2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Artificial Intelligence and Features Investigating to Detect Neuropsychiatric Symptoms in Patients with Dementia: A Pilot Study

Abeer Badawi[1], Samira Choudhury[2,3], Khalid Elgazzar[1], and Amer M. Burhan[2,3]
[1] Electrical and Computer Engineering, Ontario Tech University, Oshawa, ON, Canada
[2] Ontario Shores Centre for Mental Health Sciences, Whitby, ON, Canada
[3] Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

***Abstract*—Dementia is a chronic and irreversible condition characterized by progressive cognitive and functional decline, along with non-cognitive Neuropsychiatric Symptoms (NPS) that significantly affect patients' quality of life. This study explores the application of artificial intelligence using wearable sensor data collected from dementia patients (PwD) to develop an AI system for detecting NPS, particularly episodes of agitation, in an institutional setting. We present preliminary results from a real-world study at the Ontario Shores Center for Mental Health Sciences. The study outcomes indicate that by employing sequential feature selection, we are able to improve detection accuracy while reducing the number of features from an initial set of 198. Further, the Extra Trees classification model outperformed other algorithms in accurately classifying non-agitated normal events from agitated events. The study also shows that personalized models yielded superior results, with an average accuracy improvement of 5-10% compared to models trained on all patient data combined.**

***Index Terms*—Dementia; Agitation; Wearable sensors; Machine Learning; Neuropsychiatric Symptoms.**

I. Introduction

Dementia is a chronic, neurodegenerative disease characterized by progressive cognitive and functional decline. It is a significant cause of disability and institutionalization and is currently the seventh leading cause of death worldwide [1]. There are approximately 55 million people with dementia globally, which will grow to 78 million in 2030 and 139 million in 2050, making dementia a major global health crisis [1]. In addition to mental and functional decline, patients with dementia (PwD) often experience noncognitive neuropsychiatric symptoms (NPS) during their illness [2]. Symptoms include symptoms of psychosis, agitation, aggression, depression, anxiety, appetite, and sleep disturbances. Among NPS, agitation and aggression (AA) occur frequently, are the most challenging to manage, and are a major source of distress to PwD, caregivers, and healthcare systems [3]. These behaviors commonly occur during care and are believed to be manifestations of perceived or real unmet needs [3]. Behaviors and their poor management can accelerate cognitive decline, worsen the quality of life for patients and caregivers, and increase the rate of accidents.

AA symptoms include hitting, shouting, throwing objects, restlessness, and excessive motor activity such as pacing or wandering. Early detection of potential episodes of AA would allow the timely deployment of preventive measures or therapies, reduce the cost of care, and reduce the prevalence of critical incidents in this population [4, 5]. Recently, the use of digital technology to detect neurological conditions in a timely manner has gained great popularity; for example, the use of smartwatches to detect epileptic seizures and prevent the development of severe complications [6, 7]. There is growing evidence that digital technology may also be a solution in detecting behaviors in PwD when combined with Artificial Intelligence (AI), feature analysis, and sensory technologies. Using digital technology to develop a solution for the early detection of AA will help guide the provision of personalized interventions for PwD [8–11].

This work is a pilot study of our original system that aims to detect episodes of AA and pre-agitation using AI in PwD at Ontario Shores for Mental Health Sciences in Ontario, Canada [12]. The main focus of the original system is to detect and predict episodes of AA, explore pre-agitation and agitation events, and investigate in depth the role of exploratory data analysis in detecting episodes of AA. In this study, we focus on demonstrating the use of AI and feature exploration in detecting episodes of AA. We collect the data from two PwD who participated in a study using the multi-sensory Empatica E4 devices [13]. We compare classification methods, feature extraction, feature selection, and performance evaluation techniques to detect physiological changes during episodes of AA in PwD.

We explore 198 features from the time domain, frequency domain, and statistical features. Furthermore, we compare multiple machine learning algorithms to find the finest model and features to detect physiological changes in PwD during episodes of AA by selecting the best set of features. We select six machine learning algorithms to compare the results of each patient individually and the combined records of both patients. Our results revealed that the Extratrees model provided the best results using the sequential feature selection technique followed by the random forest and XGBoost classifiers. Our results indicate that it is possible to detect agitated behavior using sensor data with high-performance evaluation results. Furthermore, we conclude that the personalized model provided better results than the combined model with 5-10% higher performance evaluation using the Extra trees classification model. We assume that new ways of detecting agitation will help to understand complicated behaviors in PwD, reduce the risk of harm to people living with dementia and others in their environments, and improve the quality of life of these patients. The remainder of this paper is structured as follows. We describe recent related work on the detection of AA in PwD in Section II, discuss data collection, labeling methodology, preprocessing, and evaluation techniques in Section III, present the experimental results and performance evaluation in Section IV. Lastly, we summarize the work in Section V.

II. Related Work

There is proof that physiological signals, such as Actigraphy (ACT) [14], Heart Rate Variability (HRV) [15], Electrodermal Activity (EDA) [16], and body temperature [17] captured via wearable sensors are highly associated with agitation and aggression in PwD. These parameters may be combined to create predictive algorithms that predict episodes of AA. Amato et al. [18] used the Empatica E4 wristband to collect physiological and behavioral data to identify symptoms in people with Alzheimer's disease in an Italian village devoted to exploring new treatments for Alzheimer's disease. They proposed a solution to detect physiological and behavioral symptoms. However, there was a lack of details about the system's performance and results.

Moore et al. [19] mentioned that a significant problem when using sensors to detect agitation in PwD is processing and labeling the enormous amounts of data generated for each patient. They also discussed the importance of converting the data into valuable information to detect agitation accurately in a real-world situation. Goerss et al. [20] discussed that studies on physiological sensors that focus on agitation in dementia are limited to evaluations based on limited sensors or a small number of features. They believe that future research should explore the use of other sensors and use more features to detect agitation in dementia. A recent survey [21] discussed the use of wearable sensors for dementia care. They found minimal published work on wearable computing and intelligent technologies in dementia care, with approximately 300 publications. Also, they pointed out that the publications lack work investigating the classification model's training with raw data. They conclude that it is essential to investigate new algorithms or models applied in real life to identify PwD behavior.

Furthermore, Spasojevic et al. [22] showed that the use of useful features and multimodal sensor data helped identify behavioral indications of PwD. Iaboni et al. [23] discovered that personalized models improved the results of detecting agitation events in PwD. However, both research investigations of feature extraction, feature selection techniques, and classification algorithms are limited. Furthermore, the systematic review by Khan et al. [10] indicated that most of the papers discussing the relationship between sensors and agitation used simple analysis techniques. They concluded that the available research lacks focus on the data analysis component.

III. Methodology

A. E4 Wristband Data Acquisition

We selected the Empatica E4 [13] wristband for this study as it is lightweight, portable, and accurate. It has the highest precision level compared to other devices [25]. It is a smartwatch configured with the necessary sensors to monitor the following physiological parameters: Heart rate measurements calculated by a Photoplethysmography (PPG) sensor at 64 Hz frequency, movement captured using a 3-axis accelerometer at 32 Hz frequency, skin temperature and electrical properties of the skin using an Electrodermal Activity (EDA) sensor at 4 Hz frequency. The wristband collects these physiological signals and can be connected to a smartphone via wireless Bluetooth to analyze real-time signals. The recorded data from the E4 device is uploaded to the Empatica website. The data can then be downloaded from the cloud.

B. Sample Data Collection of the Study

To demonstrate the functioning of the sensors, we conducted a pilot study using PwD data, who participated in a study conducted using Empatica E4 devices at the Ontario Shores Center for Mental Health Sciences [12]. We collected data from two male participants. Participant one was 83 years old, and participant two was 85 years old. Participants in the study had a clinical diagnosis of dementia - Alzheimer's or mixed type using the criteria of the Diagnostic and Statistical Manual of Mental Disorders [24]. For privacy concerns, any data collected is confidential, ensuring all participants' anonymity. We store the data in a password-protected computer, and access to the data is restricted to the Research Team only

The participants wore the Empatica E4 device for 48 to 72 hours on three occasions during an eight-week study period. Clinical staff monitored participants' behaviors and made notes of any AA. The data from the E4 devices were compared to the clinical observational notes to confirm the validity of the recordings. Participant One had 72.13 hours, 70.38 hours, and 50.41 hours during their first, second, and third recordings, respectively. Participant 2 had 47.22 hours, 72.50 hours, and 58.56 hours during their first, second, and third recordings, respectively. Using the start times of the episodes of AA, the episodes were identified in the E4 data and labeled with "0" for normal behavior and "1" for agitated behavior. Our data structure followed two approaches: each participant's data in an individual dataset and combining both participants' data in one dataset. The goal was to investigate whether the personalized model better detects AA episodes than merging all patient data.

C. Preprocessing and Feature Extraction

Before analyzing the data, we pre-process it to obtain reliable information. We use Flirt [26] for data processing and feature extraction in the proposed system. Flirt is a toolkit

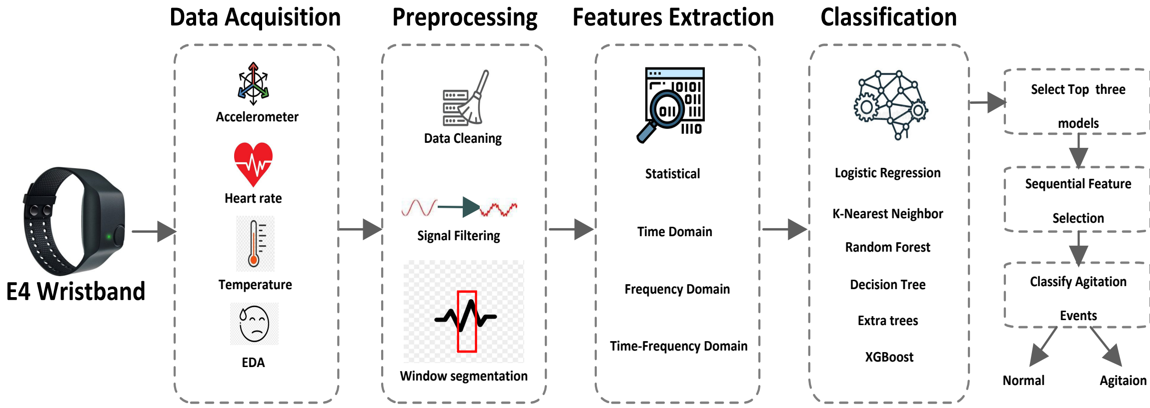


Fig. 1: E4 wristband classification system architecture.

to generate features for wearable devices, and it is an open-source package using Python that concentrates on processing physiological data. We use this open-source package as it is well-studied, uses different techniques we can compare, and provides a wide range of features to extract all necessary information. The main goal is to perform data cleaning and feature extraction of each attribute the E4 device collects with the most recent techniques and comprehensive features.

We start with a file reader to read the E4 device data and convert it to the correct format with the timestamps. The flirt function then uses several pre-processing steps to generate our features, such as filtering for noise removal and sliding windows. For the heart rate signal, the flirt function calculates the inter-beat interval (IBI), which is the time distance between two heartbeats. IBI generates heart rate variability (HRV) measures and removes artifacts within the IBI series. For the accelerometer signal, the flirt function applies a low-pass filter with a 10 Hz cut-off frequency. The EDA signal first gets artifact removal and noise filtering, and then EDA signal decomposition is included to generate the phasic and tonic components. We then apply a sliding-window approach with a window size of 1 minute and no overlapping to extract features from the datasets and calculate a feature vector from the four signals. Finally, we use the flirt function to extract features from each window from the time domain, the frequency domain, and statistical features with a total of 198 features [26].

D. Feature selection

Feature selection is generally utilized for dimensional reduction. Investigating a useful feature subset involves discovering features highly correlated with the selected feature set, but not correlated with each other. This study examines the impact of a feature selection technique called sequential feature selection (SFS) algorithms. The outstanding merit of the SFS methods is their straightforward implementation. This group of algorithms sequentially counts one feature at a time and includes that feature if it yields a more satisfactory classification accuracy. Sequential feature selection chooses features from a set of features and based on their correlation with the output, so the model performs better with minimal features. SFFS and SBFS add forward and backward steps at every iteration to ensure that the chosen feature best matches the previous and the following features. We use sequential forward floating selection since it includes the extra step of forward iteration that ensures the best set of features [27].

E. Classification Techniques and Performance Evaluation

The main purpose is to compare all algorithms, find the optimal set of parameters for each algorithm, and find the best algorithm for classification. We use Logistic Regression, Decision Tree, K-Nearesr Neighbor, Random Forest, XGBoost, and Extra Trees classifiers for classification. We then select the top three models and apply feature selection to reduce features and enhance performance. We set the number of features to the optimal features that provide the highest performance evaluation results with a minimum set of features to minimize the running time. We also apply cross-validation with k = 10 to protect against over-fitting.

Once we find the optimal features, we fit the new features to the previously selected three models and generate the results. For performance evaluation, we use several techniques to evaluate our results. Since the data is imbalanced between normal and agitation events, we use the balanced accuracy function in Sklearn. The function is built for binary classification problems to handle imbalanced data and calculates the average recall obtained from each class. We also calculate recall, F1-score, specificity, precision, geometric mean, and index balanced accuracy for evaluation. We use the imbalanced classification report which provides a classification report based on metrics used with imbalanced data.

IV. Results

A. E4 Wristband Signals Classification

We classify our dataset using six different machine learning algorithms with 198 features and two labels (Normal and agitation). We split the data into 70% training and 30% testing, and we use a standard scaler to scale our features between zero and one. We also use stratified folds with the same normal and agitation sets for training and testing. Table I shows the balanced accuracy for the different machine learning algorithms on three different datasets; patient one individually, patient two individually, and patient one and two combined. We notice that the personalized models for Participant One and Participant Two individually show better performance than patients one and Two combined with an average of 5-10% higher. These results confirm the proposed theory that each patient's aggressive behavior differs from others and that combining the patient's datasets will not show any obvious pattern to recognize normal and aggressive behavior. We also found that the top three models are Random Forest Classifier, XGBoost Classifier, and Extra Trees Classifier.

TABLE I: The balanced accuracy for the different machine learning algorithms.

Classification Model / Participant	Participant 1	Participant 2	Participant 1&2
Logistic Regression	80.803 %	82.587 %	73.221 %
Decision Tree	76.839 %	81.785 %	71.863 %
K-Neighbors	85.103 %	90.266 %	80.006 %
Random Forest	91.657 %	94.147 %	86.235 %
XGBoostoost	95.479 %	96.124 %	91.783 %
Extra Trees	87.822 %	91.094 %	85.021 %

B. Sequential Feature Selection Technique

We use the sequential feature selection technique with the top three models discussed in the previous section. We set the cross-validation to k = 10 and the float and forward features to true. Figures 2 (a), 2 (b), and 2 (c) show the Random Forest, XGBoost, and Extra trees model performance after using the sequential feature technique for participant one. The figure shows the performance for adding each feature until we reach 30 features and how it doesn't show improvement after adding more features. We conclude that approximately 30-35 features from the 198 were enough to classify the normal and agitation labels since we noticed a minimal improvement in performance for more features.

Table II shows the performance evaluation for the top three machine learning algorithms for Participant One. The results show that 2-10% enhances the accuracy after using sequential feature selection. We also notice that the extra trees model presents the best results with 96.19% accuracy, 98.54% precision, 98.55% recall, 93.82% specificity, 98.64% Fl-score, 96.13% geometric mean, and 92.84% index accuracy. For participant two, the results are shown in Table III. We notice that 2-3% enhances the accuracy after using sequential feature selection. We also notice that the extra trees model shows the highest results with 95.16% accuracy, 99.65% precision, 99.65% recall, 90.67% specificity, 99.65% Fl-score, 95.04% geometric mean, and 91.14% index accuracy.

Lastly, Table IV shows the performance evaluation for participants one and two combined after using sequential feature selection. After using sequential feature selection, the results show improvement with 2-4% accuracy. Also, the extra trees model shows the highest results with 94.15% accuracy, 98.85% precision, 98.85% recall, 89.47% specificity, 98.82% Fl-score, 93.99% geometric mean, and 89.16% index accuracy. Furthermore, features investigation showed that the heart rate features were the most selected to classify agitation events, followed by EDA and accelerometer with approximately the same features, and then the temperature features.

V. Conclusion

This paper presents initial analyses from a sample study using the Empatica E4 devices to detect episodes of AA in PwD from The Ontario Shores Centre for Mental Health Sciences. Our objective is to demonstrate the use of AI and feature engineering in detecting episodes of aggression and challenging behavior. We used up-to-date pre-processing techniques and extracted time, frequency, and statistical domain features. We compare six classification techniques, with Random Forest, XGBoost, and Extra Trees yielding the best results. Personalized models trained on individual's data outperform general models trained in aggregate data by 5-10%. This demonstrates that personalized models better understand the patient's behavior and classify it more efficiently.

We also used sequential feature selection to find the optimal set of features and showed that it improved results with fewer features with an average of 2-10% higher. Utilizing the Sequential feature selection technique reduces the running time and number of features, and provides better performance. The Extra Tree model achieved an average accuracy of 95.67% on individual data and 94.15% on aggregated data. In the future, we plan to collect more data and further investigate the features for more accurate results.

References

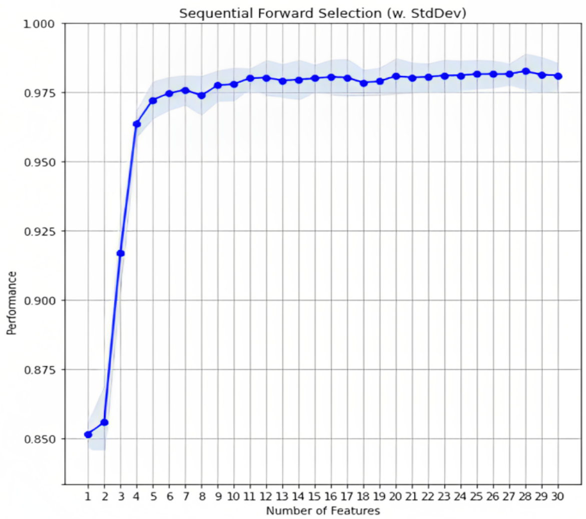
[1] "Dementia," World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia. [Accessed: 20-Sep-2023].

[2] C. G. Lyketsos, O. Lopez, B. Jones, A. L. Fitzpatrick, J. Breitner, and S. DeKosky, "Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment," *JAMA*, vol. 288, no. 12, p. 1475, 2002.

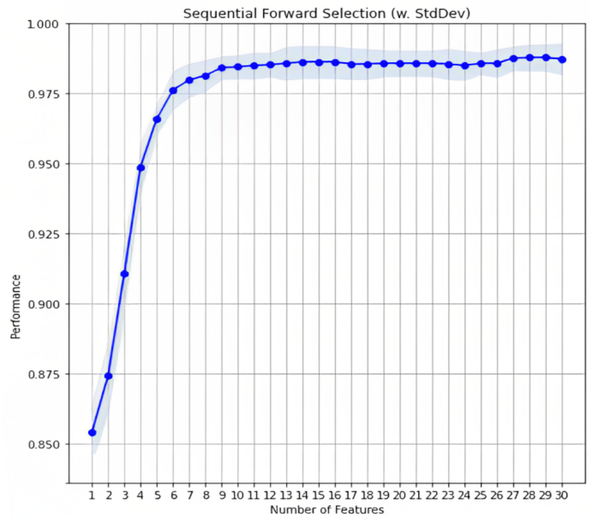
[3] C. Ballard and A. Corbett, "Agitation and aggression in people with Alzheimer's disease," *Current Opinion in Psychiatry*, vol. 26, no. 3, pp. 252-259, 2013, doi: 10.1097/YCO.0b013e32835f414b.

[4] M. Schnaider Beeri, P. Werner, M. Davidson, and S. Noy, "The cost of behavioral and psychological symptoms of dementia (BPSD) in community dwelling Alzheimer's disease patients," *International journal of geriatric psychiatry*, vol. 17, no. 5, pp. 403-408, 2002.

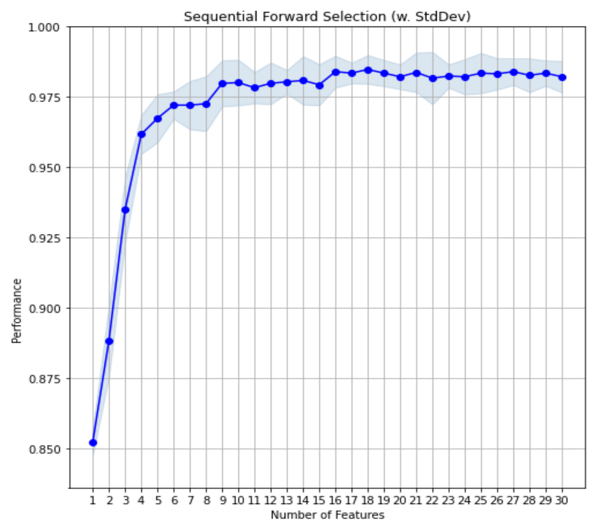
[5] G. o. Canada. [Online]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/dementia-highlights-canadian-chronic-disease-surveillance.html. [Accessed: 20-Sep-2023].



(a) Random Forest model performance.



(b) XGBoost model performance.



(c) Extratrees model performance.

Fig. 2: The three classification models' performance with sequential features technique for participant one.

TABLE II: The performance evaluation for the different machine learning algorithms for participant one.

Classification model/ performance evaluation	Balanced Accuracy	Balanced Precision	Balanced Recall	Balanced Specificity	Balanced F1-score	Geometric Mean	Index Balanced Accuracy
Random Forest Classifier	94.29 %	98.14 %	98.13 %	90.46 %	98.09 %	94.13 %	89.29 %
XGBoostoost Classifier	95.59 %	98.36 %	98.37 %	92.80 %	98.35 %	95.50 %	91.72 %
ExtraTrees Classifier	96.19 %	98.54 %	98.55 %	93.82 %	98.54 %	96.13 %	92.84 %

TABLE III: The performance evaluation for the different machine learning algorithms for participant two.

Classification model/ performance evaluation	Balanced Accuracy	Balanced Precision	Balanced Recall	Balanced Specificity	Balanced F1-score	Geometric Mean	Index Balanced Accuracy
Random Forest Classifier	94.17 %	99.53 %	99.54 %	88.80 %	99.53 %	94.00 %	89.31 %
XGBoostoost Classifier	93.87 %	98.17 %	98.54 %	87.80 %	99853 %	93.50 %	88.75 %
ExtraTrees Classifier	95.16 %	99.65 %	99.65 %	90.67%	99.65 %	95.04 %	91.14 %

TABLE IV: The performance evaluation for the different machine learning algorithms for participants one and two combined.

Classification model/ performance evaluation	Balanced Accuracy	Balanced Precision	Balanced Recall	Balanced Specificity	Balanced F1-score	Geometric Mean	Index Balanced Accuracy
Random Forest Classifier	92.63 %	98.69 %	98.70 %	88.56 %	98.67 %	93.43 %	88.17 %
XGBoostoost Classifier	94.13 %	98.78 %	98.79 %	90.36 %	98.77 %	94.43 %	89.93 %
ExtraTrees Classifier	94.15 %	98.85 %	98.85 %	89.47 %	98.82 %	93.99 %	89.16 %

[6] M. Amengual-Gual, A. Ulate-Campos, and T. Loddenkemper, ”Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients,” *Seizure*, vol. 68, pp. 31-37, 2019.

[7] R. Shegog, L. Braverman, and J. D. Hixson, ”Digital and technological opportunities in epilepsy: Toward a digital ecosystem for enhanced epilepsy management,” *Epilepsy Behavior*, vol. 102, p. 106663, 2020.

[8] A. J. Bharucha et al., ”Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges,” *The American journal of geriatric psychiatry*, vol. 17, no. 2, pp. 88-104, 2009.

[9] A. Badawi, K. Elgazzar, B. Ye, K. Newman, A. Mihailidis, A. Iaboni , and S. Khan “Investigating Multimodal Sensor Features Importance to Detect Agitation in People with Dementia,” *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2023.

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

[11] C. Fabrizio, A. Termine, C. Caltagirone, and G. Sancesario, ”Artificial intelligence for Alzheimer's disease: promise or challenge?,” *Diagnostics*, 2021.

[12] ”Ontario Shores Centre for Mental Health Sciences.” [Online]. Available: https://www.ontarioshores.ca/. [Accessed: 20-Sep-2023].

[13] “E4 wristband,” E4 for Developers. [Online]. Available: https://developer.empatica.com/index.html. [Accessed: 20-Sep-2023].

[14] A. Knuff, R. H. Leung, Dne. P. Seitz, L. Pallaveshi, and A. M. Burhan, ”Use of actigraphy to measure symptoms of agitation in dementia,” *The American Journal of Geriatric Psychiatry*, vol. 27, no. 8, pp. 865-869, 2019.

[15] W.-T. M. Au-Yeung, L. M. Miller, Z. Beattie, C. L. Reynolds, and J. Kaye, “Identifying digital biomarkers of dementia-related agitation: Preliminary results from the moderate study,” *Innovation in Aging*, vol. 3, no. Supplement_1, 2019.

[16] A. Hein, F. Krüger, S. Bader, P. Eschholz, and T. Kirste, ”Challenges of collecting empirical sensor data from people with dementia in a field study,” *in 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017: IEEE, pp. 22-25.

[17] A. W. van der Kooi, T. H. Kappen, R. J. Raijmakers, I. J. Zaal, and A. J. C. Slooter. “Temperature variability during delirium in ICU patients: an observational study”. In: Public Library of Science (PLoS). 2013

[18] F. Amato et al., ”CLONE: a promising system for the remote monitoring of Alzheimer's patients: an experimentation with a wearable device in a village for Alzheimer's care,” *in Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, 2018, pp. 255-260.

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

[20] D. Goerss et al., ”Automated sensor-based detection of challenging behaviors in advanced stages of dementia in nursing homes,” *Alzheimer's Dementia*, 2019.

[21] P. Yang, G. Bi, J. Qi, X. Wang, Y. Yang, and L. Xu, “Multimodal Wearable Intelligence for Dementia Care in Healthcare 4.0: A survey,” *Information Systems Frontiers*, 2021.

[22] S. Spasojevic et al., ”A pilot study to detect agitation in people living with dementia using multi-modal sensors,” *Journal of Healthcare Informatics Research*, vol. 5, no. 3, pp. 342-358, 2021.

[23] A. Iaboni et al., ”Wearable multimodal sensors for the detection of behavioral and psychological symptoms of dementia using personalized machine learning models,” *Alzheimer's Dementia: Diagnosis, Assessment Disease Monitoring*, vol. 14, no. 1, p. e12305, 2022, doi: https://doi.org/10.1002/dad2.12305.

[24] A. P. A. C. o. N. a. Statistics, Diagnostic and statistical manual: Mental disorders. Washington, D.C: American Psychiatric Association, 1952.

[25] S. Ollander, C. Godin, A. Campagne, and S. Charbonnier, ”A comparison of wearable and stationary sensors for stress detection,” *in 2016 IEEE International Conference on systems, man, and Cybernetics (SMC)*, 2016: IEEE, pp. 004362-004366.

[26] S. Föll et al., ”FLIRT: A feature generation toolkit for wearable data,” *Computer Methods and Programs in Biomedicine*, vol. 212, p. 106461, 2021.

[27] A. A. Badawi, A. Al-Kabbany, and H. A. Shaban, ”Sensor type, axis, and position-based fusion and feature selection for multimodal human daily activity recognition in wearable body sensor networks,” *Journal of Healthcare Engineering*, vol. 2020, 2020.