sciendo
ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: https://content.sciendo.com

Early Diagnosis of Dementia Patients by SPADE Activity Prediction Algorithm

Khairun Nisa' Minhad[1], Araf Farayez[2], Mamun Bin Ibne Reaz[2], Mohammad Arif Sobhan Bhuiyan[1], Siti Balqis Samdin[1,*], Mahdi H. Miraz[3,4,*]

[1]*School of Electrical Engineering and Artificial Intelligence, Xiamen University Malaysia, 43900, Sepang, Selangor, Malaysia,* sitibalqis.samdin@xmu.edu.my
[2]*Centre of Advanced Electronic and Communication Engineering, Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia, 43900, Bangi, Selangor, Malaysia*
[3]*School of Computing and Data Science, Xiamen University Malaysia, 43900, Sepang, Selangor, Malaysia,* m.miraz@ieee.org
[4]*School of Computing, Faculty of Arts, Science and Technology, Wrexham Glyndwr University, Wrexham, UK.*

Abstract: Dementia is not a specific disease, but a general term for age-related decline or loss of memory, cognitive abilities including problem solving and decision-making, and one's own language, which significantly interfere with daily life. Researchers around the world have developed ways to automate the diagnosis of dementia through the use of machine learning and data mining approaches. The aim of this research project is to design and develop a day-to-day activity prediction algorithm in order to accurately identify and differentiate the dementia affected patients from the healthy subjects, to ensure early diagnosis of dementia development. This research advocates a novel algorithm called 'Sequence Prediction via All Discoverable Episodes (SPADE)' as a statistical tool to map activities of daily life (ADLs) in different groups of people in order to develop a unique parameter for precise diagnosis. The results of our experiment demonstrated a significant difference (i.e. 11 %) in the sequence prediction peak accuracy between the healthy subjects and the residents with dementia. SPADE demonstrated an adequate accuracy (i.e. 80 % on average), with an improvement of about 12 % compared to the performance of M-SPEED in inferring future occurrences of activities. It is thus evident that the algorithms for activity predictions show promise for early detection of dementia symptoms without the use of any expensive clinical procedure.

Keywords: Dementia, cognitive disorder, activity prediction, sequence prediction, smart home.

1. Introduction

Dementia, sometimes referred to as Major Neurocognitive Disorder (MND), is an age-related disease associated with a decline in memory or cognitive abilities large enough to affect a person's ability to carry out daily activities. World Alzheimer Report 2022[i] reveals that people over the age of 60 represent 16 % of the total population of Brazil, the fifth largest country in the world with a great diversity of cultural and socio-economic profiles. A total of approximately 1.8 million Brazilian residents were suffering from dementia in 2019; with this number increasing every year [1]. We can expect a similar scenario all over the world. Patients with dementia are more likely to forget to accomplish their Activities of Daily Living (ADL) due to their acute memory loss, necessitating caregiver support. In order to provide appropriate care and treatment, it is crucial to identify and diagnose persons suffering from dementia. Early diagnosis of neurocognitive disorders can greatly delay the progression of dementia in elderly patients and in some cases contribute to recovery [2]-[3].

Various machine learning and artificial intelligence strategies have been used in the development of proper smart systems for the early diagnosis of cognitive impairment. In a research done by So et al., they put forward a dual-layer machine learning model inspired by the clinical methods adopted in dementia support centres [4]. Mansourian et al. also proposed a systematic and comprehensive data mining modelling approach to improve dementia evaluation [5]. Bratić et al. predicted Clinical Dementia Rating (CDR) scores using neuropsychological and demographic data by implementing machine learning algorithms [6]. In another study, decision tree, Naive Bayes and logistic regression algorithms were used to classify patients with cognitive impairment [7].

The Clinical Assessment using Activity Behaviour (CAAB) approach developed by Dawadi et al. [8] employs

DOI: 10.2478/msr-2023-0011

**Corresponding authors: m.miraz@ieee.org (M. H. Miraz); sitibalqis.samdin@xmu.edu.my (S. B. Samdin)*

statistical characteristics of resident behaviour for training the machine learning models and for making clinical diagnoses. A classification model was developed to spot changes in the daily functioning of smart home residents [9].

The most promising results were found in the experiments of Dawadi et al. where a smart home named 'Center for Advanced Studies in Adaptive Systems (CASAS)' was considered for a study with 179 participants comprising 145 healthy residents, 2 dementia patients and 32 patients with a mild cognitive disorder [8]. The researchers used several machine learning algorithms to automate the classification of the patients with cognitive disorders. The implemented methods were able to explain nearly 62 % variations in the observational scores.

Although machine learning approaches provide better results for a large number of training samples, smart homes are gradually adopting data compression algorithms due to their improved reactivity to patterns of behaviour. Algorithms for activity prediction such as Active LeZi (ALZ) [10] and Sequence Prediction via Enhanced Episode Discovery (SPEED) [11] are such methods that use various data compression methods to predict user behaviour. The time component was introduced in SPEED to devise an advanced version called 'Modified SPEED (M-SPEED)', which achieves even higher accuracy, thus making it more suitable than other algorithms that use data compression tools [12]. However, M-SPEED and SPEED both employ a decision tree generation technique in which probability assignment and subtree generation are performed simultaneously. As a result, the created tree becomes inconsistent, which reduces its sensitivity to repeated input. The odds in a recently created technique called Sequence Prediction through All Discoverable Episodes (SPADE) are changed to resolve this contradiction. In this algorithm, tree generation is separated from probability assignment, resulting in more accurate probabilities for the different activity episodes [13].

Patients with dementia frequently miss chores, while a regular smart home resident exhibits repeating patterns in their daily routine. Their behavioural pattern becomes irregular because of this. Activity prediction tools can effectively detect these patterns of irregularities and diagnose a probable cognitive disorder.

This research uses a data compression technique based on an activity prediction algorithm, to diagnose cognitive disorders in smart home residents. For this purpose, we introduce the SPADE algorithm [13]. Our proffered algorithm uses the Prediction by Partial Matching (PPM) data compression technique such as SPEED and M-SPEED, but addresses the problems related to probability inconsistency, as they appeared in the previous ones [11]. The presented method will be used to test datasets containing real-world data on healthy and dementia affected residents and classify them based on prediction accuracy. The ability of activity prediction methods to serve as diagnostic tools for dementia in its early stages can be derived from a consistently reliable classification method.

2. METHODOLOGY

This work aims to determine the developmental stages of dementia, where residents' behavioural analysis is essential to identify behavioural activity patterns and predict future activity. A sequence prediction algorithm SPADE is introduced that can use the activity patterns of the users to detect the cognitive disorder symptoms. SPADE works by detecting activities and generating episodes from them, which then form a decision tree. From this decision tree, it will assess and designate the events that are most likely to occur, by using PPM.

In the context of a home environment, residents have a strong propensity to perform daily tasks repeatedly in an identifiable pattern. The system can identify these patterns of actions in relation to home appliance usage and anticipate probable future actions or activities. Consider a situation in a kitchen where multiple different appliances are available. Before anyone starts working in a kitchen, they switch ON the lights and only then proceed to the next task. Assume next, they may want to make coffee, so they switch ON the coffee machine for 8 minutes before switching it back OFF. Similarly, in the time they make coffee, they may open the refrigerator's door, thus switching ON its light, to grab some ingredients, or even turn ON the TV in the living room to pass the time. Finally, after eating in the living room for 20 minutes, they may come back into the kitchen and switch OFF the lights to go to another room. From the situation depicted, we will elaborate on how human activity can be recognized through appliance usage to derive a specific behavioural pattern. SPADE would generate episodes that take into account that specific pattern of actions and activities. Consider an example that contains the appliance's state and its opposite state as 'AdCBbaDgBiI'. The capital letters here characterise the appliance to be in the ON state, while lowercase letters indicate that the appliance is OFF. An ON and an OFF state of an appliance signify the formation of a single episode. For instance, a bulb being switched ON and then again switched OFF after a while means that it is an episode of activities. The pseudo-codes for converting the raw action or activity sequence into a new one so that the episode is formed are shown in Fig. 1.

```
loop
   take first element
   find next opposite event
   define episode
   loop
      search all events of the episode
      eliminate duplicate events if they
      do not follow the episode conditions
      end

   store episodes and duration of time in
   array
   end of data
 forever
```

Fig. 1. To generate the episode, pseudo code to convert the raw sequence into a new sequence.

The legitimate series of tasks, also called events, as well as the related duration time information serve SPADE as input data. The variable 'length' is strictly related to the maximum length of the episode. The initial value of 'seq_queue' is NULL. Once an event's first element, let us assume it to be 'c', is entered into SPADE, the algorithm enters it into the 'seq_queue' and seeks for its opposite state, i.e. 'C' in this case. The event enters a waiting state if the opposite state is not identified.

To determine the opposite state of each event, the algorithm will first look through each event as it is added one at a time. The method will shorten the sequence and create an episode if the opposite state is observed, i.e. c…C in this case, and increase the value of 'Max_Window_Length'. The size of the window and the number of possible contexts rely on the value of 'Max_Window_Length'. Essentially, an episode is formed only if an event and its opposite event are observed by the algorithm. The algorithm is well aware that there may be sub-episodes within an episode and these episodes do not necessarily have to start with an ON event. The important aspect is to have the event with its opposite state together. Since the SPADE algorithm includes timing information of the events, the following digits of an event are temporarily stored in a separate time variable. This temporarytime variable is imperative because the total duration of an event needs to be stored along with the valid episodes found. Essentially, the algorithm looks for any episode and then stores its duration. After generating an episode, SPADE will produce each potential context along with its associated frequency of occurrences, as shown in Table 1. Next, a decision tree is created to position the episodes and to determine the probability that each of the activities will take place. The decision or activity with the highest probability is then selected as the future event. Using the analysis of this decision tree and the PPM technique, future occurrences can be predicted. This is due to the fact that PPM calculates the episode's weighted probability by shortening the length of the episode.

Table 1. Possible contexts with their frequencies.

Sequence	Current Sequence	seq_queue After current event	Maximum episode length	Episode	Max Episode length after Episode extraction	seq_queue After Episode extraction	Possible contexts from the Episode	All possible contexts with Frequency of Occurrence
1	A	A	1	Not found	1	A		A(1), a(3), B(4), b(4), C(2), c(0), D(2), d(2), G(0), g(1), I(1), i(1), Ad(1), dC(1), CB(2), ba(3), aD(2), Bb(3), Dg(1), gB(1), iI(1), Adc(1), dCB(2), CBb(2), Bba(2), baD(2), aDg(1), DgB(1), AdCB(1), dCBb(2), Cbba(2), BbaD(1), baDg(1), aDgB(1), AdCBb(1), dCBba(1), CBbaD(1), baDgB(1), AdCBba(1), dCBbaD(1)
2	d	Ad	1	Not found	1	Ad		
3	C	AdC	1	Not found	1	AdC		
4	B	AdCB	1	Not found	1	AdCB		
5	b	AdCBb	1	Not found	1	AdCBb		
6		AdCBb	1	Bb	2	AdCBb	B,b,Bb	
7	a	AdCBba	5	Not found	5	AdCBba		
8		AdCBba	6	AdCBba	6	AdCBba	A, d, C, B, b, a, Ad, dC, CB, Bb, ba, AdC, dCB, CBb, Bba, AdCB, dCBb, Cbba, AdCBb, dCBba, AdCBba	
9	D	dcBbaD	6	Not found	6	dcBbaD		
10		dcBbaD	6	dCBbaD	6	dcBbaD	d, c, B, b, a, D, dC, CB, Bb, ba,aD, dCB, CBb, Bba,baD, dCBb, Cbba, BbaD, dCBba, CbbaD, dCBbaD	
11	g	CbbaDg	6	Not found	6	CbbaDg		
12	B	CbbaDgB	6	Not found	6	CbbaDg		
13		CbbaDgB	6	baDgB	6	baDgB	b, a, D, g, B, ba, aD, Dg, gB, baD, aDg, DgB, baDg, aDgB, baDgB	
14	i	CbbaDgBi	6	Not found	6	CbbaDgBi		
15	I	CbbaDgBiI	6	iI	6	iI	i, I, iI	

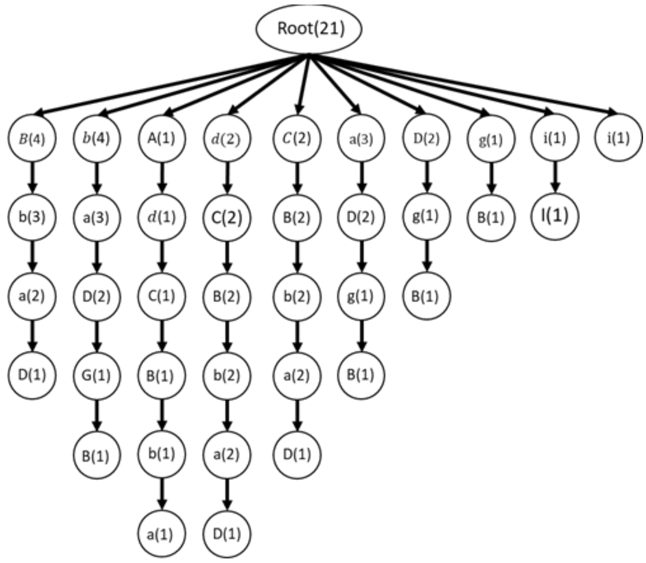

Fig. 2. Decision tree formed by SPADE for 'AdCBbaDgBiI'.

Consider 'AdCBbaDgBiIG' to be a sequence obtained from the dataset and 'AdC' be the present window state. SPADE will use every potential phase, i.e. 'A', 'd', 'C', 'Ad', 'dC', 'AdC' as well as null to form the decision tree to calculate the probability of the future events. For instance, to determine the probability that event 'b' will occur next, we need to use the decision tree depicted in Fig. 2. As can be seen form the decision tree, the probability of event 'b' occurring in the context of 'AdC' is ½ (once out of twice). Next, the lowered order probability needs to be calculated. As stated, the probability of occurrence of 'b' after the occurrence of 'Ad' is zero (0). This indicates that no occurrences of 'b' are observed after the occurrence of 'Ad', whereas one-third (1/3) is the probability of the occurrence of null. If we repeat this operation for the entire remaining order, we find that the probability of occurrence of event 'b' after event 'A' is zero (0). The probability of occurrence of null after event 'A' is one-sixth (1/6). The last phase of the calculation is null. According to the tree; the probability of occurrence of event 'b' is 7/43. Therefore, by using (1), the PPM helps to calculate the probability of 'b' as:

$$\frac{1}{2}+\frac{1}{2}\left(\frac{0}{3}+\frac{1}{3}\left(\frac{0}{6}+\frac{1}{6}\left(\frac{7}{43}\right)\right)\right)=0.5045. \qquad (1)$$

The probability of other symbols can also be calculated and compared using this method. For example, the likelihood that 'd' will appear following the same order, i.e. 'AdC', is as demonstrated in (2).

$$\frac{0}{2}+\frac{1}{2}\left(\frac{0}{3}+\frac{1}{3}\left(\frac{0}{6}+\frac{1}{6}\left(\frac{4}{43}\right)\right)\right)=0.026. \qquad (2)$$

On this basis it is also possible to determine the probabilities of the remaining events or symbols. Finally, SPADE selects the event with the maximum probability of occurrence as the next event. If the maximum probability is shared, then the time agent plays a crucial part. This is because an event that occurs over a comparatively longer period of time should be favoured as the most probable event.

An accuracy test is performed using the processed SPADE data analysed to determine how accurate it is at predicting the future events. This test is not only important to verify the validity and potential of the algorithm itself, but can also be used for analyses and comparison that help measure the phases of residents' dementia development.

3. RESULTS AND DISCUSSION

In this study, the Center for Advanced Studies in Adaptive Systems (CASAS) dataset was tested using the proposed method to analyse users' ADL. Users with mild cognitive problem, dementia patients and healthy residents' activity records are all included in this dataset [14]. We randomly grouped 3 sets of healthy subjects and corresponding 3 sets of residents with cognitive anomalies. The prediction accuracies for the respective episode lengths are depicted in Fig. 3, Fig. 4, Fig. 5. The calculations show that in group 1 the dementia subjects had a peak accuracy of 66.70 %, while that of the healthy subjects was 78.26 %, as shown in Fig. 3. In group 2, dementia subjects achieved a value of 58.04 %, while healthy subjects achieved a value of 75.74 % (Fig. 4). Lastly, group 3 recorded a value of 75.67 % for dementia patients and 92.50 % for healthy subjects (Fig. 5).

The resulting graphs show that SPADE consistently scores lower for the dementia-affected residents in terms of correctly predicting the residents' future activities. It is evident that the healthy subjects' peak accuracy is significantly higher than that of the dementia subjects. Peak accuracy differences between the two patients were estimated to be at least 11% on average. The rationale is that the algorithm can more easily examine the more predictable and recognised patterns of events that healthier residents make. In contrast, dementia patients frequently exhibit memory loss symptoms that make it challenging for them to perform daily duties sequentially. This can be in the form of either not completing a certain task or skipping an activity that they would normally do on a daily basis. Taking this factor into account, it would mean that a dementia patient's activity data would have a lot of randomness and inaccuracy, to begin with. Since the fundamental principle of SPADE is to recognize events and build a decision tree of possible outcomes and select the most likely ones, it would repeatedly score low for patients with neurocognitive disorder.

Using the MavLab dataset, a comparison between SPADE and the earlier sequence prediction method, namely M-SPEED, was also done. The 1675 chronological events in this training dataset represent a single resident's behaviour over time. In Fig. 6, the accuracy results are displayed in relation to the episode length. For M-SPEED, the average accuracy is around 68 %, while for SPADE a value of 80 % was obtained.

The findings demonstrate that SPADE outperforms M-SPEED with regard to its ability to accurately predict future events. In the simulation, the peak accuracy of M-SPEED is approximately 68 %, while for SPADE it is 80 %. This shows that SPADE's accuracy has been improved by 12 % for correctly predicting probable future events.

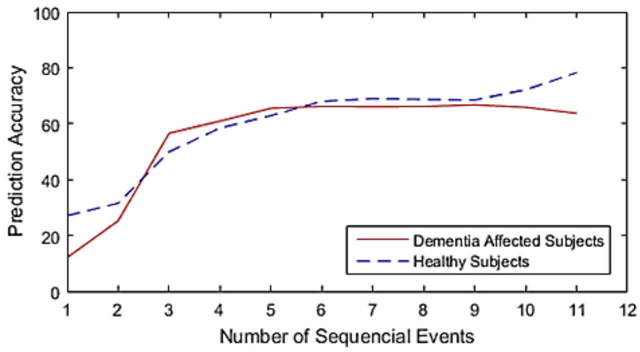

Fig. 3. Prediction accuracy for the healthy vs. the dementia-affected residents in group 1.

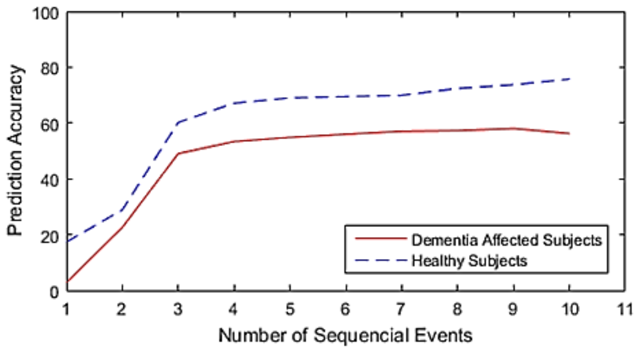

Fig. 4. Prediction accuracy for the healthy vs. the dementia-affected residents in group 2.

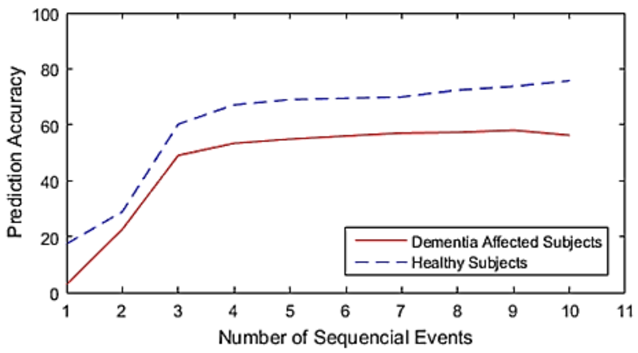

Fig. 5. Prediction accuracy for the healthy vs. the dementia-affected residents in group 3.

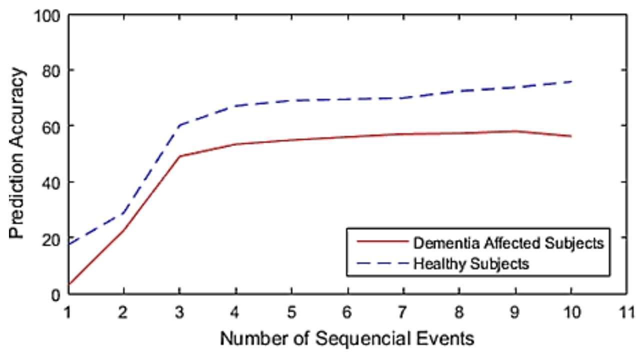

Fig. 6. Comparison of accuracy of SPADE vs. M-SPEED using MavLab dataset without time verification.

Based on the results obtained by examining its prediction accuracy, we conclude that SPADE can successfully distinguish between the healthy and the dementia-affected subjects. A higher prediction accuracy reflects a healthy subject's data pattern, while a lower prediction accuracy reflects a dementia-affected subject's data pattern. This shows that SPADE's ability to recognise behavioural patterns makes it highly effective at identifying the early stages of dementia.

4. Conclusion and future work

Detection of early symptoms of dementia and successful diagnosis play a vital role in maintaining good cognitive health in the population. Smart home algorithms are widely used to analyse resident interaction in order to look for behavioural anomalies. Activity prediction algorithms can provide us with useful tools for this purpose. In this research, we introduced SPADE for smart home application that can autonomously monitor resident actions and detect activity patterns. The algorithm was tested on a dataset of smart homes and was successful in differentiating between healthy individuals and people with cognitive disorders by observing residents' activity patterns. With a maximum accuracy difference of 11 %, it is clear that activity prediction algorithms have great potential as a diagnostic aid for dementia.

In the future, further studies can be conducted with larger datasets and different populations to validate the findings of this study. The study can also be extended to include other forms of dementia and the algorithm can be tested on different datasets. Additionally, the algorithm can be optimised to improve its accuracy even further. The findings of this study can be used to develop a practical and reliable diagnostic tool for the early detection of dementia.

Acknowledgment

This research was supported by the Xiamen University Malaysia through XMUMRF/2022-C10/IECE/0038 and XMUMRF/2021-C8/IECE/0025 grants. It is partially supported by the ICTP through the Affiliated Centres Programme.

References

[1] Oliveira, D., da Mata, F.A.F., Brucki, S., Ferri, C.P. (2022). Post-diagnostic support for people living with dementia and their family carers in Brazil. In *World Alzheimer Report 2022. Life after Diagnosis: Navigating Treatment, Care and Support*. London, England: Alzheimer's Disease International, 103-104. https://www.alzint.org/u/World-Alzheimer-Report-2022.pdf

[2] Mekuria, D.N., Sernani, P., Falcionelli, N., Dragoni, A.F. (2021). Smart home reasoning systems: A systematic literature review. *Journal of Ambient Intelligence and Humanized Computing*, 12 (4), 4485-4502. https://doi.org/10.1007/s12652-019-01572-z

[3] Tanveer, M., Richhariya, B., Khan, R.U., Rashid, A.H., Khanna, P., Prasad, M., Lin, C.T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 16 (1s), 1-35. https://doi.org/10.1145/3344998

[4] So, A., Hooshyar, D., Park, K., Lim, H. (2017). Early diagnosis of dementia from clinical data by machine learning techniques. *Applied Sciences*, 7 (7), 651. https://doi.org/10.3390/app7070651

[5] Mansourian, M., Khademi, S., Marateb, H.R. (2021). A comprehensive review of computer-aided diagnosis of major mental and neurological disorders and suicide: A biostatistical perspective on data mining. *Diagnostics*, 11 (3), 393. https://doi.org/10.3390/diagnostics11030393

[6] Bratić, B., Kurbalija, V., Ivanović, M., Oder, I., Bosnić, Z. (2018). Machine learning for predicting cognitive diseases: Methods, data sources and risk factors. *Journal of Medical Systems*, 42, 243. https://doi.org/10.1007/s10916-018-1071-x

[7] Kang, M.J., Kim, S.Y., Na, D.L. et al. (2019). Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. *BMC Medical Informatics and Decision Making*, 19, 231. https://doi.org/10.1186/s12911-019-0974-x

[8] Dawadi, P.N., Cook, D.J., Schmitter-Edgecombe, M. (2016). Automated cognitive health assessment from smart home-based behavior data. *IEEE Journal of Biomedical and Health Informatics*, 20 (4), 1188-1194. https://doi.org/10.1109/JBHI.2015.2445754

[9] Javed, A.R., Fahad, L.G., Farhan, A.A., Abbas, S., Srivastava, G., Parizi, R.M., Khan, M.S. (2021). Automated cognitive health assessment in smart homes using machine learning. *Sustainable Cities and Society*, 65, 102572. https://doi.org/10.1016/j.scs.2020.102572

[10] Gopalratnam, K., Cook, D.J. (2007). Online sequential prediction via incremental parsing: The active LeZi algorithm. *IEEE Intelligent Systems*, 22 (1), 52-58. https://doi.org/10.1109/MIS.2007.15

[11] Alam, M.R., Reaz, M.B.I., Ali, M.M. (2012). SPEED: An inhabitant activity prediction algorithm for smart homes. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42 (4), 985-990. https://doi.org/10.1109/TSMCA.2011.2173568

[12] Farayez, A., Reaz, M.B.I., Arsad, N. (2018). Computational enhancement of all possible context generation in modified-SPEED algorithm. In *2018 International Conference on Advances in Computing, Communications and Informatics*. IEEE, 1424-1428. https://doi.org/10.1109/ICACCI.2018.8554387

[13] Farayez, A., Reaz, M.B.I., Arsad, N. (2018). SPADE: Activity prediction in smart homes using prefix tree-based context generation. *IEEE Access*, 7, 5492-5501. https://doi.org/10.1109/ACCESS.2018.2888923

[14] Cook, D.J., Schmitter-Edgecombe, M. (2009). Assessing the quality of activities in a smart environment. *Methods of Information in Medicine*, 48 (5), 480-485. https://doi.org/10.3414/me0592

Received November 13, 2022
Accepted March 13, 2023

[i] https://www.alzint.org/resource/world-alzheimer-report-2022/