Received: 25 October 2021 | Revised: 24 February 2022 | Accepted: 27 February 2022 | Published online: 22 April 2022

DOI: 10.1002/dad2.12305

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Wearable multimodal sensors for the detection of behavioral and psychological symptoms of dementia using personalized machine learning models

Andrea Iaboni[1,2] | **Sofija Spasojevic**[1,3] | **Kristine Newman**[4] | **Lori Schindel Martin**[4] | **Angel Wang**[4] | **Bing Ye**[1,3] | **Alex Mihailidis**[1,3] | **Shehroz S. Khan**[1,5]

[1]KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada

[2]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

[3]Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, Ontario, Canada

[4]Daphne Cockwell School of Nursing, Ryerson University, Toronto, Ontario, Canada

[5]Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada

Correspondence
Andrea Iaboni, KITE Research Institute, Toronto Rehabilitation Institute, 550 University Ave, 5–105
Toronto, ON M5G 2A2, Canada.
Email: andrea.iaboni@uhn.ca

Funding information
The Alzheimer Society of Canada; The Walter and Maria Schroeder Institute for Brain Innovation and Recovery

Abstract

Introduction: Behavioral and psychological symptoms of dementia (BPSD) signal distress or unmet needs and present a risk to people with dementia and their caregivers. Variability in the expression of these symptoms is a barrier to the performance of digital biomarkers. The aim of this study was to use wearable multimodal sensors to develop personalized machine learning models capable of detecting individual patterns of BPSD.

Methods: Older adults with dementia and BPSD (n = 17) on a dementia care unit wore a wristband during waking hours for up to 8 weeks. The wristband captured motion (accelerometer) and physiological indicators (blood volume pulse, electrodermal activity, and skin temperature). Agitation or aggression events were tracked, and research staff reviewed videos to precisely annotate the sensor data. Personalized machine learning models were developed using 1-minute intervals and classifying the presence of behavioral symptoms, and behavioral symptoms by type (motor agitation, verbal aggression, or physical aggression).

Results: Behavioral events were rare, representing 3.4% of the total data. Personalized models classified behavioral symptoms with a median area under the receiver operating curve (AUC) of 0.87 (range 0.64–0.95). The relative importance of the different sensor features to the predictive models varied both by individual and behavior type.

Discussion: Patterns of sensor data associated with BPSD are highly individualized, and future studies of the digital phenotyping of these behaviors would benefit from personalization.

KEYWORDS
behavioral and psychological symptoms, dementia, machine learning, neuropsychiatric symptoms, personalization, wearable sensors

1 | INTRODUCTION

People living with dementia can experience behavioral and psychological symptoms of dementia (BPSD). These symptoms are related to the neuropsychiatric manifestations of their disease, can arise in the context of unmet needs, and are influenced by interactions with the environment and caregivers.[1] While motor agitation and verbal or physical aggression are commonly used categories to describe these behaviors, the specific expression of these behaviors varies greatly depending on the individual.[2,3] The clinical significance of these symptoms is tied to the distress they cause to people with dementia and their caregivers, the risk that they can present to the safety of the individual and those around them, and their impact on the health-care system.[4,5]

With advances in wearable technologies, there has been interest in developing digital biomarkers of behavioral symptoms using data collected from a variety of wearable sensors.[6,7] To date, most studies in this area have focused on using wearables to assess the severity of symptoms by summarizing wearable data over periods from days to months and finding correlations with observational rating scales.[8–10] For example, studies have found a moderate correlation between activity levels (accelerometers) and clinical measures of wandering or agitation,[11–15] and some association between sensor measures of stress (electrodermal activity, galvanic skin response) and ratings of agitation.[16] These studies have been limited by small sample sizes, brief periods of data collection (< 1 month), and reliance on observational clinical rating scales with moderate reliability.[6]

The next frontier is the use of wearable sensor data to develop algorithms that can detect the onset and presence of the behaviors in real time.[17] This has been termed digital phenotyping, or the moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices.[18] The value of a digital phenotype for behavioral symptoms in dementia lies in its uses for clinical care and research for monitoring symptoms over time and supporting intervention and treatment.[19] As a first step toward this goal, we have previously demonstrated that wearable sensor data can be used to detect behavioral symptoms minute-to-minute with an area under the receiver operating curve (AUC) of 0.82 and that incorporating data from multiple sensors improves algorithm performance over a single sensor.[20] While these algorithms achieved a fair performance in classifying the presence of agitation, there were several important limitations. These generic models combined the sensor data from all participants and combined all subtypes of behaviors under the label of "agitation": these generic models thus are missing important information about the individual expression of behaviors. Given the large inter-individual differences in the expression of BPSD, the digital biomarkers for these symptoms are also likely to be heterogeneous, as are the biomarkers for different types of behaviors.

Toward an understanding of inter-individual differences in the digital biomarkers of agitation, the aims of this study are to: (1) develop personalized models able to detect the presence or absence of behavioral symptoms, (2) examine behavioral symptom–specific personalized models, and (3) explore which sensor features are most important for modeling agitation within individuals and for specific behaviors. To achieve this, we used wearable multimodal sensor data collected from people with dementia with BPSD admitted to a tertiary dementia care unit to develop and evaluate personalized models that classify each minute as "agitation" or "non-agitation" behavior. We then developed separate personalized models for each of three categories of behavior (motor agitation, verbal aggression, physical aggression), and examined the most important features in each of these models.

RESEARCH IN CONTEXT

1. **Systematic Review**: There are few systematic studies that examine the accuracy of digital biomarkers for detecting behavioral symptoms. The most commonly examined behaviors are motor agitation or apathy and use a single sensor (accelerometer). Sensor data are often summarized over days to weeks and across individuals, and correlated with clinical scales rather than directly observed behaviors. These limitations have impacted the accuracy of digital biomarker studies thus far and are an important barrier to advances in this field.
2. **Interpretation**: Using a multisensor wearable device we can accurately detect behavioral symptoms minute-to-minute. Personalization significantly improves the accuracy of the model to detect behavioral symptoms. Feature ranking is valuable to demonstrate the variability across individuals.
3. **Future Directions**: Our findings lend optimism to the search for digital biomarkers for behavioral symptoms in dementia and will help drive future studies in this area toward tuning or active learning to improve the personalization of machine learning models.

2 | METHODS

2.1 | Study overview

An overview of the study is provided in Figure 1. The study took place on the Specialized Dementia Unit (SDU) at Toronto Rehabilitation Institute (TRI), University Health Network, Toronto, Ontario, Canada, and was approved by the designated research ethics boards at University Health Network and Ryerson University, Toronto, Ontario, Canada. Details about the study and setting can be found in Ye et al.[21]

2.2 | Study participants

Participants were 20 older adults admitted to the SDU from long-term care homes for the assessment and treatment of BPSD in the context of advanced dementia. Inclusion criteria were age greater than 55 years

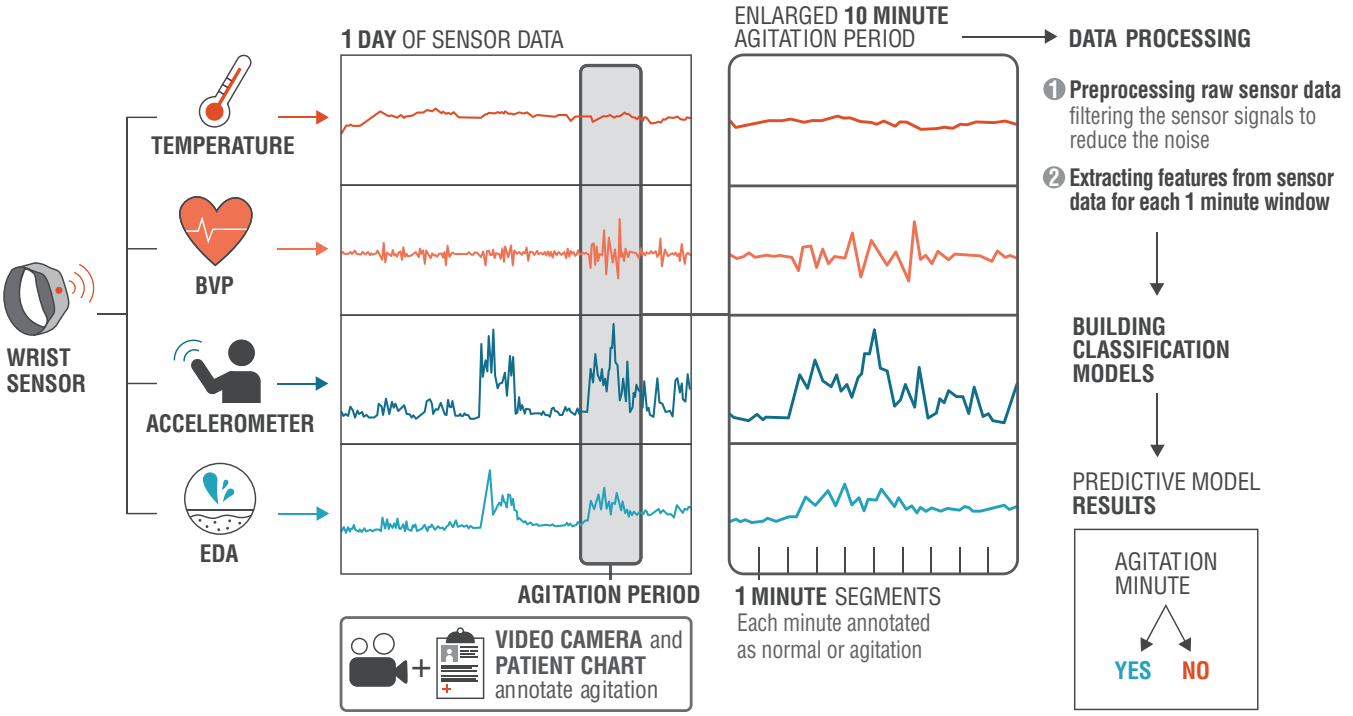

FIGURE 1 Study overview. BVP, blood volume pulse; EDA, electrodermal activity

old, diagnosis of dementia, and history of agitation or aggressive behaviors. Consent was provided by the substitute decision-makers in all cases, and individuals were withdrawn from the study if they showed dissent to wearing the device. Data collection was stopped when one of these conditions was met: no behavioral symptoms documented for 1 week, repeated removal of the wristband, on discharge from the unit, or after 2 months of data collection.

2.3 | Data collection

The wearable device used was the Empatica E4 (Empatica)[22] which collects motion data (through accelerometer [ACC]) and physiological indicators (blood volume pulse [BVP], electrodermal activity [EDA], and skin temperature [TEMP]). The Empatica E4 device has previously been evaluated in stress detection studies among other applications.[23–26] The E4 was applied to the dominant wrist each morning after the individual received their morning care (usually between 0700 and 1000), and was removed when preparing for bed at night (usually between 1700 and 2000), which was necessary for daily device charging and data download. Fifteen video cameras were installed throughout the SDU-TRI in the common areas, such as hallways, dining, and recreational areas.[21]

Nursing staff were provided training on the documentation of behavioral events, and asked to note the start and end times, the location, and the context behind the event. Each event was marked with a green sticker in the margin of the chart. Research staff would review the charts weekly and gather information about each event. The video recordings corresponding to 30 minutes before and after the charted period of agitation were then watched by clinical research staff to fine-tune the start and end times of the events and to gather further descriptive information about the event. Finally, the description of each event was used to divide them into one of three categories (motor agitation, verbal aggression, or physical aggression) using the definitions as provided by the Behavioral Supports Ontario Dementia Observation System (BSO-DOS).[27,28] If at any point during the event, an aggressive act such as hitting, pushing, or kicking was noted, the entire event was categorized as physical aggression. Where there was a combination of verbal aggression and motor agitation, the event was categorized based on the behavior that was predominant. To be included in an analysis by behavior category, a participant needed to have at least two agitation events in the given category. Any data which fell outside of an labeled agitation event was categorized as non-agitation, and this could include the participant at rest, engaged in activity, or walking.

2.4 | Data processing and feature extraction

Agitation events are relatively infrequent compared to non-agitation, thus the data included in this analysis was restricted to only include days in which at least one agitation event took place and was able to be fully labeled with start and end time (a total of 125 days of data collection). We extracted 37 features from filtered sensor readings after applying non-overlapping windows of 1-minute length. Chosen features represent the most commonly used time and frequency-domain characteristics of each sensor signal.[20,29–36] The extracted features are listed in Table S1 in supporting information along with their definitions.

2.5 | Data analysis

Personalized models were developed to classify agitation versus non-agitation minutes. To directly compare generic versus personalized models, we analyzed our dataset using two different approaches. The first was a leave-one-subject-out approach, where the model was trained with data from all participants but one and tested on the last participant. The second approach involved using the data from one participant at a time, with half of the participant's data used for training and half for testing. We thus built two models for each participant—a generic one based on the leave-one-subject-out approach and a personalized model. To compare these two approaches, for each participant, the models are tested on the same dataset. For example, the personalized model for participant 1 was trained on half of its data and tested on the second half. The leave-one-subject-out model for participant 1 is trained on data from the other 16 participants (excluding participant 1) and tested the same half of the data that was used for testing of the personalized model. Personalized classification models were also developed separately for each participant to classify non-agitation versus each separate category of behavioral symptom.

In our previous study,[20] we identified that the random forest (RF) classifier with cost had the best performance among all tested classifiers; therefore, we chose to use it in this study to classify using 1-minute data windows. The cost matrix was introduced into the RF model to address data imbalance by assigning a higher cost to the classification of agitation event as non-agitation than the classification of non-agitation as an agitation event.

Classifiers were built based on the 37 time and frequency-domain features extracted from sensor data using window approach (non-overlapping windows with 1-minute length) as described in Spasojevic et al.[20] Classifier parameters were tuned using internal two-fold cross-validation on the training set. There were two parameters for RF classifier—the number of trees with the range 10, 30, 50, 70, 90 and the number of predictors with values in the range [f/5, 2f/5, 3f/5, 4f/5], where f is the number of features in the feature set (f = 37). After inner cross-validation, the selected parameters were used to retrain the classifiers on the training set and performance was evaluated based on test set. We performed 2-fold cross-validation for all classifiers to evaluate the data sets. AUC was used as a performance metric.

To examine the relative importance of different sensors and features for agitation detection, we applied *rankfeatures()* function in Matlab that ranks features by class separability (agitation vs. non-agitation) using different criteria to assess the significance of every feature for separating two labeled groups. We used the "roc" criterion as a non-parametric test due to non-normal distributed classes. This criterion represents the area between the empirical receiver operating characteristic curve and the random classifier slope. Feature ranking was performed for each mode,l which provided weights reflecting the feature importance in the model. These weights were then normalized within each category in the range of 0 to 1 (where 1 is the highest weighted feature) by dividing all weights by the maximum weight within each category.

TABLE 1 Description of participants and data

	Total N = 17
Demographics	
Age (years), mean (SD)	78.9 (8.9)
Age (years), range	65–93
Sex (% female)	58.8%
Clinical characteristics	
Clinical Dementia Rating score: number (%)	0: 0
	1: 0
	2: 3 (18%)
	3: 14 (82%)
Neuropsychiatric Inventory total score, mean (SD)	55.1 (23.7)
Neuropsychiatric Inventory agitation subscale, mean (SD)	8.2 (4.4)
Neuropsychiatric Inventory aggression subscale, mean (SD)	7.8 (3.7)
Neuropsychiatric Inventory motor disturbance subscale, mean (SD)	8.9 (3.9)
Data characteristics	
Days of data included per participant, median (interquartile range)	4 (2–11)
Hours of data per day included per participant, median (interquartile range)	10.2 (8.8–11.3)
Labeled agitation events per participant, median (IQR range)	9 (5–30)

Abbreviations: IQR, interquartile range; SD, standard deviation.

3 | RESULTS

3.1 | Description of participants and behaviors

Twenty participants consented to the study, but three did not exhibit any behavioral events during the data collection period and were excluded from this analysis. The included 17 participants' demographic details are listed in Table 1. Over the 125 days of data included in this analysis, there were 305 fully labeled agitation events. The duration of agitation events varied from 1 minute up to 3 hours, with a mean duration of 14.9 ± 23.5 minutes and a median duration of 6 minutes (interquartile range 2–20).

3.2 | Personalized models: all agitation

Combining all events as "agitation," 17 personalized models were built to classify agitation minutes versus non-agitation minutes within each individual participant. Model performance was variable across participants, with a median AUC of 0.87, and a range of AUC values from 0.64 to 0.95 (Figure 2A). Compared to generic models, personalized models significantly outperformed the generic models for all but one

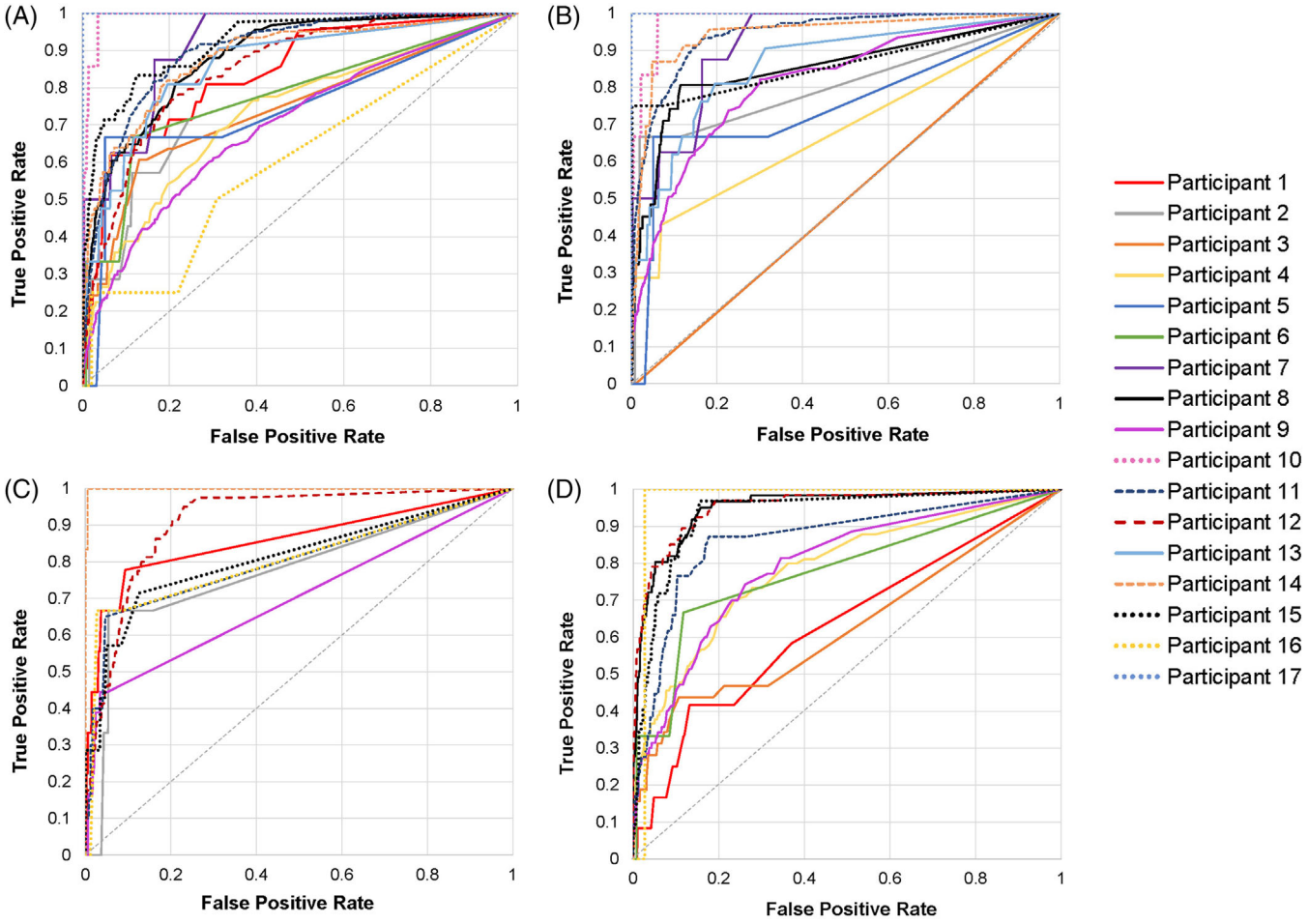

FIGURE 2 Receiver operating characteristic (ROC) curves demonstrating the predictive power of the personalized models by plotting the true positive rate (sensitivity) versus false positive rate (1-specificity) at different thresholds for classifying agitation/non-agitation. Each curve represents a different participant (see legend). The dashed gray line represents an ROC curve for a random classifier. ROC curves are shown for (A) all agitation events, (B) motor agitation, (C) verbal aggression, and (D) physical aggression

participant (Figure 3). The mean absolute difference in AUC score between the personalized and "leave-one-out" models was 0.28 ± 0.19.

3.3 | Personalized models: behavior categories

Personalized models were built for each participant to distinguish each separate category of behavior (motor, verbal, physical) and non-agitation minutes. In total, there were 32 sets of data analyzed across 17 participants as not all participants had the minimum required two behavioral events in each category (Table 2). The median AUC for the motor agitation model was 0.90 (range 0.49–0.98), for verbal agitation 0.86 (0.53–0.98) and for physical aggression 0.82 (0.71–1.0). Three personalized models across the three behavior categories had poor performance (AUC < 0.7), all of which had small number of agitation events on which to train the models (< 7 agitation events each), while 11 models had AUC values > 0.9 (Figure 2B, C, D).

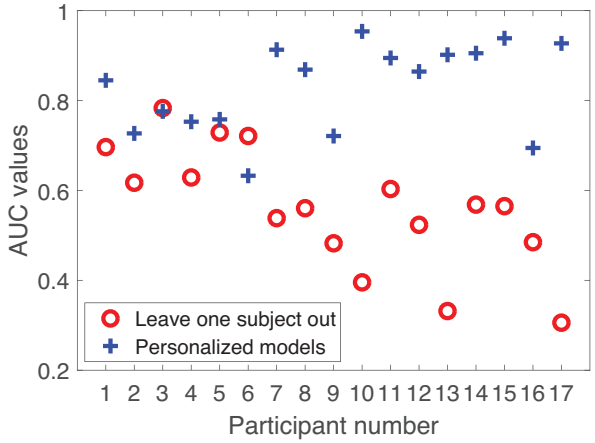

FIGURE 3 Comparison by participant of the performance of personalized models to generic leave-one-subject-out models for the classification of agitation minutes versus non-agitation minutes. AUC, area under the receiver operating characteristic curve

TABLE 2 Number of non-agitation and agitation 1-minute windows by agitation type across participants

Participant	Non-agitation minutes	Total agitation minutes	Motor	Verbal	Physical	Ratio non-agitation: agitation	Participant
1	2470	2429	41		18	23	59:1
2	2004	1991	13	6	6	1	153:1
3	9977	9910	67	2		65	148:1
4	13393	13196	197	15	1	181	67:1
5	507	501	6	6			84:1
6	1944	1938	6			6	323:1
7	560	544	16	16			34:1
8	5619	5437	182	61		121	30:1
9	10485	10094	391	215	36	140	26:1
10	1138	1124	14	13		1	80:1
11	8868	8049	819	601	124	94	10:1
12	4245	3951	294		160	134	13:1
13	1463	1421	42	42			34:1
14	2821	2698	123	47	76		22:1
15	2858	2775	83	7	13	63	33:1
16	1338	1330	8		6	2	166:1
17	634	582	52	52			11:1
Total	**70324**	**67970**	**2354**	**1083**	**440**	**831**	**29:1**

3.4 | Feature ranking

To determine which features and sensors were important markers for agitation behaviors, feature ranking was performed for each personalized model. The relative weight of each feature in each personalized model is shown visually in heatmaps in Figure 4 and numerically in the supporting information. There was considerable variability between participants in the most important sensor features for classifying agitation versus non-agitation (Figure 4A). Looking across participants, the accelerometer features were most heavily weighted, in particular the mean, minimum, and standard deviation of the ACC signal, as well as the spectral entropy of the ACC signal. Phasic EDA measures, maximum heart rate, and mean temperature were also weighted as important in a subset of participants. Examining the differences between participants in the top 10 most heavily weighted features, participants appeared in two groupings: 8 participants had a median of 6 (range 5–8) ACC features in the top 10, while in the remaining 9 participants, there was a median of 1 (range 0–2) ACC features in their top 10. These nine participants had a combination of EDA features (median 3, range 1–6), BVP features (median 4, range 3–7), and TEMP features (median 1, range 0–2) in their top 10.

Examining the personalized models by behavior category, features from across all four sensors were associated with motor agitation including the standard deviation, mode, and entropy of the EDA signal, possibly representing motion artifacts in the EDA signals (Figure 4B). Interestingly, EDA features were weighted similarly to the ACC movement features and BVP features, with participants having a median of 3 (range 0–7) EDA features in their top 10, a median of 2 (range 0–7) ACC features, and a median of 3 (range 0–7) of BVP features. In six participants, the mean temperature was heavily weighted, while the maximum heart rate was heavily weighted in three participants.

Events consisting primarily of verbal aggression were most strongly associated with BVP features including the maximum and standard deviation of the heart rate, and the mean temperature (Figure 4C), although there was considerable individual variability in the feature pattern. For example, six of the eight participants with verbal aggression had few ACC features (median of 1, range 0–2), while two had at least seven ACC features in their top 10.

Overall, the ACC features were most important for classifying events involving physical aggression (Figure 4D), including the standard deviation and several Teager energy features of the signal. Across all participants, there were a median of 6 (range 0–8) ACC features in their top 10, although two groupings are again seen, with six participants having a median of seven ACC features (range 5–8) and four participants having a median of one (range 0–2). Those four participants with fewer ACC features had a predominance of BVP features (median 6, range 4–8) such as heart rate variability and interquartile range.

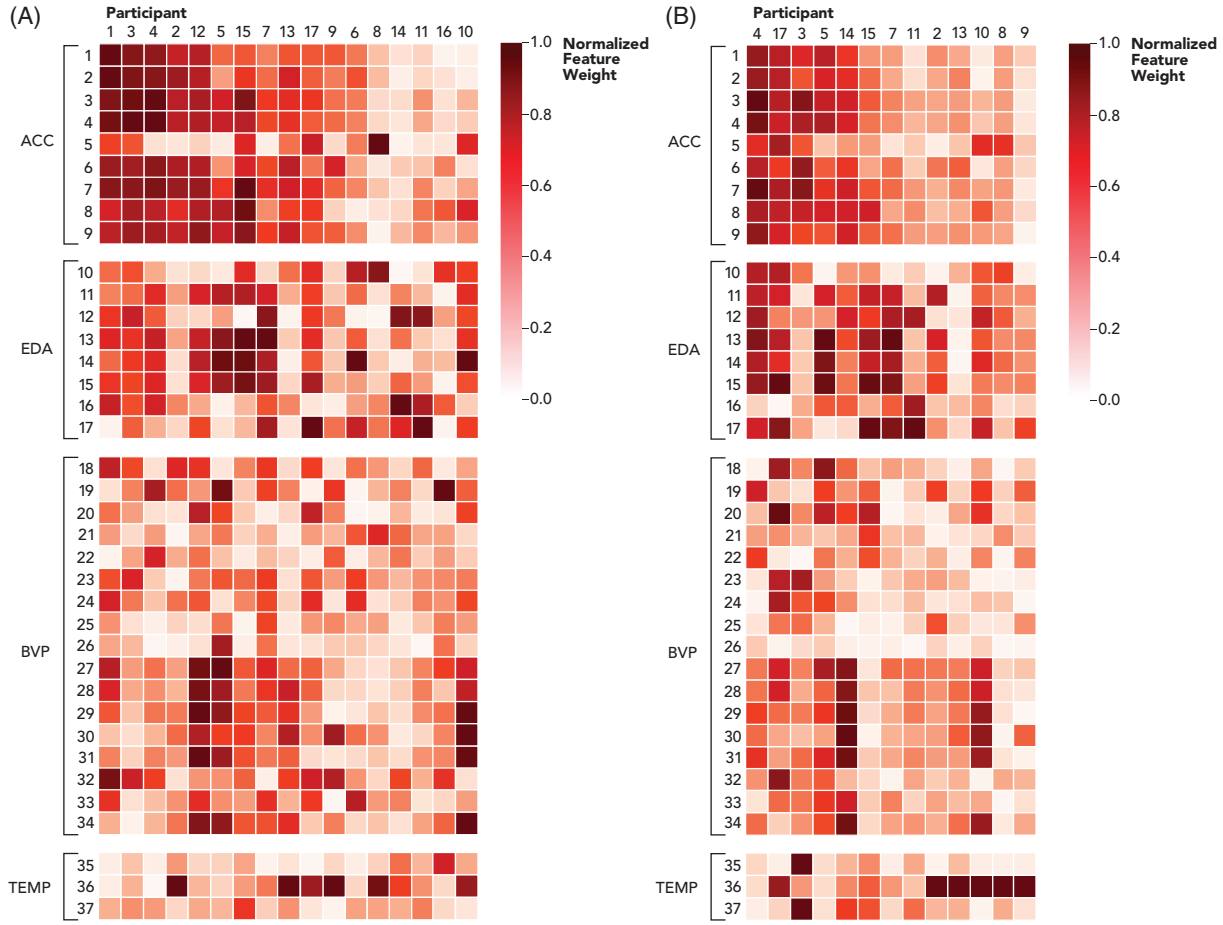

FIGURE 4 Feature ranking for personalized models for (A) all agitation events, (B) motor agitation, (C) verbal aggression, and (D) physical aggression. Refer to legend for the label for each feature extracted from the accelerometer (ACC), electrodermal activity (EDA), blood volume pulse (BVP), and temperature (TEMP) sensors

ACC	1	Teager Energy Maximum	BVP	18	Heart Rate Variability
	2	Teager Energy Minimum		19	Mean Inter-Beat-Interval
	3	Teager Energy Mean		20	Power Spectral Density
	4	Teager Energy Simple Square Integral		21	Spectral Energy Ratio
	5	Mean		22	Tachogram Power Low
	6	Minimum		23	Tachogram Power Medium
	7	Standard Deviation		24	Tachogram Power High
	8	Inter-Quartile Range		25	Tachogram Energy Ratio
	9	Spectral Entropy		26	Mean
	10	DC Power		27	Maximum
EDA	11	Tonic Max of the Signal Derivative		28	Minimum
	12	Tonic Mode of the Signal		29	Standard Deviation
	13	Phasic Area under the Signal		30	Inter-Quartile Range
	14	Phasic Number of Peaks		31	Total Average Power
	15	Standard Deviation		32	Spectral Entropy
	16	Spectral Entropy		33	Energy
	17	Entropy		34	DC power
			TEMP	35	Slope Angle
				36	Mean
				37	Standard Deviation

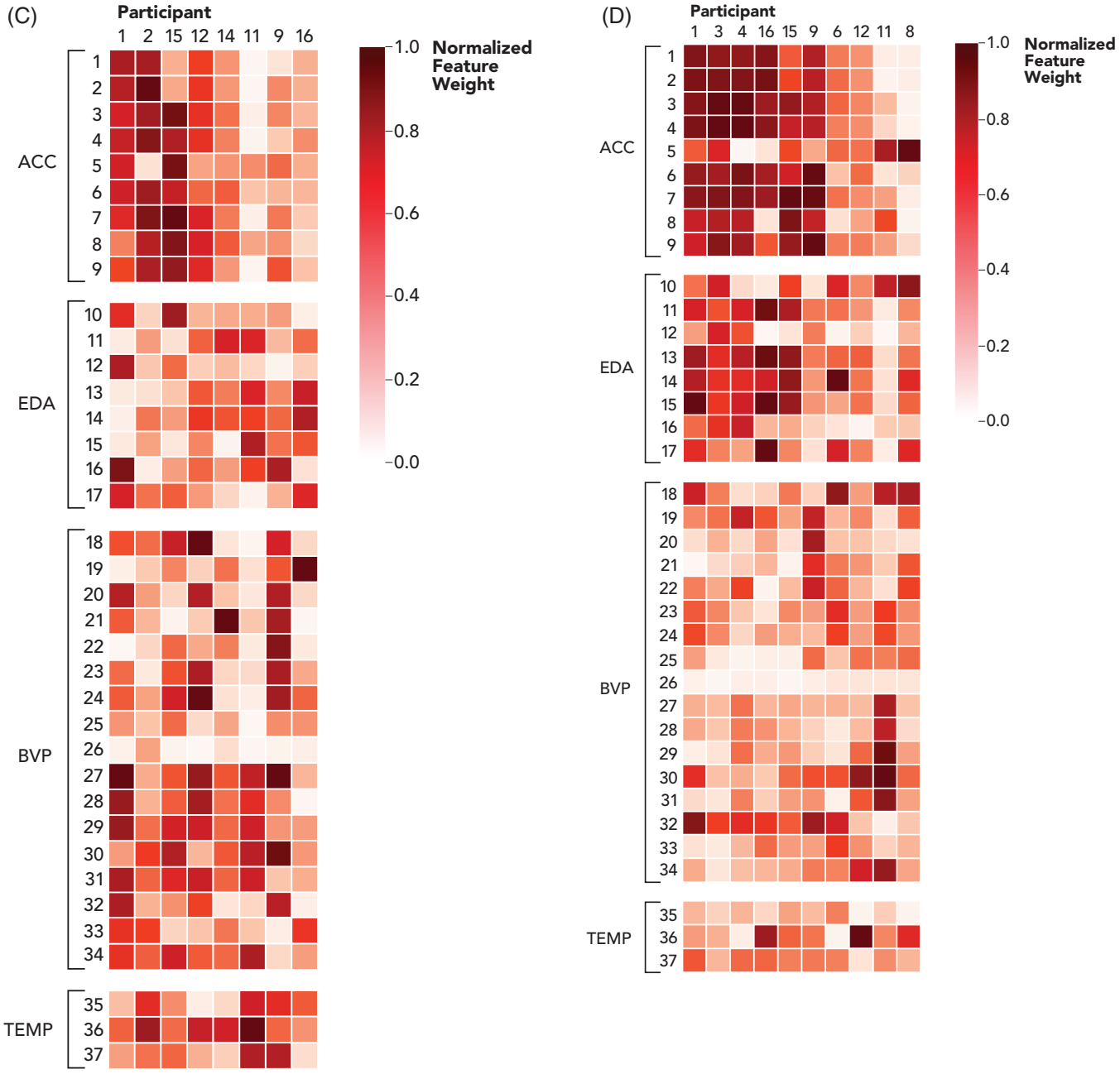

FIGURE 4 Continued

4 | DISCUSSION

In this study, we make use of rigorously annotated multimodal sensor data to detect episodes of agitation in people with dementia. Rather than looking for common signals across all participants, we have developed models that are personalized based on the individual and that are based upon the type of behavior being exhibited. Our results demonstrate that it is possible to classify behaviors as agitated or not agitated with fair to good performance in 1-minute windows. We have confirmed that personalized models significantly improve upon the performance of generic models. To further support the value of personalized and behavior-specific models, we confirm that the sensor features weighted by the machine learning models varies both by individual and by type of behavior.

These findings are an important step to move beyond correlation and toward predictive analytics for behavioral symptoms in dementia. While there are discernable patterns in wearable sensor signals moment to moment that are characteristic of agitated behaviors, these patterns are largely individual. Unlike studies that have used commercial-grade wearable devices with proprietary data processing algorithms, we used machine learning to delve into a large and complex raw sensor dataset in 1-minute time frames. This approach has the dual advantages of making use of the full richness and detail of the data and of improving signal recognition within the data.[37–39] One limitation of

this approach was that it required a research-grade multimodal sensor device with a short battery life. We were thus unable to examine behaviors occurring at night. Wearable technologies are rapidly advancing and affordable, clinical-grade multi-modal devices with advanced data processing and synchronization capabilities and week-long battery life are on the horizon.

By using interpretable models in this study, our results provide some insights as to which features of the sensor data are most important in predicting the presence of agitation. As expected, many of the most strongly weighted features were related to movement. Physical activity in people with severe dementia has previously been shown to be correlated with degree of agitation.[40] However, any algorithms that use motor activity alone to detect agitation risks either misclassifying healthy physical activity, such as dancing, or pathologizing wandering behaviors. Similarly, a focus on motor activity alone may miss other types of agitation, such as verbal aggression, or agitation in people with dementia who are non-ambulatory. There is also a need to tease apart "non-agitation," a category in this study which encompasses positive behaviors (social engagement, pleasurable activities), neuropsychiatric symptoms such as apathy, and periods of rest or sleep. Identifying patterns consistent with these behaviors would allow us to go beyond markers of agitation or distress to track positive engagement. In this study, we have not specifically proposed a threshold or cut-off for defining agitation or non-agitation. Striking a balance between the sensitivity of the classification algorithm to correctly detect agitation and the risk of false positives depends largely on the application of the algorithm and an examination of the costs of false positives versus false negatives.

Our findings suggest that features derived from multiple different sensors are important for the development of digital biomarkers for behavioral symptoms. Vulnerability to stress is considered a determinant of BPSD and expression of behavioral symptoms can be conceptualized as a stress response.[1] As such, digital biomarkers of stress are also relevant in identifying and monitoring BPSD. Previous studies have demonstrated that electrodermal activity and heart rate are reliable digital biomarkers of stress in young, healthy populations, and that there is a large inter-individual variability in expression of stress response.[41] In people with dementia, studies have found a correlation between heart rate and agitation,[42] and electrodermal activity and agitation.[15,16] In further support of the need of multimodal sensors, we have previously shown that incorporating data from several different sensors outperforms a single sensor for detecting agitation.[20]

The development and clinical validation of algorithms to classify behaviors are challenging and require that data be accurately annotated moment to moment as to the presence or absence of the behavior of interest. An important strength of this study is a well-annotated sensor dataset. Clinically important episodes of agitation were flagged by staff and reviewed in videos to establish as best as possible the start and end time. However, one limitation is that it is possible that agitation events took place which were not captured on video or observed by staff. Future work can make use of positive-unlabeled learning algorithms,[43] an approach in which models are trained on both positive data (agitation events) and unlabeled data (combination of normal and unreported agitation events) to address this limitation.

It is important to note that most of the poorly performing personalized models were due to an inadequate amount of agitation events for training the supervised classification models. Even in this cohort with severe BPSD, episodes of clinically important agitation are relatively rare events. Given the degree of imbalance in the data between agitation and normal events, future studies should consider alternatives to supervised classification, such as the use of unsupervised learning methods, such as anomaly detection and/or one-class classification approaches.[44,45] The small sample size in this study is an important limitation. Our sample size was limited by the large data storage requirements for the videos and the resource-intensive annotation process. Larger data sets are needed to validate and ensure the generalizability of the algorithms. It is possible that generic models using a larger dataset may approach the performance of personalized models. In this study, models were built using sensor data alone, but we found that participants could be grouped based on their patterns of sensor features associated with agitation. There may be clinical or demographic variables, such as dementia subtype or severity, that distinguishes these groups. Including these variables in the models would be expected to help generic model performance. To address these issues, future large studies are required in clinically well-characterized cohorts, and using efficient methods for labeling of the sensor data with behavioral symptoms in real time.

With the revolution in digital medicine underway comes the need for appropriately validated digital biomarkers.[46] There is increasing recognition of the value of personalized metrics for health and well-being based on longitudinal monitoring of symptoms. The most likely clinical use of BPSD digital biomarkers is for n = 1 monitoring, such as tracking symptoms or behaviors over time and in response to treatment, thus there is a clear need for agitation detection algorithms to be accurate within individuals over time. Our results support that there are individualized patterns of digital biomarkers for agitation in dementia, and that to improve the accuracy of agitation detection algorithms, future studies are needed to characterize these individual patterns.

ACKNOWLEDGMENTS

This study was funded by the Alzheimer Society of Canada and the Walter and Maria Schroeder Institute for Brain Innovation and Recovery.

CONFLICTS OF INTEREST

Outside of this work, AI has received research grants from the Canadian Institutes of Health Research; AGE-WELL NCE; Province of Ontario; and the Centre for Aging + Brain Health Innovation and an Academic Scholars Award, Department of Psychiatry, University of Toronto. SS has received postdoctoral funding from the University Health Network, AGE-WELL NCE, and Royal Bank of Canada. KN has received grants from the Social Sciences and Humanities Research Council. AW received a research grant from the Nursing Research Interest Group, Ryerson University. LSM is president of the not-for-profit Canadian Gerontological Nursing Association. SSK has received a research grant from the Alzheimer's Society and a honorarium from

the International Conference on Gerontechnology, Hong Kong. BY and AM have no disclosures.

ORCID

Andrea Iaboni https://orcid.org/0000-0003-4268-6832

REFERENCES

1. Kolanowski A, Boltz M, Galik E, et al. Determinants of behavioral and psychological symptoms of dementia: a scoping review of the evidence. *Nurs Outlook* 2017;65:515-529. https://doi.org/10.1016/J.OUTLOOK.2017.06.006.
2. Kales HC, Gitlin LN, Lyketsos CG, Detroit Expert Panel on Assessment and Management of Neuropsychiatric Symptoms of Dementia. Management of neuropsychiatric symptoms of dementia in clinical settings: recommendations from a multidisciplinary expert panel. *J Am Geriatr Soc* 2014;62:762-769. https://doi.org/10.1111/jgs.12730.
3. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. *BMJ*(Clinical Research Ed) 2015;350:h369. https://doi.org/10.1136/bmj.h369.
4. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. *JAMA* 2014;311:1052-1060. https://doi.org/10.1001/jama.2014.304.
5. Herrmann N, Lanctôt KL, Sambrook R, et al. The contribution of neuropsychiatric symptoms to the cost of dementia care. *Int J Geriatr Psychiatry* 2006;21:972-976. https://doi.org/10.1002/gps.1594.
6. Khan SS, Ye B, Taati B, Mihailidis A. Detecting agitation and aggression in people with dementia using sensors-A systematic review. *Alzheimer's Dementi* 2018;14:824-832. https://doi.org/10.1016/j.jalz.2018.02.004.
7. Teipel S, König A, Hoey J, et al. Use of nonintrusive sensor-based information and communication technology for real-world evidence for clinical trials in dementia. *Alzheimer's Dementi* 2018;14:1216-1231. https://doi.org/10.1016/j.jalz.2018.05.003.
8. Nagels G, Engelborghs S, Vloeberghs E, van Dam D, Pickut BA, de Deyn PP. Actigraphic measurement of agitated behaviour in dementia. *Int J Geriatr Psychiatry* 2006;21:388-393. https://doi.org/10.1002/gps.1483.
9. Knuff A, Leung RH, Seitz DP, Pallaveshi L, Burhan AM. Use of actigraphy to measure symptoms of agitation in dementia. *AmJ Geriatr Psychiatr* 2019;27:865-869. https://doi.org/10.1016/j.jagp.2019.02.013.
10. Mahlberg R, Walther S. Actigraphy in agitated patients with dementia. Monitoring treatment outcomes. *Z Gerontol Geriatr* 2007;40:178-184. https://doi.org/10.1007/s00391-007-0420-z.
11. Cohen-Mansfield J, Werner P, Culpepper WJ, Wolfson M, Bickel E. Assessment of ambulatory behavior in nursing home residents who pace or wander: a comparison of four commercially available devices. *Dement Geriatr Cogn Disord* 1997;8:359-365. https://doi.org/10.1159/000106656.
12. Valembois L, Oasi C, Pariel S, Jarzebowski W, Lafuente-Lafuente C, Belmin J. Wrist actigraphy: A simple way to record motor activity in elderly patients with dementia and apathy or aberrant motor behavior. *J Nutr Health Aging* 2015;19:759-64. https://doi.org/10.1007/S12603-015-0530-Z.
13. Goerss D, Hein A, Bader S, et al. Automated sensor-based detection of challenging behaviors in advanced stages of dementia in nursing homes. *Alzheimer's Dement* 2019;16:672-80. https://doi.org/10.1016/J.JALZ.2019.08.193.
14. Kikhia B, Stavropoulos TG, Meditskos G, et al. Utilizing ambient and wearable sensors to monitor sleep and stress for people with BPSD in nursing homes. *J Ambient Intelligence Humanized Comput* 2018;9:261-273. https://doi.org/10.1007/s12652-015-0331-6.
15. Kikhia B, Stavropoulous TG, Andreadis S, et al. Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Basel, Switzerland)* 2016;16:1989. https://doi.org/10.3390/S16121989.
16. Melander CA, Kikhia B, Olsson M, Wälivaara B-M, Sävenstedt S. The impact of using measurements of electrodermal activity in the assessment of problematic behaviour in dementia. *Dement Geriatr Cognit Disord Extra* 2018;8:333-347. https://doi.org/10.1159/000493339.
17. Nesbitt C, Gupta A, Jain S, Maly K, Okhravi HR. Reliability of Wearable Sensors to Detect Agitation in Patients with Dementia. Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology – ICBBT '18, New York, New York, USA: ACM Press; 2018, p. 73–77. https://doi.org/10.1145/3232059.3232067.
18. Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacol* 2016;41:1691-1696. https://doi.org/10.1038/npp.2016.7.
19. Insel TR. Digital Phenotyping: technology for a new science of behavior. *JAMA* 2017;318:1215-1216. https://doi.org/10.1001/jama.2017.11295.
20. Spasojevic S, Nogas J, Ye B, et al. A pilot study to detect agitation in people living with dementia using multi-modal sensors. *J Healthc Inform Res* 2021;5:342-58.
21. Ye B, Khan SS, Chikhaoui B, et al. Challenges in collecting big data in a clinical environment with vulnerable population: lessons learned from a study using a multi-modal sensors platform. *Sci Eng Ethics* 2019;25:1447-1466. https://doi.org/10.1007/s11948-018-0072-y.
22. Empatica: E4 Wristband from hepatica 2018. https://www.empatica.com/en-eu/research/e4 (accessed March 24, 2021).
23. Ollander S, Godin C, Campagne A, Charbonnier S. A comparison of wearable and stationary sensors for stress detection. 2016. IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE; 2016, p. 004362–004366. https://doi.org/10.1109/SMC.2016.7844917.
24. van Lier HG, Pieterse ME, Garde A, et al. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behavior Res Methods* 2020;52:607-629. https://doi.org/10.3758/s13428-019-01263-9.
25. Ghandeharioun A, Fedor S, Sangermano L, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE; 2017, p. 325–332. https://doi.org/10.1109/ACII.2017.8273620.
26. Pietilä J, Mehrang S, Tolonen J, et al. *Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities*, Springer, Singapore; 2018, p. 145–148. https://doi.org/10.1007/978-981-10-5122-7_37.
27. Penko M, Quirt H, Schindel Martin L, Iaboni A. Behaviour assessment tools in long-term care homes in Canada: a survey. *Aging Ment Health* 2020:1-12. https://doi.org/10.1080/13607863.2020.1793904.
28. Behavioural Supports Ontario. Behavioural Supports Ontario-Dementia Observation System (BSO-DOS) n.d. https://brainxchange.ca/BSODOS (accessed April 19, 2021).
29. Begum S. Sensor signal processing to extract features from finger temperature in a case-based stress classification scheme. 2009 IEEE International Symposium on Intelligent Signal Processing, IEEE; 2009, p. 193–198. https://doi.org/10.1109/WISP.2009.5286562.
30. Chen Y-C, Hsiao C-C, Zheng W-D, Lee R-G, Lin R. Artificial neural networks-based classification of emotions using wristband heart rate monitor data. *Medicine (Baltimore)* 2019;98:e16863. https://doi.org/10.1097/MD.0000000000016863.
31. Soleymani M, Villaro-Dixon F, Pun T, Chanel G. Toolbox for Emotional feAture extraction from physiological signals (TEAP). *Frontiers in ICT* 2017;4:1. https://doi.org/10.3389/fict.2017.00001.

32. Greco A, Valenza G, Scilingo EP. *Advances in Electrodermal Activity Processing with Applications for Mental Health*. Cham: Springer International Publishing; 2016. https://doi.org/10.1007/978-3-319-46705-4.
33. Posada-Quintero HF, Chon KH. Innovations in electrodermal activity data collection and signal processing: a systematic review. *Sensors* 2020;20:479. https://doi.org/10.3390/s20020479.
34. Fowles DC, Christie MJ, Edelberg R, Grings WW, Lykken DT, Venables PH. Committee report. Publication recommendations for electrodermal measurements. *Psychophysiology* 1981;18:232-239. https://doi.org/10.1111/j.1469-8986.1981.tb03024.x.
35. Bankole A, Anderson M, Smith-Jackson T, et al. Validation of noninvasive body sensor network technology in the detection of agitation in dementia. *Am J Alzheimer's Dis Other Dement* 2012;27:346-354. https://doi.org/10.1177/1533317512452036.
36. Bankole A, Anderson M, Knight A, et al. Continuous, non-invasive assessment of agitation in dementia using inertial body sensors. Proceedings of the 2nd Conference on Wireless Health – WH '11, New York, New York, USA: ACM Press; 2011, p. 1. https://doi.org/10.1145/2077546.2077548.
37. Obermeyer Z, Emanuel EJ. Predicting the future – Big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216-1219. https://doi.org/10.1056/NEJMp1606181.
38. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton D, Clifford GD. Machine Learning and Decision Support in Critical Care. *Proc IEEE* 2016;104:444-466. https://doi.org/10.1109/JPROC.2015.2501978.
39. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiol Meas* 2018;39:05TR01. https://doi.org/10.1088/1361-6579/aabf64.
40. Ishimaru D, Tanaka H, Nagata Y, Takabatake S, Nishikawa T. Physical activity in severe dementia is associated with agitation rather than cognitive function. *Ame J Alzheimer's Dis Dement* 2020;35:1533317519871397. https://doi.org/10.1177/1533317519871397.
41. Smets E, Rios Velazquez E, Schiavone G, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *NPJ Digital Medicine* 2018;1:67. https://doi.org/10.1038/s41746-018-0074-9.
42. Favela J, Cruz-Sandoval D, Morales-Tellez A, Lopez-Nava IH. Monitoring behavioral symptoms of dementia using activity trackers. *J Biomed Inform* 2020;109:103520. https://doi.org/10.1016/j.jbi.2020.103520.
43. Bekker J, Davis J. Learning from positive and unlabeled data: a survey. *Machine Learn* 2020;109:719-760. https://doi.org/10.1007/s10994-020-05877-5.
44. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. *Knowled Eng Rev* 2014;29:345-374. https://doi.org/10.1017/S026988891300043X.
45. Khan SS, Mishra PK, Javed N, et al. Unsupervised deep learning to detect agitation from videos in people with dementia. *IEEE Access* (Accepted) 2021.
46. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digit Med* 2019;2. https://doi.org/10.1038/s41746-019-0090-4.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Iaboni A, Spasojevic S, Newman K, et al. Wearable multimodal sensors for the detection of behavioral and psychological symptoms of dementia using personalized machine learning models.. *Alzheimer's Dement*. 2022;14:e12305. https://doi.org/10.1002/dad2.12305