11-2022

Predictive self-organizing neural networks for in-home detection of Mild Cognitive Impairment

Seng Khoon TEH

Iris RAWTAER

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Citation

TEH, Seng Khoon; RAWTAER, Iris; and TAN, Ah-hwee. Predictive self-organizing neural networks for in-home detection of Mild Cognitive Impairment. (2022). *Expert Systems with Applications*. 205, 1-14.
Available at: https://ink.library.smu.edu.sg/sis_research/7215

Predictive Self-Organizing Neural Networks for In-Home Detection of Mild Cognitive Impairment

Seng-Khoon Teh[a] (sk.teh.2020@smu.edu.sg), Iris Rawtaer[b] (iris.rawtaer@singhealth.com.sg), Ah-Hwee Tan[c] (ahtan@smu.edu.sg)

[a] Postgraduate Program, School of Computing and Information Systems, Singapore Management University, Singapore
[b] Department of Psychiatry, Sengkang General Hospital, Singhealth Duke NUS Academic Medical Center, Singapore
[c] School of Computing and Information Systems, Singapore Management University, Singapore

Corresponding Author:
Seng-Khoon Teh
School of Computing and Information Systems, Singapore Management University, Singapore
Tel: (65) 9386-2367
Email: sk.teh.2020@smu.edu.sg

Predictive Self-Organizing Neural Networks for In-Home Detection of Mild Cognitive Impairment

Seng-Khoon Teh[a], Iris Rawtaer[b], Ah-Hwee Tan[c,*]

[a]*Postgraduate Program, School of Computing and Information Systems, Singapore Management University, Singapore*
[b]*Department of Psychiatry, Sengkang General Hospital, Singhealth Duke NUS Academic Medical Center, Singapore*
[c]*School of Computing and Information Systems, Singapore Management University, Singapore*

Abstract

In-home sensing of daily living patterns from older adults coupled with machine learning is a promising approach to detect Mild Cognitive Impairment (MCI), a potentially reversible condition with early detection and appropriate intervention. However, the number of subjects involved in such real-world studies is typically limited, posing the so-called small data problem to most predictive models which rely on a sizable number of labelled data. In this work, a predictive self-organizing neural network known as fuzzy Adaptive Resonance Associate Map (fuzzy ARAM) is proposed to detect MCI using in-home sensor data collected from a unique Singapore cross-sectional study. Specifically, mean and standard deviation of nine in-home behavioural attributes of 49 subjects over two months were derived for each subject from the raw sensor data. We first applied fuzzy ARAM to the 49-subject data set with missing data, and achieved a F1-score of 58.3% to detect MCI from cognitive healthy. To eliminate the effect of missing data, we next conducted our study using an even smaller 25-subject data set with no missing values, of which fuzzy ARAM achieved a F1-score of 63.6%. To derive concise rules for prediction and interpretation, antecedent pruning was subsequently employed. For the 25-subject data set, the F1-score improved to 76.2%, while the symbolic IF-THEN rules revealed that behaviour metrics such as variation of forgetfulness and sleep contained notable predictive utility. Compared with Support Vector Machines (SVM), Decision Tree, Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), our benchmark experiments show that fuzzy ARAM provided the highest predictive performance and yielded unique rules for MCI detection. These results demonstrate the potential of fuzzy ARAM to detect MCI using in-home monitoring sensor data.

Keywords: Predictive Self-Organizing Neural Networks, Adaptive Resonance Associative Map, Fuzzy ARAM, In-home Monitoring, Mild Cognitive Impairment

*Corresponding author.
Email addresses: sk.teh.2020@smu.edu.sg (Seng-Khoon Teh), iris.rawtaer@singhealth.com.sg (Iris Rawtaer), ahtan@smu.edu.sg (Ah-Hwee Tan)

1. Introduction

Dementia is a neurodegenerative disease with prevalence rate of approximately 50 million worldwide presently and projected to triple by 2050 (World Health Organization, 2021). Early detection of at-risk stage of dementia known as Mild Cognitive Impairment (MCI) coupled with multidomain, multicomponent interventions may halt or even reverse the progression of MCI to become dementia (McMaster et al., 2020). To facilitate early detection, healthcare professionals and researchers are actively investigating different novel approaches to detect MCI in the community (Tierney & Lermer, 2010). An objective and accurate technique that can be deployed long-term within the community for MCI detection would represent a significant advancement in the field of dementia, impacting the world.

In-home sensor monitoring technology to capture data related to daily living behaviours of older adults is a promising technique to detect MCI (Hayes et al., 2008). For instance, Akl et al. (2015) demonstrated the potential of in-home walking speed data and in-room activity distribution as predictive features for MCI detection. Khan & Jacobs (2021), on the other hand, has shown that complex in-home movement parameters related to cognitive confusion and forgetfulness can be used to identify MCI from cognitive healthy subjects. As the number of subjects involved in such studies are typically limited (sample size <300), classical machine learning model such as Support Vector Machine (SVM) has thus far been explored to differentiate MCI from cognitive healthy (Akl et al., 2015; Khan & Jacobs, 2021). Clearly, most in-home studies centered their attention on extraction of novel data features from the raw sensor data to identify MCI from cognitive healthy. Investigation of novel machine learning models that may potentially be more suited for such small data challenge to identify MCI from cognitive healthy remains elusive.

Adaptive Resonance Associative Map (ARAM), a predictive self-organizing neural network (Tan, 1995) (Carpenter & Grossberg, 1992), has been successfully applied to tackle a wide range of pattern recognition problems (Tan & Pan, 2005). In particular, it has been shown to achieve predictive performance to be equivalent or even superior to many machine learning models including Backpropagation Neural Network for small and noisy data set of less than 300 samples (Tan, 1995). Besides the capability to offer superior predictive performance, ARAM possesses an unique capability to address missing data problem (Granger et al., 2000), a problem which especially exacerbates machine learning performance in an already small and noisy data set. Specifically, due to the employment of complement coding in ARAM for preserving amplitude information from input data attribute whilst preventing category proliferation problem, complement coding can be further set to fill in for any missing values of an input data attribute (Granger et al., 2000); this unique property enables ARAM to retain the original data sample size during machine learning, circumventing missing data problem. On top of all these, ARAM can also furnish symbolic IF-THEN rules for interpretation (Tan & Pan, 2005); interpretability of prediction model is critical in healthcare domain (Amann et al., 2020), especially when using it for clinical decision support

(Zihni et al., 2020) such as in the application of in-home sensor monitoring system for detection of MCI from cognitive healthy. Evidently, ARAM inherently possesses multiple properties (i.e., high accuracy for small data problem, missing data handling, compatibility to symbolic IF-THEN rules) that could be applied in healthcare domain. However, to the best of our knowledge, it has not been investigated for use on in-home monitoring sensor data for early detection of MCI.

This study primarily aims to investigate the utility of ARAM to identify MCI from cognitive healthy based on in-home behaviour patterns such as duration of sleep derived from in-home monitoring sensors. In specific, univariate testing was first carried out to assess potential diagnostic utility of 18-sensor data derived in-home behaviour patterns. Next, we investigate the correlation of different combinations of behaviour patterns with the cognitive status using ARAM to evaluate its feasibility for MCI detection. As there were missing values in the multi-model behaviour patterns, we employed complement coding strategy (Granger et al., 2000) during ARAM modeling on the original 49-subject data set; we also utilized a 25-subject data set with no missing values (Kang, 2013) for ARAM modeling. Consequently, exhaustive search of different hyperparameters including learning rates was conducted on the 49- and 25-subject data sets with different combinations of behaviour to build different ARAM models for identifying MCI from cognitive healthy. To derive concise rules for prediction and interpretation, different pruning strategies were also investigated. Using the model with the highest predictive performance, symbolic IF-THEN rules from ARAM were then extracted to elicit mechanistic insights for MCI prediction from cognitive healthy. The selected ARAM model was subsequently benchmarked against models developed by SVM, Decision Tree, Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to evaluate the advantage of ARAM for MCI detection using in-home sensor monitoring data.

The major contributions of this paper are summarized below. First, to the best of our knowledge, this is the first research investigation on novel machine learning model that may be more suited to tackle small data problem in the field of in-home sensor monitoring for MCI detection using real-world data. Secondly, this work represents amongst the first in-home sensor monitoring application that reports on strategies to address missing data value issue for machine learning modeling. The presence of missing data value complicates the already small data problem faced by machine learning; the missing data value problem is set to grow further in this field when researchers start to integrate different types of emerging in-home sensor data (e.g. wearable data such as daily heart rate) with the conventional passive infrared (PIR) motion sensor data for machine learning. Thirdly, unlike previous machine learning classification model application on in-home sensor detection of MCI, this work, for the first time, provides novel symbolic IF-THEN interpretable rules to elucidate possible factors to differentiate MCI from cognitive healthy.

The rest of this paper is organized as follows: Section II provides a literature review on the related work. Section III then introduces the proposed ARAM model, its missing data handling strategy, rule algorithm, and classical pruning strategy with capability to reduce complexity of fuzzy ARAM rules. Subsequently, Section IV describes the real-world data set and the data processing method. Section V and VI next present the experimental results and discussion, respectively. Finally, Section VII concludes the paper.

2. Related Work

Broadly, in the field of in-home monitoring sensor for MCI detection in a real world setting, different types of prediction models on different derived data features mainly from PIR motion sensor have been investigated. With this, this section provides the overview of related prediction model work as follows: (1) Individual-specific norm approach; (2) Group-norm unsupervised approach; (3) Group-norm supervised approach.

2.1. Individual-specific norm approach

To date, there had been extensive research on investigation of models that can track trajectory changes in functional outcomes of a subject (Akl & Mihailidis, 2015; Akl et al., 2014, 2017). For instance, Akl et al. (2017) pioneered the use of inhomogeneous Poisson regression to characterize in-room activity distribution of a subject in different rooms throughout a day, coupled with the use of Kullback-Leibler (KL)-divergence to quantify the changes of an activity distribution with time (e.g. weeks) and demonstrated feasibility to detect MCI of an individual with time. This type of model basically leverages on deviation of a measure with time from each subject baseline measurement to detect possible correlation with MCI. It does not leverage on group-norm (comparing group norm of MCI vs. cognitive healthy such as the mean walking speed of MCI subjects vs. cognitive healthy) for establishing the model.

2.2. Group-norm unsupervised approach

Broadly, there are two methods that used group-norm method to detect MCI from cognitive healthy using data from in-home monitoring sensors. One method is via unsupervised machine learning technique. Specifically, Akl et al. (2016) introduced a combination of k-means and affinity propagation to create exemplars of room activity distribution (also using inhomogeneous Poisson regression to characterize room activity distribution) to different cognitive status, followed by classifying room activity distributions belonging to a test subject to an exemplar via KL-divergence similarity measure. As each exemplar is associated with a cognitive status, the classification of the test subject to an exemplar will enable determination of the cognitive status of the test subject. Note that this method, thus far, has only leveraged on a simple input data attribute (i.e., in-room activity for one type of room such as either bedroom or living room) to build a classifier model.

2.3. Group-norm supervised approach

Another group-norm approach is via use of supervised machine learning technique that can effectively employ multiple input data attributes to predict MCI from cognitive healthy. For example, Akl et al. (2015) had reported on a study using Support Vector Machine (SVM) on data from 18 MCI and 79 cognitive healthy subjects. Their work experimented on sliding windows of one week to 24 weeks on each subject longitudinal data (which can range from roughly one year to three years) to extract data up to 13 features including walking speed changes from each sliding window, general room activity in the house, age and gender as data input attributes for training of SVM model to detect MCI. In their work, they achieved a receiver operating curve-area under the curve (ROC-AUC) of 0.97, signifying high predictive accuracy potential to detect MCI.

Most recently, with the same in-home sensor data source as Akl et al. (2015), Khan & Jacobs (2021) also investigated on a new data approach to detect MCI. Specifically, Khan & Jacobs (2021) explored on the extraction of up to 40 novel complex movement parameters related to cognitive confusion and forgetfulness as input data attributes for SVM to detect MCI. In contrast, Akl et al. (2015)'s supervised machine learning study mainly focused on use of walking speed data and in-room activity distribution. One key point to highlight for Khan & Jacobs (2021)'s study is that though Khan & Jacobs (2021) had only selected 22 MCI and 22 cognitive healthy subjects, they had used each month of the subject data as one data record for modeling, enabling their group to use 395 MCI and 264 cognitive healthy data points for model training to predict onset of MCI. With this data set, they reported predictive accuracy of approximately 80% to detect MCI. Clearly, these two research studies are focused on different use of movement input data attributes for MCI detection. Investigation of alternative machine learning models that may yield improved predictive performance particularly for a small data set has yet to be carried out.

3. Proposed Machine Learning Model

To differentiate MCI from cognitive healthy using in-home sensor monitoring data especially in a small data set, we propose the use of a self-organizing neural network model called the fuzzy Adaptive Resonance Associative Map (fuzzy ARAM) (Tan, 1995; Tan & Pan, 2005). Figure 1 shows the fuzzy ARAM architecture. At high level, fuzzy ARAM comprises of two overlapping adaptive resonance theory (ART) networks sharing a single category field. With this architecture, it synchronizes the unsupervised categorization of two pattern sets for learning of a supervised mapping between the pattern sets (Tan, 1995). Specifically, an fuzzy ARAM system consists of an input field F_1^a, an output field F_1^b, and a category field F_2. With the input feature vectors at F_1^a coupled with their corresponding class vectors at F_1^b, fuzzy ARAM learns to associate combinations of key input features to their respective classes through the encoding of recognition nodes in category field F_2 (Tan & Pan, 2005). Note that each recognition node j is associated with two adaptive weight templates w_j^a and w_j^b. When a recognition node's weight templates have not encoded any input patterns, it means the recognition node has not learnt any new pattern. In other words, it is in an uncommitted state. For such weight templates, the weight templates are initiated to 1's. During the start of learning, only one such uncommitted recognition node in the F_2 field will be created (Tan & Pan, 2005).

To yield committed recognition nodes which indicates that learning has occurred, fuzzy ARAM requires the setting of the following parameters (Tan, 1995): contribution parameter $\gamma \in [0,1]$, the choice parameters $\alpha_a > 0$ and $\alpha_b > 0$; the vigilance parameters $\rho_a \in [0,1]$ and $\rho_b \in [0,1]$; and the learning rates $\beta_a \in [0,1]$ and $\beta_b \in [0,1]$. In brief, the contribution parameter γ and the choice parameters α_a and α_b are firstly utilized in a minimax learning operation. Specifically, the contribution parameter and the choice parameters are used with a fuzzy minimum operation on weight templates and an input pattern to compute a choice function $\gamma\frac{|x^a \wedge w_j^a|}{\alpha_a+|w_j^a|} + (1-\gamma)\frac{|x^b \wedge w_j^b|}{\alpha_b+|w_j^b|}$ for each F_2 node. This is followed by the code competition and selection process, wherein the F_2 node with the maximum choice function value is identified. Subsequently, during the top-down priming process, the vigilance parameter ρ_a and ρ_b are employed as criteria in vigilance tests $\frac{|x^a \wedge w_j^a|}{|x^a|} \geq \rho_a$ and $\frac{|x^b \wedge w_j^b|}{|x^b|} \geq \rho_b$ to determine whether F_2 node with the highest choice function value or its weight templates are sufficiently close to their respective feature and class vectors. If the vigilance criterion is fulfilled, *resonance* occurs. Else, if the vigilance constraints is violated, a match tracking process will be activated. This mechanism basically adjusts the vigilance criterion, then selects another F_2 node under the revised criterion until a *resonance* is achieved. This search and test process is guaranteed to end as fuzzy ARAM will either find a committed node that satisfies the vigilance criterion or activate an uncommitted node which would definitely satisfy the criterion due to its initial weight values of 1' s. Once *resonance* occurs, template learning ensues. The learning rate β_a and β_b are then leveraged to control the rate of adaptive weight to converge to equilibrium in response to each input pattern with the following formulas: $w_J^{a(new)} = (1\text{-}\beta^a)w_J^{a(old)} + \beta^a(x^a \wedge w_J^{a(old)})$, $w_J^{b(new)} = (1\text{-}\beta^b)w_J^{b(old)} + \beta^b(x^b \wedge w_J^{b(old)})$. With this network architecture, it thus creates a dynamic number of committed F_2 nodes in response to the incoming patterns, creating a learned fuzzy ARAM model. After the model completes learning from all input patterns, it can then be used to predict the class vector based on use of only the input feature vector. For more detail information of the fuzzy ARAM learning and prediction processes, please refer to Tan (1995).

3.1. Missing value handling techniques

In real-world situations, collection of a complete data set is often challenging, resulting in missing input attribute values which necessitates to be handled during machine learning and prediction phases. In fuzzy ARAM, we have at least two approaches to address missing input attribute values (Granger et al., 2000; Li & Parker, 2008), namely the replacement by "0" method and the replacement by "1" approach. In brief, replacement by "0" method is the most simplistic approach as it involves just the replacement of the complement-coded input vector where the i^{th} record is missing with (0,0). On the other hand, replacement by "1" method not only requires the replacement of the complement-coded input vector where the i^{th} record is missing with (1,1), the denominator of vigilance test $\frac{|x^a \wedge w_j^a|}{|x^a|} \geq \rho_a$ needs to be replaced with a fixed value to increase the chances for the vigilance test to pass (Granger et al., 2000).

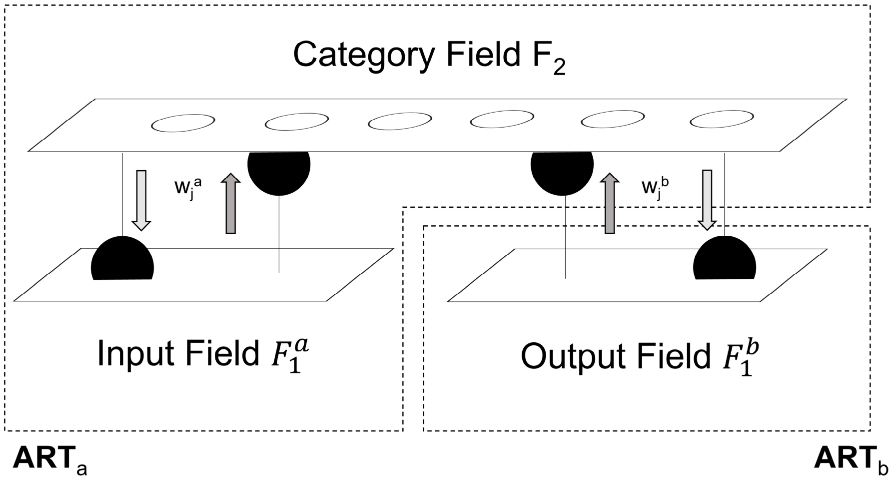

Figure 1: The Adaptive Resonance Associative Map (ARAM) architecture primarily comprises of a category field F_2, an input field F_1^a, and an output field F_1^b which are connected by bidirectional conditional pathways indicated by links with semi-circle heads. The directions of the activity propagation of the pathways are indicated by light and dark grey arrows. Note that the light grey arrows indicate the bottom-up propagation process followed by code competition and selection process, while the dark arrows indicate the top-down priming (resonance or reset) process followed by the template learning process.

Overall, either approach can be employed during fuzzy ARAM learning and prediction phase to handle missing input attribute values.

3.2. Rule extraction approach

As a fuzzy ARAM network learns, each node in the F_2 field encodes a group of input patterns that is associated with an output prediction (Tan & Pan, 2005). In this way, rules can be extracted for interpretable of a fuzzy ARAM network. Specifically, we define that a pair of learned weight vectors for each node represents a set of rule associating antecedents to consequences (Tan & Pan, 2005). Consequently, for any committed F_2 node j, there will be a pair of corresponding weight template vectors w_j^a and w_j^b, which we can derive an IF-THEN rule of the form given below:

$$\mathcal{C} : \mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n \tag{1}$$

where $\mathcal{C}$ is the class indicated by the non-zero attribute value in w_j^b, while $\mathcal{A}_1$, $\mathcal{A}_2$, ...,$\mathcal{A}_n$ are the antecedents or conditions corresponding to the non-zero input attribute values in w_j^a. Using complementary coding, a pair of complement coded weight values (w_{ji}^a, $\bar{w}_{ji}^a$) for a feature i translates into a value range of [w_{ji}^a, 1 - $\bar{w}_{ji}^a$]. For instance, a pair of weight values (0.7, 0.0) for feature i indicates a value of [0.7, 1.0], i.e. the normalized feature value $a_i \geq 0.7$. The obtained value ranges may subsequently be mapped back to the original scale of the expression values for human interpretation. For example, the normalized feature value range of $a_i < 0.47$ may correspond to the heart rate (beats per minute) $x_i < 12$ in absolute terms.

3.3. Pruning strategies

To reduce the complexity of fuzzy ARAM rules which may improve generalizability and eventual model interpretation (Carpenter & Tan, 1993), rule pruning can be employed to select a concise set of rules from trained ARAM networks (Carpenter & Tan, 1995). For the rule pruning algorithm, pruning of the rules are based on the confidence factors associated with each rule. A confidence factor for each F_2 node is derived from its usage frequency in a *training* set and its predictive accuracy on a *predicting* set. Formally, we define the confidence factor CF_j as follows:

$$CF_j = \delta Usage_j + (1 - \delta) Accuracy_j, \tag{2}$$

where $Usage_j$ is the usage of node j, $Accuracy_j$ is its accuracy, and $\delta \in [0,1]$ is a weighting factor. For a small data set similar as this study, we typically set $\delta = 1$, resulting in computation of confidence factors solely based on usage in the training set (Carpenter & Tan, 1995). After confidence factors are determined, F_2 nodes can be generally pruned from the network using *threshold pruning* where F_2 nodes with confidence factors below a given threshold τ are removed from the network. Typically, τ is set to be 0.1 for small data sets. Alternatively, an *iterative rule-pruning* approach can be implemented. This method essentially removes an recognition category with the lowest confidence factor from a learned fuzzy ARAM network during a training and prediction set, provided that the removal of the recognition category does not lead to a deterioration of predictive performance on the same training and prediction set. Either approach will yield a fuzzy ARAM systems with a smaller set of rules, potentially improving generalizability and providing ease for interpretation (Tan & Pan, 2005).

4. Experimentation on Real-World Data

We conducted a proof of concept feasibility study to evaluate whether the in-home behaviour pattern differences between a group of cognitive healthy and a group of MCI subjects can be used to used to construct a fuzzy ARAM model for detection of MCI. Formally, we define that given the in-home sensor data and cognitive status of a group of cognitive healthy and a group of MCI subjects at a cross-sectional time point, we seek to differentiate MCI subjects from cognitive healthy subjects using fuzzy ARAM. We also aim to derive symbolic IF-THEN rules from fuzzy ARAM to interpret the final MCI detection model.

4.1. Data source

We employed data from a cross-sectional study conducted over a period of two months. The study commenced in March 2016 and was completed in August 2018. Institutional ethics review board approvals were obtained (reference number: 2015/ 01076) (Rawtaer et al., 2020). Informed consent was obtained from participants before screening them for eligibility. 49 subjects were eligible and each individual subject's home was instrumented with an in-home sensor network system for a duration of two months (Rawtaer et al., 2020). In brief, the in-home sensor network system consists of three PIR motion sensors located in living room, kitchen and bedroom, two magnetic contact sensors placed on main house door and medicine box, a pressure sensor placed on bed, a smart-plug for monitoring TV usage, Bluetooth beacon sensors tagged on key and wallet, and a wearable hand-wrist sensor that can yield heart rate (beats per minute) and daily steps. As each type of sensors operates with a different mechanism, data from each sensor type are recorded at different frequencies. For instance, the motion sensors and contact sensors only recorded data whenever they are triggered. On the other hand, the pressure sensors, the smart-plug, Bluetooth beacon sensors and wearable hand-wrist sensor captured data at every 5 minutes, 10 minutes, 4 minutes and 60 minutes intervals, respectively (Chen et al., 2019). The sensed data from the different sensors were subsequently wirelessly transmitted to the gateway. The gateway then aggregated and transmitted the data to the backend server via secure cellular communications (e.g. 3G) for monitoring and processing. Each data point was identifiable only via the sensor node identifier; the mapping between the sensor node identifier and the home was securely stored and accessible only by the study investigators (Rawtaer et al., 2020).

The participants' cognitive status of MCI or cognitive healthy was established at baseline with the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), Clinical Dementia Rating (CDR) and detailed neuropsychological test batteries and via a consensus panel (Rawtaer et al., 2020). Overall, there were 21 MCI and 28 cognitive healthy participants. This data field is also available to this study for correlation to the sensor data for data analysis and machine learning modeling.

4.2. In-home behaviors captured and data preprocessing

Different combinations of sensor data were used to specifically measure nine types of in-home behaviors, namely forgetfulness of medicine taking, forgetfulness of bringing key when traveling out, forgetfulness of bringing wallet when traveling out, number of outdoor activities, duration of time away from home, duration of television use, duration of sleep, number of steps each day, and heart beat per minute (Rawtaer et al., 2020). For instance, participants were provided with a sensor-equipped medication box to store all their prescription medication; data were generated whenever the box was opened. These data, taken together with the expected medication frequency information obtained at baseline, allowed us to determine the number of times a participant forgot to take their medication at the prescribed time. For more details, please refer to Rawtaer et al. (2020) for an elaborated description of the specific combinations of sensors used to measure the different behaviors. Overall, with the various types of raw sensor data readings, these were first converted into a common format and aggregated into a database. Purging of the data was performed to remove erroneous data and periods where the system was down/partially down. After data purging, sensor-specific data cleaning or validation was performed to ensure that only valid sensor data are processed. After the process of cleaning data, metrics such as mean and standard deviation of frequency of forgetting medication per month per user were computed.

4.3. Preliminary data exploration

4.3.1. Data completeness

With the cleaned data from each participant (subject), we preliminary explored the data set. As data completeness is pivotal towards construct of an accurate machine learning model, we first examine the amount of missing data on the nine in-home behavior metrics. Table 1 shows the statistics for missing value from all subjects for each behavior metric. Clearly, three behavior metrics utilizing only unobtrusive monitoring sensors such as PIR sensors and smart-plug experienced no missing values, while the six other behavior metrics mostly involving the use of wearable sensors suffered from missing values ranging from 4.3% (2/49) to 32.4% (12/49) of total number of subjects. For more detail information on the missing values for each of the nine in-home behavior metrics in relation to the different subjects, please refer to Supplementary Table 1. With these results, we further summarized the missing values statistics by per subject basis (Table 2). We first define subject with missing values as any subject with at least one missing behavior metric value. Using this definition, 49.0% (24/49) of subjects suffered from missing values. However, by analyzing the average number of behavior metric with missing values per subject, approximately 1.0 (46/49) behavior metric is missing per subject. These results indicate that even though many subjects suffered from at least one missing values issue, each subject on average only experienced about one missing in-home behavior metric value.

4.3.2. Diagnostic utility of the in-home behavior metrics

With the different number of non-missing values from each of the in-home behavior metrics, we next explored the diagnostic utility of these various behavior metrics for differentiation of MCI from cognitive healthy. Specifically, we employed the unpaired Student t-test to determine the predictive utility of each behavior metric collected over a two-month period. This univariate test was selected as it excels in testing the differences of population means belonging to two sample groups with continuous variables. This statistical test was, in fact, used in one of our prior studies (Rawtaer et al., 2020), and also employed in another related work by a different group Wu et al. (2021). Figure 2 and Figure 3 summarize the comparison of nine in-home behaviour metrics for MCI subjects against cognitive healthy when using two months-mean and two-months-standard deviation measures, respectively. Overall, there is no significant statistical differences in any of the nine in-home behaviour metrics when using the mean measure. For the standard deviation measure, only sleep duration was found to be statistically

Table 1: Statistics of subjects with missing values per behavior metric.

Behavior metrics	**Wearable sensors***	**Subjects with missing values**	
		Number	Percentage
Television use daily (min)	No	0	0.0%
Time away from home daily (min)	No	0	0.0%
No. of outings daily	No	0	0.0%
Frequency of forgetting keys/month	Yes	2	4.3%
Frequency of forgetting wallet/month	Yes	2	4.3%
Sleep duration daily (min)	No	10	25.6%
Heart rate (bpm)	Yes	10	25.6%
Steps (daily)	Yes	12	32.4%
Frequency of forgetting medicine/month	Yes	0	0.0%

*Note: Wearable sensors include Bluetooth beacon sensors and hand-wrist sensor.

Table 2: Overall missing values summary.

Data summary parameters	**Statistical values**
Total unique subjects	49
Number of unique subjects with missing values	24
Overall % subjects with $\geq$ 1 missing value	49.0%
Average $\pm$ standard deviation (SD) number of missing behavior metric value per subject	1$\pm$1

significant (p<0.05). Clearly, this preliminary univariate data analysis demonstrated that the standard deviation measure of sleep duration contains statistically significant predictive utility for MCI detection, while each of the 17 other behaviour metrics contains subtle individual diagnostic information to detect MCI, suggesting that it may be necessary to combine behavior metrics to provide notable predictive utility for MCI detection.

4.4. Fuzzy ARAM modeling using data set with missing values

From our preliminary data exploration result, we observed that each subjects on average only suffered about one missing behavior metric value out of a total of nine behavior metrics. As fuzzy ARAM possesses the capability to handle missing values for learning and prediction (Granger et al., 2000), we first investigated the use of a full 49-subject data set for ARAM modeling to detect MCI from cognitive healthy. A min-max normalization operation was first carried out to scale non-missing values from each of the data input attribute to be bounded between 0 and 1 (Tan & Pan, 2005). For each of the data attributes, we then created a corresponding vector by subtracting min-max normalized non-missing values from 1. Hence, each of the data attributes was associated with two values, which is known as complement coding (Tan, 1995). For data input attribute with missing value, complement coding of (1,1) was

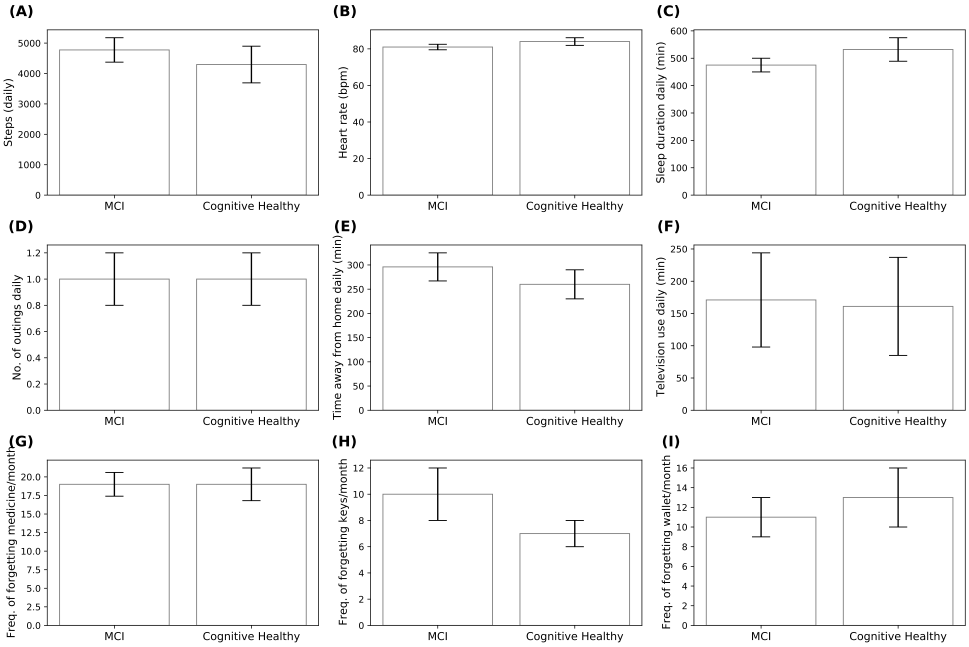

Figure 2: Comparison of mean values of nine in-home behavior metrics measured over two months between the cognitively healthy group and MCI. Note that the error bar represents standard error (standard deviation normalized by square root of sample size).

utilized to replace any missing value from any specific data input attribute so that we can employ the full 49-subject data set (Granger et al., 2000). To evaluate the performance on this small sample data set, the leave-one-out, cross-validation method was employed. In brief, in the first round of leave-one-out, cross-validation, one data point (which represented one subject in this study) was left out, while the remaining were used for model training. For example, when using the 49-subject data set, one data point from one subject was withheld, while the 48 subjects were used for modeling building. The built model was then used to classify the withheld data point from the one subject. This process was repeated until all withheld data were classified. In other words, this process was repeated 49 times. In each run, the training patterns are presented in a random order. Predictive performance metrics were then computed based on results from all withheld data points. As this study may potentially suffer from imbalanced class data set problem (Saito & Rehmsmeier, 2015), F1-score was employed as the main performance metric. Accuracy was, nevertheless, also reported in this report as a secondary performance metric for evaluation.

For all fuzzy ARAM learning experiments, we fixed choice parameter α_b, learning rate β_b, baseline vigilance ρ_b, contribution parameter γ, and number of voting ARAM

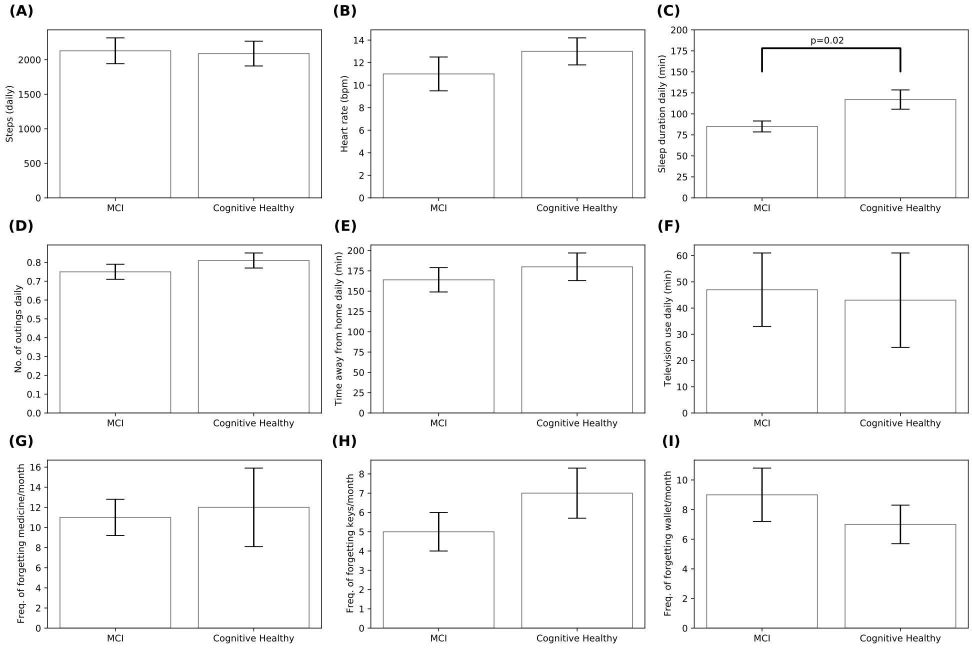

Figure 3: Comparison of standard deviation values of nine in-home behavior metrics measured over two months between the cognitively healthy group and MCI. Note that the error bar represents standard error, and p-value is shown for any pairwise comparison (i.e., MCI and cognitively healthy group) which is statistically significant at 5% level.

models to be 0.001, 1, 1, 1, and 5, respectively. For maximal generalization, match-tracking is used for all experiments. We then varied different combinations of choice parameter α_a between 0.001 to 0.1, learning rate β_a between 0.1 to 1.0, baseline vigilance ρ_a between 0.1 to 1.0, and with different combination of pruning strategies including no pruning, threshold pruning at 0.1, and the iterative-rule pruning. Note that pruning approach in this small data set study is solely based on computation of usage in the training set (Tan & Pan, 2005). This variation of different combinations of hyperparameters and different pruning strategies was repeated for each set of data input attributes. As our preliminary data exploration work suggested that there were limited diagnostic utility in 17 out of 18 behavior metrics for MCI detection when used individually, the combining of multiple behavior metrics into one set of data input attributes for fuzzy ARAM modeling would be necessary. Thus, we grouped nine in-home behavior metrics related to two months-mean measure, all nine in-home behavior metrics associated with two-months-standard deviation measure, and all 18 in-home behavior metrics that are related to both two months-mean and two-months-standard deviation measures into three combinations of data input attributes for fuzzy ARAM modeling. The variation of different combinations of hyperparameters and different pruning strategies were thus repeated for each of these three combinations of data input attributes.

Table 3 provides the performance of fuzzy ARAM models using different pruning methods for mean, standard deviation, and mean+standard deviation measures of selected in-home behaviours measured over two-months from 49 subjects. For the mean measure, high vigilance ($\rho_a = 0.9$), slow learning ($\beta_a = 0.1$), and iterative rule-based pruning strategy were found to be pivotal to build the classifier with the highest F1-score of 56.5%. Choice parameter (α_a), however, seemingly did not play a key role. In contrast, choice parameter of $\alpha_a = 0.1$, low vigilance ($\rho_a = 0.0$) and fast learning ($\beta_a = 1.0$) were discovered to be instrumental to build the classifier with the highest F1-score (58.3%) using the standard deviation measure. Note that iterative rule-based pruning strategy was found again to provide the highest predictive utility amongst the different pruning methods when using the standard deviation input attribute measure. For the mean+standard measure, choice parameter of $\alpha_a = 0.1$, low vigilance ($\rho_a = 0.0$), and a range of learning rate starting from slow to fast learning ($\beta_a = 0.5$ or 1.0) were found to be important for building a classifier with the highest F1-rate of 53.7%. Clearly, the different pruning strategies apparently does not affect the predictive performance. Note that though leave-one-out,cross-validation is an unbiased performance error estimator, it is well-known to suffer from significant variability (Efron, 1983). To estimate this variability is non-trivial (Bengio & Grandvalet, 2004). As a proxy, we employed the bootstrapping sampling method to derive an approximate measure of variance for our cross-validation approach (Efron, 1983). Our result shows that the standard deviations of our performance error estimations of F1-score and accuracy rate are in the range of 4-7%. Overall, these predictive performance results demonstrate that fuzzy ARAM can be applied on data set with missing values to build a model with reasonable predictive utility for MCI detection.

Table 3: Performance of fuzzy ARAM models using different pruning methods for mean, standard deviation, and mean+standard deviation measures of selected in-home behaviours measured over two-months from 49 subjects.

				No Pruning		Threshold Pruning		Rule-Pruning	
Measures	α_a	ρ_a	β_a	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**
Mean	0.001	0.9	0.1	52.2% (48.0%)	55.1% (57.1%/53.6%)	52.2% (48.0%)	55.1% (57.1%/53.6%)	56.5% * (52.0%)	59.2% (61.9%/57.1%)
	0.01	0.9	0.1	52.2% (48.0%)	55.1% (57.1%/53.6%)	52.2% (48.0%)	55.1% (57.1%/53.6%)	56.5% * (52.0%)	59.2% (61.9%/57.1%)
	0.1	0.9	0.1	52.2% (48.0%)	55.1% (57.1%/53.6%)	52.2% (48.0%)	55.1% (57.1%/53.6%)	56.5% * (52.0%)	59.2% (61.9%/57.1%)
Std	0.1	0.0	1.0	55.3% (50.0%)	57.1% (61.9%/53.6%)	55.3% (50.0%)	57.1% (61.9%/53.6%)	58.3% * (51.9%)	59.2% (66.7%/53.6%)
Mean+Std	0.1	0.0	1.0	53.7% * (55.0%)	61.2% (52.4%/67.9%)	53.7% * (55.0%)	61.2% (52.4%/67.9%)	50.0% (52.4%)	59.2% (47.6%/67.9%)
	0.1	0.0	0.5	53.7% * (55.0%)	61.2% (52.4%/67.9%)	53.7% * (55.0%)	61.2% (52.4%/67.9%)	53.7% * (55.0%)	61.2% (52.4%/67.9%)

* denotes the fuzzy ARAM model with the highest F1-score performance respectively for the mean, standard deviation, and mean+standard deviation measures of selected in-home behaviors. Note that recall is not provided as sensitivity is the equivalent of recall.

From the F1-score and accuracy rate results, there is approximately 40.0% - 50.0% prediction error rate. One plausible reason for the higher error rate could be because the patterns with complement coding of (1,1) for missing values (from 24 subject data) are mostly misclassified. We, thus, investigated on the relationship between missing values and prediction errors. From the fuzzy ARAM modeling with different combinations of parameters and different sets of behavior metrics, we randomly selected 38 sets of prediction results with 49 prediction labels in each set of results. Retrospectively, we linked each of the 1,862 predictions labels with whether it was misclassified and whether it was associated with missing values. A contingency table was used to show the relationship between predictions labels which contain missing values and prediction errors (Table 4). We observed that 46.0% (431/937) of prediction labels with missing values were incorrectly predicted, whilst 54.4% (504/925) of prediction labels without missing values were incorrectly predicted. Pearson's Chi-square confirmed that prediction labels with missing values was associated with lower error rate than those without missing values at 5% significance level. Overall, this experiment indicates that the compromised predictive performance of fuzzy ARAM due to the use of complement coding of (1,1) is most likely caused by reasons not directly related to increased misclassification from patterns with complement coding of (1,1) for missing values.

Table 4: Relationship between missing values and prediction error.

Prediction labels with missing input attribute	**Incorrect prediction**	**Correct prediction**	**Total**	**p-value**
Yes	431	506	937	0.0002
No	504	421	925	
Total	935	927	1862	

4.5. Fuzzy ARAM modeling using data set with no missing value

To further assess the efficacy of fuzzy ARAM to handle small data, we next investigated the modeling performance of fuzzy ARAM to detect MCI from cognitive healthy using the data set with no missing values. The final data set sample size with no missing values used in this experiment was thus 25, with 10 subjects' data belonging to MCI and 15 subjects' belonging to cognitive healthy. The min-max normalization operation was again performed to scale non-missing values from each of the data input attribute to be bounded between 0 and 1 (Tan & Pan, 2005). With the normalized 25-subject data set, it was then subjected to the same experimental conditions as in the 49-subject data set with fixing and varying the same fuzzy ARAM parameters. Table 5 provides the performance of fuzzy ARAM models using different pruning methods for mean, standard deviation, and mean+standard deviation measures of selected in-home behaviours measured over two-months from 25 subjects. For the mean measure, choice parameter of $\alpha_a = 0.1$, low vigilance ($\rho_a = 0.0$) and fast learning ($\beta_a = 1.0$) were found to be critical towards building the classifier with the highest F1-score of 54.5%. The different pruning strategies seemingly did not affect the predictive performance. For the standard deviation measure, the different pruning strategies were found again not to affect the predictive performance. A choice function of $\alpha_a = 0.01$, low vigilance ($\rho_a = 0.5$) and slow learning rate ($\beta_a = 0.5$) were elucidated to be important for the highest F1-score of 63.6%. For the mean+standard deviation measure, a choice function of $\alpha_a = 0.01$, low vigilance ($\rho_a = 0.0$), fast learning ($\beta_a = 1.0$) and iterative rule-based pruning strategy were revealed to be important to build the classifier with the highest F1-rate of 63.6%. Clearly, application of fuzzy ARAM on the 25-subject data set yielded an improved predictive performance as compared to the 49-subject data set (highest F1-rate for 25-subject data set: 63.6%; highest F1-rate for 25-subject data set: 58.3%). Note that the standard deviation index of our performance error estimations was estimated to be approximately between 5-11% after we had employed the bootstrap sampling technique as a proxy. Overall, this demonstrates that fuzzy ARAM can be employed on data set with no missing values to build a model with improved predictive utility (as compared to models used on data with missing data) for MCI detection.

Table 5: Performance of fuzzy ARAM models using different pruning methods for mean, standard deviation, and mean+standard deviation measures of selected in-home behaviours over measured two-months from 25 subjects.

				No Pruning		Threshold Pruning		Rule-Pruning	
Measures	α_a	ρ_a	β_a	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**	**F1-score (Precision)**	**Accuracy (Sensitivity/Specificity)**
Mean	0.1	0.0	1.0	54.5% * (50.0%)	60.0% (60.0%/60.0%)	54.5% * (50.0%)	60.0% (60.0%/60.0%)	54.5% * (50.0%)	60.0% (60.0%/60.0%)
Std	0.01	0.5	0.5	63.6% * (58.3%)	68.0% (70.0%/66.7%)	63.6% * (58.3%)	68.0% (70.0%/66.7%)	63.6% * (58.3%)	68.0% (70.0%/66.7%)
Mean+Std	0.01	0.0	1.0	60.9% (53.8%)	64.0% (70.0%/60.0%)	60.9% (53.8%)	64.0% (70.0%/60.0%)	63.6% * (58.3%)	68.0% (70.0%/66.7%)

* denotes the fuzzy ARAM model with the highest F1-score performance respectively for the mean, standard deviation, and mean+standard deviation measures of selected in-home behaviors. Note that recall is not provided as sensitivity is the equivalent of recall.

4.6. Fuzzy ARAM modeling with antecedent pruning

From fuzzy ARAM modeling results using 49- and 25-subject data sets in Table 3 and 5, it is clear that iterative rule-pruning approach has often been found to be pivotal towards building fuzzy ARAM models with high F1-score performance. This suggests that rule-pruning strategy can improve generalizability. Besides rule pruning to reduce model complexity, antecedent pruning can also be used to derive concise rules for prediction and interpretation. Essentially, this techniques relies on calculation of the error factor for each antecedent in each rule based on its performance on the training and predicting sets. When a rule makes a predictive error, each antecedent of the rule that also appears in the current input has its error factor increased in proportion to the smaller of its magnitudes in the rule and in the input vector. After the error factor for each antecedent is determined, an iterative pruning strategy, similar to the one for rules, removes redundant antecedents by setting the corresponding weight to zero (Carpenter & Tan, 1995). This technique will reduce the number of antecedents employed in the prediction model, thereby only presenting the antecedents with the most predictive utility for interpretation. In this work, we also experimented fuzzy ARAM modeling with antecedent pruning. Table 6 presents the hyperparameters and the predictive performance of fuzzy ARAM models with the highest F1-score within each type of input attribute measure (i.e., mean, standard deviation, and mean+standard deviation) for both 49- and 25-subject data sets. For the 49-subject data set, the standard deviation measure with choice parameter of $\alpha_a = 0.1$, high vigilance ($\rho_a = 0.9$) and slow learning ($\beta_a = 0.5$) were found to be important towards building a classifier with the highest F1-score predictive performance of 60.0%. For the 25-subject data set, choice parameter of $\alpha_a = 0.1$ and slow learning ($\beta_a = 0.5$) when applied on the standard deviation measure were found again to be critical towards building a classifier with the highest F1-score of 76.2%. Clearly, the 25-subject data set with standard deviation measure using choice parameter of $\alpha_a = 0.1$, low vigilance ($\rho_a = 0.5$) and slow learning ($\beta_a = 0.5$) were found to provide the highest predictive F1-score amongst all modeling experiments. To validate whether the fuzzy ARAM model had converged to a stabilized value, we plotted the training and testing loss curves with respect to the number of epochs. Figure 4 shows the convergence of the loss curves starting from approximately the third epoch, indicating a successful learning process. Collectively, these results demonstrate the utility of fuzzy ARAM with antecedent pruning to differentiate MCI from cognitive healthy with high predictive performance.

Table 6: Performance of fuzzy ARAM models using antecedent pruning methods for mean, standard deviation, and mean+standard deviation measures of selected in-home behaviours over two-months from 49- and 25- subjects.

Subject sample size	**Measure**	α_a	ρ_a	β_a	**F1-score**	**Precision**	**Accuracy**	**Sensitivity**	**Specificity**
49	Mean	0.01	0.0	0.1	52.2%	48.0%	55.1%	57.1%	53.6%
	Std	0.1	0.9	0.5	60.0% *	63.2%	67.3%	57.1%	75.0%
	Mean+Std	0.001	0.5	0.5	56.4%	61.1%	65.3%	52.4%	75.0%
25	Mean	0.01	0.9	0.1	57.1%	54.5%	64.0%	60.0%	66.7%
	Std	0.1	0.5	0.5	76.2% *	72.7%	80.0%	80.0%	80.0%
	Mean+Std	0.01	0.0	0.5	63.6%	58.3%	68.0%	70.0%	66.7%

* denotes the highest F1-score performance from each sample size.Note that recall is not provided as sensitivity is the equivalent of recall.

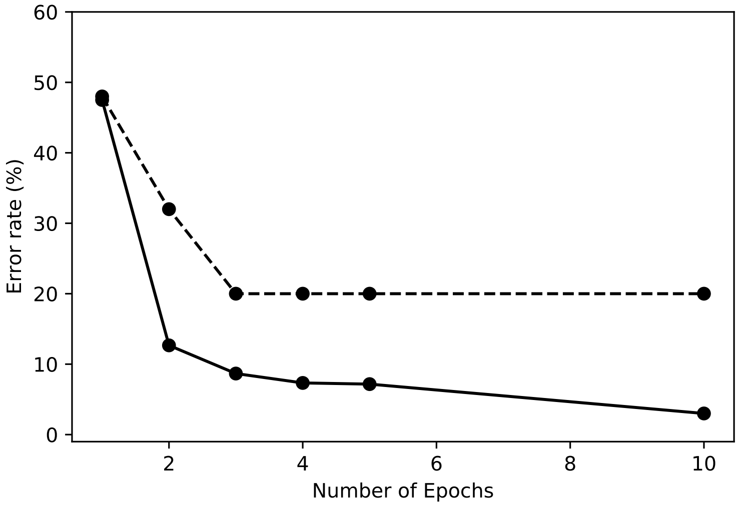

Figure 4: Loss curve: the image shows loss curves for the training (solid line) and testing (dotted line) data. Convergence for both curves started from approximately from the third epoch, indicating a successful learning process.

4.7. Fuzzy ARAM derived prediction rules

As fuzzy ARAM possesses the capability to provide interpretable information (Tan & Pan, 2005), we employed this specific model to extract symbolic rules for interpretation. In this experiment, we extracted the symbolic rules using iterative-rule pruning method from the 25-subject data set with standard deviation measure for insight comparison between the two different pruning strategies. In specific, we employed the model with the lower average number of committed F_2 nodes using the iterative-rule pruning method as shown in Table 5 (i.e., of $\alpha_a = 0.001$, low vigilance ($\rho_a = 0.5$), and slow learning ($\beta_a = 0.5$)). Note that iterative rule-pruning method was used as benchmark as it was found most frequently to be critical towards building fuzzy ARAM models with high F1-score performance. Table 7 first presents the complete list of fuzzy ARAM rules to detect MCI and cognitive healthy using iterative-rule pruning approach. Clearly, three sets of rules based on nine antecedents with confidence level ranging from 0.17 to 1.00 were derived to identify MCI, while another three sets of nine rules based on nine antecedents with confidence level ranging from 0.14 to 1.00 were constructed to identify cognitive healthy. Overall, 54 antecedents (6 rules x 9 attributes) were employed from fuzzy ARAM with iterative-rule pruning approach to identify MCI from cognitive healthy. We next investigated the fuzzy ARAM rules using antecedent pruning method. Table 8 shows a summary of the complete list of fuzzy ARAM rules to detect MCI and cognitive healthy using antecedent pruning approach. Similar as the iterative-rule pruning approach, three rules with confidence level ranging from 0.17 to 1.00 were derived to identify MCI, while another three sets with confidence level ranging from 0.14 to 1.00 were constructed to identify cognitive healthy.

Upon examining the value ranges in each rule, we observed that value ranges from the antecedent pruning approach are also similar as those from the iterative-rule pruning approach. For instance, for the rule to detect MCI with confidence level of 1.00, the value ranges for the antecedent **Variation** (Heart rate (bpm)) using the iterative-rule pruning method was ≥ 4 and < 12, while those using the antecedent-pruning technique was < 12. Thus, it appears that the antecedent-pruning technique retains as much as possible the value ranges of the antecedent derived from the iterative-rule pruning approach. We next examined the number of antecedents in each rule. We confirmed that the number of antecedent used has been considerably reduced from 54 to 21, with 12 antecedents utilized to detect MCI and 9 antecedents found to be useful to detect cognitive healthy. Clearly, symbolic rules derived using antecedent-pruning presents a concise set of antecedents for prediction and interpretation.

Based on the above analysis, consequently, Table 8 based on fuzzy ARAM with antecedent pruning method was utilized for further interpretation of the symbolic IF-THEN rules to differentiate MCI from cognitve healthy. From the three derived rules for MCI, we observed that MCI is consistently associated with behavior metric of forgetfulness such as frequency of forgetting wallet and medicine. In addition, we also found that behavior metric related to physiologic parameters which consists of steps and heart rate were also consistently found in all three derived rules for MCI. Though these behavior metrics (or antecedents) related to forgetfulness and physiologic were found in all three derived rules for MCI, these metrics were also found in the three rules for recognizing cognitive healthy. This suggests that these behavior metrics related to forgetfulness and physiologic were sensitive but not specific to recognize MCI. We further examined any behavior metric that is unique to identify MCI. We confirmed that the sleep duration behavior metric is one specific marker that can be used to recognize MCI, but not for cognitive healthy. On the other hand, we found that the behavior metric related to time away from home daily was specific to recognition of cognitively healthy. Collectively, these rules derived from fuzzy ARAM model with antecedent pruning method provide novel insights on possible sensitive and specific casual factors that can be used to detect MCI from cognitive healthy.

4.8. Performance benchmark

In addition to evaluation of the capability of fuzzy ARAM to detect MCI from cognitive healthy using in-home sensor monitoring data, we carried out a benchmark experiment to confirm the advantage of fuzzy ARAM when confirmed to other classical or well-known models used in this field. Since SVM has been routinely applied by Akl et al. (2015) and Khan & Jacobs (2021) for detection of MCI from cognitive healthy using in-home sensors, SVM is selected as one benchmark algorithm (Cortes & Vapnik, 1995). As fuzzy ARAM belongs to a type of "white-box" model, we also select a classical "white-box" model for comparison with fuzzy ARAM, which is Decision Tree (Breiman et al., 1984). In totality, we selected SVM and Decision Tree to benchmark against fuzzy ARAM. We first compared the three models in terms of their predictive performance. Mean Average Precision (mAP) and Receiver Operating Characteristics-Area Under the Curve (ROC-AUC) were chosen as the main evaluation metrics (He & Ma, 2013). In specific, the fuzzy ARAM model using 25-subject

Table 7: Iterative pruning-based fuzzy ARAM rules to detect MCI and cognitive healthy.

Predict	**MCI (Confidence 1.00, usage = 1.00**)
IF	**Variation** (Time away from home daily (min)) $\geq 81, < 224$
AND	**Variation** (Heart rate (bpm))$\geq 4, < 12$
AND	**Variation** (Steps (daily)) $\geq 892, < 2369$
AND	**Variation** (No. of outings daily) $\geq 1, < 18$
AND	**Variation** (Freq. of forgetting keys/month) < 17
AND	**Variation** (Freq. of forgetting wallet/month) $\geq 3, < 15$
AND	**Variation** (Sleep duration daily (min)) $\geq 52, < 98$
AND	**Variation** (Freq. of forgetting medicine/month) $\geq 1, < 12$
AND	**Variation** (Television use daily (min)) < 13
Predict	**MCI (Confidence 0.33, usage = 0.33)**
IF	**Variation** (Time away from home daily (min)) $\geq 258, < 326$
AND	**Variation** (Heart rate (bpm)) $= 6$
AND	**Variation** (Steps (daily)) $\geq 1663, < 2570$
AND	**Variation** (No. of outings daily) $\geq 6, < 11$
AND	**Variation** (Freq. of forgetting keys/month) $\geq 5, < 12$
AND	**Variation** (Freq. of forgetting wallet/month) $\geq 5, < 9$
AND	**Variation** (Sleep duration daily (min)) $\geq 140, < 163$
AND	**Variation** (Freq. of forgetting medicine/month) $\geq 3, < 7$
AND	**Variation** (Television use daily (min)) $\geq 17, < 25$
Predict	**MCI (Confidence 0.17, usage = 0.17)**
IF	**Variation** (Time away from home daily (min)) $= 77$
AND	**Variation** (Heart rate (bpm)) $= 11$
AND	**Variation** (Steps (daily)) $= 355$
AND	**Variation** (No. of outings daily) $= 3$
AND	**Variation** (Freq. of forgetting keys/month) $= 1$
AND	**Variation** (Freq. of forgetting wallet/month) $= 3$
AND	**Variation** (Sleep duration daily (min)) $= 61$
AND	**Variation** (Freq. of forgetting medicine/month) $= 27$
AND	**Variation** (Television use daily (min)) $= 28$
Predict	**Cognitive healthy (Confidence 1.00, usage = 1.00)**
IF	**Variation** (Time away from home daily (min)) $\geq 81, < 166$
AND	**Variation** (Heart rate (bpm))$\geq 9, < 19$
AND	**Variation** (Steps (daily)) $\geq 1362, < 3610$
AND	**Variation** (No. of outings daily) $\geq 5, < 13$
AND	**Variation** (Freq. of forgetting keys/month) $\geq 1, < 2$
AND	**Variation** (Freq. of forgetting wallet/month) < 10
AND	**Variation** (Sleep duration daily (min)) $\geq 73, < 178$
AND	**Variation** (Freq. of forgetting medicine/month) $\geq 1, < 55$
AND	**Variation** (Television use daily (min)) < 17
Predict	**Cognitive healthy (Confidence 0.71, usage = 0.71)**
IF	**Variation** (Time away from home daily (min)) $\geq 190, < 248$
AND	**Variation** (Heart rate (bpm))$\geq 5, < 13$
AND	**Variation** (Steps (daily)) $\geq 1798, < 3040$
AND	**Variation** (No. of outings daily) $\geq 9, < 22$
AND	**Variation** (Freq. of forgetting keys/month) $\geq 1, < 14$
AND	**Variation** (Freq. of forgetting wallet/month) $\geq 2, < 16$
AND	**Variation** (Sleep duration daily (min)) $\geq 84, < 115$
AND	**Variation** (Freq. of forgetting medicine/month) $\geq 1, < 23$
AND	**Variation** (Television use daily (min)) $\geq 9, < 254$
Predict	**Cognitive healthy (Confidence 0.14, usage = 0.14)**
IF	**Variation** (Time away from home daily (min)) $= 418$
AND	**Variation** (Heart rate (bpm))$\geq 5, = 10$
AND	**Variation** (Steps (daily)) $\geq 1798, = 1026$
AND	**Variation** (No. of outings daily) $\geq 9, = 5$
AND	**Variation** (Freq. of forgetting keys/month) $= 1$
AND	**Variation** (Freq. of forgetting wallet/month) $= 2$
AND	**Variation** (Sleep duration daily (min)) $= 201$
AND	**Variation** (Freq. of forgetting medicine/month) $= 5$
AND	**Variation** (Television use daily (min)) $= 384$

Table 8: Antecedent pruning-based fuzzy ARAM rules to detect MCI and cognitive healthy.

Predict	**MCI (Confidence 1.00, usage = 1.00)**
IF	**Variation** (Heart rate (bpm)) < 12
AND	**Variation** (Steps (daily)) < 2369
AND	**Variation** (No. of outings daily) < 18
AND	**Variation** (Freq. of forgetting wallet/month) < 15
AND	**Variation** (Sleep duration daily (min)) < 98
AND	**Variation** (Freq. of forgetting medicine/month) < 12
AND	**Variation** (Television use daily (min)) < 13
Predict	**MCI (Confidence 0.33, usage = 0.33)**
IF	**Variation** (Heart rate (bpm)) < 6
AND	**Variation** (Freq. of forgetting wallet/month) < 9
Predict	**MCI (Confidence 0.17, usage = 0.17)**
IF	**Variation** (Steps (daily)) $= 355$
AND	**Variation** (Freq. of forgetting medicine/month) < 27
AND	**Variation** (Television use daily (min)) < 28
Predict	**Cognitive healthy (Confidence 1.00, usage = 1.00)**
IF	**Variation** (No. of outings daily) < 13
AND	**Variation** (Freq. of forgetting keys/month) < 2
AND	**Variation** (Freq. of forgetting wallet/month) < 10
AND	**Variation** (Television use daily (min)) < 17
Predict	**Cognitive healthy (Confidence 0.71, usage = 0.71)**
IF	**Variation** (Time away from home daily (min)) ≥ 190, < 248
AND	**Variation** (Heart rate (bpm)) < 13
AND	**Variation** (Steps (daily)) ≥ 1798
AND	**Variation** (Television use daily (min)) ≥ 9
Predict	**Cognitive healthy (Confidence 0.14, usage = 0.14)**
IF	**Variation** (Freq. of forgetting medicine/month) < 5

data set with antecedent pruning method built using nine behavior metrics related to standard deviation measure ($\alpha_a = 0.1$, low vigilance ($\rho_a = 0.5$), and slow learning ($\beta_a = 0.5$)) was selected throughout this experiment for comparison with SVM and Decision Tree. We then utilized the same 25-subject data set containing nine behavior metrics related to standard deviation measure for construct of multiple SVM models. Leave-one-out, cross-validation was used to evaluate the predictive performance of the different models, out of which we selected the SVM model with the highest F1-score for comparison with fuzzy ARAM. In brief, we conducted a grid-search of hyperparameters of C ranging [0.001, 1000], γ ranging from [0.001, 1000], and amongst different kernels including linear and Radial-Basis Function (RBF). The SVM model with γ of 0.01, C value of 100 and with RBF kernel was eventually selected as the final model for comparison. For Decision Tree, we utilized the same approach as SVM to select a model eventually for comparison with fuzzy ARAM and SVM. Briefly, we conducted a grid-search of hyperparameters of maximum tree depth ranging [3, 20], minimum number of data points required in a leaf node ranging from [1, 10], and amongst different node splitting criterion including Gini and entropy. The Decision Tree with a maximum tree depth of 5, minimum number of data points required in a leaf node of 10 and with Gini node splitting criterion was eventually selected as the final model for comparison. As Decision Tree can provide interpretable information, we also compare fuzzy ARAM against Decision Tree on the types of interpretable information that can be yielded.

Figure 5 shows the mAP and ROC-AUC for the three selected classifiers each belonging to fuzzy ARAM, SVM and Decision Tree, respectively. In specific, mAPs of 0.70, 0.68, and 0.70 (Figure 5 (A)), and ROC-AUC of 0.78, 0.71, and 0.53 (Figure 5 (B)), respectively were achieved by fuzzy ARAM, SVM and Decision Tree. Clearly, fuzzy ARAM achieved the highest predictive performance amongst the three models regardless of mAP or ROC-AUC metric. To confirm fuzzy ARAM achieved the highest predictive performance amongst the three models, we next evaluated fuzzy ARAM against SVM and Decision Tree on F1-scores. With F1-scores of 76.2%, 73.7% and 63.6% from fuzzy ARAM, SVM and Decision Tree, respectively, we confirm that fuzzy ARAM can achieve improved predictive performance as compared to classical models in the area of in-home sensor monitoring for MCI detection field using small data. Using the Decision Tree with the optimal predictive performance, we also extracted interpretable rules from the model for comparison against fuzzy ARAM. The Decision Tree model revealed that if a subject's **Variation** (Sleep duration daily (min)) < 99, a subject could be suffering from MCI. If not, the subject should be cognitive healthy. This rule is similar to one of the rules found by fuzzy ARAM to detect MCI as shown in Table 8, though there are more rules and more antecedents discovered by fuzzy ARAM to detect MCI and cognitive healthy. Overall, with the benchmark investigation, our result demonstrated that fuzzy ARAM provides improved predictive performance for differentiating MCI from cognitive healthy as compared to classical models including SVM and Decision Tree, and yields unique interpretable insight that Decision Tree inadequately provides.

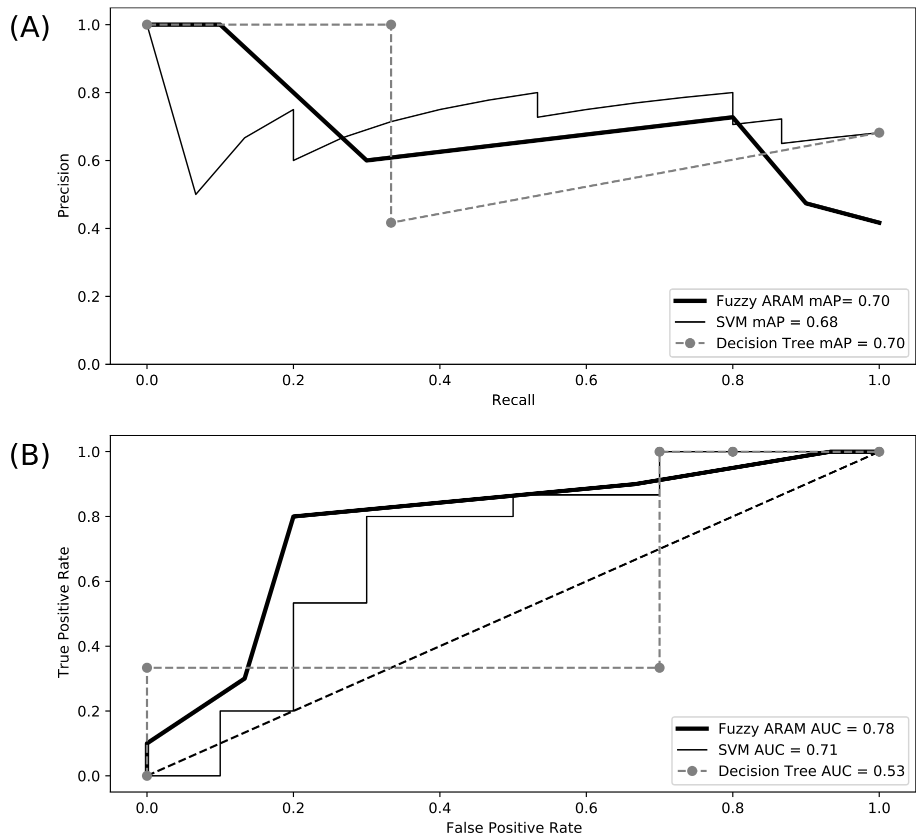

Figure 5: (A) Mean Average Precision (mAP) and (B) Receiver Operating Characteristics-Area Under the Curve (ROC-AUC) predictive performance comparison of fuzzy ARAM against SVM and Decision Tree using the 25-subject data set containing nine behavior metrics related to standard deviation measure.

To further validate the performance efficacy of the predictive self-organizing neural networks for use in the small data sample size problem, we also benchmarked fuzzy ARAM against the state-of-the-art deep learning models including Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) neural network models with different hyperparameters including different hidden layers (e.g. 3-11 layers) and different neurons in each layer (e.g. 12-128) after 300 epochs with Adam optimizer using fixed learning rate of 0.001. Table 9 provides the predictive performance of the highest F1-score model from each of the neural network architecture (i.e., fuzzy ARAM, MLP, CNN, and LSTM) for the 25-subject data set. Clearly, fuzzy ARAM outperformed the different neural network architectures in a small data set situation.

Table 9: Predictive performance of the highest F1-score model from each of the neural network architecture (i.e., ARAM, MLP, CNN, LSTM) for MCI detection.

Models	F1-score	Precision	Accuracy	Sensitivity	Specificity
ARAM	76.2%	72.7%	80.0%	80.0%	80.0%
MLP	57.1%	54.5%	64.0%	60.0%	66.7%
CNN	63.6%	58.3%	68.0%	70.0%	66.7%
LSTM	58.3%	50.0%	60.0%	70.0%	53.3%

Note that recall is not provided as sensitivity is the equivalent of recall.

5. Discussion

5.1. Overcoming small sample size and missing data challenges

In-home sensor monitoring of older adults permits longitudinal study of activities of daily living that could be correlated to different age-related potentially reversible conditions including MCI (Lussier et al., 2019). As these studies are typically resource intensive to manage, so far only small subject sample size studies typically less than 300 are carried out, and out of which only a fraction of these subjects are the MCI cases (Khan & Jacobs, 2021). For instance, in the ORCATECH study over the span of 10 years , only 205 subjects had participated, out of which only 34 subjects developed MCI. Within the 34 participants, only 22 did not change their residence during the 10 years period (Khan & Jacobs, 2021). Hence, Khan & Jacobs (2021) eventually only used 22 MCI and 22 cognitive healthy subjects' data for a balanced data set to build a machine learning model. As most machine learning algorithms necessitate at least a reasonable large sample size to construct an accurate prediction model (Qi & Luo, 2020), Khan & Jacobs (2021) employed each month of the subject data as one data record for modeling, resulting in an eventual total sample size of 395 MCI and 264 cognitive healthy data points for model training, which is still considered as a fairly small data set for machine learning (Tan, 1995). Clearly, in the field of in-home monitoring sensor data for MCI detection, it is necessary to investigate machine leaning model that can be employed in a small data problem. As a result, in this work, we preliminary explored a self-organizing neural networks, which had been previously shown to perform well in a small data problem (Tan, 1995), on a small in-home sensor data set consisting of 49 subjects to build a machine learning model that can be used to accurately detect MCI from cognitive healthy.

Though missing data is a common issue in any real-world longitudinal study, this issue will further reduce a already small sample size data set, affecting the yielding of a reasonable data set sample for machine learning. In this work, we have not only introduced a new algorithm model for MCI prediction, we have also investigated different techniques to handle missing values for machine learning (Tlamelo et al., 2021). In specific, we have explored the list-wise deletion technique which enables us to use a data subset without any missing data (Li & Parker, 2008). Our pilot study indicated that list-wise deletion technique (i.e., the data set without any missing data) can provide a high predictive performance. One possible reason for the ability of list-wise deletion technique to achieve reasonably good predictive performance could be because the missing data are missing at random (Kang, 2013). As machine learning models typically can model such randomness in subject selection, the machine learning algorithm is expected to be able to be trained well and provide good prediction outcome. However, if the missing data is biased to begin with (Kang, 2013), the prediction model may not be able to tolerate dropping of subjects' data with missing data, potentially resulting in a compromised prediction model. Note that though data set without any missing data may be used if the missing values are missing at random, the reduced sample size will result in decrease of statistical power (Kang, 2013), leading to the challenges of detection of MCI with low effect size. To retain the original sample size for maintaining statistical power, imputation approaches should be used if appropriate values can replace the missing values.

ARAM complementary coding scheme for tackling missing values is a unique imputation technique (Granger et al., 2000). Besides list-wise deletion method, we have therefore investigated this technique for replacement of missing values. From our experiment, predictive performance of ARAM complementary coding of missing values degraded roughly 15% as compared to the use of data set without any missing data, which is in good agreement with Granger et al. (2000)'s result. Upon checking on the origin of the increased misclassified cases, our results indicated that most of the misclassification cases arise from subjects without any missing data instead of subjects with missing data. Empirically, this suggests that the ARAM complementary coding of (1,1) for missing values may still affect the template matching process, though the denominator of vigilance test for such missing value cases has been modified to permit higher odds of vigilance to pass (Granger et al., 2000). To unravel the details of the cause for increased misclassification, a more in-depth investigation is warranted in future study. One notes that missing value is a prevalent issue for practical in-home sensor monitoring implementation. As a result, this issue needs to be addressed for exploitation of all data collected for building of a powerful prediction model to detect MCI. Tackling this is non-trivial, as it can come from two dimensions, namely from combining multi-modal sensor data into one complete data set in a cross-sectional study design where it is challenging to ensure the availability of all sensor devices at one time point (Liu et al., 2018); and also from missing data with just one sensor along different time points during a longitudinal study (Nguyen et al., 2020) as it is practically difficult to collect all repeated measurements over time. In this study, we have only started to investigate the problem of combining multi-modal sensor data into one complete data set front. More powerful imputation techniques such as indicator vector, spatio-temporal imputation algorithm, and CLUSTIMP should be explored in the future for ARAM related models (Granger et al., 2000) (Li & Parker, 2008) (Karesiddaiah & Savarimuthu, 2021). With more powerful strategy to overcome missing values, it is expected that predictive performance can be improved instead of degraded, as what we have faced here and what Granger et al. (2000) and Li & Parker (2008) had reported previously. In summary, this work represents the start of investigation for handling missing values from data collected during in-home multi-modal sensor monitoring. With more powerful and suitable imputation approaches, only then fuzzy

ARAM can maximize use of all data to construct a powerful machine learning model that can be used to detect MCI from cognitive healthy.

5.2. Findings from fuzzy ARAM modeling

Fuzzy ARAM is particularly attractive for healthcare application as it permits extraction of interpretable rules for healthcare professionals and researchers to comprehend the logic of the constructed prediction model. In this work, from univariate testing, the variability of sleep measure was elicited to be statistically different for MCI as compared to cognitive healthy subjects. Subsequently, Fuzzy ARAM model discovered that the variability of sleep measure could be used as a predictive feature for MCI detection, supporting the findings from the univariate test that the variability of sleep measure contained significant and specific predictive utility for MCI detection. This is in good agreement with existing data that suggests poorer sleep duration and quality in individuals with MCI (Diem et al., 2016). Additionally, from fuzzy ARAM alone, the frequency of forgetting medicine/wallet and physiological parameters such as step counts and heart rate were uncovered to be a sensitive predictive indicator for MCI detection. From clinical viewpoint, these are common symptoms that are elicited in a clinician's approach to evaluating of cognitive decline. Therefore, our symbolic IF-THEN rules insight corroborates with clinical understanding. Additionally, the discovery of forgetfulness behavior metrics in our work is also consistent with Khan & Jacobs (2021) study where they reported statistical significance of confusion and forgetfulness parameters related to monthly-standard deviation measure. The good agreement of various observations from this study as compared to Hayes et al. (2014), Khan & Jacobs (2021) and clinical understanding indicated that clinically meaningful predictive information could be contained in the time variation of in-home behaviour for early detection of MCI from cognitive healthy.

Collectively, it is evident that variation of in-home behaviour of interests with time contain important diagnostic utility that can be used to detect MCI. Potentially, a different time window to extract variation of in-home behaviour of interests such as one week or even six month-standard deviation can provide higher diagnostic power to detect MCI. Exploration of different time window period ranging from single day to weeks and even to months should be explored to determine the optimal time window for extraction of input data attributes for a classifier model building. Algorithms such as variant of ART can be investigated to discover and learn the optimal time window, especially since ART has been demonstrated to be able to learn a range of time in a continuous space due to its ability to permit time to be encoded as the activation of input nodes (Gao et al., 2021) (Gao & Tan, 2014). ART is a promising candidate due to its attractiveness in high predictive accuracy, intrinsic capability to handle missing values and ability to provide interpretability (Tan & Pan, 2005), unlike deep learning neural network models such as CNN and LSTM which cannot inherently handle missing data, and can only promise high predictive accuracy albeit with black box understanding of the model (Amann et al., 2020).

Regardless of the different types of neural network architecture, hyperparameter tuning remains the most challenging issue to address for achieving a stable and accurate prediction model. For instance, in this work, only after exhaustive search of hyperparameters including α_a, ρ_a, β_a, and different rule-pruning strategies, we then observed that rule-pruning strategy can augment the predictive performance of ART models to a level that is comparable to the MLP result. In fact, the observation that rule-pruning of fuzzy ARAM can improve predictive performance was first reported previously by Carpenter & Tan (1993); In specific, Carpenter & Tan (1993) reported that rule-pruning based on thresholding at 0.5 of fuzzy ARAM on continuous variables can improve test performance. Hence, the phenomenon we observed in this work, which also uses continuous variables, is largely consistent with Carpenter & Tan (1993). In this work, besides the observation that rule-pruning strategy can improve predictive performance, we also discovered that the concise rules derived from antecedent pruning can further augment predictive performance. As the subject sample size used in this work is small but the number of input attributes are relatively large, overfitting may occur within fuzzy ARAM (Matias et al., 2021). Hence, we postulated that antecedent pruning may remove relatively unimportant antecedents from the rules (which could be "noise" instead of rare events), thereby improving generalizability of fuzzy ARAM for prediction. We draw parallel of this phenomenon with L1-norm regularization (Tibshirani, 1996). The findings that a more concise set of rules can improve predictive generalizability was first reported by Pourpanah et al. (2016) where genetic algorithm was employed on top of fuzzy ARAM to select antecedent with the highest diagnostic utility. Hence, our findings corroborated with Pourpanah et al. (2016)'s study. To further confirm this hypothesis, an experiment to characterize noise resilience ability of fuzzy ARAM using different pruning strategies is definitely warranted in the future.

In this study, from observing predictive utility hidden in time variation of different in-home behaviour of interest, and relating to Khan & Jacobs (2021)'s study which reported on the increasing trend of MCI prediction accuracy rate from six months preceding onset of MCI diagnosis, we inferred the possibility to detect MCI up to six months prior to its onset using machine learning on in-home sensor data. Prediction of when an event (in this case it is MCI) can occur can be a classification machine learning problem if the time of event prediction is defined prior such as prediction of the occurrence of an event six months ahead. Otherwise, this can be modeled as a time-to-event statistical learning problem (Lee & Wang, 2013) or a time-to-event deep learning problem (Ranganath et al., 2016) where the time of event prediction is not defined prior. In time-to-event prediction, the objective is to model and predict the time when an event will occur for an entity. To date, to the best of our knowledge, time-to-event prediction utilizing in-home sensors derived behavioural data has not been explored for longitudinal MCI detection.

6. Conclusion

In conclusion, we demonstrated the feasibility of fuzzy ARAM model to detect MCI from cognitive healthy with a F1-score of 76.2% using a small sample size of 25-subject data set related to variation of activities of daily living. Though fuzzy ARAM has demonstrated high predictive performance in a small data sample size, it excels using big data. To employ big data from a real-world in-home sensor monitoring longitudinal study, missing data challenge needs to be tackled; fuzzy ARAM needs to be equipped with a robust missing value approach to leverage of all data collected for model building. With big data that potentially may also contain significant amount of noise as what we have preliminary experienced in this work with small data, development of new ART-based strategy to generalize well on unseen data set will be another critical area to look into. In addition to the small data sample size limitation faced in this work, we also wish to highlight that we have only explored the use of in-home sensor monitoring data at an daily aggregation level for each household, despite having collected data at a finer temporal resolution within a day and at different spatial locations within a house Rawtaer et al. (2020). With the recent findings from Wu et al. (2021) that unravelled the predictive utility of spatiotemporal paterns of daily routines for MCI detection, we see potential to extend our earlier work using spatiotemporal activities of daily living ART (Gao et al., 2021) for MCI detection. On top of all these, due to the initial promises of Khan & Jacobs (2021)'s study which indicated the feasibility to predict MCI around six months ahead of initial onset of MCI, time-to-event modeling could be more appropriate to be applied for this problem instead of modeling this as a classification problem. Overall, time-to-survival based modeling of variance of activities of daily living on a temporal basis using fuzzy ARAM or a modified ARAM model for MCI detection from cognitively healthy can be a future direction starting from this work.

7. Acknowledgement

This research was supported by SingHealth Duke NUS Academic Medical Centre, Academic Medicine Programme Grant and the Jubilee Technology Fellowship awarded to Ah-Hwee Tan by Singapore Management University.

References

Akl, A., Chikhaoui, B., Mattek, N., Kaye, J., Austin, D., & Mihailidis, A. (2016). Clustering home activity distributions for automatic detection of mild cognitive impairment in older adults. *Journal of ambient intelligence and smart environments*, *8 4*, 437–451. doi:10.3233/AIS-160385.

Akl, A., & Mihailidis, A. (2015). Estimating in-home walking speed distributions for unobtrusive detection of mild cognitive impairment in older adults. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5175–5178). doi:10.1109/EMBC.2015.7319557.

Akl, A., Snoek, J., & Mihailidis, A. (2014). Generalized linear models of home activity for automatic detection of mild cognitive impairment in older adults. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 680–683). doi:10.1109/EMBC.2014.6943682.

Akl, A., Snoek, J., & Mihailidis, A. (2017). Unobtrusive detection of mild cognitive impairment in older adults through home monitoring. *IEEE Journal of Biomedical and Health Informatics*, *21*, 339–348. doi:10.1109/JBHI.2015.2512273.

Akl, A., Taati, B., & Mihailidis, A. (2015). Autonomous unobtrusive detection of mild cognitive impairment in older adults. *IEEE Transactions on Biomedical Engineering*, *62*, 1383–1394. doi:10.1109/TBME.2015.2389149.

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*, 310. doi:10.1186/S12911-020-01332-6.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.*, *5*, 1089–1105.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks. doi:10.1201/9781315139470.

Carpenter, G. A., & Grossberg, S. (1992). A self-organizing neural network for supervised learning, recognition, and prediction. *IEEE Communications Magazine*, *30*, 38–49. doi:10.1109/35.156802.

Carpenter, G. A., & Tan, A.-H. (1993). Rule extraction, fuzzy artmap, and medical databases. In *Proceedings of the world congress on neural networks* (pp. 501–506). volume 56.

Carpenter, G. A., & Tan, A.-H. (1995). Rule extraction: From neural architecture to symbolic representation. *Connection Science*, *7*, 3–27. doi:10.1080/09540099508915655.

Chen, B., Tan, H.-P., Rawtaer, I., & Tan, H.-X. (2019). Objective sleep quality as a predictor of mild cognitive impairment in seniors living alone. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 1619–1624). doi:10.1109/BigData47090.2019.9005629.

Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning*, *20*, 273–297. doi:10.1007/BF00994018.

Diem, S., Blackwell, T., Stone, K., Yaffe, K., Tranah, G., Cauley, J., Ancoli-Israel, S., Redline, S., Spira, A., Hillier, T., & Ensrud, K. (2016). Measures of sleep-wake patterns and risk of mild cognitive impairment or dementia in older women. *American Journal of Geriatric Psychiatry*, *24*, 248–258. doi:10.1016/j.jagp.2015.12.002.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, *78*, 316–331. doi:10.1080/01621459.1983.10477973.

Gao, S., & Tan, A.-H. (2014). User daily activity pattern learning: A multi-memory modeling approach. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 1542–1548). doi:10.1109/IJCNN.2014.6889908.

Gao, S., Tan, A.-H., & Setchi, R. (2021). Learning adl daily routines with spatiotemporal neural networks. *IEEE Transactions on Knowledge and Data Engineering*, *33*, 143–153. doi:10.1109/TKDE.2019.2924623.

Granger, E., Rubin, M., Grossberg, S., & Lavoie, P. (2000). Classification of incomplete data using the fuzzy artmap neural network. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000.* (pp. 35–40). volume 6. doi:10.1109/IJCNN.2000.859369.

Hayes, T. L., Abendroth, F., Adami, A., Pavel, M., Zitzelberger, T. A., & Kaye, J. A. (2008). Unobtrusive assessment of activity patterns associated with mild cognitive impairment. *Alzheimer's Dementia*, *4*, 395–405. doi:10.1016/J.JALZ.2008.07.004.

Hayes, T. L., Riley, T., Mattek, N., Pavel, M., & Kaye, J. A. (2014). Sleep habits in mild cognitive impairment. *Alzheimer Disease & Associated Disorders.*, *28*, 145–150. doi:10.1097/WAD.0000000000000010.

He, H., & Ma, Y. (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications*. (1st ed.). Wiley-IEEE Press. doi:10.1002/9781118646106.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*, 402–406. doi:10.4097/KJAE.2013.64.5.402.

Karesiddaiah, S. K., & Savarimuthu, N. (2021). Clustering based imputation algorithm using unsupervised neural network for enhancing the quality of healthcare data. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 1771–1781. doi:10.1007/s12652-020-02250-1.

Khan, T., & Jacobs, P. G. (2021). Prediction of mild cognitive impairment using movement complexity. *IEEE Journal of Biomedical and Health Informatics*, *25*, 227–236. doi:10.1109/JBHI.2020.2985907.

Lee, E. T., & Wang, J. W. (2013). *Statistical Methods for Survival Data Analysis*. (4th ed.). Wiley Publishing. doi:10.1002/0471458546.

Li, Y., & Parker, L. (2008). Classification with missing data in a wireless sensor network. In *IEEE SoutheastCon 2008* (pp. 533 – 538). doi:10.1109/SECON.2008.4494352.

Liu, M., Gao, Y., Yap, P. T., & Shen, D. (2018). Multi-Hypergraph Learning for Incomplete Multimodality Data. *IEEE Journal of Biomedical and Health Informatics*, *22*, 1197–1208. doi:10.1109/JBHI.2017.2732287.

Lussier, M., Lavoie, M., Giroux, S., Consel, C., Guay, M., Macoir, J., Hudon, C., Lorrain, D., Talbot, L., Langlois, F., Pigot, H., & Bier, N. (2019). Early detection of mild cognitive impairment with in-home monitoring technologies using functional measures: A systematic review. *IEEE Journal of Biomedical and Health Informatics*, *23*, 838–847. doi:10.1109/JBHI.2018.2834317.

Matias, A. L. S., da Rocha Neto, A. R., Mattos, C. L. C., & Gomes, J. P. P. (2021). A novel fuzzy artmap with area of influence. *Neurocomputing*, *432*, 80–90. doi:10.15837/ijccc.2009.2.2418.

McMaster, M., Kim, S., Clare, L., Torres, S. J., Cherbuin, N., D'Este, C., & Anstey, K. J. (2020). Lifestyle risk factors and cognitive outcomes from the multidomain dementia risk reduction randomized controlled trial, body brain life for cognitive decline (bbl-cd). *Journal of the American Geriatrics Society*, *68*, 2629–2637. doi:10.1111/JGS.16762.

Nguyen, M., He, T., An, L., Alexander, D. C., Feng, J., & Yeo, B. T. T. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. *Neuroimage*, *222*, 117203. doi:10.1016/J.NEUROIMAGE.2020.117203.

Pourpanah, F., Lim, C. P., & Saleh, J. M. (2016). A hybrid model of fuzzy artmap and genetic algorithm for data classification and rule extraction. *Expert Systems with Applications*, *49*, 74–85. doi:10.1016/j.eswa.2015.11.009.

Qi, G., & Luo, J. (2020). Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (pp. 1–1). doi:10.1109/TPAMI.2020.3031898.

Ranganath, R., Perotte, A., Elhadad, N., & Blei, D. (2016). Deep survival analysis. In *Proceedings of the 1st Machine Learning for Healthcare Conference* (pp. 101–114). volume 56.

Rawtaer, I., Mahendran, R., Kua, E., Tan, H.-P., Tan, H., Lee, T.-S., & Ng, T. (2020). Early detection of mild cognitive impairment with in-home sensors to monitor behavior patterns in community-dwelling senior citizens in singapore: Cross-sectional feasibility study. *Journal of Medical Internet Research*, *22*, e16854. doi:10.2196/16854.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROc plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, *10*, e0118432. doi:10.1371/JOURNAL.PONE.0118432.

Tan, A.-H. (1995). Adaptive resonance associative map. *Neural Networks*, *8*, 437–446. doi:10.1016/0893-6080(94)00092-Z.

Tan, A.-H., & Pan, H. (2005). Predictive neural networks for gene expression data analysis. *Neural Networks*, *18*, 297–306. doi:10.1016/J.NEUNET.2005.01.003.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Tierney, M. C., & Lermer, M. A. (2010). Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. *Journal of Alzheimer's disease*, *20*, 823–832. doi:10.3233/JAD-2010-091672.

Tlamelo, E., Thabiso, M., Dimane, M., Thabo, S., Banyatsang, M., & Oteng, T. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*, 140. doi:10.1186/s40537-021-00516-9.

World Health Organization (2021). *2021 Alzheimer's disease facts and figures*. Technical Report 3. doi:10.1002/ALZ.12328.

Wu, C.-Y., Dodge, H. H., Gothard, S., Mattek, N., Wright, K., Barnes, L. L., Silbert, L. C., Lim, M. M., Kaye, J. A., & Beattie, Z. (2021). Unobtrusive Sensing Technology Detects Ecologically Valid Spatiotemporal Patterns of Daily Routines Distinctive to Persons With Mild Cognitive Impairment. *The Journals of Gerontology: Series A*, . doi:10.1093/gerona/glab293.

Zihni, E., Madai, V. I., Livne, M., Galinovic, I., Khalil, A. A., Fiebach, J. B., & Frey, D. (2020). Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. *PLOS ONE*, *15*, 1–15. doi:10.1371/JOURNAL.PONE.0231166.

Appendix 1

Supplementary Table 1. Missing values (without tick and shaded in grey) for each of the nine in-home behavior metrics in relation to the different subjects

Subject No.	Television use daily (min)	Time away from home daily (min)	No. of outings daily	Freq. of forgetting keys/month	Freq. of forgetting wallet/month	Sleep duration daily (min)	Heart rate (bpm)	Steps (daily)	Freq. of forgetting medicine/month
1	✓	✓	✓	✓	✓	✓			✓
2	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓			
4	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓	
6	✓	✓	✓	✓	✓	✓	✓	✓	
7	✓	✓	✓	✓	✓	✓	✓	✓	✓
8	✓	✓	✓	✓	✓	✓			
9	✓	✓	✓			✓			✓
10	✓	✓	✓	✓	✓		✓	✓	
11	✓	✓	✓	✓	✓		✓	✓	
12	✓	✓	✓	✓	✓	✓	✓	✓	✓
13	✓	✓	✓	✓	✓				✓
14	✓	✓	✓	✓	✓	✓	✓	✓	✓
15	✓	✓	✓			✓	✓	✓	✓
16	✓	✓	✓	✓	✓		✓	✓	✓
17	✓	✓	✓	✓	✓	✓	✓	✓	
18	✓	✓	✓	✓	✓		✓	✓	✓
19	✓	✓	✓	✓	✓		✓	✓	
20	✓	✓	✓	✓	✓		✓	✓	✓
21	✓	✓	✓	✓	✓	✓	✓	✓	✓
22	✓	✓	✓	✓	✓	✓	✓	✓	✓
23	✓	✓	✓	✓	✓	✓	✓	✓	✓
24	✓	✓	✓	✓	✓		✓		✓
25	✓	✓	✓	✓	✓	✓	✓	✓	✓
26	✓	✓	✓	✓	✓	✓	✓	✓	✓
27	✓	✓	✓	✓	✓	✓	✓	✓	✓
28	✓	✓	✓	✓	✓	✓	✓	✓	✓
29	✓	✓	✓	✓	✓	✓	✓	✓	✓
30	✓	✓	✓	✓	✓	✓	✓	✓	✓
31	✓	✓	✓	✓	✓	✓	✓	✓	✓
32	✓	✓	✓	✓	✓	✓	✓	✓	✓
33	✓	✓	✓	✓	✓	✓			✓
34	✓	✓	✓	✓	✓	✓	✓	✓	✓
35	✓	✓	✓	✓	✓	✓	✓	✓	✓
36	✓	✓	✓	✓	✓	✓	✓	✓	✓
37	✓	✓	✓	✓	✓		✓	✓	
38	✓	✓	✓	✓	✓	✓	✓	✓	
39	✓	✓	✓	✓	✓	✓			✓
40	✓	✓	✓	✓	✓	✓	✓	✓	
41	✓	✓	✓	✓	✓	✓	✓	✓	✓
42	✓	✓	✓	✓	✓				✓
43	✓	✓	✓	✓	✓	✓	✓	✓	✓
44	✓	✓	✓	✓	✓	✓	✓	✓	✓
45	✓	✓	✓	✓	✓	✓			
46	✓	✓	✓	✓	✓	✓	✓	✓	✓
47	✓	✓	✓	✓	✓	✓	✓	✓	✓
48	✓	✓	✓	✓	✓	✓	✓	✓	✓
49	✓	✓	✓	✓	✓	✓			✓