Supplementary Material

Altered Motor Activity Patterns within 10-Minute Timescale Predict Incident Clinical Alzheimer's Disease

SUPPLEMENTARY METHOD

Details for detrended fluctuation analysis (DFA)

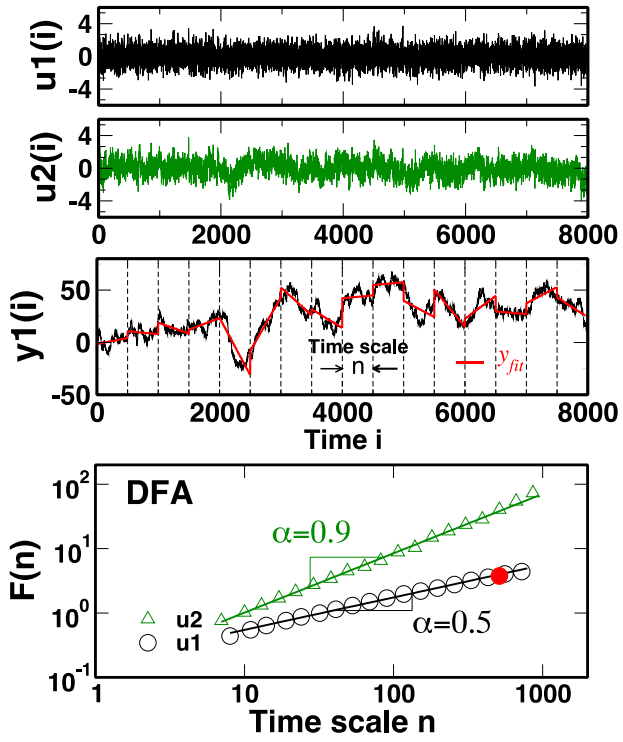
For a signal with mathematically 'ideal' fractal temporal correlations at different timescales, F(n) follows a power-law form: $F(n) = n^{\alpha}$ where the scaling exponent α quantifies temporal correlations. There are 10 sampling points for n from 2^1 to 2^2 minutes, 2^2 to 2^3 minutes, etc. $\alpha = 0.5$ suggests no correlation in the fluctuations ("white noise"); $\alpha > 0.5$ suggests positive correlations, where large values are more likely to be followed by large values (and vice versa); $\alpha < 0.5$ suggests negative correlations, where large values are more likely to be followed by small values (and vice versa). $\alpha = 1.0$ corresponds to the most complex fluctuation patterns in physical systems, and α values close to 1 have been observed in many physiological outputs under healthy conditions (note, 1 is not the upper limit of α). However, for real physiological signals even under healthy conditions, fractal patterns persist only within certain limited timescales such that the power-law form of F(n) and the value of α may be different in different regions. The behavior of F(n) might become more complicated in physiological outputs under pathological conditions. See Supplementary Figure 1 for a graphical illustration.

Model development with nested 5-fold cross validation

We used nested 5-fold cross validation (CV) to fit all models, where the inner CV loop was designed to determine the best hyperparameters and the outer CV loops was used to estimate unbiased out-of-sample performance. To achieve this, the participants were first randomly split into 5 folds (outer CV loop). Each of these outer folds was used as the testing set and other four folds were used as the corresponding training set. Then, in each training set, 5 folds (inner CV loop) were created, and each of these inner folds was used as the validation set and the other four inner folds as the corresponding inner training set. Every time before fitting, the features with continuous value were standardized so that they had zero mean and unit standard deviation. A model was fitted on each inner training set with specific hyperparameters. Within each outer CV

loop, the combination of hyperparameters that achieved the highest averaged performance across validation folds was selected as the best hyperparameters; a model was fitted with those best hyperparameters and applied to the testing fold in this outer CV loop to obtain a testing performance; and the final out-of-sample performance was based on the average of the 5 testing folds in the outer CV loop.

The hyperparameters for DeepSurv included the number of hidden nodes in the hidden layer from either 20 or 50, and dropout rate from either 0.1 or 0.2. The hyperparameters for the Cox model included the relative strength of encouraging coefficient sparsity from a list of [0.1,0.2,…,0.9]; overall penalty from a list of [0.001, 0.01, 0.1, 1, 10, 100, 1000]. The hyperparameters for the RSF included the number of trees selected from a list of [20, 50, 100]; the max depth selected from [3, 5, 10]; and the minimum number of samples required to be at a leaf node from either 30 or 50.



Supplementary Figure 1. An illustration of detrended fluctuation analysis (DFA).

Step 1: Given a time series $u(i)$ with length N, where i = 1, 2, …, N. We first remove its mean to get $u(i) - \bar{u}$, and then accumulate its values to get another time series $y(i) = \sum_{j=1}^{i} u(j) - \bar{u}$.

Step 2: $y(i)$ is divided into time windows of length n samples.

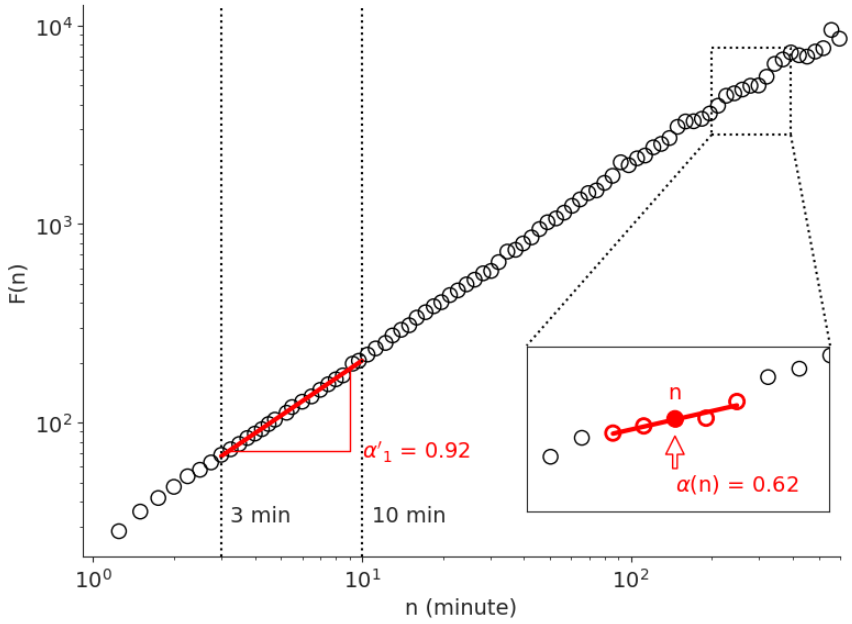
Step 3: A local least squares is calculated by minimizing the mean square error within each time window. Let $y_{fit}(i)$ denote the resulting piecewise straight-line fits.

Step 4: The detrended fluctuation is computed as the root mean square error: $F(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y(i) - y_{fit}(i)\right)^2}$.

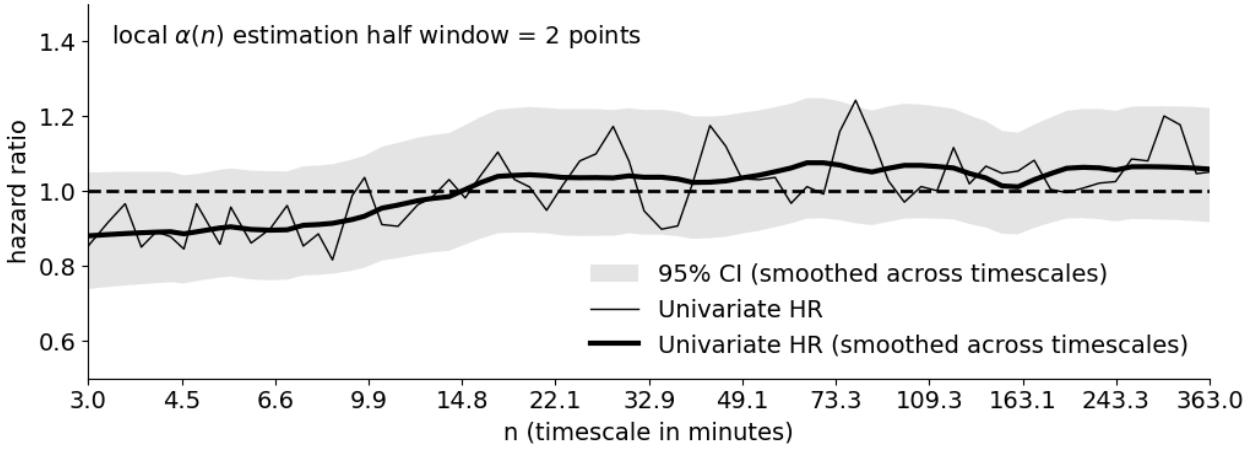
Step 5: Repeat steps 2,3,4 for a range of different window sizes n.

Step 6: The DFA curve is the function of n vs. F(n), where $F(n) = n^{\alpha}$, and α is the DFA scaling exponent.

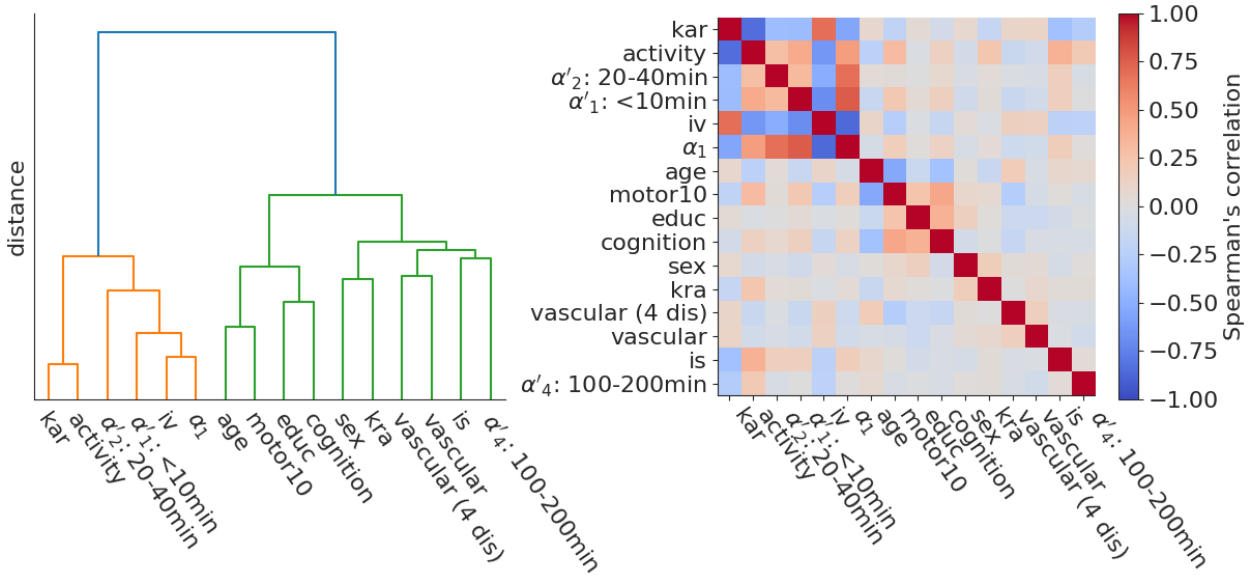
Note: As shown in the bottom panel, the black actigraphy (u1) has α = 0.5 and the green actigraphy (u2) has α = 0.9.



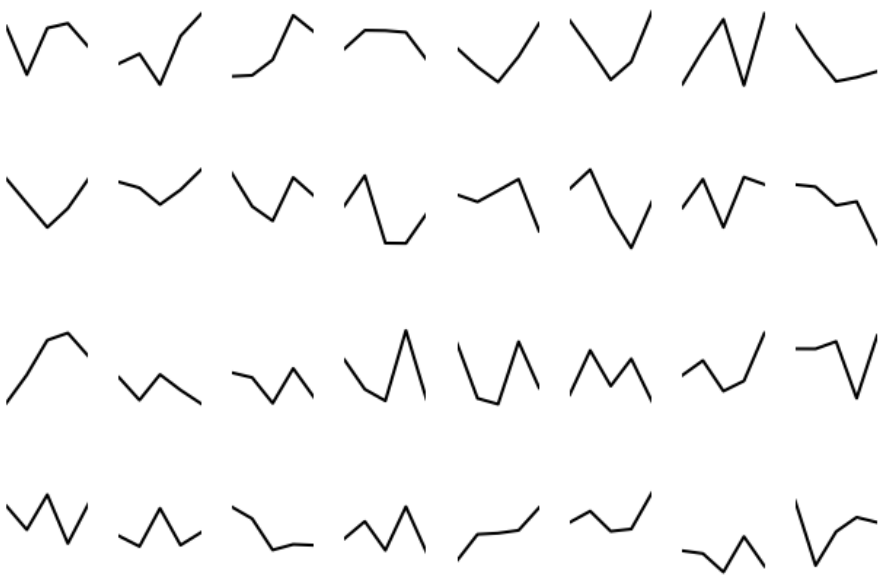
Supplementary Figure 2. α'_1 is the scaling exponent for timescale <10 min, which is the slope of the detrended fluctuation analysis (DFA) curve within the 3 to 10 min range (<3 min is unstable), as indicated by the two vertical dotted lines. Note that we first obtained the DFA curve using the entire actigraphy recording, and then took the <10 min range in DFA. There is no segmentation of the actigraphy into short segments. The local scaling exponents α(n) is illustrated on the right side. For a particular n, we take 2 additional points on the left and 2 on the right, and the slope of these 5 points is α(n).



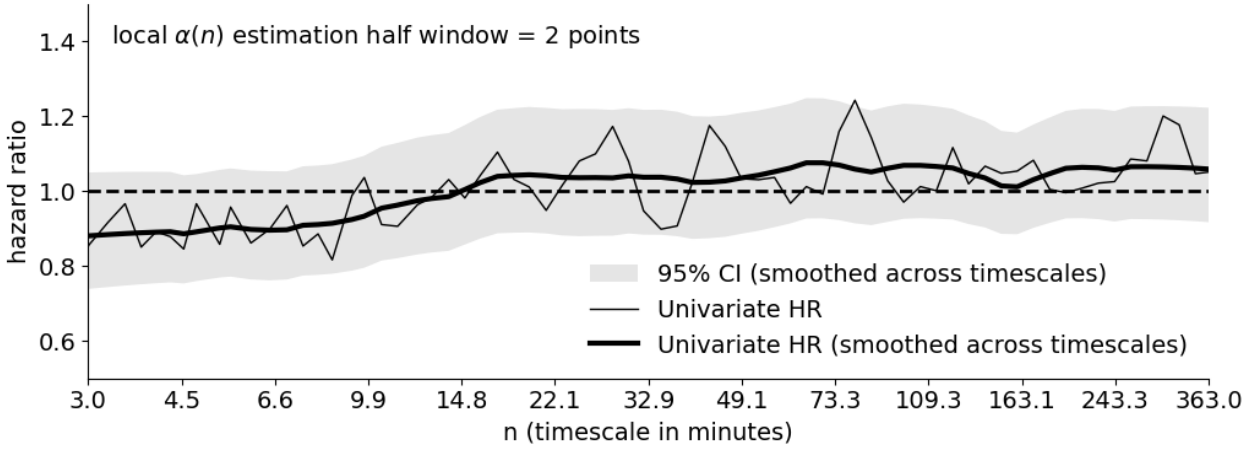
Supplementary Figure 3. Univariate hazard ratio of local DFA scaling exponents α(n) as a function of the estimation half window used to estimate the local α(n). For example, when the half window is 2 points, for each n, the local α(n) are estimated using two points at both sides plus one middle point. As the half window increases, the magnitude of univariate hazard ratio increases; the noise in univariate hazard ratio decreases; and the timescale less than around 7 minutes becomes significant, but not for the longer timescale.



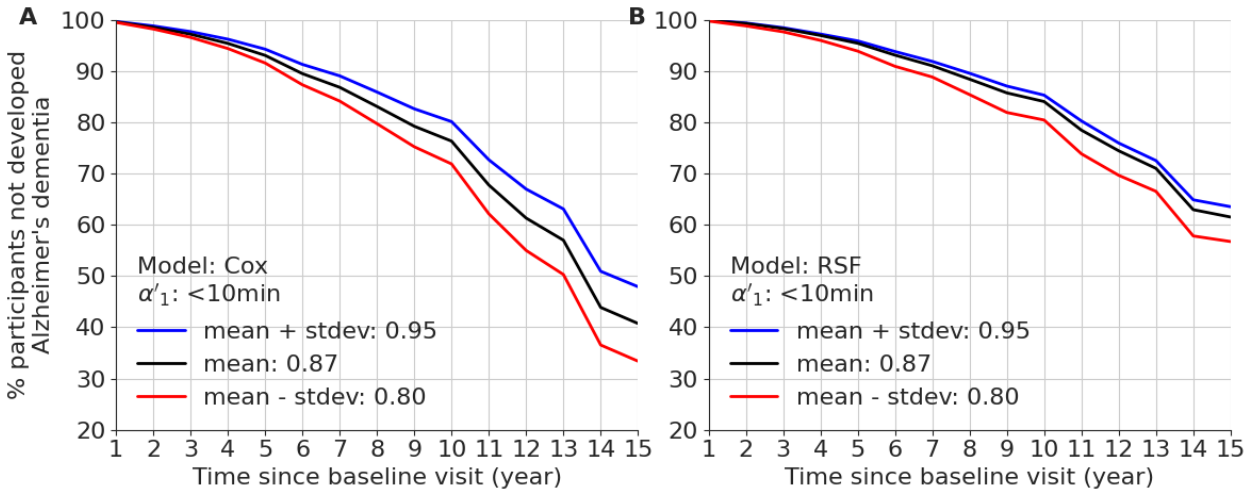
Supplementary Figure 4. A) Dendrogram of the hierarchical clustering result of the features. The distance on the y-axis is based on absolute Spearman's correlation, but they only make relative sense. B) The correlation matrix showing the pairwise Spearman's correlation, where red means positive and blue means negative. They are rearranged so that the clustered features are close to each other.



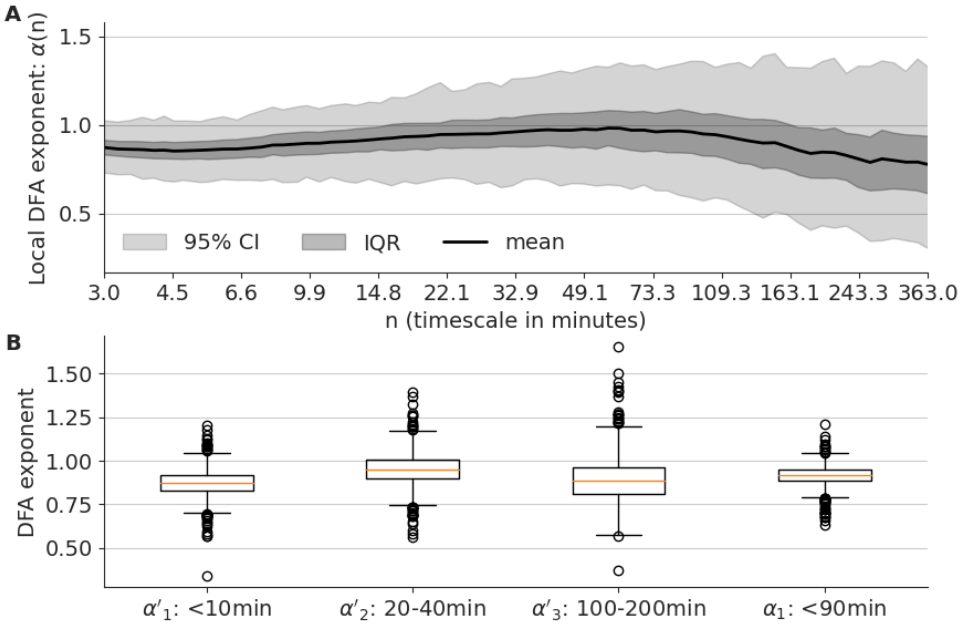
Supplementary Figure 5. The kernels learned by the first layer of the convolutional neural network in DeepSurv.



Supplementary Figure 6. Univariate hazard ratio of local DFA scaling exponents α(n) as a function of timescale n. Each value is obtained by fitting a Cox proportional hazard model with α(n) and covariates as input, therefore it is not affected by collinearity among the α(n) at different timescales.



Supplementary Figure 7. The predicted survival curve for different values of α'_1. A) The Cox model. B) The RSF model. In both models, results were obtained for the three values of α'_1: mean plus one standard deviation (blue, good group), mean (black, average group), and mean minus standard deviation (red, bad group). The other variables were selected to be mean across the cohort. Survival curve was a function of time t vs. the probability of not having the onset of clinical AD (dementia) up to time t.



Supplementary Figure 8. A) The local scaling exponent as a function of timescale. The solid line is for cohort mean; the dark grey for 95% percentile confidence interval (95% CI); and light grey for the interquartile range (IQR). The cohort-level mean of the local scaling exponent α(n) reached its maximum value around n = 60 min and kept decreasing at larger time scales. In addition, the inter-individual variation of α(n) increased as time scale increased. Although temporal correlations had a dramatic change at timescale of 90 min, we found the dependence of temporal activity correlations on timescale was much more complex in the elderly. For instance, even within the region of <90 min, temporal correlations were not stable. B) The mean scaling exponents in four specific timescale regions: α'_1 <10 min; α'_2 20-40 min; α'_3 100-200 minutes; and α_1 <90 min. Note the first three exponents were newly defined based on the DeepSurv results, and the last one was commonly used in previous studies. The upper, middle, and lower bounds of the boxes represent 25% (Q1), median, 75% percentile (Q3) of the distribution. The upper error bar represents Q3 + 1.5 × (Q3-Q1); and the lower error bar represents Q1 - 1.5 × (Q3-Q1). The circles represent outliers that are higher than the upper error bar or lower than the lower error bar.

Note that longer education was associated with higher hazard ratio in the Cox model (Supplementary Table 1), but the effect appeared to be not so important in the RSF model (Supplementary Table 2). The seeming discrepancy was likely due to the collinearity between the education duration and cognition, which affected the results of the Cox model more than its influence on the RSF.

Supplementary Table 1. Adjusted hazard ratios of DFA exponents and other significant features in Cox model (*p<0.05)

Category	**Feature**	**Adjusted hazard ratio 15 years after baseline when increased by 1 stdev**
Associate with higher hazard	Age at baseline visit	1.60 [1.34 -- 1.91]*
	Years of education	1.24 [1.09 -- 1.42]*
Associate with lower hazard	Cognition at baseline visit	0.35 [0.31 -- 0.41]*
	Motor function	0.69 [0.58 -- 0.82]*
	α'_1: < 10 min	0.80 [0.67 -- 0.96]*
Insignificant	α'_3: 100 to 200 min	1.09 [0.95 -- 1.25]
	Interdaily stability [is]	1.06 [0.93 -- 1.22]
	Transition probability from activity to resting [kar]	1.05 [0.86 -- 1.28]
	Intradaily variability [iv]	1.05 [0.85 -- 1.30]
	Daily activity level	1.05 [0.87 -- 1.26]
	α'_2: 20 to 40 min	1.03 [0.88 -- 1.22]
	Male sex	0.95 [0.83 -- 1.08]
	Transition probability from resting to activity [kra]	0.94 [0.85 -- 1.03]
	Vascular disease burden	0.91 [0.81 -- 1.02]
	Vascular disease risk factor	0.90 [0.80 -- 1.03]

Error! Reference source not found.**Supplementary Table 2.** Feature importance from RSF ranked in descending order

Feature	**RSF's feature importance (reduction in Gini impurity)**	**Direction of impact on hazard**
Cognition at baseline visit	29.3 [19.7 – 38.9]	-
Age at baseline visit	20.6 [12.7 – 28.4]	+
Motor function	19.5 [12.3 – 26.7]	-
α'_1: < 10 min	10.4 [7.4 – 13.5]	-
Intradaily variability [iv]	6.9 [4.5 – 9.3]	+
Daily activity level	5.0 [2.9 – 7.1]	-
Interdaily stability [is]	2.7 [0.7 – 4.7]	-
α'_3: 100 to 200 min	2.1 [0.9 – 3.3]	+
Transition probability from activity to resting [kar]	2.0 [0.2 – 3.7]	-
Years of education	1.3 [0.4 – 2.1]	0
Male sex	1.1 [0.8 – 1.3]	0
Vascular disease burden	0.6 [0.1 – 1.2]	+
Vascular disease risk factor	0.6 [0.1 – 1.0]	-
Transition probability from resting to activity [kra]	0.45 [-0.2 – 1.1]	-
α'_2: 20 to 40 min	0.1 [-0.9 – 1.1]	+