-step feature extraction procedure. In the first step, we extract a number of statistical accelerometer features (i.e., mean, variance, standard deviation, maximum and minimum of each axis readings, magnitudes of each reading, Goertzel coefficients of 1-5 Hz, MFCCs etc.) from each frame to construct frame-level features. In the second step, we calculated the change-point score of each feature.

C. Feature Selection

We follow a two-step feature extraction technique which generates a total of 33 features. As we implement the overall feature extraction and classification method on limited resource device (i.e., small computational memory and low battery powered devices) and wearable platform (i.e., earring), we aim to build our system more computationally efficient excluding the consideration of unnecessary features. Therefore, the goal is to select a minimum number of features that achieve reasonably good classification performance. We use the correlation feature selection (CFS) algorithm to select the subset of features [30]. The CFS algorithm evaluates the

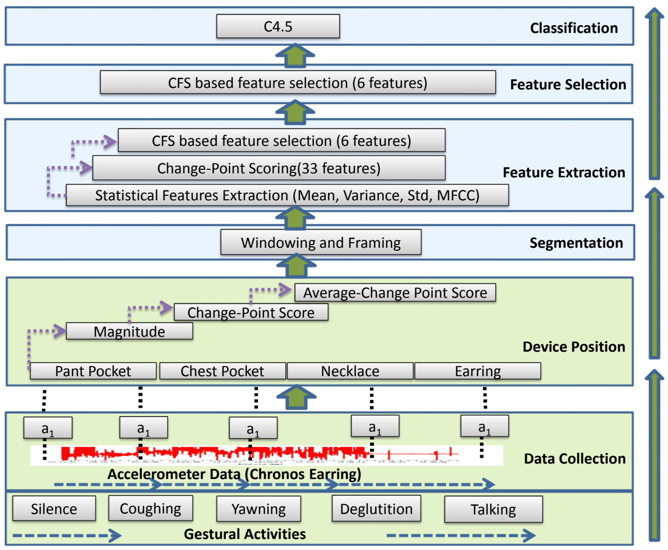

Fig. 8. Flowchart of proposed model

TABLE II. The TP rate, FP rate, Precision, Recall and F-measure for each class from the LOPO experiment using C4.5 as classifier and change-point score as frame-level features

Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure
Silence	95.7%	5.8%	96.4%	95.7%	96.0%
Coughing	86.0%	00.3%	84.0%	86.0%	85.0%
Yawning	90.5%	0.1%	93.8%	90.5%	92.1%
Deglutition	88.5%	1.9%	85.8%	88.5%	87.1%
Talking	96.5%	1.2%	95.9%	96.5%	96.2%
Weighted Avg.	**94.8%**	**4.1%**	**94.8%**	**94.8%**	**94.8%**

TABLE III. Comparison with prior works' classification recall measure

Methods	Hayley Hung [16] 2013	BodyBeat [24] 2014	Our Model
Silence	N/A	74.38%	95.7%
Coughing	N/A	80.0%	86.0%
Yawning	24%	75.0%	90.5%
Deglutition	21%	72.09%	88.5%
Talking	82.0%	81.06%	96.5%
Weighted Avg.	**N/A**	**71.2 %**	**94.8%**

worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. To identify locally predictive attributes we apply forward Best First Search (BFS). It iteratively adds attributes with the highest correlation in the class as long as there is no other combinations of the attributes which generate a better correlation. Finally we identify six features based on change-point score of (variance of x, variance of y, variance of z, standard deviation of Y, maximum x and minimum x) as the most optimized feature set for the target classifier.

D. *Classification Results*

We use C4.5 as the classification algorithm. We choose C4.5 over other classification algorithms because it is both computationally efficient and lightweight to be implemented in resource-constrained devices. We use six statistical features and change-point score as frame level features with a frame size of 21 samples and window size of 2 seconds on our classifier. To validate our classifier's performance, we used k-fold cross validation [31] using k-value as 10. Table II shows the class level true positive rate (TP rate), false positive rate (FP rate), recall, precision and F-measure from the 10-fold cross validation experiment of our classifier. Silence and talking have been detected with a 95% and 96% accuracy respectively as shown in Table II which is significantly higher than the prior proposed classification methods [16], [24] ([16] could not detect silence, but reported 82% accuracy for talking. [24] reported 74.38% and 81.06% accuracy respectively). From Table III we see that our model outperforms other existing solutions in detecting and recognizing different types of gestural activities achieving an average of 94.8% accuracy.

VII. Discussion

The correlation between the body motion and social behavior of the people has been well-established by the social psychologists [25], [26], [27]. Existing research in social psychology also highlights a strong correlation between the speech and body gestures among the speaker and listener [26], [27]. In this work, we propose a novel approach for gestural activity recognition using only a single energy efficient sensor, accelerometer embedded in an ubiquitous earring. GeSmart attest significant energy savings and higher detection accuracy compared to the existing methods (Fig. 9).

Privacy and energy efficiency: The larger group deployment reinforced the importance of considering privacy aspects of data logging, collection and analysis. Collecting sensor data, particularly from microphone or camera, involves recording people in unconstrained and unpredictable situations, both in public and private space. It may include the recoding of unnecessary audio or video information without proper consent of the users which is unethical and often illegal. Hence, most of the people are reluctant of wearing some devices which capture audio or video of ADLs. Our system conforms user's privacy by avoiding audio or video recording. Meanwhile, only using accelerometer can reduce significant amount of energy drainage which is always a bottleneck. Fig. 9 (b) shows a simple measure of battery power drainage of different sensors in Google Nexus 4 smartphone. It shows that that accelerometer sensor help improve the battery life of smartphone 2.3 and 3.3 times respectively compared to an audio and audio cum accelerometer sensor based activity recognition approach.

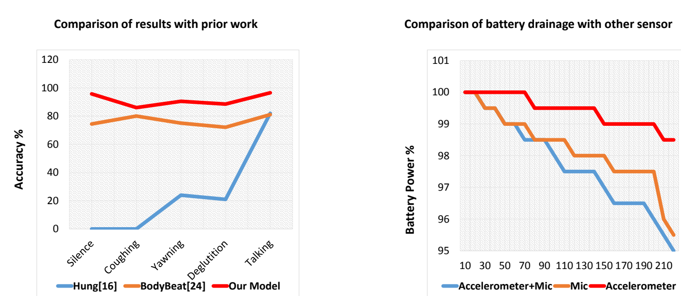

Fig. 9. (a) Comparison of Results with prior works (b) Comparison of battery power drop of sensors

VIII. Conclusion and Future Work

Older adults' health safety assurance has become increasingly important as the number of elderly people living worldwide and average life expectancy of them increases. In this paper, we have exploited the significance of GAs on elderly health-care and presented GeSmart, an energy efficient infrequent GAR model to predict the chronic behavioral conditions. We advocate that microscopic gestural activity recognition can provide useful insights for long-term care and behavioral health. We propose a hybrid classification approach based on change-point detection method which outperforms the previous GA detection method's accuracy by over 23.6% [24]. We have also shown through extensive experimentations with a variety of GAs that the position ear is a viable option to consider for capturing slightest perturbation of gestural signal in presence of regular ADLs.

We plan to explore the possible less energy consuming classifiers (such as Dynamic Bayesian Network) for designing energy-efficient smart devices in the form of jewelries. We also plan to test our model on real target age group such as older adults in an uncontrolled environment. Finally, based on the early potential results reported in this work, we plan to bring GeSmart to life by using it for several healthcare applications such as agitation detection for Alzheimer's patients or tremor detection for Parkinson's patients. The ability to recognize GAs using just the motion sensor opens up the potential for recognizing and analyzing people's activities of daily living (ADL) without explicitly capturing other costly sensor data and paying for computational overhead. For minimizing the ground truth collection and large scale deployment, we plan to apply active learning and transfer learning techniques along with our proposed approaches.

Acknowledgment

We thank Joseph Taylor and H M Sajjad Hossain for their contributions.

References

[1] Nirmalya Roy, Archan Misra, Diane Cook, Infrastructure-assisted smartphone-based ADL recognition in multi-inhabitant smart environments, PerCom, 2013, 38–46

[2] JH Bergmann, AH McGregor, Body-worn sensor design: What do patients and clinicians want?, Annals of Biomedical Engineering, 2011, 39, 9, 2299–2312

[3] S. M. Shahriar Nirjon, Robert F. Dickerson, Qiang Li, Philip Asare, John A. Stankovic, Dezhi Hong, Ben Zhang, Xiaofan Jiang, Guobin Shen, Feng Zhao, MusicalHeart: a hearty way of listening to music, SenSys, 2012, 43–56

[4] S. S. Intille, Using a live-in laboratory for ubiquitous computing research, PERVASIVE, 2006, LNCS , 3968

[5] N. Gyorbiro, A. Fabian, G. Homanyi, An activity recognition system for mobile phones, Mobile Networks and Applications, 2008, 14, 82–91

[6] J. Kwapisz, G. Weiss, S. Moore, Activity recognition using cell phone accelerometerse, International Workshop on Knowledge Discovery from Sensor Data, 2010, 10–18

[7] Suman Nath, ACE: exploiting correlation for energy-efficient and continuous context sensing, MobiSys, 2012, 28, 29–42

[8] M. Coulson, Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence, Journal of Nonverbal Behavior, 1992, 28, 2, 117–139

[9] H. Gunes, M. Piccardi, Bi-modal emotion recognition from expressive face and body gestures, Journal of Network and Computer Applications, 2007, 30, 4, 1334–1345

[10] Ginevra Castellano, Loic Kessous, George Caridakis, Multimodal emotion recognition from expressive faces, body gestures and speech, Artificial Intelligence and Innovations 2007: from Theory to Applications IFIP The International Federation for Information Processing , 2007, 247, 375–388

[11] R. Cowie, E Douglas-Cowie, Automatic statistical analysis of the signal and prosodic signs of emotion in speech, Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings, 1996, 3, 1989–1992

[12] Xia Mao, Bing Zhang, Yi Luo, Multi-level Speech Emotion Recognition Based on HMM and ANN, WRI World Congress on Computer Science and Information Engineering, 2009, 7, 225–229

[13] Xin Min Cheng, Pei Ying Cheng,Li Zhao, A Study on Emotional Feature Analysis and Recognition in Speech Signal, Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2009, 1, 418–420

[14] K. H. Kim, S. W. Bang, S. R. Kim, Emotion recognition system using short-term monitoring of physiological signals, Medical and Biological Engineering and Computing , 2004, 42, 419–427

[15] R. W. Picard , E. Vyzas , J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23, 1175–1191

[16] Hayley Hung, Gwenn Englebienn, Jeroen Kools, Classifying social actions with a single accelerometer, UbiComp '13 Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, 2013, 207–210

[17] I.A. Essa, A.P. Pentland, Coding, analysis, interpretation, and recognition of facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence , 1997, 19, 757–763

[18] Ying-Li Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence , 2001, 23, 97–115

[19] T. Choudhury, A. Pentland, Characterizing Social Networks using the Sociometer, NAACOS, 2004

[20] C.-C. Lian and J. Y.-j. Hsu, Probabilistic Models for Concurrent Chatting Activity Recognition, IJCAI, 2009, 1138-1143

[21] F. S. Wei, L. C. D. Silva,, Speech based emotion classification, International Conference on Electrical Electronic Technology,, 2001, 1, 297–301

[22] Oh-Wook Kwon, Kwokleung Chan, Jiucang Hao, Te-Won Lee, Emotion Recognition by Speech Signals, In proceeding of: 8th European Conference on Speech Communication and Technology, EUROSPEECH, 2003, 125–128

[23] L.C De Silva, Pei Chi Ng, Bimodal emotion recognition, Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings. , 2000, 332–335

[24] T. Rahman, A. Adams, E. Carroll, B. Zhou, H. Peng, Mi Zhang, T. Choudhury, BodyBeat: A Mobile System for Sensing Non-Speech Body Sounds, International Conference on Mobile Systems, Applications and Services (MobiSys), 2014

[25] T. L. Chartrand and J. A. Bargh, The chameleon effect: the perception-behavior link and social interaction, Journal of Personality and Social Psychology, 1999, 76, 6, 893-910

[26] A. Kendon, Conducting Interaction: Patterns of Behavior in Focused Encounters, Cambridge University Press, 1990

[27] D. McNeill, Language and Gesture, Cambridge University Press New York, 2000

[28] McGarvey LP, Patterns of cough in the clinic,Pulm Pharmacol Ther. 2011 Jun;24(3):300-3

[29] Song Liu, Makoto Yamada, Nigel Collier, Masashi Sugiyama, Change-point detection in time-series data by relative density-ratio estimation, Neural Networks. 2013; 72-83, http://dx.doi.org/10.1016/j.neunet.2013.01.012

[30] Hall, M. A. Correlation-based Feature Selection for Machine Learning. PhD Thesis (April 1999).

[31] Kohavi, Ron. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995 (San Mateo, CA: Morgan Kaufmann) 2 (12): 11371143.

[32] Fehmi Ben Abdesslem, Andrew Phillips, and Tristan Henderson.Less is more: energy-efficient mobile sensing with senseless. MobiHeld, page 61-62. ACM, (2009)

[33] Y. Kawahara, T. Yairi, and K. Machida. Change-point detection in time-series data based on subspace identification. In Proceedings of the 7th IEEE International Conference on Data Mining, pages 559564, 2007

[34] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. Journal of the Royal Statistical Society, Series B, 28(1):131142, 1966.

[35] K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 50:157175, 1900.

[36] M. Yamada, T. Suzuki, T. Kanamori, H. Hachiya, and M. Sugiyama. Relative densityratio estimation for robust distribution comparison. Neural Computation, 2013. to appear.

[37] http://processors.wiki.ti.com/index.php/EZ430-Chronos

[38] http://sugiyama-www.cs.titech.ac.jp/ song/change_detection

[39] http://www.bosch-sensortec.com/de/homepage/products_3/3_axis _sensors/acceleration_sensors/bma250_1/bma250