scientific reports

OPEN

Machine learning-based predictive models for the occurrence of behavioral and psychological symptoms of dementia: model development and validation

Eunhee Cho[1], Sujin Kim[2], Seok-Jae Heo[3], Jinhee Shin[4], Sinwoo Hwang[5], Eunji Kwon[5], SungHee Lee[6], SangGyun Kim[6] & Bada Kang[1]✉

The behavioral and psychological symptoms of dementia (BPSD) are challenging aspects of dementia care. This study used machine learning models to predict the occurrence of BPSD among community-dwelling older adults with dementia. We included 187 older adults with dementia for model training and 35 older adults with dementia for external validation. Demographic and health data and premorbid personality traits were examined at the baseline, and actigraphy was utilized to monitor sleep and activity levels. A symptom diary tracked caregiver-perceived symptom triggers and the daily occurrence of 12 BPSD classified into seven subsyndromes. Several prediction models were also employed, including logistic regression, random forest, gradient boosting machine, and support vector machine. The random forest models revealed the highest area under the receiver operating characteristic curve (AUC) values for hyperactivity, euphoria/elation, and appetite and eating disorders; the gradient boosting machine models for psychotic and affective symptoms; and the support vector machine model showed the highest AUC. The gradient boosting machine model achieved the best performance in terms of average AUC scores across the seven subsyndromes. Caregiver-perceived triggers demonstrated higher feature importance values across the seven subsyndromes than other features. Our findings demonstrate the possibility of predicting BPSD using a machine learning approach.

The number of people living with dementia is estimated to be more than 50 million worldwide; with the increasing aging population, this number is expected to triple by 2050[1]. Behavioral and psychological symptoms of dementia (BPSD), also known as neurocognitive symptoms, are a heterogeneous group of non-cognitive symptoms and behaviors such as agitation, aggression, apathy, and depression that manifest in individuals diagnosed with dementia[2]. Almost 90% of individuals with dementia at all stages and etiologies are affected by these symptoms[3]. They are increasingly recognized as the most complex, challenging, and costly aspects of dementia care[4], and result in decreased independence in activities of daily living and quality of life[5], nursing home placement[6], healthcare utilization[7], and increased caregiver burden and depression[8] for individuals with dementia.

The factors that have been identified as contributing to BPSD can be categorized as (1) those related to individuals with dementia, such as dementia-related neurobiological factors, unmet needs, and premorbid personality; (2) those related to caregivers, such as communication approach; and (3) environmental triggers such as lack of stimuli and environmental change[9]. They cause symptoms independently or in combination with other aspects[9]. Disturbances in circadian rhythm, including impaired sleep and inadequate physical activity, have also been identified as stressors that can trigger BPSD[10,11]. Recent empirical research has identified an association

[1]Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, 50-1, Yonsei-Ro, Seodaemun-gu, Seoul 03722, Republic of Korea. [2]Department of Nursing, Yong-In Arts and Science University, Gyeonggi-do, Korea. [3]Division of Biostatistics, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea. [4]College of Nursing, Woosuk University, Jeollabuk-do, Korea. [5]Korea Armed Forces Nursing Academy, Daejeon, Korea. [6]BRFrame Inc, Seoul, Korea. ✉email: bdkang@yuhs.ac

between disruption of circadian rhythms and BPSD, including depression, anxiety, agitation[12], and sundowning-related aggressive symptoms[13]. However, given that systematic and continuous observation of sleep and activity levels is challenging to assess using traditional methods that rely on retrospective reports of caregiver observation, only a few methodologically rigorous studies have examined the influence of sleep and activity levels on such symptoms[14].

Although a downward trajectory for the cognitive and functional decline over the course of dementia is expected, the manifestation of BPSD varies among individuals[9]. Current evidence suggests that person-centered non-pharmacological interventions that match the needs of persons with dementia and their abilities are a first-line treatment for managing BPSD. The most appropriate interventions for those with dementia should be selected individually after considering the causes and patterns of BPSD[15]. Predicting BPSD by identifying contributing factors and monitoring triggers is the first step in selecting and implementing individually customized non-pharmacological interventions to prevent and manage target symptoms[15].

Machine learning and wearable technology have immense potential to overcome the methodological limitations of existing dementia research through the precise analysis of clinical information derived from digital devices[16]. Wearable technology such as actigraphy allows for continuous biometric monitoring, including levels of sleep and activity in everyday conditions, and for connecting them with clinical symptoms[16,17]. Machine learning facilitates the identification of underlying patterns and relationships between variables directly from data and the development of data-driven prediction models[18]. Although machine learning has been employed to develop predictive models for the incidence and detection of Alzheimer's disease[19], this analytic technique has rarely been applied to research on BPSD.

Hence, machine learning models are leveraged in this study to predict the occurrence of BPSD among community-dwelling older adults living with dementia based on various factors, including actigraphy-measured sleep and physical activity levels and diary-based caregiver-perceived symptom triggers.

Methods

Study design.
This study utilized a prospective observational design with three-wave data collection to build predictive models for BPSD subsyndromes. The second wave of data collection was conducted after the first wave, with repeated measures from participants in the first wave who had agreed to participate in the second wave. A detailed description of the first and second waves of data collection is reported elsewhere[20]. In the third wave of data collection, a validation dataset was collected from new participants independently of the first and second wave data. We employed the first and second wave data for model training (i.e., the training dataset) and the third wave data for external validation (i.e., the test dataset).

We employed a standard mining methodology that comprised four steps: (1) data acquisition, (2) data preprocessing (e.g., data cleaning, class imbalance training, and dataset class optimization), (3) model learning, and (4) model evaluation[21].

Recruitment and data collection.
The first wave of data collection was conducted between June 2018 and June 2019. Eligible older adults with dementia living at home were recruited via on-site visits from outpatient neurological clinics at two tertiary hospitals and daycare centers in Seoul and the broader Gyeonggi region in Korea. The second wave of data collection, which involved first-wave participants who agreed to continue the study, was administered between July 2019 and June 2020. For external validation, eligible participants were recruited between July 2020 and November 2020 from an outpatient neurological clinic, where the first and second waves of data collection were conducted. The inclusion criteria applied to the three-wave data collection were (1) being at least 65 years old, (2) having a diagnosis of dementia, and (3) having a score of less than 24 on the Korean version of the Mini-Mental State Examination (K-MMSE)[22].

Eligibility screening and data collection were performed by trained research staff.

After eligibility was established, trained research staff collected demographic and health data through interviews with family caregivers and older adults with dementia. Furthermore, chart reviews and standardized scales for physical, functional, and neuropsychological assessments were administered. Following the baseline assessment, the participants wore an actigraphy device on their wrists continuously for two weeks, and primary caregivers logged BPSD in the symptom diary daily for 14 consecutive days.

Ethical considerations.
All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the Declaration of Helsinki (1964) and its later amendments or comparable ethical standards. Institutional review board approval was obtained from the Yonsei University Health System Severance Hospital (IRB 4-2018-0348, 4-2019-0314, 4-2020-0454) and Ilsan Hospital (IRB 2018-10-002-001, 2019-08-012-001). Legal representatives of all the participants provided written informed consent before enrollment after receiving a full explanation of the study procedures. The participants also provided verbal assent and written informed consent was obtained when possible.

Features.
Demographic and health data. At baseline, demographic and health data comprised age, sex, marital status, education level, dementia diagnosis, and neurological and psychiatric medications.

Cognitive and functional status. Scores (range 0–30) on the K-MMSE were used to assess cognitive functioning, with lower scores indicating greater cognitive deficits[22]. For the K-MMSE, Cronbach's α was 0.91 in older Korean adults with dementia[23]. The severity of dementia was measured via the Korean version of the expanded Clinical Dementia Rating (CDR) scale, which assesses six functional domains—memory, orientation, judgment

and problem-solving, community affairs, home and hobbies, and personal care[24]. The summed score ranged from 0 (none) to 5 (terminal dementia), and good inter-rater reliability for the overall CDR ratings was confirmed in Korean patients with dementia (kappa value range: 0.86–1.00)[25]. Functional independence was evaluated using the Korean version of Activities of Daily Living (K-ADL), which consists of seven items rated on a 3-point Likert-type scale, with higher scores indicating more severe levels of dependency[26]. The K-ADL was validated for older Korean adults with dementia, with good reliability (Cronbach's α = 0.94)[26].

Personality type.   The family caregiver informant-rated premorbid personality traits of older adults with dementia were assessed using the Korean version of the Big Five Inventory (BFI-K)[27]. The BFI-K constitutes 15 items rated on a 5-point Likert-type scale that measures 5 domains of personality traits: openness, conscientiousness, neuroticism, extraversion, and agreeableness. The internal consistency of the BFI-K was good; Cronbach's α ranged from 0.67 to 0.82[27].

Actigraphy data: nighttime sleep and physical activity.   Older adults with dementia were fitted with a wrist-worn actigraphy device (ActiGraph wGT3X-BT, ActiGraph Corporation, Pensacola, FL, USA), which they wore all day for 14 consecutive days. The participants were instructed to remove the device when bathing or for a few minutes as needed. Previous validity studies have demonstrated that wrist actigraphy is a reliable and suitable method for objectively measuring sleep–wake cycles in older adults with dementia[28,29]. Raw acceleration data were collected along the three axes. We used ActiLife (version 6.13.3, Pensacola, FL, US) software to export the data and process raw acceleration data to sleep and physical activity parameters using vector magnitude count in 60-s epoch data (i.e., counts per minute). The vector magnitude is calculated as the square root of the sum of the squares of acceleration for each of the three axes (x, y, z). The Cole-Kripke algorithm was applied to score a one-minute epoch as asleep or awake[30]. Moreover, the previous night's sleep parameters were employed to predict BPSD the following day. In this study, nighttime sleep was defined as the period between 20:00 (8:00 pm) and 08:00 (8:00 am). The following nighttime sleep parameters were generated: total sleep time, wake time after sleep onset, sleep efficiency, defined as the ratio of sleep duration over the assumed sleep period (total sleep time/ [total sleep time + wake time after sleep onset] × 100), number of awakenings, and mean awake length (wake time after sleep onset/number of awakenings). The following physical activity parameters were also generated: energy expenditure (calories burned) in kcal per day, metabolic equivalents per day, total time spent in moderate-to-vigorous physical activity per day, percentage of time spent in moderate-to-vigorous physical activity per day, and the number of steps per day. We employed physical activity parameters measured the same day, which reflected the physical conditions during the day when BPSD occurred.

Symptom diary data: BPSD and caregiver-perceived symptom triggers.   A symptom diary that comprised a structured, easy-to-use checklist modeled on the Neuropsychiatric Inventory (NPI) was developed to assess the presence and severity of BPSD (i.e., delusions, hallucinations, agitation/aggression, depression/dysphoria, anxiety, elation/euphoria, apathy/indifference, disinhibition, irritability/lability, aberrant motor behaviors, sleep and nighttime behaviors, and appetite and eating disorders) daily[31]. It also included a checklist that assessed caregiver-perceived triggers of BPSD (i.e., hunger/thirst, urination/bowel movement, pain/discomfort, sleep disturbance, noise, light, temperature), interpersonal triggers (i.e., factors related to the person(s) who were present), and changes in the environment. Caregivers were also instructed to check "other causes" in the symptom diary if the perceived trigger was a factor that could not be categorized under any of the options listed in the checklist, and then, list the factors. Family caregivers were instructed to check all options that were perceived as triggers of BPSD on the same day when the symptoms had occurred. The symptom diary was designed to overcome recall bias (e.g., the NPI is based on the caregiver's two-week retrospective rating), enable daily monitoring of the occurrence of BPSD, and link symptoms to triggers daily[32].

Recent studies have established that clustering several individual BPSD that are highly correlated and co-occur enhances the clinical utility of the assessment of BPSD, thus allowing for a more meaningful interpretation of the study findings and increasing power by raising the number of participants who endorsed the symptom cluster rather than the individual symptoms alone[2,33,34]. Based on previous NPI factor analysis studies, we clustered certain individual symptoms into three subsyndromes: *psychotic* symptoms (hallucination and delusion), *affective* symptoms (depression, anxiety, and apathy), and *hyperactivity* symptoms (agitation/aggression, disinhibition, and irritability)[34–37]. As prior studies have demonstrated that euphoria/elation, aberrant motor behaviors, sleep and nighttime behaviors, and appetite and eating disorders do not load into any clusters[33,34,36–38], we analyzed them as individual subsyndromes consisting of only one symptom.

Data preprocessing.
Missing actigraphy data were encountered for two main reasons: the improper wearing of the device and lack of participant compliance (e.g., constant removal of the device or not wearing the device). The number of participants with missing actigraphy data was 81/225 (36%). The mean number of days per person with missing actigraphy data was 0.9. The occurrence rates of BPSD were similar regardless of missing actigraphy data (Supplementary Table 1). Therefore, multivariate imputation was applied using chained equations[39] to address the missing actigraphy data. Before training the models, we applied min–max normalization for continuous features. For categorical features, target encoding was employed instead of one-hot encoding. Target encoding reduces feature dimensions by converting categorical features to numerical values derived from target variables, assuming that a categorical feature is related to the outcomes[40]. There was an issue of outcome class imbalance for BPSD subsyndromes. While 26.8% of the participants exhibited affective symptoms, only 4.4% and 4.7% exhibited aberrant motor behaviors and euphoria/elation, respectively (Table 2). Researchers in various disciplines have prioritized the class imbalance problem and suggested strategies to address the

issues of imbalanced data sets[41–43]. This study applied a synthetic minority oversampling technique to address the outcome class imbalance issue[44].

Predictive modeling.   Multiple machine learning approaches were selected for this study, including logistic regression, random forest[45], gradient boosting machine[46], and support vector machine[47]. We investigated each of these machine learning methods with a specific learning algorithm to gauge their effectiveness, and then selected the best-performing model that could predict each subsyndrome of BPSD[18]. Using logistic regression, the most common and well-established binary classifier[48], as the baseline model, we evaluated the degree to which the machine learning models improved performance over the baseline model.

To avoid overfitting, hyperparameter tuning through random search[49] was implemented with five-fold cross-validation for each machine learning method. Binary cross-entropy was employed as the evaluation criterion for five-fold cross-validation. The hyperparameters for tree complexity were considered for the random forest and gradient boosting machine models. The gradient boosting machine model was iteratively trained to minimize the loss function using stochastic gradient boosting. Thus, we considered the learning rate and number of trees for the gradient boosting machine. Various kernel functions such as linear, polynomial, and radial basis kernels can be utilized for the support vector machine[50]. Linear instead of nonlinear kernels, such as the radial basis function kernel, were used in this study to prevent overfitting in small datasets and calculate feature importance. For the support vector machine models, only the regularization hyperparameter was employed to determine the optimal model. All the selected hyperparameters are described in Supplementary Table 2. Feature importance analysis was performed to investigate the contribution of a range of features in predicting the seven subsyndromes of BPSD and to sort the importance of the top 10 influential features for prediction.

Statistical analysis.   Categorical variables are summarized as the number of participants with percentages and continuous variables as means with standard deviations. Furthermore, two-sample independent t-tests and Fisher's exact tests were used to compare the training and test dataset differences, respectively. The performances of the prediction models were compared and evaluated using several indices—accuracy, precision, sensitivity (recall), specificity, F1 score, and area under the receiver operating characteristic curve (AUC).

Statistical significance was set at $p < 0.05$, and all analyses were performed using R, version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) and Python, version 3.7 (Python Software Foundation, Wilmington, USA).

Results

Participant characteristics.   Table 1 presents the participant characteristics of both the training ($N = 187$ older adults with dementia) and test sets ($N = 35$ older adults with dementia) at baseline. The mean age of the participants was 80.4 years (SD 7.4) for the training dataset and 80.7 (SD 5.6) for the test dataset. The majority of participants were women (59% for the training dataset; 63% for the test dataset), married (62% for the training dataset; 63% for the test dataset), and with an educational level of middle school or below (57% for both the training and the test datasets). Table 1 summarizes the prevalence of BPSD among the participants.

Summary of actigraphy and symptom diary data.   Symptom diary and actigraphy data were collected over an average of 13.30 days (SD 2.08) per person (training set: mean [SD] = 13.35 [2.11]; test set: mean [SD] = 13.00 [1.91]).

Affective symptoms were the subsyndromes of BPSD that occurred most frequently (training set: 26.8%; test set: 17.1%), followed by hyperactivity symptoms (training set: 17.8%; test set: 12.7%). Aberrant motor behaviors, euphoria/elation, and appetite and eating disorders had lower frequency rates for both the training and test datasets. The average total sleep times per day were 7.7 h (for the training dataset) and 6.3 h (for the test dataset), and the average total nighttime sleep times per day were 5.9 h (training dataset) and 5.1 h (test dataset). The average energy expenditure was 459.9 kcal/day (training dataset) and 442.2 kcal/day (test dataset). The participants averaged 5908.7 steps (training dataset) and 5750.5 steps (test dataset) daily. Sleep disturbance (training dataset: 13.5%; test dataset: 8.4%) and interpersonal triggers (training dataset: 8.5%; test dataset: 10.8%) were the most frequently reported triggers. Table 2 provides the univariate analyses of actigraphy and symptom diary data.

Performance comparison for prediction models.   Table 3 presents the prediction performance of all prediction models based on the training dataset with five-fold cross-validation. Gradient boosting machine models showed higher AUC values compared to other prediction models for predicting hyperactivity (0.706), affective symptoms (0.747), and appetite and eating disorders. (0.816). While the support vector machine model demonstrated the highest AUC value (0.706) for psychotic symptoms, the random forest model exhibited the highest AUC value (0.942) for sleep and nighttime behavior. The logistic regression models denoted the highest AUC values for aberrant motor behaviors (0.822) and euphoria/elation (0.696).

Table 4 presents the prediction performance of all prediction models based on the test dataset, and Supplementary Fig. 1 illustrates the receiver operating characteristics and precision-recall curves for the test dataset. Compared with the logistic regression models, the machine learning models revealed better performance for all seven subsyndromes. Specifically, the random forest and gradient boosting machine models performed better than the logistic regression and support vector machine models for most subsyndromes. The random forest models exhibited higher AUC values than the other prediction models for predicting hyperactivity (0.835), euphoria/elation (0.968), and appetite and eating disorders (0.888). The gradient boosting machine models presented higher AUC values than the other prediction models for predicting psychotic symptoms (0.801), affective

Participant characteristic	Training set (*n* = 187)	Test set (*n* = 35)	*p*-Value
Age (years), mean (*SD*)[a]	80.41 (7.40)	80.69 (5.56)	0.801
Gender, *n* (%)			0.851
Male	76 (40.6)	13 (37.1)	
Female	111 (59.4)	22 (62.9)	
Marital status, *n* (%)			> 0.999
Married	115 (61.5)	22 (62.9)	
Widowed or divorced	72 (38.5)	13 (37.1)	
Education, *n* (%)			> 0.999
Elementary school or below	86 (46.0)	16 (45.7)	
Middle school	20 (10.7)	4 (11.4)	
High school	48 (25.7)	9 (25.7)	
College or above	33 (17.6)	6 (17.1)	
BFI[b], mean (*SD*)			
Openness	8.58 (2.90)	8.03 (2.77)	0.297
Conscientiousness	11.58 (2.77)	11.60 (2.77)	0.965
Neuroticism	7.56 (2.80)	7.40 (3.41)	0.771
Extraversion	8.53 (1.89)	8.83 (2.72)	0.536
Agreeableness	10.93 (2.96)	11.46 (2.37)	0.317
ADL[c], Sum of score, mean (*SD*)	11.27 (3.88)	8.94 (2.85)	< 0.001
Companionship and mental support	1.81 (0.75)	1.51 (0.66)	0.029
Transportation and shopping	1.65 (0.77)	1.31 (0.47)	< 0.001
Preparing meals	2.03 (0.85)	1.66 (0.76)	0.016
Managing a person's household	1.51 (0.63)	1.17 (0.38)	< 0.001
Managing medications	1.28 (0.55)	1.11 (0.32)	0.015
Communicating with others	1.32 (0.58)	1.09 (0.28)	< 0.001
Managing finances	1.65 (0.67)	1.26 (0.44)	< 0.001
MMSE[d] score, mean (*SD*)	16.16 (6.13)	16.09 (4.71)	0.946
CDR[e] score, *n* (%)			0.064
Questionable	26 (13.9)	12 (34.3)	
Mild	59 (31.6)	7 (20.0)	
Moderate	57 (30.5)	10 (28.6)	
Severe	35 (18.7)	6 (17.1)	
Profound	7 (3.7)	0 (0.0)	
Terminal	3 (1.6)	0 (0.0)	
Sedative (yes), *n* (%)	115 (61.5)	23 (65.7)	0.707
Dementia (yes), n (%)			
Alzheimer disease	86 (46.0)	14 (40.0)	0.581
Lewy body dementia	68 (36.6)	19 (54.3)	0.060
Vascular dementia	32 (17.2)	6 (17.1)	> 0.999
Other dementia	50 (26.9)	9 (25.7)	> 0.999
BPSD Subsyndrome[f], *n* (%)			
Hyperactivity symptoms	88 (47.1)	19 (54.3)	0.548
Psychotic symptoms	63 (33.7)	8 (22.9)	0.288
Affective symptoms	105 (56.1)	21 (60.0)	0.813
Aberrant motor behaviors	28 (15.0)	1 (2.9)	0.093
Euphoria/elation	44 (23.5)	3 (8.6)	0.078
Appetite and eating disorders	44 (23.5)	6 (17.1)	0.542
Sleep and nighttime behaviors	68 (36.4)	13 (37.1)	1.000

Table 1. Summary statistics of study participants and prevalence of subsyndromes of behavioral and psychological symptoms of dementia (BPSD). [a]*SD*: standard deviation. [b]BFI: Big Five Inventory. [c]ADL: Activities of daily living. [d]MMSE: Mini-Mental State Examination. [e]CDR: Clinical Dementia Rating scale. [f]BPSD: Behavioral and psychological symptoms of dementia.

symptoms (0.936), and aberrant motor behaviors (0.498). Moreover, the support vector machine model showed the highest AUC value (0.929) for sleep and nighttime behavior.

Characteristics	Training set ($n = 2497$ days)	Test set ($n = 455$ days)	p-Value
Sleep, mean (SD)[a]			
TST[b] (min)	464.54 (260.94)	379.51 (253.26)	< 0.001
WASO[c] (min)	54.05 (42.18)	27.90 (26.06)	< 0.001
Efficiency (min)	0.89 (0.08)	0.93 (0.06)	< 0.001
NoA[d] (min)	19.39 (13.96)	11.69 (9.81)	< 0.001
MAL[e] (min)	2.72 (1.10)	2.20 (1.05)	< 0.001
TST at night (min)	353.68 (171.60)	308.58 (183.47)	< 0.001
WASO at night (min)	41.55 (32.38)	22.44 (21.46)	< 0.001
Efficiency at night (min)	0.88 (0.08)	0.93 (0.05)	< 0.001
NoA at night (min)	15.32 (10.50)	9.76 (8.28)	< 0.001
MAL at night (min)	2.60 (1.15)	2.12 (1.03)	< 0.001
Activity, mean (SD)			
Kcals/day	459.92 (338.66)	442.18 (298.39)	0.254
METs[f]/day	1.15 (0.13)	1.14 (0.11)	0.186
MVPA[g]/day	79.70 (73.97)	75.80 (67.00)	0.261
% of MVPA/day	5.70 (5.19)	5.40 (4.71)	0.217
Steps/day	5908.73 (4501.06)	5750.51 (4151.32)	0.461
BPSD[h] subsyndrome, n (%)			
Hyperactivity symptoms	444 (17.8)	58 (12.7)	0.008
Psychotic symptoms	318 (12.7)	31 (6.8)	< 0.001
Affective symptoms	668 (26.8)	78 (17.1)	< 0.001
Aberrant motor behaviors	111 (4.4)	1 (0.2)	< 0.001
Euphoria/elation	117 (4.7)	4 (0.9)	< 0.001
Appetite and eating disorders	209 (8.4)	11 (2.4)	< 0.001
Sleep and nighttime behaviors	309 (12.4)	42 (9.2)	0.059
BPSD triggers, n (%)			
Hunger/thirst	179 (7.2)	16 (3.5)	0.003
Urination/bowel movement	264 (10.6)	8 (1.8)	< 0.001
Pain/discomfort	221 (8.9)	43 (9.5)	0.656
Sleep disturbance	337 (13.5)	38 (8.4)	0.002
Noise	93 (3.7)	6 (1.3)	0.007
Light	84 (3.4)	5 (1.1)	0.007
Temperature	126 (5.0)	12 (2.6)	0.022
Interpersonal triggers	212 (8.5)	49 (10.8)	0.126
Environmental change	120 (4.8)	7 (1.5)	< 0.001
Other causes[i]	269 (10.8)	15 (3.3)	< 0.001

Table 2. Descriptive statistics for actigraphy and symptom diary. [a]SD: Standard deviation. [b]TST: Total sleep time. [c]WASO: Wake after sleep onset. [d]NoA: Number of awakenings. [e]MAL: Mean awake length. [f]METs: Metabolic equivalents. [g]MVPA: Moderate-to-vigorous physical activity. [h]BPSD: Behavioral and psychological symptoms of dementia. [i]Other causes: Any other caregiver-perceived BPSD trigger that could not be categorized as one of the options provided in the symptom diary (e.g., medical treatment, hospital visits, and nightmare).

The gradient boosting machine model achieved the best performance in terms of the average AUC scores across the seven subsyndromes for both the training and test datasets. Tables 3 and 4 present the findings for the other performance indices.

Feature importance. The top 10 most significant features of the gradient boosting machine models, which achieved the best performance in terms of average AUC scores across the seven subsyndromes, were ranked using the permutation feature importance method. Figure 1 illustrates the relative importance of the predictors included in the seven subsyndromes. Caregiver-perceived triggers reported in the symptom diary, including interpersonal triggers, sleep disturbance, urination/bowel movement, and pain/discomfort, exhibited higher feature importance values than the other features across the seven subsyndromes. The CDR score was the most influential feature for psychotic symptoms and ranked in the top 10 for hyperactivity, euphoria/elation, and appetite and eating disorders. The features for sleep and activity levels, such as waking after sleep onset at night and percentage of moderate-to-vigorous physical activity per day, were in the top ranks of feature importance for

Subsyndrome	Model	AUC[a]	Accuracy	Precision	Sensitivity	Specificity	F1 Score
Hyperactivity symptoms	LR[b]	0.682	0.652	0.589	0.731	0.459	0.591
	RF[c]	0.682	0.680	0.524	0.912	0.108	0.483
	GBM[d]	0.706	0.695	0.590	0.890	0.216	0.548
	SVM[e]	0.689	0.652	0.594	0.720	0.486	0.597
Psychotic symptoms	LR	0.682	0.809	0.556	0.889	0.226	0.557
	RF	0.692	0.875	0.439	0.996	0.000	0.467
	GBM	0.631	0.762	0.478	0.853	0.097	0.476
	SVM	0.706	0.824	0.561	0.911	0.194	0.556
Affective symptoms	LR	0.745	0.719	0.731	0.876	0.514	0.696
	RF	0.728	0.668	0.708	0.924	0.333	0.612
	GBM	0.747	0.691	0.708	0.883	0.441	0.659
	SVM	0.737	0.730	0.739	0.869	0.550	0.712
Aberrant motor behaviors	LR	0.822	0.832	0.546	0.846	0.500	0.547
	RF	0.534	0.961	0.480	1.000	0.000	0.490
	GBM	0.609	0.961	0.480	1.000	0.000	0.490
	SVM	0.679	0.816	0.492	0.846	0.100	0.470
Euphoria/ elation	LR	0.696	0.785	0.544	0.808	0.438	0.539
	RF	0.641	0.918	0.541	0.975	0.063	0.522
	GBM	0.627	0.910	0.589	0.958	0.188	0.580
	SVM	0.632	0.816	0.546	0.846	0.375	0.550
Appetite and eating disorders	LR	0.619	0.883	0.511	0.893	0.250	0.500
	RF	0.709	0.980	0.492	0.996	0.000	0.495
	GBM	0.816	0.973	0.492	0.988	0.000	0.493
	SVM	0.629	0.895	0.514	0.905	0.250	0.507
Sleep and nighttime behaviors	LR	0.889	0.777	0.663	0.776	0.784	0.680
	RF	0.942	0.906	0.895	0.991	0.405	0.752
	GBM	0.895	0.918	0.869	0.977	0.568	0.810
	SVM	0.888	0.770	0.663	0.763	0.811	0.677

Table 3. Performance comparison of the prediction models for subsyndromes of behavioral and psychological symptoms of dementia (BPSD) for the training dataset with five-fold cross-validation. [a]AUC: Area under the receiver operating characteristic curve. [b]LR: Logistic regression. [c]RF: Random Forest. [d]GBM: Gradient boosting machine. [e]SVM: Support vector machine.

aberrant motor behaviors and sleep and nighttime behaviors. Except for two symptoms (aberrant motor behaviors and euphoria/elation), premorbid personality types were among the top 10 influential features.

Discussion

This study constructed predictive models for seven subsyndromes of BPSD, with overall good prediction accuracy, through a supervised machine learning technique based on actigraphy measures of sleep and activity levels, symptom diary entries of caregiver-perceived symptom triggers, standardized measures of cognitive and functional status and premorbid personality type, and a medical chart review. Through the feature importance approach, we also identified the relative importance of predicting each subsyndrome. The machine learning algorithms developed and validated in our study will inform timely and appropriate preventative approaches that identify individuals at high risk for specific subsyndromes of BPSD, and thus, provide customized interventions that address the underlying causes and triggers of the target symptoms.

Although the predictive machine learning models varied among the subsyndromes with regard to different evaluation metrics, our findings demonstrated the feasibility and usefulness of the machine learning approach for predicting BPSD by accounting for multifaceted factors. Although several studies have applied machine learning algorithms with a focus on predicting the future onset of or detecting undiagnosed dementia[19,51], few have used machine learning models to predict the various symptoms observed in older adults with dementia. A recent study employed the deep learning approach to forecast agitation episodes based on environmental stimuli data, including ambient noise level, room temperature, and relative humidity, that was collected up to 30 min before the occurrence of an actual episode[52]. While it demonstrated the feasibility and efficacy of deep learning models in predicting agitation episodes (accuracy: 98.6%; sensitivity: 84.8%), the models developed in our study were more comprehensive and practical because they examined diverse BPSD, which often co-occurs in real care settings. Besides, a machine learning approach can account for a range of multifaceted contributing factors, including actigraphy data and symptom diary data measured over time, which might not have been feasible in conventional statistical modeling owing to highly complex relationships among the variables. In recent years, big healthcare data have emerged along with advances in assistive technologies, including wearable and

Subsyndrome	Model	AUC[a]	Accuracy	Precision	Sensitivity	Specificity	F1 Score
Hyperactivity symptoms	LR[b]	0.788	0.869	0.707	0.946	0.365	0.675
	RF[c]	0.835	0.889	0.806	0.984	0.261	0.660
	GBM[d]	0.820	0.867	0.698	0.950	0.322	0.657
	SVM[e]	0.828	0.854	0.681	0.917	0.443	0.681
Psychotic symptoms	LR	0.770	0.821	0.580	0.844	0.516	0.594
	RF	0.827	0.959	0.979	1.000	0.419	0.785
	GBM	0.801	0.936	0.774	0.985	0.290	0.679
	SVM	0.786	0.823	0.591	0.842	0.581	0.609
Affective symptoms	LR	0.866	0.814	0.697	0.856	0.619	0.713
	RF	0.927	0.869	0.776	0.918	0.645	0.779
	GBM	0.936	0.885	0.801	0.920	0.723	0.810
	SVM	0.888	0.844	0.759	0.824	0.935	0.789
Aberrant motor behaviors	LR	0.128	0.966	0.499	0.968	0.000	0.491
	RF	0.339	0.998	0.499	1.000	0.000	0.499
	GBM	0.498	0.998	0.499	1.000	0.000	0.499
	SVM	0.106	0.952	0.499	0.954	0.000	0.488
Euphoria/ elation	LR	0.765	0.877	0.512	0.880	0.500	0.494
	RF	0.968	0.991	0.497	0.998	0.000	0.498
	GBM	0.955	0.989	0.497	0.995	0.000	0.497
	SVM	0.858	0.874	0.517	0.875	0.667	0.500
Appetite and eating disorders	LR	0.712	0.896	0.549	0.907	0.455	0.562
	RF	0.888	0.975	0.487	1.000	0.000	0.494
	GBM	0.862	0.968	0.632	0.988	0.182	0.603
	SVM	0.740	0.865	0.526	0.878	0.364	0.523
Sleep and nighttime behaviors	LR	0.911	0.925	0.776	0.951	0.679	0.794
	RF	0.912	0.933	0.903	0.995	0.333	0.723
	GBM	0.900	0.938	0.848	0.982	0.506	0.785
	SVM	0.929	0.935	0.800	0.956	0.728	0.819

Table 4. Performance comparison of the prediction models for subsyndromes of behavioral and psychological symptoms of dementia (BPSD) for the test dataset. [a]AUC: Area under the receiver operating characteristic curve. [b]LR: Logistic regression. [c]RF: Random Forest. [d]GBM: Gradient boosting machine. [e]SVM: Support vector machine.

mobile technologies in healthcare for older adults[53], which has increased the number of candidate features[54]. Machine learning is a promising tool in predicting BPSD arising from interactions among a web of personal, interpersonal, and environmental factors captured by various digital health technologies[54]. Moreover, machine learning algorithms that predict BPSD subsyndromes can be incorporated into a mobile app to support health-care providers' decision-making in selecting individually customized non-pharmacological interventions that address the underlying causes of the target symptoms[54].

Our outcomes concerning feature importance suggested that not all features contributed consistently to the prediction of each subsyndrome. Caregiver-perceived symptom triggers, including interpersonal triggers, sleep disturbance, and pain/discomfort, were consistently ranked within the five most influential features across the seven subsyndromes. The features for sleep and activity levels, such as waking after sleep onset at night and percentage of moderate-to-vigorous physical activity per day, ranked highly in terms of feature importance for aberrant motor behaviors and sleep and nighttime behaviors. Age, severity of dementia, cognitive function, education level, and premorbid personality type were also ranked among the 10 most influential features. While activity data were relatively less influential than other features, metabolic equivalents per day were ranked in the top 10 important features for sleep and nighttime behaviors, as was wake after sleep onset at night and the percentage of moderate-to-vigorous physical activity per day for aberrant motor behaviors. The clinical implications of the results are that healthcare providers and caregivers need to consider BPSD as heterogeneous and having different important predictors. Hence, effective interventions to prevent and manage target predictors and symptoms should be employed.

While caregiver-perceived symptom triggers evaluated by the symptom diary were among the most influential features, activity data were less influential than expected for most subsyndromes in our study, except for aberrant motor behaviors. This notion is somewhat consistent with Valembois et al.'s observational study[55], which investigated the association between motor activity and sleep duration measured by actigraphy with different types of BPSD. They found a significant increase in motor activity among those with aberrant motor behavior between 21:00 (9:00 pm) and midnight. They explained the phenomenon in terms of the sundown syndrome, referring to an increase in BPSD from late afternoon to night; otherwise, no relationship between sleep duration
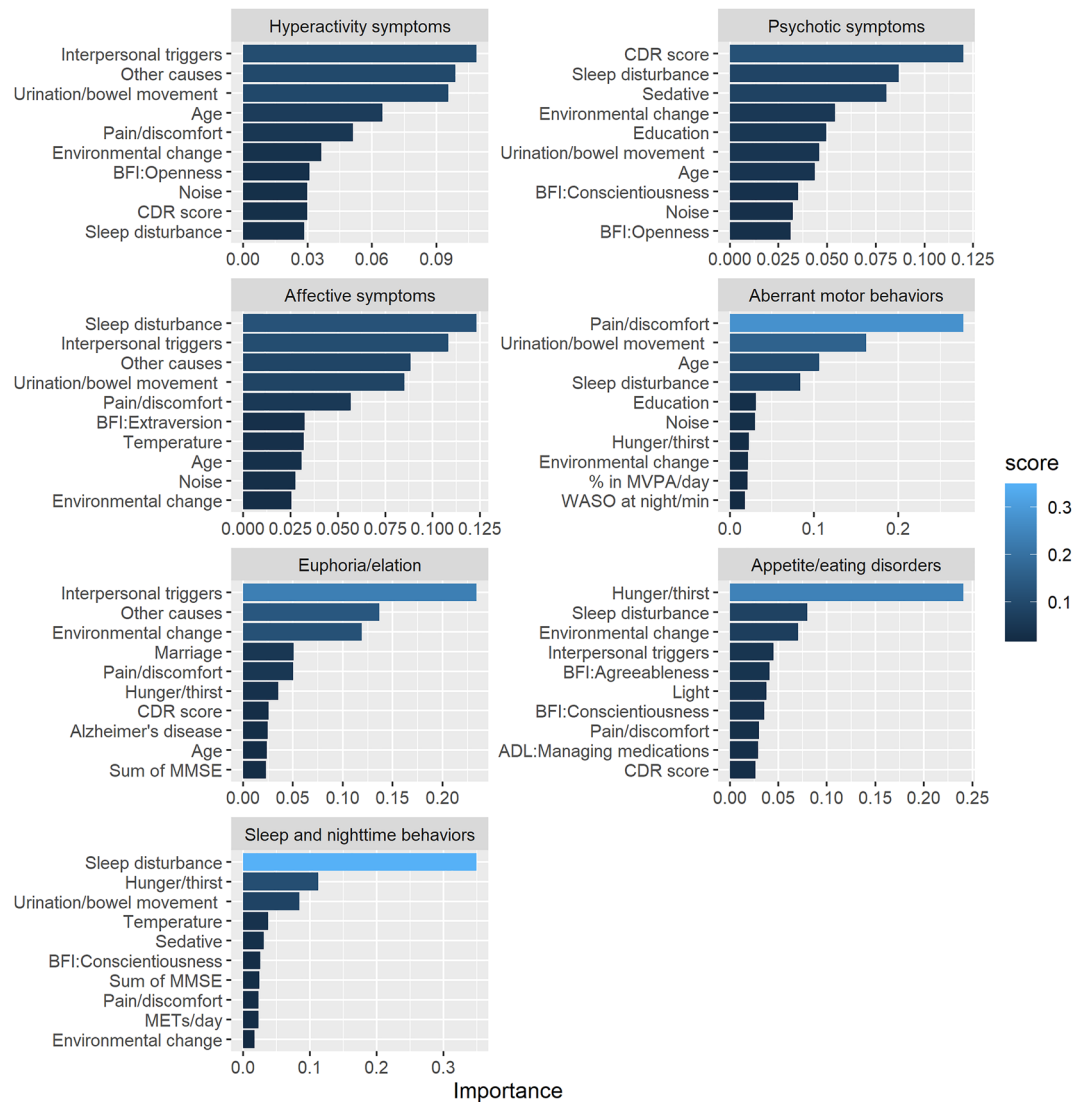
Figure 1. Importance of the top 10 features calculated by the gradient boosting machine for each subsyndrome of the behavioral and psychological symptoms of dementia (BPSD). "Other causes" indicates any other caregiver-perceived BPSD trigger that cannot be categorized as one of the options provided in the symptom diary (e.g., medical treatment, hospital visits, and nightmare). *ADL* activities of daily living; *BFI* Big Five Inventory; *CDR* Clinical Dementia Rating scale; *MMSE* Mini-Mental State Examination; *WASO* wake after sleep onset; *MVPA* moderate-to-vigorous physical activity; *METs* metabolic equivalents.

or night awakening episodes and any type of BPSD was found[55]. Another study utilized the machine learning approach to predict mobility, cognitive, and depressive symptoms related to Alzheimer's disease from activity-aware smart home behavior data (e.g., activities of daily living, sleep, outings, and global routines)[56]. While statistically significant predictors were observed for mobility and cognitive symptoms, depression measured by the Geriatric Depression Scale (GDS) was weakly correlated with the global set of smart home behavioral data[56]. This outcome somewhat aligns with our finding that affective symptoms were predictable using machine learning models, but its classification (occurrence vs. absence) was mostly not influenced by actigraphy data. Additionally, the inclusion of a variety of predictors in the machine learning model might have decreased the importance of features derived from actigraphy data in our study. For clinical implications, a comprehensive monitoring system that utilizes a symptom diary written by direct caregivers, in addition to wearable sensor technologies, is required for accurate prediction.

A strength of our study is the inclusion of a variety of input data, objectively measured by actigraphy, and caregiver-perceived symptom triggers assessed using a symptom diary to develop the most powerful predictive machine learning models. Furthermore, external validation was performed using an independent test dataset to prevent overestimation of the results[57,58]. Consistent results from the training and test datasets that differed substantially in terms of severity of functional impairment, characteristics of sleep and activity levels, caregiver-perceived symptom triggers, and frequency of occurrence of symptoms can be regarded as confirmation of

generalizability[58]. Finally, the symptom diary, from which the most influential features were derived, was easy to measure daily in real life. The feature importance results will facilitate the development of a digital diary that tracks BPSD using devices such as smartphones and tablets. A digital diary enables caregivers to log the symptom manifestation and circumstances, including diverse triggers, in real time, and the accumulated data can be analyzed to provide an individualized approach to symptom management.

Our study had certain limitations. First, the symptom diary was written daily, and caregivers checked one or more types of symptoms that were observed and one or more factors that were perceived as triggers for the symptoms each day. Hence, we could not disentangle which caregiver-perceived trigger contributed to which type of individual symptom observed that day. Future research needs to investigate the link between caregiver-perceived triggers and specific symptoms on an episodic rather than a daily basis. Episodic prediction algorithms can provide more precise and accurate information regarding specific symptoms. Second, the prediction performances for certain subsyndromes were poor, particularly in the case of aberrant motor behaviors (AUC = 0.498 in the gradient boosting machine model). The findings regarding the prediction performance need to be interpreted cautiously considering the small sample size of the test dataset. The frequency of occurrence of certain symptoms, particularly aberrant motor behaviors and euphoria/elation, was very low (e.g., 2/871 days of aberrant motor behavior). The overall small sample size of the test dataset might lead to lower power for pattern recognition[59]. Future replication studies with larger sample sizes are needed to obtain more accurate predictions for the subsyndromes that were less observed in our study and to prevent model overfitting and biased machine learning performance[59]. As it was challenging to collect accurate data on comorbid conditions for older adults who were recruited from other outpatient neurological clinics (i.e., daycare centers), we did not account for any comorbidities that might have affected their BPSD. Additionally, caregivers' daily written reports were used rather than direct observation of BPSD, which could result in an observer and recall bias. While dementia care has relied on caregivers' subjective assessment reports in most cases[4], future studies need to employ technology such as wearable sensors, non-wearable motion sensors, or assistive and smart home technologies to objectively monitor BPSD in real time, thus revolutionizing precision dementia care[60].

Conclusions

This study developed and validated prediction models for BPSD subsyndromes using a machine learning approach. Overall, four classification models were trained to identify the optimal prediction model for each subsyndrome. To our knowledge, this study is the first to employ machine learning to predict BPSD using a wide variety of data, including actigraphy data. It suggests that machine learning-based prediction models can classify the manifestations of subsyndromes of BPSD. This study also identified influential predictors for specific subsyndromes that can be employed to prevent and manage target symptoms. Based on the outcomes of this study, algorithms that predict BPSD can be clinically applied to monitor and predict BPSD subsyndromes. Delivery of person-centered dementia care can be achieved through the early prediction of target subsyndromes and the provision of individually tailored non-pharmacological interventions that address the underlying causes of BPSD. Machine learning algorithms can also be embedded into smartphone applications to increase their clinical utility. Accordingly, this study is the first step toward personalized care for BPSD management using digital health technologies.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to the data security requirements of the hospital but are available from the corresponding author on reasonable request.

References

1. Nichols, E. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8 (2022).
2. Cerejeira, J., Lagarto, L. & Mukaetova-Ladinska, E. B. Behavioral and psychological symptoms of dementia. *Front. Neurol.* **3**, 73. https://doi.org/10.3389/fneur.2012.00073 (2012).
3. Ballard, C. & Corbett, A. Management of neuropsychiatric symptoms in people with dementia. *CNS Drugs* **24**, 729–739. https://doi.org/10.2165/11319240-000000000-00000 (2010).
4. Kales, H. C., Kern, V., Kim, H. M. & Blazek, M. C. Moving evidence-informed assessment and management of behavioral and psychological symptoms of dementia into the real world: Training family and staff caregivers in the DICE Approach. *Am. J. Geriatr. Psychiatry* **28**, 1248–1255. https://doi.org/10.1016/j.jagp.2020.08.008 (2020).
5. Dyer, S. M., Harrison, S. L., Laver, K., Whitehead, C. & Crotty, M. An overview of systematic reviews of pharmacological and non-pharmacological interventions for the treatment of behavioral and psychological symptoms of dementia. *Int. Psychogeriatr.* **30**, 295–309. https://doi.org/10.1017/S1041610217002344 (2018).
6. Toot, S., Swinson, T., Devine, M., Challis, D. & Orrell, M. Causes of nursing home placement for older people with dementia: A systematic review and meta-analysis. *Int. Psychogeriatr.* **29**, 195–208. https://doi.org/10.1017/S1041610216001654 (2017).
7. Herrmann, N. *et al.* The contribution of neuropsychiatric symptoms to the cost of dementia care. *Int. J. Geriatr. Psychiatry* **21**, 972–976. https://doi.org/10.1002/gps.1594 (2006).
8. Baharudin, A. D., Din, N. C., Subramaniam, P. & Razali, R. The associations between behavioral-psychological symptoms of dementia (BPSD) and coping strategy, burden of care and personality style among low-income caregivers of patients with dementia. *BMC Public Health* **19**, 447. https://doi.org/10.1186/s12889-019-6868-0 (2019).
9. Kales, H. C., Gitlin, L. N. & Lyketsos, C. G. Assessment and management of behavioral and psychological symptoms of dementia. *BMJ* **350**, h369. https://doi.org/10.1136/bmj.h369 (2015).
10. Richards, K. C. & Beck, C. K. Progressively lowered stress threshold model: Understanding behavioral symptoms of dementia. *J. Am. Geriatr. Soc.* **52**, 1774–1775. https://doi.org/10.1111/j.1532-5415.2004.52477.x (2004).

11. Volicer, L. & Hurley, A. C. Management of behavioral symptoms in progressive degenerative dementias. *J. Gerontol. A Biol. Sci. Med. Sci.* **58**, 837–845. https://doi.org/10.1093/gerona/58.9.M837 (2003).
12. Kolberg, E. *et al.* The effects of bright light treatment on affective symptoms in people with dementia: A 24-week cluster randomized controlled trial. *BMC Psychiatry* **21**, 377. https://doi.org/10.1186/s12888-021-03376-y (2021).
13. Todd, W. D. Potential pathways for circadian dysfunction and sundowning-related behavioral aggression in Alzheimer's disease and related dementias. *Front. Neurosci.* **14**, 910. https://doi.org/10.3389/fnins.2020.00910 (2020).
14. Hjetland, G. J. *et al.* An actigraphy-based validation study of the Sleep Disorder Inventory in the nursing home. *Front. Psychiatry* **11**, 173. https://doi.org/10.3389/fpsyt.2020.00173 (2020).
15. Scales, K., Zimmerman, S. & Miller, S. J. Evidence-based nonpharmacological practices to address behavioral and psychological symptoms of dementia. *Gerontologist* **58**, S88–S102. https://doi.org/10.1093/geront/gnx167 (2018).
16. Cho, C. H. *et al.* Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the Circadian Rhythm: Prospective observational cohort study. *J. Med. Internet Res.* **21**, e11029. https://doi.org/10.2196/11029 (2019).
17. Asgari Mehrabadi, M. *et al.* Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: Instrument validation study. *JMIR Mhealth Uhealth* **8**, e20465. https://doi.org/10.2196/20465 (2020).
18. Mufti, H. N., Hirsch, G. M., Abidi, S. R. & Abidi, S. S. R. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: Models development and validation study. *JMIR Med. Inform.* **7**, e14993. https://doi.org/10.2196/14993 (2019).
19. Escudero, J., Zajicek, J. P. & Ifeachor, E. Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2011**, 6470–6473. https://doi.org/10.1109/IEMBS.2011.6091597 (2011).
20. Cho, E. *et al.* Factors associated with behavioral and psychological symptoms of dementia: Prospective observational study using actigraphy. *J. Med. Internet Res.* **23**, e29001. https://doi.org/10.2196/29001 (2021).
21. Luo, W. *et al.* Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J. Med. Internet Res.* **18**, e323. https://doi.org/10.2196/jmir.5870 (2016).
22. Kang, Y., Na, D. L. & Hahn, S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *J. Korean Neurol. Assoc.* **15**, 300–308 (1997).
23. Song, J., Park, J. & Kim, H. Impact of behavioral and psychological symptoms of dementia on caregiver burden in nursing homes. *J. Korean Gerontol. Nurs.* **15**, 62–74 (2013).
24. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A. & Martin, R. L. A new clinical scale for the staging of dementia. *Br. J. Psychiatry* **140**, 566–572. https://doi.org/10.1192/bjp.140.6.566 (1982).
25. Choi, S. H. *et al.* Estimating the validity of the Korean version of expanded Clinical Dementia Rating (CDR) Scale. *J. Korean Neurol. Assoc.* **19**, 585–591 (2001).
26. Won, C. W., Rho, Y. G., Kim, S. Y., Cho, B. R. & Lee, Y. S. The validity and reliability of Korean Activities of Daily Living (K-ADL) Scale. *J. Korean Geriatr. Soc.* **6**, 98–106 (2002).
27. Kim, J.-H., Kim, B.-H. & Ha, M.-S. Validation of a Korean version of the Big Five Inventory. *J. Hum. Understand. Couns.* **32**, 47–65 (2011).
28. Camargos, E. F., Louzada, F. M. & Nóbrega, O. T. Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: Application, usefulness, and challenges. *Sleep Med. Rev.* **17**, 475–488. https://doi.org/10.1016/j.smrv.2013.01.006 (2013).
29. Figueiro, M. G. *et al.* Tailored lighting intervention for persons with dementia and caregivers living at home. *Sleep Health* **1**, 322–330. https://doi.org/10.1016/j.sleh.2015.09.003 (2015).
30. Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J. & Gillin, J. C. Automatic sleep/wake identification from wrist activity. *Sleep* **15**, 461–469. https://doi.org/10.1093/sleep/15.5.461 (1992).
31. Cummings, J. L. The Neuropsychiatric Inventory: Assessing psychopathology in dementia patients. *Neurology* **48**, S10-16. https://doi.org/10.1212/WNL.48.5_Suppl_6.10S (1997).
32. Morganti, F., Soli, A., Savoldelli, P. & Belotti, G. The Neuropsychiatric Inventory-Diary Rating Scale (NPI-Diary) is a method for improving stability in assessing neuropsychiatric symptoms in dementia. *Dement. Geriatr. Cogn Disord. Extra* **8**, 306–320. https://doi.org/10.1159/000490380 (2018).
33. Liew, T. M. Symptom clusters of neuropsychiatric symptoms in mild cognitive impairment and their comparative risks of dementia: A cohort study of 8530 older adults. *J. Am. Med. Dir. Assoc.* **20**(1054), e1051-1054.e1059. https://doi.org/10.1016/j.jamda.2019.02.012 (2019).
34. Liew, T. M. Neuropsychiatric symptoms in cognitively normal older persons, and the association with Alzheimer's and non-Alzheimer's dementia. *Alzheimers Res. Ther.* **12**, 35. https://doi.org/10.1186/s13195-020-00604-7 (2020).
35. Aalten, P. *et al.* Behavioral problems in dementia: Factor analysis of the neuropsychiatric inventory. *Dementia Geriatr. Cogn. Disord.* **15**, 99–105. https://doi.org/10.1159/000067972 (2003).
36. van der Linde, R. M. *et al.* Longitudinal course of behavioral and psychological symptoms of dementia: A systematic review. *Br. J. Psychiatry* **209**, 366–377. https://doi.org/10.1192/bjp.bp.114.148403 (2016).
37. van der Linde, R. M., Dening, T., Matthews, F. E. & Brayne, C. Grouping of behavioral and psychological symptoms of dementia. *Int. J. Geriatr. Psychiatry* **29**, 562–568. https://doi.org/10.1002/gps.4037 (2014).
38. Canevelli, M. *et al.* Behavioral and psychological subsyndromes in Alzheimer's disease using Neuropsychiatric Inventory. *Int. J. Geriatr. Psychiatry* **28**, 795–803. https://doi.org/10.1002/gps.3904 (2013).
39. Buuren, S. V. & Oudshoorn, C. G. M. Multivariate imputation by chained equations (MICE) V1.0 User's manual; https://stefvanbuuren.name/publications/MICE%20V1.0%20Manual%20TNO00038%202000.pdf (2000).
40. Micci-Barreca, D. Preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *ACM SIGKDD Explor. Newsl.* **3**, 27–32. https://doi.org/10.1145/507533.507538 (2001).
41. Choi, S.-H. Estimation of the validity of the Korean version of the Expanded Clinical Dementia Rating (CDR) scale. *J. Korean Neurol. Assoc.* **19**, 585–591 (2001).
42. Japkowicz, N. Class imbalance problem: Significance and strategies. Proceedings of the International Conference on Artificial Intelligence (2000).
43. Japkowicz, N. Class imbalances: We focus on the correct issue. Workshop on Learning from Imbalanced Dataset II (2003).
44. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority oversampling technique. *J. Artif. Intell. Res.* **16**, 321–357. https://doi.org/10.1613/jair.953 (2002).
45. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
46. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232. https://doi.org/10.1214/aos/1013203451 (2001).
47. Suthaharan, S. Support vector machine. In *Machine learning models and algorithms for big data classification. Integrated Series in Information Systems* Vol. 36 207–235 (Springer, 2016). https://doi.org/10.1007/978-1-4899-7641-3_9.
48. Hyun, S., Moffatt-Bruce, S., Cooper, C., Hixon, B. & Kaewprag, P. Prediction model for hospital-acquired pressure ulcer development: Retrospective cohort study. *JMIR Med. Inform.* **7**, e13785. https://doi.org/10.2196/13785 (2019).
49. Bergstra, J. & Bengio, Y. Random search for hyperparameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012).

50. Maji, S., Berg, A. C. & Malik, J. Efficient classification for additive kernel SVMs. *IEEE Trans. Pattern. Anal. Mach. Intell.* **35**, 66–77. https://doi.org/10.1109/TPAMI.2012.62 (2013).
51. Hane, C. A., Nori, V. S., Crown, W. H., Sanghavi, D. M. & Bleicher, P. Predicting onset of dementia using daily notes and machine learning: A case-control study. *JMIR Med. Inform.* **8**, e17819. https://doi.org/10.2196/17819 (2020).
52. HekmatiAthar, S., Goins, H., Samuel, R., Byfield, G. & Anwar, M. Data-driven forecasting of agitation for persons with dementia: A deep learning-based approach. *SN Comput. Sci.* **2**, 326. https://doi.org/10.1007/s42979-021-00708-3 (2021).
53. Sapci, A. H. & Sapci, H. A. Innovative assisted living tools, remote monitoring technologies, artificial intelligence-driven solutions, and robotic systems for aging societies: A systematic review. *JMIR Aging* **2**, e15429. https://doi.org/10.2196/15429 (2019).
54. Verdonk, C., Verdonk, F. & Dreyfus, G. How can machine learning be used in clinical practice during an epidemic?. *Crit. Care* **24**, 265. https://doi.org/10.1186/s13054-020-02962-y (2020).
55. Valembois, L. *et al.* Wrist actigraphy: A simple way to record motor activity in elderly patients with dementia, apathy, or aberrant motor behavior. *J. Nutr. Health Aging* **19**, 759–764. https://doi.org/10.1007/s12603-015-0530-z (2015).
56. Alberdi, A. *et al.* Smart home-based prediction of multi-domain symptoms related to Alzheimer's disease. *IEEE J. Biomed. Health Inform.* **22**, 1720–1731. https://doi.org/10.1109/JBHI.2018.2798062 (2018).
57. Toll, D. B., Janssen, K. J., Vergouwe, Y. & Moons, K. G. Validation, updating and impact of clinical prediction rules: A review. *J. Clin. Epidemiol.* **61**, 1085–1094. https://doi.org/10.1016/j.jclinepi.2008.04.008 (2008).
58. Ho, S. Y., Phua, K., Wong, L. & Bin Goh, W. W. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns* **1**, 100129. https://doi.org/10.1016/j.patter.2020.100129 (2020).
59. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS One* **14**, e0224365. https://doi.org/10.1371/journal.pone.0224365 (2019).
60. Husebo, B. S. *et al.* Sensing technology to monitor behavioral and psychological symptoms and assess treatment response in people with dementia. A systematic review. *Front. Pharmacol.* **10**, 1699. https://doi.org/10.3389/fphar.2019.01699 (2020).

Author contributions

E.C.: Conceptualization, methodology, funding acquisition; S.J.K., J.S., S.H., E.K.: Acquisition of data; E.C., B.K., S.J.H., S.H.L., and S.G.K.: Data analysis and interpretation; E.C, B.K., S.J.H., S.J.K., J.S., S.H., E.K., S.H.L., S.G.K.: Drafting and revising the article.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-35194-5.

Correspondence and requests for materials should be addressed to B.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.