TATC: Predicting Alzheimer's Disease with Actigraphy Data

Jia Li
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
lijia@se.cuhk.edu.hk

Yu Rong
Tencent AI Lab
Shenzhen
royrong@tencent.com

Helen Meng
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
hmmeng@se.cuhk.edu.hk

Zhihui Lu
Department of Medicine and Therapeutics
The Chinese University of Hong Kong
anjolu@link.cuhk.edu.hk

Timothy Kwok
Department of Medicine and Therapeutics
The Chinese University of Hong Kong
tkwok@cuhk.edu.hk

Hong Cheng
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
hcheng@se.cuhk.edu.hk

ABSTRACT

With the increase of elderly population, Alzheimer's Disease (AD), as the most common cause of dementia among the elderly, is affecting more and more senior people. It is crucial for a patient to receive accurate and timely diagnosis of AD. Current diagnosis relies on doctors' experience and clinical test, which, unfortunately, may not be performed until noticeable AD symptoms are developed. In this work, we present our novel solution named time-aware TICC and CNN (TATC), for predicting AD from actigraphy data. TATC is a multivariate time series classification method using a neural attention-based deep learning approach. It not only performs accurate prediction of AD risk, but also generates meaningful interpretation of daily behavior pattern of subjects. TATC provides an automatic, low-cost solution for continuously monitoring the change of physical activity of subjects in daily living environment. We believe the future deployment of TATC can benefit both doctors and patients in early detection of potential AD risk.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Computing methodologies** → *Neural networks*;

KEYWORDS

Alzheimer's Disease; actigraphy data; circadian rhythm; attention

ACM Reference Format:
Jia Li, Yu Rong, Helen Meng, Zhihui Lu, Timothy Kwok, and Hong Cheng. 2018. TATC: Predicting Alzheimer's Disease with Actigraphy Data. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3219819.3219831

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219831

1 INTRODUCTION

The world is witnessing a dramatic increase of elderly population. It is reported by UN[1] that 1 in 5 on this planet will age 60 or older by 2050. Alzheimer's Disease (AD), the most common cause of dementia among the elderly, is an irreversible and progressive neurodegenerative disease that destroys memory and other important mental functions, resulting in the loss of intellectual and social skills [14]. It is estimated that AD will double its frequency in the next 20 years [1]. AD symptoms are gradually developed and in many cases are not easily recognized by patients or caregivers until severe behavioural and cognitive changes happen. It is critical to diagnose the cognitive status of a patient related to AD in an accurate and timely manner, so that effective strategies can be implemented to prevent cognitive decline of the patient. Unfortunately, current diagnosis of AD relies on doctors' experience and clinical test such as Montreal Cognitive Assessment (MoCA) or magnetic resonance imaging (MRI), which is very costly. Patients may lose the best opportunity of timely diagnosis and treatment.

A recent study [17] has identified physical activity as one of the modifiable risk factors for AD. Meta-analysis of prospective studies [24] has also reported that physical activity has a significant protective effect against AD. In this context some studies [28, 29] begin to use wrist-worn devices such as actigraphs to assess objective and continuous physical activity records in free-living environment, in the hope of understanding the characteristics of physical activity among people with different cognitive status related to AD. Motivated by these findings, this paper aims to address the following problem using a data mining approach: *can actigraphy data be used to predict Alzheimer's Disease?* To this end, we launched the project of Hong Kong Alzheimer's Disease Study in 2016, in which 560 Chinese men and 500 Chinese women aged 65 years and older were recruited. Subjects are further categorized into three cognitive groups related to AD and are required to wear an actigraph for recording their physical activity. As the recorded actigraphy data is a time series, we formulate our problem as a time series classification problem.

In the literature, many methods (e.g., [3, 4, 23]) have been proposed on time series classification. Most of them only consider the

[1] http://www.un.org/esa/population/publications/worldageing19502050/pdf/90chapteriv.pdf

relative temporal order between the series of data points, but ignore the absolute clock time associated with them. Two time series sequences that have identical values but happen at differen time are regarded the same. This assumption does not hold in our scenario, as the absolute time matters in our problem. As an example, a subject with a low level of physical activity at 01:00am is clearly different from another subject with a similar level of activity at 10:00am, since the former can be interpreted as *sleeping at night* whereas the latter as *physically inactive in daytime*. In medicine, the absolute clock time in monitoring patients' behavior is closely related to the *effect of biological clock/circadian rhythm*, which has been proved to be very important in the Nobel Prize in Medicine 2017[2]. Thus the absolute clock time should be considered in our classification model.

In our model design, another important consideration is its ability of providing meaningful interpretation of the actigraphy data, which corresponds to the daily behavior pattern of subjects, e.g., sleeping, exercising, light activity, etc. This motivates us to design a multivariate time series classification model called time-aware TICC and CNN (TATC). It is a neural attention-based deep learning model that provides meaningful interpretation. In TATC, we take a composite representation learning approach to extract discriminative features using both unsupervised and supervised learning. TICC [12] is the unsupervised component that infers hidden states, e.g., sleeping and exercising, from time series. CNN is the supervised component to learn discriminative temporal features from time series. Then we utilize a time-aware attention mechanism to capture the *effect of circadian rhythm* by learning context weights for different time intervals.

Our contributions are summarized as follows.

- We take a data mining approach to predicting AD from actigraphy data collected under the project of Hong Kong Alzheimer's Disease Study. Our proposed TATC method provides an automatic, low-cost solution for continuously monitoring the change of physical activity of subjects in daily living environment. We believe the future deployment of TATC can benefit both doctors and patients in early detection of potential AD risk.
- We design and implement TATC, a neural attention-based deep learning model for multivariate time series classification. We use a time-aware attention mechanism to model the effect of circadian rhythms. TATC outperforms three baselines for time series classification with promising prediction performance. It also provides meaningful interpretation of the inferred hidden states, which corresponds to the daily behavior pattern of subjects.
- We report our experiences and insights gained from this study, in particular, on data collection and its practical value to clinical diagnosis.

The remainder of this paper is organized as follows. Section 2 describes our data collection procedure and the data format. Section 3 presents the design of TATC. Section 4 reports the experimental results. Section 5 summarizes our lessons and insights gained in this study. Section 6 reviews related work. Finally, Section 7 concludes the paper.

[2]https://www.nobelprize.org/nobel_prizes/medicine/laureates/2017/

2 DATA COLLECTION

Our study targets subjects who are 65 years and older in Hong Kong and may suffer from AD. We use actigraphy data which records their physical activity to predict AD. This motivates the design of TATC, yet our methodology is potentially applicable to some other diseases such as Parkinson's disease. In this section, we first provide a description of participants enrolled in this project, and then describe the format of the data for our analysis.

2.1 Participants

The project of Hong Kong Alzheimer's Disease Study was initiated by Prince of Wales Hospital[3] in 2016. Subjects aged 65 years and older were recruited under the project. 560 Chinese men and 500 Chinese women were enrolled, among who some subjects have been diagnosed early stage of AD in memory clinics or geriatrics clinics and taken AD drugs for at least 3 months. During 2016 – 2017, the cohort was invited for a repeated questionnaire interview and measurement of physical performance.

To evaluate the subjects' cognitive status, their cognitive function was assessed by the validated Hong Kong version of Montreal Cognitive Assessment (HK-MoCA) [27]. The whole scale has a total score of 30, and 21/22 is the cutoff score to differentiate cognitive impairment from cognitive normal. Participants who scored 22 or below were suggested for further diagnosis by the clinical doctor. Based on their MoCA score and doctor's diagnosis, participants were categorized into three cognition groups: normal control (NC), mild cognitive impairment (MCI) and Alzheimer's Disease (AD). For the group of NC, no symptom of AD is exhibited. For the group of MCI, the symptoms related to the thinking ability may start to be noticeable to caregivers but they do not affect the daily life of subjects [1]. For the group of AD, the dementia may affect the subjects' daily life. Out of 1,060 subjects enrolled in this project, 729 had their cognitive status evaluated and actigraphy data properly recorded. Among the 729 subjects, 441 of them were classified as NC, 103 were MCI, and 185 were AD. There were still 331 subjects who did not complete the procedure either because they did not come to hospital to see the doctor or because their actigraphy data was not valid. The criterion of judging whether the collected actigraphy data is valid or not is illustrated in Section 2.3.

2.2 Personal Particulars of Subjects

In the interview questionnaire, three types of subject information are collected. The first type is the personal information of subjects, including gender, age, years of education (MASCH), and body mass index (BMI). The second type is the clinical history of subjects. Specifically, subjects were asked whether they had diabetes before (MHDIAB) and whether they had heart disease before (MHMI). The third type records physical test results of subjects, including their maximum grip strength (GRIPAM) and average walking speed (GAITSPEED). The three types of features are listed in Table 1. The statistics of these feature values for subjects in NC, MCI and AD groups are displayed in Table 2.

[3]http://www3.ha.org.hk/pwh/index_e.asp

Table 1: Features of personal particulars

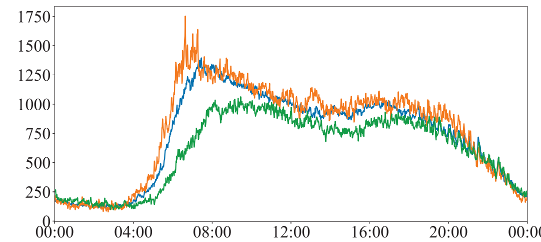
Type	Notation	Meaning
Personal information	Gender	M=male, F=female
	Age	65 years and older
	MASCH	years of education
	BMI	body mass index
Clinical history	MHDIAB	medical history of diabetes: Y=Yes, N=No
	MHMI	medical history of heart disease: Y=Yes, N=No
Physical test	GRIPAM	maximum grip strength (kg)
	GAITSPEED	walking speed (m/s)

2.3 Physical Activity Records

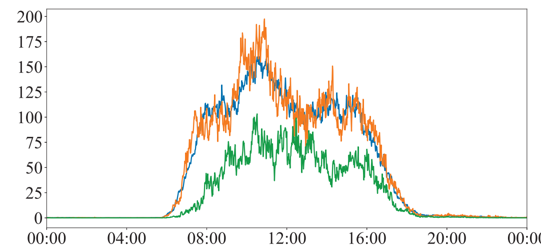
Besides the information collected in the interview questionnaire, participants were invited to wear an actigraph GT3X (Pensacola, Florida, USA) on their non-dominant wrist for 7 consecutive days, except when bathing or swimming. The actigraph GT3X is a small device used for capturing and recording continuous, high resolution physical activity. It contains a 3-axis accelerometer and a light sensor. The accelerometer monitors general arm movement and yields acceleration in three axes in 60-second epochs, which we denote as X_{acc}, Y_{acc}, and Z_{acc} thereafter. The light sensor records ambient light in the same frequency, which we denote as Lux. Thus we get a 4-dimensional vector as $I = [X_{acc}, Y_{acc}, Z_{acc}, Lux]$ every minute for each subject. Through 7 consecutive days, we obtain a time series in the form of $\langle (I_1; t_1), (I_2; t_2), \ldots, (I_k; t_k) \rangle$, where I_k is the vector $[X_{acc}, Y_{acc}, Z_{acc}, Lux]$ recorded at time t_k. We exemplify a time series fragment generated by the actigraph in a 3-minute window in Table 3.

Since physical activity is closely related to the 24-hour biological clock, we generate an average circadian activity of a subject in a 24-hour time span. Specifically, for every minute of a day, e.g., 23:59, we collect the records of a subject at that minute through 7 days and compute a 7-day average. Thus, for each subject, the average circadian activity is represented as a 4-dimensional multivariate time series consisting of 1,440 timestamps, one for each minute in 24 hours. In case subjects forgot to put on the actigraph after bathing or swimming, there are some non-wear periods. We apply methods proposed in [8] to detect these periods. We ignore such non-wear periods when the average function is applied. For example, given a subject's X_{acc} values at timestamp 23:59 of 7 consecutive days as $6, 3, Null, Null, 4, 12, 1$, where $Null$ represents no record in a non-wear period, the average X_{acc} of that subject at timestamp 23:59 is calculated by $(6 + 3 + 4 + 12 + 1)/5 = 5.2$. If the entries of a subject at the same minute through 7 days are all $Null$, we denote the minute as *missing timestamp*, and fill in the average values of all the subjects at the same minute. We consider a time series to be valid if the number of *missing timestamps* is less than 720 out of 1,440 (i.e., half a day).

To illustrate the behavior of each group, we take the time series data of all subjects within the same group and compute a group average. Figure 1a is a plot of the group average circadian activity in terms of X_{acc} of the three groups NC (plotted in blue), MCI (plotted in orange) and AD (plotted in green) in 24 hours, respectively. Figure 1b is a plot of the group average circadian activity in terms of Lux of the three groups in 24 hours.



(a) Average circadian activity by X_{acc} of NC (blue), MCI (orange) and AD (green)



(b) Average circadian activity by Lux of NC (blue), MCI (orange) and AD (green)

Figure 1: The group average circadian activity per minute by three groups. The main gaps between each pair of groups appear in the wake-up time, morning and afternoon.

2.4 Observations

Based on the group average circadian activity, we can make the following observations.

- MCI subjects are very similar to NC subjects from both motor movement measured by X_{acc} and ambient light measured by Lux, while AD subjects are well separated from NC and MCI subjects in the daytime.
- The difference between groups varies over time. In the evening the difference is not obvious. The first big gap appears in the wake-up time (i.e., 4:00 – 8:00am) in which NC and MCI subjects become quite active while AD subjects are not. Compared with NC, MCI subjects show bigger variance in both X_{acc} and Lux.

Similar observations can be made from the measurements of Y_{acc} and Z_{acc}. From the above observations, we conclude that different time intervals have different degree of importance, thus we should pay more attention to some specific time intervals, such as wake-up time, morning, and afternoon.

3 THE PROPOSED TATC MODEL

TATC is designed to predict Alzheimer's Disease based on the collected actigraphy data. We formulate this problem as a classification problem whose outcome is a probability that the subject has the disease. Specifically, we construct two binary classifiers: one on AD versus NC, and the other on MCI versus NC. Formally, suppose that we are given a set of subjects' data denoted as $D = \{(X_i, y_i)\}_{i=1}^{m}$, where X_i is the multivariate time series representing the average circadian activity of subject i, y_i is the true class label, and m is the

Table 2: Statistics of personal information of subjects in NC, MCI and AD groups

Group	No. of subjects	Age	MASCH	BMI	MHDIAB	MHMI	GRIPAM	GAITSPEED
NC	441 (M/F: 287/154)	82.4±3.6	7.7±4.9	23.5±11.2	Y/N: 89/352	Y/N: 50/391	24.9±4.7	0.8±0.2
MCI	103 (M/F: 57/46)	83.3±3.6	4.3±4.4	23.4±10.8	Y/N: 21/82	Y/N: 9/94	18.4±6.1	0.8±0.2
AD	185 (M/F: 68/117)	80.6±5.8	6.5±6.0	23.5±12.5	Y/N: 50/135	Y/N: 20/165	14.2±6.1	0.6±0.2

Age, MASCH, BMI, GRIPAM, GAITSPEED are listed as mean±standard deviation.

Table 3: A multivariate time series example generated by actigraph GT3X

SubjectID	Date	Time	X_{acc}	Y_{acc}	Z_{acc}	Lux
HK0001	20161001	23:59	23	8	5	0
HK0001	20161002	00:00	32	15	9	4
HK0001	20161002	00:01	7	5	2	0

number of subjects, the target is to train a classification model to predict the class label for any subject given his/her actigraphy data.

In the literature, many methods on time series classification ignore the absolute clock time associated with the series of data points. Two time series sequences that have identical values but happen at different time are regarded the same. But this assumption does not hold in our scenario. A subject with a low level of activity at 01:00am is clearly different from another subject with a similar level of activity at 10:00am. According to the observations in Section 2, the actigraphy data exhibits a clear circadian rhythm that is controlled by the human biological clock. Different time intervals have different degree of importance to differentiate the behavior of the three groups of subjects. This motivates the design of TATC which integrates the absolute clock time into the model. In TATC, we first take a composite representation learning approach to extract meaningful features using both unsupervised and supervised learning. Then we utilize a time-aware attention mechanism to model the *effect of circadian rhythm*. Based on the composite feature representation and time-aware attention, we use RNN to model the whole time series and capture the global temporal dependencies in the time series. Finally, we take the output of the RNN model as input, and construct a binary classifier for prediction. Figure 2 depicts the overall framework of TATC.

In this section, we first introduce the composite feature representation learning approach, and describe the time-aware attention mechanism. Then we present temporal dependency modeling methods.

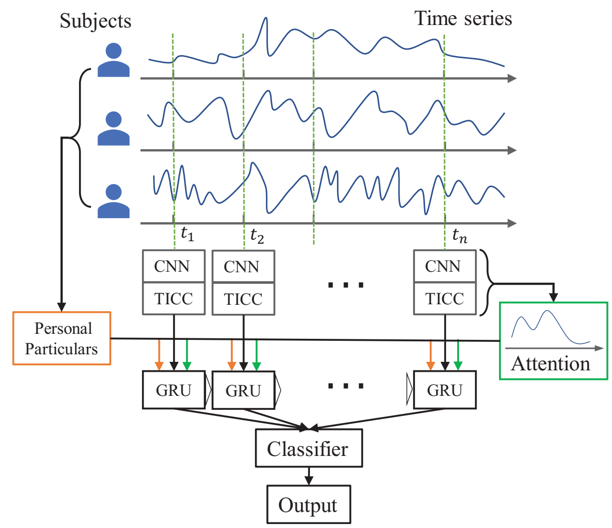


Figure 2: Framework of the proposed TATC model

3.1 Composite Feature Representation Learning

To build a classifier for AD prediction, we need to extract discriminative features from the actigraphy data. We define two types of features that are complementary to each other. The first type of features describes the daily behavior pattern of subjects, e.g., sleeping, exercising, etc., which provides an interpretable representation of the raw, high-dimensional actigraphy data. This type of features is learned by Toeplitz Inverse Covariance-based Clustering (TICC) [12], in an unsupervised learning approach. The second type of features is learned by Convolutional Neural Network (CNN), in a supervised learning approach by leveraging the class labels of the training data. We use the composite representation of these two types of features to characterize a time series.

3.1.1 TICC Representation. The actigraphy data can be expressed as a timeline of a small number of hidden states, which correspond to the daily behavior pattern of subjects, e.g., sleeping, exercising, etc. Such hidden states, once inferred from the time series, can be used as an interpretable representation of the raw, high-dimensional actigraphy data, and leveraged to learn the classification model. To this end, we apply Toeplitz Inverse Covariance-based Clustering (TICC), recently proposed by Hallac et al. [12], on the actigraphy data to infer the hidden state for each timestamp, and construct the TICC representation based on the hidden states.

Given a subject's average circadian activity time series $X = \langle x_j \rangle_{j=1}^{T}$ where $T = 1,440$ and x_j is a 4-dimensional vector, to infer the hidden state for each timestamp, TICC assigns each timestamp to one Gaussian inverse covariance Θ_k by minimizing the following objective function:

$$\min \sum_{k=1}^{\kappa} \sum_{x_j \in P_k} -\zeta(x_j, \Theta_k) + \beta \mathbb{1}\{x_{j-1} \notin P_k\}. \quad (1)$$

In (1), κ is the number of hidden states. P_k is the corresponding cluster of hidden state k. $\mathbb{1}\{x_{j-1} \notin P_k\}$ is the indicator function indicating that the previous hidden state is different from the current hidden state. β is the penalty parameter. $\zeta(x_j, \Theta_k)$ is the log likelihood that x_j belongs to P_k. The Gaussian inverse covariance Θ_k inferred by (1) represents the daily behavior pattern of subjects, such as sleeping and exercising. It is worth pointing out that the penalty term $\mathbb{1}\{x_{j-1} \notin P_k\}$ is vital to ensure the temporal consistency of subjects' behaviors. This penalty makes sense since a subject cannot jump from sleeping to exercising and then back to sleeping again in several minutes. By solving the optimization problem in (1), we can infer the hidden state of each minute.

To construct the TICC representation, we propose to abstract the minute-level time series $X = \langle x_j \rangle_{j=1}^{T}$ into a coarse-grained representation, in the unit of a longer interval, e.g., an hour. There are two considerations for this interval-level representation: (1) The minute-level representation yields high-dimensional features. This may pose a challenge considering the number of subjects participating in this project is relatively small. Using a longer interval as the representation unit reduces the dimensionality; and (2) It is unnecessary to characterize the daily behavior pattern of subjects with such a fine granularity as minute. For example, it suffices to know that a subject usually wakes up during 5:00 – 5:30am, but it does not make a big difference, be it 5:00am or 5:01am. With these considerations, we segment a time series X into n equal-length subsequences, i.e., $X = \langle X_j \rangle_{j=1}^{n}$, $X_j = \langle x_j, x_{j+1}, \ldots, x_{j+T/n-1} \rangle$. In each subsequence X_j, we infer the hidden state for each minute by (1) and calculate the distribution of hidden states. For example, suppose $n = 24$, i.e., we segment a time series of a day into 24 intervals and each interval lasts one hour. We then calculate the distribution of hidden states in an hour. Suppose there are totally $\kappa = 5$ hidden states, and in an hour the counts for each hidden state are 15, 15, 12, 12, 6, then the hidden state distribution for that hour is 25%, 25%, 20%, 20%, 10%. After calculating the distribution of hidden states in each interval, we obtain the temporal representation $\boldsymbol{TICC} \in \mathbb{R}^{n \times \kappa}$ for a subject over n intervals.

3.1.2 CNN Representation. Besides the TICC representation, we also use Convolutional Neural Network (CNN) to learn discriminative temporal features from the time series in a supervised approach. In CNN, a filter is used to do convolution operations with different temporal subsequences within the timespan of the filter. Similar to the segmentation in Section 3.1.1, we first segment a time series X into n equal-length subsequences $\langle X_j \rangle_{j=1}^{n}$. In each subsequence, we calculate γ statistics for each univariate time series separately. Specifically, we consider $\gamma = 4$ types of statistics: mean, standard deviation, max value and cross zero rate (CZR). The definition of CZR for a subsequence X_j is:

$$CZR(X_j) = \frac{1}{T/n - 1} \sum_{t=1}^{T/n-1} \mathbb{1}\{(x_{j+t} - \mu)(x_{j+t-1} - \mu) < 0\}. \quad (2)$$

In (2), μ represents the mean value of the subsequence, and $\mathbb{1}\{(x_{j+t} - \mu)(x_{j+t-1} - \mu) < 0\}$ is the indicator function indicating that for any two adjacent timestamps one value is larger than μ while the other one is smaller. Suppose we still segment a time series into 24 intervals and each interval lasts one hour. We then calculate 4 statistics for $X_{acc}, Y_{acc}, Z_{acc}, Lux$, respectively. Thus we get $4 \times 4 = 16$ statistics every hour. The statistics calculated from the j-th subsequence of a subject, X_j, can be represented by a vector:

$$\boldsymbol{c}_j = [f_1(X_j), f_2(X_j), f_3(X_j), f_4(X_j)], \quad (3)$$

where f_1, f_2, f_3, f_4 represent the mean, standard deviation, max value and CZR functions, respectively. Over n intervals, the statistics for a subject are denoted by $\boldsymbol{C} = \langle \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n \rangle$, and $\boldsymbol{C} \in \mathbb{R}^{n \times 4\gamma}$.

It is known that hierarchically integrated multi-layer CNN is very powerful in learning non-linear discriminative features when sufficient training samples are provided. But given the small number of subjects in our problem, the multi-layer CNN would suffer from serious overfitting. To alleviate this issue, we use a single-layer CNN which takes the statistics in $\boldsymbol{C}$ as input to learn the feature representation. It is worth mentioning that since we are working with temporal data, the CNN filters are all 1-dimensional. Specifically, we rewrite $\boldsymbol{C} \in \mathbb{R}^{n \times 4\gamma}$ as $\boldsymbol{C} = [\bar{\boldsymbol{c}}_1, \bar{\boldsymbol{c}}_2, \ldots, \bar{\boldsymbol{c}}_{4\gamma}]$, where $\bar{\boldsymbol{c}}_k \in \mathbb{R}^{\boldsymbol{n}}$. To detect one kind of temporal features from $\boldsymbol{C}$ for interval j, we apply the following calculation:

$$CNN_j = \text{ReLU}(b + \sum_{k=1}^{4\gamma} [\sum_{l=1}^{L} \boldsymbol{F}_k(l) \bar{\boldsymbol{c}}_k(j - l)]), \quad (4)$$

where b is the bias, L is the length of filter and $\boldsymbol{F}_k$ is the filter working on dimension k. By detecting one kind of temporal features with (4), we get 1-dimensional CNN representation for the current channel. Thus for α kinds of temporal features, the final CNN representation of a subject can be written as $\boldsymbol{CNN} \in \mathbb{R}^{n \times \alpha}$.

By combining the TICC representation and CNN representation, we obtain the composite representation $\boldsymbol{TC} \in \mathbb{R}^{n \times (\kappa + \beta)}$ over n intervals:

$$\boldsymbol{TC} = [\boldsymbol{TICC}, \boldsymbol{CNN}]. \quad (5)$$

3.2 Time-aware Attention

In this subsection, we propose a time-aware attention mechanism to deal with the *effect of circadian rhythm*. In general, the basic assumption of the attention mechanism is that only selective parts of input features are informative for the end learning task. Specifically, in our problem setting, different time intervals in a day have different degree of importance to differentiate the three groups of subjects, which can be realized by the time-aware attention mechanism. In the following, we first describe how to learn the attention weights for different time intervals based on the composite representation $\boldsymbol{TC}$. Then we transform $\boldsymbol{TC}$ into the time-aware representation with the attention weights.

To learn the attention weights for different time intervals, we first embed the multivariate $\boldsymbol{TC}$ into a univariate time series by a linear model. There are two considerations for this embedding: (1) we do not need to differentiate variates within the same time interval; and (2) we can reduce the number of parameters compared with learning the attention weights directly from $\boldsymbol{TC}$. Formally, given $\boldsymbol{TC} = \langle \boldsymbol{tc}_j \rangle_{j=1}^{n}$, where each $\boldsymbol{tc}_j$ is a $(\kappa + \beta)$-dimensional vector for the composite representation of the j-th interval, we embed $\boldsymbol{tc}_j$ into a scalar by:

$$S_j = \boldsymbol{E}^{\mathrm{T}} \boldsymbol{tc}_j + b_0. \quad (6)$$

Here $\boldsymbol{E} \in \mathbb{R}^{\kappa+\beta}$ is the shared weight parameter of all time intervals, $b_0 \in \mathbb{R}$ is the bias. Hence the required parameter number is $(\kappa+\beta+1)$.

After getting a univariate time series $\boldsymbol{S} = \langle S_j \rangle_{j=1}^{n}$, we learn an attention weight for a specific time interval j by:

$$\alpha_j = \boldsymbol{W}_j^{\top} \boldsymbol{S}, \tag{7}$$

where $\boldsymbol{W}_j \in \mathbb{R}^n$ is an n-dimensional vector of parameters. To obtain the attention weights $\boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^{n}$ for n time intervals, we need n^2 parameters. Together with the number of embedding parameters, the total number of parameters needed is $(n^2 + \kappa + \beta + 1)$.

After that, a softmax operator is applied on $\boldsymbol{\alpha}$ and we get the time-aware attention mechanism specially for limited training samples. For each time interval in $\boldsymbol{TC} = \langle \boldsymbol{tc}_j \rangle_{j=1}^{n}$, we combine the feature representation $\boldsymbol{tc}_j$ with the corresponding attention weight α_j, and get the time-aware representation as $\boldsymbol{TC}_{\boldsymbol{\alpha}} = \langle [\alpha_j, \boldsymbol{tc}_j] \rangle_{j=1}^{n}$. Note that this time-aware attention is learned using backpropagation, thus it is a data-driven approach to evaluating how important a specific time interval is in contributing to the prediction of AD.

3.3 Temporal Dependency Modeling

Recurrent Neural Networks (RNN), such as Long Short-Term Memory recurrent neural network (LSTM) and Gated Recurrent Unit (GRU) [9], are famous for their excellent performance in modeling the long-term temporal dependencies in time series data. In our solution, we use GRU to capture the temporal dependencies, since GRU can achieve the same level of performance but requires fewer parameters compared with LSTM. The consideration to take the temporal dependencies into modeling is that subjects have some long-term patterns and these patterns cannot be represented by short-term representations such as $\boldsymbol{TC}_{\boldsymbol{\alpha}}$. For example, [10] finds that some severe AD subjects have some disturbances in sleep-wake cycles, and these kinds of patterns cannot be captured by $\boldsymbol{TC}_{\boldsymbol{\alpha}}$.

GRU can capture the long-term temporal dependencies by keeping the recurrent hidden states inside. The hidden state h_j is updated upon the previous hidden state h_{j-1} with the input $\boldsymbol{TC}_{\boldsymbol{\alpha}}$ and the personal particulars of a subject, denoted as $\boldsymbol{d}$, as defined in Section 2.2:

$$h_j = \phi(h_{j-1}, \boldsymbol{TC}_{\boldsymbol{\alpha}}, \boldsymbol{d}), \tag{8}$$

where ϕ is a nonlinear function such as composition of a logistic sigmoid with an affine transformation.

4 EXPERIMENTS

In this section, we validate the effectiveness of TATC for predicting AD and MCI with actigraphy data.

4.1 Experimental Setup

4.1.1 Baselines and Metrics. We use three methods on time series classification as baselines: (1) Dynamic Time Warping (DTW) [4], implemented as a sum of squared DTW distances in each dimension; (2) BOSS [23]. As BOSS works on univariate time series only, we train a base classifier on each univariate time series and build an ensemble of four base classifiers; and (3) SMTS [3] which ranked second in the gesture recognition competition organized by 2nd ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data[4]. For fair comparison with these baselines, personal particulars of subjects are excluded from TATC and only actigraphy time series is used.

[4]https://aaltd16.irisa.fr/challenge/

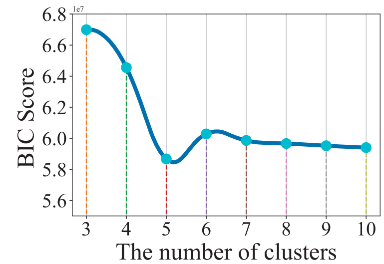


Figure 3: BIC score corresponding to different number of clusters

Table 4: Number of parameters in the TATC architecture

Component	**Input**	**Output**	**#Parameters**
CNN representation	24×16	12×2	64
max-pooling	12×2	6×2	0
TICC representation	24×5	24×5	0
max-pooling	24×5	6×5	0
TC representation	$6 \times (2 + 5)$	6×7	0
attention	6×7	6×1	44
time-aware representation	$6 \times (1 + 7)$	6×8	0
GRU	6×8	6×2	66
fully connected	6×2	2	26
logistic regression	2	1	3
			total: 203

Two binary classifiers are constructed: one on AD versus NC, and the other on MCI versus NC. To handle the imbalanced class distribution, we perform oversampling by SMOTE [6] on the AD and MCI samples in the training set.

We perform 5-fold cross validation and report the average results. In the medical domain, *sensitivity*, *specificity*, and Area Under the receiver operating characteristics Curve (*AUC*) are most commonly used metrics for evaluating the classification performance. In our problem, sensitivity measures the recall of positives (i.e., AD or MCI), and specificity measures the recall of negatives (i.e., NC).

4.1.2 Implementation details. We use minibatch based Adam [15] to minimize the binary cross-entropy loss. He-normal [13] is used as the initializer for CNN. The drop-out strategy is used in the full connected layer with a rate of 0.3.

We use Bayesian Information Criterion (BIC) to decide the optimal number of clusters in TICC, which is set to 5 as shown in Figure 3. According to BIC, we also discover the optimal penalty β to be 400. The number of temporal features α in CNN is set to 2. The number of time intervals n is set to 24. Detailed information of TATC's parameter size is listed in Table 4.

4.2 Results

The experimental results for predicting AD and MCI are listed in Table 5 and Table 6 respectively. For predicting AD, TATC achieves the best performance among 4 approaches, with a good balance

Table 5: Quantitative comparison of different classifiers to predict AD

Approach	Sensitivity	Specificity	AUC
DTW	90.3%	47.5%	68.9%
BOSS	38.7%	91.3%	76.1%
SMTS	45.2%	92.5%	84.5%
TATC	80.6%	86.3%	86.2%

Table 6: Quantitative comparison of different classifiers to predict MCI

Approach	Sensitivity	Specificity	AUC
DTW	70.0%	48.3%	59.1%
BOSS	5.7%	95.5%	58.8%
SMTS	5.0%	91.0%	58.5%
TATC	42.3%	81.3%	61.7%

between sensitivity and specificity. It shows great promise of being put into practice for early detection of AD. SMTS comes second, with a high specificity but a low sensitivity of 45.2%. DTW is biased towards AD, leading to a high sensitivity, but a low specificity and AUC.

For predicting MCI, TATC still achieves the best performance. We note that the classification performance is not as good as that for predicting AD. In the literature of MCI research, MCI is further categorized into stable MCI (sMCI) and progressive MCI (pMCI) [25]. Subjects who convert to AD within 36 months are classified as pMCI, and those who do not convert to AD are classified as sMCI. sMCI subjects are physically as active as NC subjects. We also observe in Figure 1 that the circadian activity of MCI and NC subjects is very close, making it hard to differentiate these two groups based on their physical activity. To improve the performance of detecting MCI, we plan to explore the possibility of incorporating other measurements besides actigraphy data.

We further evaluate the effectiveness of different components of TATC in predicting AD. In this group of experiments, personal features are included. Specifically we apply chi-squared test on the features listed in Table 1 for the purpose of feature selection, and select MASCH, MHDIAB, and GRIPAM as discriminative personal features. The classification performance is listed in Table 7. In *simple classifier*, only personal features are used, which achieves 65.6% in AUC. *cnn* shares the same architecture with TATC, but does not use TICC representation or time-aware attention. *cnn* achieves 79.7% in AUC. *ticc+cnn* uses both TICC and CNN representation, but lacks the time-aware attention mechanism. It achieves 83.6% in AUC. TATC achieves the best performance by including all components, demonstrating the effectiveness of the composite feature representation with TICC and CNN, the time-aware attention mechanism, as well as the personal features.

4.3 Hidden State Interpretation

In this subsection, we interpret the hidden states learned by TATC. As the optimal hidden state number is 5 according to BIC, 5 Markov

Table 7: Quantitative comparison of different components of TATC to predict AD

Approach	Sensitivity	Specificity	AUC
simple classifier	22.5%	92.5%	65.6%
cnn	64.5%	86.2%	79.7%
ticc+cnn	71.0%	81.3%	83.6%
TATC	73.5%	94.3%	88.2%

In *simple classifier*, only personal features are used. In *cnn*, there is no TICC or time-aware attention. In *ticc+cnn*, there is no time-aware attention.

Table 8: PageRank and mean values for five hidden states

State	Interpretation	Measure	X_{acc}	Y_{acc}	Z_{acc}	Lux
*1	**good sleep**	PageRank	0	0	0	0
		mean	0	0	0	0
*2	**sedentary activity**	PageRank	0.37	0.37	0.26	0
		mean	704	746	917	0
*3	**light activity**	PageRank	0.22	0.34	0.20	0.24
		mean	932	972	1240	23
*4	**moderate activity**	PageRank	0.31	0.30	0.25	0.14
		mean	1293	1281	1600	118
*5	**exercising**	PageRank	0	0	0	0
		mean	2563	2238	2270	628

random field (MRF) are generated, each corresponding to a hidden state. An MRF is a weighted undirected graph which consists of 4 vertices, and each vertex represents a variable from $I = [X_{acc}, Y_{acc}, Z_{acc}, Lux]$. If there is a partial correlation between two variables within a hidden state, there is an edge connecting them in the corresponding MRF. A large edge weight indicates that the two variables are heavily dependent on each other. Notice for each hidden state, its MRF will not change throughout time, thus the MRF can be treated as a unique signature for each hidden state. PageRank [21] is a commonly used graph analytic measure to quantify the relative importance of each vertex inside a graph. We apply PageRank to each of the 5 MRF to measure how influential a variable is inside a hidden state. If a variable has a very high PageRank value in an MRF, it means it has a strong capacity to influence other variables. In addition to PageRank, mean value of each variable reflects the intensity of body movement and ambient light. We list the calculated PageRank and mean values in Table 8.

We infer an interpretation for each hidden state as follows. For *1, as the PageRank and mean of *Lux* are both zero, we can infer that ambient light in this hidden state does not change along with body movement, meaning the circumstances do not change. Together with the fact that the mean of all variables is zero, we interpret *1 as *good sleep*. For *2, with the same observation on *Lux*, we infer the circumstances do not change. We also observe that *2 has the second lowest movement as measured by the mean values, and X_{acc}, Y_{acc} and Z_{acc} have a clear dependence of each other as reflected by PageRank. Thus we infer *2 as *sedentary activity* such as disturbed sleep. As for *5, all mean values reach the maximum indicating intensive body movement and high ambient light, whereas

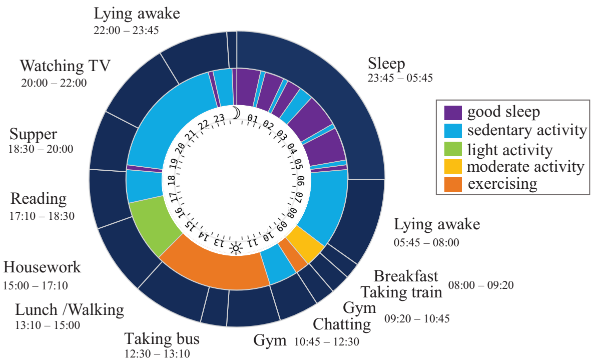
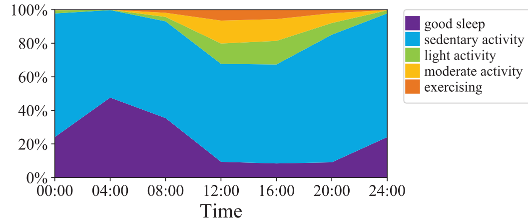
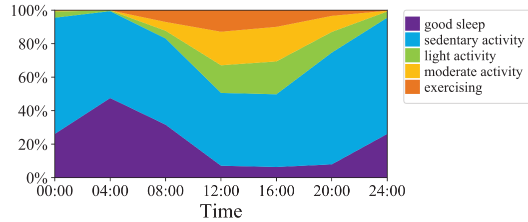


Figure 4: Self-report daily activities along with inferred hidden states. Inner circle represents hidden states inferred by TATC. Outer circle represents self-report activities.



(a) Distribution of hidden states for AD subjects



(b) Distribution of hidden states for NC subjects

Figure 5: Circadian activity comparison between AD and NC subjects

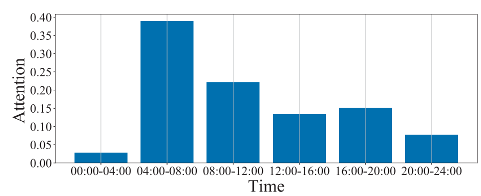


Figure 6: Average attention weight learned from NC and AD subjects

all PageRank values are 0 indicating no clear dependence between the variables. Thus we interpret it as *exercising*. For hidden states *3 and *4, their patterns are quite similar. The difference between these two is that acceleration and *Lux* are smaller for *3, which is interpreted as *light activity*, whereas *4 is interpreted as *moderate activity*. These can be regarded as a reasonable interpretation of the 5 hidden states learned by TATC.

We have designed a detailed self-report questionnaire for subjects to record their daily activities. We validate our interpretation of the hidden states by aligning their reported activities along with the hidden states in 24 hours. One NC case is exemplified in Figure 4. With reference to the self-report activities, we get the real-world interpretation of the hidden states. For example, *lying awake* and *watching TV* are both clustered into hidden state *2 *sedentary activity*, which makes sense because both of them involve low body movements and *Lux* remains stable. *Housework* belongs to hidden state *3 *light activity* since the subject needs to walk around inside the house and *Lux* may also change. *Gym* is related to vigorous body movement and is clustered into hidden state *5 *exercising*. One interesting finding is that the subject is unaware of his/her disturbed sleeping at night, while our method can capture light body movement during sleep and interpret those short periods as *sedentary activity*.

Based on the interpretation of the 5 hidden states, we compare the circadian activity of AD and NC subjects. Along 24 hours, we calculate the distribution of 5 hidden states for AD subjects (Figure 5a) and NC subjects (Figure 5b). We observe that AD subjects spend nearly 85% of their time on either *good sleep* or *sedentary activity*, versus 75% for NC subjects. Another interesting discovery is that the difference between AD and NC subjects is more obvious during 4:00am – 12:00noon. NC subjects spend 20% time on *exercising* while AD subjects spend only 9% in the same period. NC subjects spend 20% time on *moderate activity* during 8:00am – 12:00noon while AD subjects spend 13% only. This also proves that we should pay different attention to different time intervals. In the next subsection we give detailed analysis on the time-aware attention mechanism.

4.4 Interpretation of Attention Mechanism

The motivation to apply time-aware attention in TATC is that different time intervals in a day have different degree of importance to differentiate the three groups of subjects. To validate the effectiveness of the attention mechanism, we plot the average attention weight learned from the classification of AD versus NC subjects in a 24-hour time span in Figure 6. Indeed, we observe that different time intervals have different attention weights. The largest attention weight appears at 4:00 – 8:00am, which is the typical wake-up time for the elderly. This is consistent with our observation in Figure 1 that the biggest gap between AD and NC appears during 4:00 – 8:00am. It also justifies our interpretation of hidden states in Figure 5 where we notice NC group exercises more and has less sedentary activity than AD group. We also find that the time intervals 0:00 – 4:00am and 20:00 – 24:00pm have the lowest attention weights, as this is sleeping time with little body movement. It is consistent with our discovery in the hidden state interpretation that the two groups are similar at these two intervals.

5 LESSONS AND INSIGHTS

This paper presents our experiences of applying deep learning techniques to predict Alzheimer's Disease based on the collected actigraphy data, and realize our ideas in a solution called TATC. We summarize our lessons and insights gained from this project.

5.1 Data Collection

Data collection is very challenging with respect to the aged cohort. As the average age of our subjects is above 80, engaging them in repeated clinical assessments is not an easy task even though we have standardized the procedures. For data collection from actigraphy devices, the subjects were educated on the proper usage of GT3X in the interview. But many subjects still forgot to put on the device after bathing or swimming, or forgot to wear it for various reasons. In addition, we have designed a detailed self-report questionnaire for subjects to record their daily activities, but the returned pieces are of low quality. In the end we could use only 20 self-report questionnaires for validation. An important lesson is that our data collection procedure should be made simple and bring little disturbance to subjects' daily life, so that more valid and valuable data can be recorded. In future work, we shall examine the inclusion of other devices such as GT9X, which can be worn during bathing and swimming.

5.2 Practical Value to Clinical Diagnosis

Traditional cognitive status diagnosis involves lots of clinical assessments and clinic visits (see Section 2), which heavily rely on the domain knowledge of doctors. These clinical assessments bring much burden to patients and may deteriorate their cognitive status. For those who have been diagnosed MCI, the progression from MCI to AD is unpredictable and there may be years or even dozens of years before AD is developed. In such a long period, frequent clinic visits become infeasible especially for those who refuse or have difficulties in doing so, and patients may lose the best opportunity of timely diagnosis and treatment.

Our proposed TATC method provides an automatic, low-cost solution for continuously monitoring the change of physical activity of subjects in daily living environment. The actigraphy data is sent to the server on a daily basis, upon which the classification model can be applied on the incoming data for prediction. Once potential risk of AD is identified, doctors will be alerted immediately. Then they can arrange clinic visits for subjects for further diagnosis and treatment. We believe the future deployment of TATC can benefit both doctors and patients in early detection of potential AD risk, particularly for those who have been diagnosed MCI, as monitoring the trajectory of changes is of great importance to them. This is the most important contribution of this study to the medical domain.

6 RELATED WORK

This work is related to time series representation, attention-based neural network and healthcare interpretations.

Time series representation techniques can be generally classified into two groups. The first group is non-transformed techniques. Representative works include SAX [16], Shapelets [26], TSF [11], and DTW [4]. This group of techniques works in the original time domain and represents time series as a common shape (e.g., DTW), or divides time series into several segments and represents them accordingly (e.g., SAX and TSF). In contrast to these non-transformed techniques, the second group transforms time series into another space. Representative works include SVD [5], BOSS [23], SMTS [3], MFCC [18], and CNN [20]. This group, especially CNN-based representation, has become quite popular in recent years for its good performance in classification.

Attention-based neural network has been successfully used especially in machine translation [2] and sentence summarization [22]. In healthcare domain the effectiveness of this mechanism has also been demonstrated by RETAIN [7] and Dipole [19] on Electronic Health Records (EHR). To interpret how attention works, Dipole exemplifies several patient visits and analyses why some visits are more important than others.

Understanding the hidden states behind the observed time series is vital for healthcare applications. In [20], a deep neural network is trained to connect the observed data and the hidden activity. In [12] each hidden state is represented by a Markov random field, and graph analytics such as betweenness centrality is used to achieve reasonable interpretation.

7 CONCLUSION

We design a multivariate time series classification method TATC for predicting AD with actigraphy data. TATC takes a neural deep learning approach with time-aware attention for modeling the effect of circadian rhythm. It achieves promising prediction performance in terms of sensitivity, specificity and AUC. It also generates meaningful interpretation of daily behavior pattern of subjects. TATC shows great potential and practical value in continuous monitoring of physical activity of subjects and in early detection of AD risk. For future work, we plan to explore the possibility of incorporating other measurements for predicting MCI.

ACKNOWLEDGMENTS

The authors would like to thank Xixin Wu and Siyuan Zhang for discussions and suggestions. This work is supported by the Research Grants Council Earmarked Grant CUHK4101/02M, the National Institute of Health R01 Grant AR049439-01A1 and the CUHK Stanley Ho Big Data Decision Analytics Research Centre.

REFERENCES

[1] A. Alberdi, A. Aztiria, and A. Basarab. 2016. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine* 71 (2016), 1–29.
[2] D. Bahdanau, K. Cho, and Y. Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[3] M. G. Baydogan and G. Runger. 2015. Learning a symbolic representation for multivariate time series classification. *Data Mining and Knowledge Discovery* 29, 2 (2015), 400–422.
[4] D. J. Berndt and J. Clifford. 1994. Using dynamic time warping to find patterns in time series. In *KDD Workshop*. 359–370.
[5] J. A. Cadzow, B. Baseghi, and T. Hsu. 1983. Singular-value decomposition approach to time series modelling. *IEE Proceedings F (Communications, Radar and Signal Processing)* 130, 3 (1983), 202–210.
[6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16 (2002), 321–357.
[7] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. 2016. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In *NIPS*. 3504–3512.
[8] L. Choi, Z. Liu, C. E. Matthews, and M. S. Buchowski. 2011. Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and Science in Sports and Exercise* 43, 2 (2011), 357–364.
[9] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Deep Learning and Representation Learning Workshop*.

[10] A. N. Coogan, B. Schutová, S. Husung, K. Furczyk, B. T. Baune, P. Kropp, F. Häßler, and J. Thome. 2013. The circadian system in Alzheimer's Disease: Disturbances, mechanisms, and opportunities. *Biological Psychiatry* 74, 5 (2013), 333–339.
[11] H. Deng, G. Runger, E. Tuv, and M. Vladimir. 2013. A time series forest for classification and feature extraction. *Information Sciences* 239 (2013), 142–153.
[12] D. Hallac, S. Vare, S. P. Boyd, and J. Leskovec. 2017. Toeplitz inverse covariance-based clustering of multivariate time series data. In *KDD*. 215–223.
[13] K. He, X. Zhang, S. Ren, and J. Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*. 1026–1034.
[14] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen. 2016. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of Aging* 46 (2016), 180–191.
[15] D. Kingma and J. Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980
[16] J. Lin, E. Keogh, L. Wei, and S. Lonardi. 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* 15, 2 (2007), 107–144.
[17] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. C. Mansfield, C. Cooper, N. Fox, L. N. Gitlin, R. Howard, H. C. Kales, E. B. Larson, K. Ritchie, K. Rockwood, E. L Sampson, Q. Samus, L. S Schneider, G. Selbæk, L. Teri, and N. Mukadam. 2017. Dementia prevention, intervention, and care. *The Lancet* 390, 10113 (2017), 2673–2734.
[18] B. Logan. 2000. Mel frequency cepstral coefficients for music modeling. In *ISMIR*.
[19] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao. 2017. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *KDD*. 1903–1911.
[20] F. J. Ordóñez and D. Roggen. 2016. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.
[21] L. Page, S. Brin, R. Motwani, and T. Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.
[22] A. M. Rush, S. Chopra, and J. Weston. 2015. A Neural Attention Model for Abstractive Sentence Summarization. *CoRR* abs/1509.00685 (2015). arXiv:1509.00685
[23] P. Schäfer. 2015. The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery* 29, 6 (2015), 1505–1530.
[24] F. Sofi, D. Valecchi, D. Bacci, R. Abbate, G. F. Gensini, A. Casini, and C. Macchi. 2011. Physical activity and risk of cognitive decline: a meta-analysis of prospective studies. *Journal of Internal Medicine* 269, 1 (2011), 107–117.
[25] C. Y. Wee, P. T. Yap, and D. Shen. 2013. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping* 34, 12 (2013), 3411–3425.
[26] L. Ye and E. Keogh. 2011. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data Mining and Knowledge Discovery* 22, 1-2 (2011), 149–182.
[27] P. Y. Yeung, L. L. Wong, C. C. Chan, J. L. Leung, and C. Y. Yung. 2014. A validation study of the Hong Kong version of Montreal Cognitive Assessment (HK-MoCA) in Chinese older adults in Hong Kong. *Hong Kong Medical Journal* 20, 6 (2014), 504–510.
[28] J. M. Zeitzer, T. Blackwell, A. R. Hoffman, S. Cummings, S. Ancoli-Israel, and K. Stone. 2018. Daily patterns of accelerometer activity predict changes in sleep, cognition, and mortality in older men. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 73, 5 (2018), 682–687.
[29] J. M. Zeitzer, R. David, L. Friedman, E. Mulin, R. Garcia, J. Wang, J. A. Yesavage, P. H. Robert, and W. Shannon. 2013. Phenotyping apathy in individuals with Alzheimer Disease using functional principal component analysis. *The American Journal of Geriatric Psychiatry* 21, 4 (2013), 391–397.