Agitation Detection in People Living with Dementia using Multimodal Sensors

Shehroz S. Khan[1], Sofija Spasojevic[1], Jacob Nogas[2], Bing Ye [2], Alex Mihailidis[1,2], Andrea Iaboni [1,2], Angel Wang [3], Lori Schindel Martin [3] and Kristine Newman [3]

***Abstract*— People Living with Dementia (PLwD) often exhibit behavioral and psychological symptoms of dementia; with agitation being one of the most prevalent symptoms. Agitated behaviour in PLwD indicates distress and confusion and increases the risk to injury to both the patients and the caregivers. In this paper, we present the use of wearable devices to detect agitation in PLwD. We hypothesize that combining multi-modal sensor data can help in building better classifiers to identify agitation in PLwD in comparison to a single sensor. We present a unique study to collect motion and physiological data from PLwD. This multi-modal sensor data is subsequently used to build predictive models to detect agitation in PLwD. The results on Random Forest for 28 days of data from PLwD show a strong evidence to support our hypothesis and highlight the importance of using multi-modal sensor data for detecting agitation events amongst them.**

I. Introduction

As the population of older adults increase, there is a rise in the number of people living with dementia (PLwD) [1]. The current World Health Organization estimates suggest that worldwide the number of PLwD is around 50 million, with nearly 10 million new cases every year [1]. Behavioral and Psychological Symptoms of Dementia (BPSD) represent a heterogeneous group of non-cognitive type of symptoms that can affect up to 90% of PLwD [2]. Agitation is one of the most prevalent type of BPSD present in PLwD [2]. Many of the agitation events are caused due to unmet needs of PLwD and the distress experience by them [3]. When agitated, PLwD can harm themselves, other patients and the staff.

Various clinical measures have been developed to assess agitation, such Pittsburgh Agitation Scale [4] and Cohen-Mansfeld Agitation Inventory [5]. However, these assessment tools are subjective, retrospective and cross-sectional, and are not helpful at prospectively identifying patterns of behaviour or detect events of agitation [6]. Therefore, it is important to develop objective and automatic methods of detecting agitation in PLwD to provide interventions to avoid harm to the patients and staff.

In this paper, we present a novel multi-modal sensing study currently being carried out Dementia Specialized Unit at Toronto Rehabilitation Institute (SDU-TRI), Canada. To the our best knowledge, this is the first study that collects data from PLwD using multi-modal sensor network. In this paper, we present the design of the agitation study, data collection and labeling methodology, data pre-processing, and preliminary results. We further discuss the challenges and plans for future research.

[1]KITE, Toronto Rehab-University Health Network, Toronto, ON, M5G 2A2, `shehroz.khan@uhn.ca`
[2]University of Toronto, Canada
[3]Ryerson University, Canada

II. Related Work

There have been very few studies on detecting agitation and aggression in PLwD using multi-modal sensors [7]. The systematic review by Khan et al. [8] suggests that many previous studies found correlation between actigraphy (accelerometer based devices) and agitation among PLwD. Most of these studies focus on changes in person's motor activities while undergoing an agitated state. However, there may be different scenarios where the motor activities may not strongly correlate with agitation, for e.g. shouting and screaming. In these situations, other physiological parameters may be useful in identifying agitation in conjunction with actigraphy. Unfortunately, there is not much literature about the use of multi-modal sensing technologies to detect agitation in PLwD. Fowler et al. [9] show the use of smartphone sensors to detect agitation in healthy adults (acting as patients). Chikhaoui et al. [10] show the use of ensemble learning by combining the data from Kinect camera and accelerometer for detecting agitation in younger adults. Nesbitt et al. [11] present a pilot study to collect data from PLwD using smart watches and phones and discuss its feasibility. They identify the need for individual profiles of patients so that personalized deviations can be identified from such data. Moore et al. [12] mentions that regardless of the specialized sensors used to detection agitation in PLwD, a major issue is the processing and labeling of large amounts of data generated for each patient.

III. Design of the Agitation Study

In this section, we briefly discuss the multi-modal sensor platform – Detection of Agitation and Aggression (DAAD) in PLwD [13]. The DAAD system is currently installed at the SDU-TRI. This system allows researchers to collect a novel source of patient information, including their motion and physiological information, sleep quality, interaction with their surroundings, and video data. Patients were recruited in the study based on the recommendation of a clinician researcher, who monitored their behaviour through out the course of the study.

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

Fig. 1. Multimodal Sensors used in the study following the DAAD framework: (From Top Left in clockwise manner) Video Cameras, Empatica E4 wearable device, Pressure Mat, Motion Sensor, Door Sensor.

A. Multi-modal Sensors

Once a participant is identified, relevant consents were taken from their substitute decision makers [6]. Each participant may be recruited in the study for a maximum of 2 months (in consultation with the clinician). If the participant does not show any agitation behaviour during two weeks, he/she is removed from the study. The participant wears an Empatica E4 wearable device [14]. This device can capture motion (through accelerometer), and physiological indicators (blood volume pulse (BVP), electrodermal activity (EDA), and skin temperature (TEMP)). Every night before the participant sleeps, the nurse would take off the device and put it for charging. Every morning, a researcher will download the data and put the device back on the participant (with the assistance of the nurse). This device gives access to raw sensor data, on which predictive models can be built. A pressure mat is placed under the participant's bed throughout the time of study to collect sleep related data, which includes their heart rate, respiration rate and bed exits. Each participant's washroom is fixed with two motion sensors and one door sensors to monitor excessive movement and opening/closing the door. Fifteen video cameras are installed throughout the ward in the common areas, such as hallways, dining and entertainment areas. The video cameras are not installed in the participants' rooms and their audio recording is turned off due to privacy concerns, as advised by the ethics board [6]. Figure 1 shows the different sensors used in the study.

B. Data Labeling

Data labeling is very challenging task in this study since we need the start and end times of the agitation events to build classifiers. The nurses on the SDU-TRI were provided training to record a participant's behaviour in their nursing charts by highlighting the start and end times of agitation events, the location of the event and the context behind it, and sticking a 'green' color sticky dot on the nursing chart to facilitate in locating the agitation event while reading these charts [6]. Every week, a team of researchers would read these nursing charts and make entries in a word-processing file about the agitation events. A nursing research assistant would then review the video recordings corresponding to the times of agitation recorded by researchers and fine tune the start and end timings. This two-step labelling process is very important to build better predictive models by making the labels of agitation events as accurate as possible.

IV. Data Collection

This study has been approved by the designated Research Ethics Board. The data collection for this study is ongoing. So far, data from two participants (P1 and P2 - both females, 80 and 93 years old) corresponding to 28 days have been collected and fully labeled with event's start and end timings. The number of days of data collected for P1 was 15 days and P2 was 13 days. Those days may not be consecutive due to technical problems or operational issues. The number of agitation events recorded for P1 was 5 and for P2 was 9. Therefore, for both the participants, we obtained less than one agitation event per day. This highlights the inherent skew in the data for having large number of normal events in comparison to very few agitation events.

A. Data Processing

In this paper, we present analysis using only the sensor data from E4 wearable device. The data from motion and door sensors, as well as from pressure mat will be taken into account in future work. Different sensors in the E4 wearable device sample data at different sampling frequency. Therefore, as a first step, all the sensors were sampled to 64Hz to match with the maximum sampling rate given by BVP. Then, all the sensor data was combined together. A first order Butterworth low-pass filter with 20Hz was used to remove the noise from each of the raw sensor readings. As a baseline, the first 10 minutes recording of the multi-modal sensor data from the wearable device was scraped off because the device may not be instantly put on the participant after switching it on.

B. Feature Extraction

The features were extracted from all the sensor readings after applying non-overlapping windows. Non-overlapping windows were used in order to prevent information leakage from the training to the testing set. For the accelerometer, the x, y, and z values were combined to obtain their norm (ACC) and features were extracted from them. For this study, the following 10 generic features were extracted for each of the ACC, EDA and BVP [15]:

- Time Domain – mean, maximum, minimum, standard deviation and inter-quartile range.
- Frequency Domain – total average power, spectral entropy, energy, DC power and entropy.

For the TEMP sensor, only the first four time domain features were extracted. For EDA, entropy and spectral entropy gave NaN values for some time windows. Therefore, these two features were removed from building the classifiers. Hence, the total number of features used in this analysis is 32.

V. Experimental Analysis

The size of time windows used to extract features in this experiment were kept at 1, 3, and 5 minutes. This is done to understand the importance of number of samples for detecting agitation. In this paper, we only include the days when agitation occurred; they will contain both normal and agitation events. Table I shows the number of normal and agitation events for P1 and P2 by varying the number of window sizes. It can be observed that number of normal events are much greater than the agitation events.

TABLE I
Number of normal and agitation events for P1 and P2 corresponding to different window sizes.

Window Size (min)	P1		P2	
	#Normal	#Agitation	#Normal	#Agitation
1	2572	13	2682	42
3	854	3	893	13
5	512	2	534	8

We built classification models based on combined data from both participants, and also built separate models for each participant. The experimental settings are the following:

- The data from the four sensors were combined in 15 different ways (e.g. ACC, ACC+EDA, ACC+BVP, ACC+BVP+EDA,...) to understand the merit of using different sensors in detecting agitation events in PLwD.
- We used two standard classifiers - Support Vector Machines (SVM) and Random Forest (RF) from MATLAB. We have tested also Logistic Regression (LR) classifier, but since it gave unsatisfactory outcomes, we do not present those results.
- We performed two-fold cross validation to evaluate all the data sets. The predicted scores from each fold on the test set are concatenated, and area under the ROC curve (AUC) is used as a performance metric.
- Two parameters each of SVM (Box Constraint and Kernel Scale) and RF (Number of Trees and Number of Predictors) were tuned by employing an internal two-fold cross validation on the training fold.
- The range of values for which Box Constraint and Kernel Scale were set at $[0.01, 0.1, 1, 10, 100]$, the Number of Trees was in the range $[10, 30, 50, 70, 90]$, while the value for Number of Predictors was in the range $[f/5, 2f/5, 3f/5, 4f/5]$, where f is the number of features in the data for a given sensor combination.
- The best parameters' values obtained after inner cross-validation are used to train the SVM and RF and scores were calculated on the test set.

VI. Results

Table II shows the results on combined data from both the participants after applying SVM and RF classifiers. The gray cell shows the best AUC obtained across different classifiers, time windows and feature combinations. We obtain the best AUC value of 0.890 for the feature combination ACC+EDA+TEMP, for 1 minute time window. This shows that multi-modal sensing can perform better in detecting agitation across people. Table III shows the AUC values individually for P1 and P2 after employing different window sizes with RF classifier. The gray cells are the ones that give the best average AUC values (across P1 and P2) for all window sizes and feature combinations. We observe that the best average AUC values are obtained by using the feature combination BVP+EDA+TEMP (P1=0.838 and P2=0.868, average=0.853) for 1 minute time window. This result shows the usefulness of multi-modal sensing in detecting personalized agitation events. We did not show the results for SVM for individual level models because they perform worse than the RF classifiers. From both the Tables II and III, we infer that 1 minute time window with RF classifier performs the best with multi-modal sensing on both the combined and individual models. In both methods, the best performing models, EDA and TEMP are the common sensing modality.

VII. Conclusion and Future Work

In this paper, we present preliminary results of a real-world study that employs multi-modal sensors for detecting agitation in PLwD. The results suggest the feasibility of multi-modal sensors for detecting agitation in PLwD at combined data from different participants and at individual levels. As we collect more data in this study, we hope to improve the results, reduce bias introduced due to small data set and develop more insights into individual sensing modality and its effect in detecting agitation in PLwD. In future, we would extract domain specific features for different sensors and study its impact on the performance of the classifiers. The agitation events occur very less in comparison to the normal data; therefore, we will apply one-class classification methods [16] to detect agitation.

References

[1] World Health Organization, "Dementia," http://www.who.int/news-room/fact-sheets/detail/dementia, 2017, accessed: 2018-11-01.

[2] J. Cerejeira, L. Lagarto, and E. Mukaetova-Ladinska, "Behavioral and psychological symptoms of dementia," *Frontiers in neurology*, vol. 3, p. 73, 2012.

TABLE II

AUC VALUES FOR COMBINED DATA OF BOTH PARTICIPANTS FOR SVM AND RF.

Sensor Type	Feature Type	SVM			RF		
		Window Size (min)			Window Size (min)		
		1	3	5	1	3	5
Single	ACC	0.581	0.465	0.509	0.579	0.579	0.731
	BVP	0.542	0.559	0.480	0.599	0.666	0.500
	EDA	0.655	0.522	0.463	0.655	0.462	0.569
	TEMP	0.615	0.476	0.643	0.713	0.655	0.576
Multi-modal	ACC+BVP	0.557	0.491	0.467	0.616	0.613	0.490
	ACC+EDA	0.629	0.668	0.593	0.629	0.634	0.568
	ACC+TEMP	0.529	0.542	0.408	0.763	0.632	0.598
	BVP+EDA	0.568	0.528	0.469	0.732	0.617	0.533
	BVP+TEMP	0.556	0.617	0.539	0.771	0.682	0.572
	EDA+TEMP	0.603	0.482	0.594	0.664	0.639	0.569
	ACC+BVP+EDA	0.549	0.537	0.601	0.664	0.569	0.454
	ACC+BVP+TEMP	0.546	0.492	0.535	0.747	0.729	0.412
	BVP+EDA+TEMP	0.547	0.538	0.526	0.844	0.555	0.664
	ACC+EDA+TEMP	0.596	0.643	0.611	0.890	0.713	0.478
	ACC+EDA+BVP+TEMP	0.536	0.554	0.642	0.805	0.752	0.599

TABLE III

AUC VALUES FOR PERSON SPECIFIC CROSS-VALIDATION FOR DIFFERENT WINDOW SIZES FOR RF

Sensor Type	Feature Type	Window Size (min)					
		1		3		5	
		P1	P2	P1	P2	P1	P2
Single	ACC	0.509	0.603	0.485	0.704	0.463	0.690
	BVP	0.493	0.617	0.650	0.555	0.473	0.484
	EDA	0.529	0.735	0.820	0.496	0.473	0.559
	TEMP	0.628	0.738	0.489	0.555	0.489	0.572
Multi-modal	ACC+BVP	0.598	0.639	0.482	0.609	0.448	0.583
	ACC+EDA	0.476	0.770	0.479	0.666	0.465	0.477
	ACC+TEMP	0.713	0.917	0.632	0.761	0.702	0.668
	BVP+EDA	0.563	0.700	0.623	0.730	0.470	0.543
	BVP+TEMP	0.714	0.765	0.447	0.677	0.676	0.609
	EDA+TEMP	0.740	0.906	0.478	0.721	0.468	0.635
	ACC+BVP+EDA	0.569	0.583	0.467	0.645	0.469	0.515
	ACC+BVP+TEMP	0.679	0.811	0.638	0.704	0.457	0.570
	BVP+EDA+TEMP	0.838	0.868	0.479	0.871	0.701	0.803
	ACC+EDA+TEMP	0.674	0.926	0.606	0.652	0.460	0.567
	ACC+EDA+BVP+TEMP	0.687	0.779	0.457	0.622	0.459	0.636

[3] J. Cohen-Mansfield, "Measurement of inappropriate behavior associated with dementia," *Journal of Gerontological Nursing*, vol. 25, no. 2, pp. 42–51, 1999.

[4] J. Rosen, L. Burgio, M. Kollar, M. Cain, M. Allison, M. Fogleman, M. Michael, and G. S. Zubenko, "The pittsburgh agitation scale: a user-friendly instrument for rating agitation in dementia patients," *The American Journal of Geriatric Psychiatry*, vol. 2, no. 1, pp. 52–59, 1994.

[5] J. Cohen-Mansfield, "Instruction manual for the cohen-mansfield agitation inventory (cmai)," *Research Institute of the Hebrew Home of Greater Washington*, 1991.

[6] B. Ye, S. S. Khan, B. Chikhaoui, A. Iaboni, L. S. Martin, K. Newman, A. Wang, and A. Mihailidis, "Challenges in collecting big data in a clinical environment with vulnerable population: Lessons learned from a study using a multi-modal sensors platform," *Science and Engineering Ethics*, pp. 1–20, 2018.

[7] S. Teipel, C. Heine, A. Hein, F. Krüger, A. Kutschke, S. Kernebeck, M. Halek, S. Bader, and T. Kirste, "Multidimensional assessment of challenging behaviors in advanced stages of dementia in nursing homesthe insidedem framework," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 8, pp. 36–44, 2017.

[8] S. S. Khan, B. Ye, B. Taati, and A. Mihailidis, "Detecting agitation and aggression in people with dementia using sensors – a systematic review," *Alzheimer's & Dementia*, vol. 14, no. 6, pp. 824 – 832, 2018.

[9] C. Fowler, A. Gupta, K. Maly, K. Karlowicz, M. Gunnam, R. Gudipati, M. Kukunooru, and R. Rachamalla, "Detecting agitation onset in individuals with dementia using smart phone sensors," in *HEALTHINFO 2017 : The Second International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*, 2017.

[10] B. Chikhaoui, B. Ye, and A. Mihailidis, "Ensemble learning-based algorithms for aggressive and agitated behavior recognition," in *Ubiquitous Computing and Ambient Intelligence*. Springer, 2016, pp. 9–20.

[11] C. Nesbitt, A. Gupta, S. Jain, K. Maly, and H. R. Okhravi, "Reliability of wearable sensors to detect agitation in patients with dementia: A pilot study," in *Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology*. ACM, 2018, pp. 73–77.

[12] P. Moore, F. Xhafa, L. Barolli, and A. Thomas, "Monitoring and detection of agitation in dementia: Towards real-time and big-data solutions," in *P2P, Parallel, Grid, Cloud and Internet Computing, 2013 8^{th} International Conference on*. IEEE, 2013, pp. 128–135.

[13] S. S. Khan, T. Zhu, B. Ye, A. Mihailidis, A. Iaboni, K. Newman, A. H. Wang, and L. S. Martin, "Daad: A framework for detecting agitation and aggression in people living with dementia using a novel multi-modal sensor network," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 703–710.

[14] Empatica, "E4 wristband from empatica," https://www.empatica.com/en-eu/research/e4/, 2018, accessed: 2018-10-27.

[15] S. S. Khan, M. E. Karg, D. Kulić, and J. Hoey, "Detecting falls with x-factor hidden markov models," *Applied Soft Computing*, vol. 55, pp. 168–177, 2017.

[16] S. S. Khan and M. G. Madden, "One-class classification: taxonomy of study and review of techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, pp. 345–374, 2014.