

Article

Measures of Morphological Complexity of Gray Matter on Magnetic Resonance Imaging for Control Age Grouping

Tuan D. Pham ^{1,*}, Taishi Abe ², Ryuichi Oka ² and Yung-Fu Chen ^{3,4}

Received: 8 October 2015 ; Accepted: 30 November 2015 ; Published: 9 December 2015

Academic Editor: Raúl Alcaraz Martínez

¹ Department of Biomedical Engineering, Linköping University, Linköping 581 83, Sweden

² Graduate School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu 965-8580, Japan; d8161106@u-aizu.ac.jp (T.A.); oka@u-aizu.ac.jp (R.O.)

³ Department of Dental Technology and Materials Science, Central Taiwan University of Science and Technology, Taichung 40601, Taiwan; yfchen@ctust.edu.tw

⁴ Department of Health Services Administration, China Medical University, Taichung 40402, Taiwan

* Correspondence: tuan.pham@liu.se; Tel.: +46-13-28-67-78

Abstract: Current brain-age prediction methods using magnetic resonance imaging (MRI) attempt to estimate the physiological brain age via some kind of machine learning of chronological brain age data to perform the classification task. Such a predictive approach imposes greater risk of either over-estimate or under-estimate, mainly due to limited training data. A new conceptual framework for more reliable MRI-based brain-age prediction is by systematic brain-age grouping via the implementation of the phylogenetic tree reconstruction and measures of information complexity. Experimental results carried out on a public MRI database suggest the feasibility of the proposed concept.

Keywords: brain-age grouping; brain-age prediction; magnetic resonance imaging; phylogenetic tree reconstruction; measures of complexity; chaos; nonlinear dynamics; neurodegeneration

1. Introduction

The ability of computer methods that can predict healthy chronological age of the brain based on radiological images, such as magnetic resonance imaging (MRI), has recently led to a new research direction in computational neuroscience [1–7]. This new type of study is important because it holds promise of being able to train computers to identify neurodegenerative disorders at an early onset, where image samples collected for brain diseases are limited for clinical inference [8,9]. In other words, the brain imaging data of healthy or control participants can be utilized as models for machine learning, which is then applied to estimate the brain age of a subject under study based on the brain imaging data of the subject. If the predicted brain age is older than its chronological age, then there is some evidence of its accelerated aging that indicates abnormal cognitive impairment [10,11], or traumatic brain injury [12].

In fact, a normal aging process is known as the progression of gradually accumulated pathologies associated with physical decline, cognitive impairment, and brain-volume loss [13,14]. It is well perceived that the degenerations of the brain result in fast aging processes, thus causing brain disorders; and healthy brain imaging data can offer useful medical information for early detection and intervention of these neuropathological and cognitive changes using computational methods [10,15,16]. For example, Alzheimer’s disease (AD), which is not a normal part of brain aging and known as the most common form of dementia that causes problems with memory, thinking and

behavior. AD symptoms usually progress slowly and deteriorate over years to the stage at which daily tasks can be severely hindered [17]. Because AD shares many aspects of abnormal brain aging, by applying pattern recognition methods, structural MRI-based features have been discovered as promising biomarkers to identify the early setting of mild cognitive impairment to AD based on the matching of similarity between constructed computational models of healthy and pathological brain aging patterns [11,16].

Most computational methods developed for the detection of early stages of neurodegenerative disease are based on the notion of brain age estimation using the MRI of the brain. The mean age prediction accuracy using hidden Markov models was within the mean absolute error of 2.41 years as reported in [3], 4.98 years using relevance vector machine for regression (RVR) [2], 6.3 years using quantitative brain water maps [18], and within the root mean squared error of 6.5 years using the RVR [19]. Regarding the use of brain age estimate for the diagnosis of traumatic brain injury, the suggestion of brain injury was reported when the brain age is estimated to be older than its chronological age with mean errors of 4.66 years and 5.97 years for gray matter (GM) and white matter (WM), respectively [7]. While the application of computational models for estimating brain ages using structural MRI to quantify the atrophy of the brain is promising for the early diagnosis of brain diseases, different computational methods provide different results with different respective errors, making it difficult to assess the reliability of the age estimation. Furthermore, although the prediction accuracy can be achieved within an average error of about 5 years, the standard deviation can be as large as 10 years [7], setting the physiological brain age inference with large uncertainty for medical decision making and treatment. Another issue for brain age estimation is that it is difficult to obtain sufficient MRI data of control subjects, which match the age of each individual patient and other confounding factors that correlate with both dependent and independent variables [4]. Thus, there is a need to develop some alternative computational methodology for a more reliable prediction of the normal or control brain age using structural MRI data. This is the motivation of this paper that aims to explore the concept of brain age grouping in order to provide a more robust way of physiological inference from control chronological brain age models for more reliable diagnoses of brain disorders.

A methodologically sound procedure for brain age grouping is by way of the concept of phylogeny developed in genomics [20]. The hypothesis of phylogeny in molecular biology is that if genomes involve by mutations, then the quantitative difference in a nucleotide sequence between a genome pair proportionally indicates how recently those two genomes shared a common ancestor. Two genomes that diverged in the recent past would be expected to have fewer differences than a pair of genomes whose common ancestor is more ancient. The implication is that by comparing three or more genomes, it is possible to infer not just only the similarities but the evolutionary relationships between them [20]. Likewise, chronological brain ages are altered over time and can be divided into separate groups or hybridize together due the complexity of the brain structure. This is analogous to the evolutionary branching process in molecular phylogenetics, which may be depicted with a phylogenetic tree, and the place of each of the various groups of age on the tree is based on a hypothesis about the brain structure in which cerebral atrophy occurred. This hypothesis is also found in comparative linguistics, which is a branch of historical linguistics, where similar concepts are used with respect to relationships between languages [21]. Thus, by partitioning brain ages into groups instead of estimating a single chronological age of the brain, we can better reduce the prediction uncertainty, and be able to validate the result by examining the relationships within and between the models built for the brain groups.

To enable the performance of “phylogenetic” analysis for brain age grouping, we need to extract informative features of the brain on MRI so that the dissimilarity or distance matrix between brain groups can be determined. It is known that GM is an important evidence for the assessment of cerebral atrophy revealed by MRI scans of the brain, because GM regions are found to be specifically susceptible to the cognitive decline and AD process [22]. Furthermore, the importance of the

measurement of structures and changes of the brain during its development, aging, learning, disease and evolution is recognized as an emerging field of neuro-informatics known as brain morphometry, which aims to quantify anatomical features of the brain in terms of shape, mass, and volume using typically MRI data to gain insight into the pattern of the brain and its longitudinal extent across individuals or different biological species [23].

A feasible way for feature extraction is to transform the GM morphology from MRI scans into time series, which was found useful for a study of the complexity of brain folding, cortical surface structure in AD and aging [16]. In fact, the human brain is relatively larger and more wrinkled than other species. Brain wrinkles increase its surface, indicating more neurons, therefore more folds of the brain means more surface area and more neurons. Declines in brain function in the elderly are sometimes due to widening of the folds (sulci), resulting in fewer neurons, in comparison with young people [24]. The brain outer layer (cerebral cortex) is a central region in the mammalian brain that monitors complex cognitive behaviors [25]. Recent scientific evidence has suggested that the size and extent of the folded gray matter of the brain are important factors that influence cognitive abilities and sensorimotor skills [26]. Furthermore, the conversion of images into sequences for applications of time-series analysis tools has been utilized for solving several problems in image data mining [27–29]. Because time-series data of the GM morphology are inherently complex, chaos and nonlinear dynamics analyses of these time-series data are therefore suitable mathematical techniques for extracting their informative statistical properties. Moreover, chaos and nonlinear dynamics have been increasingly reported as effective computational methods for analyzing complex data in medicine and biology [30–36]. We therefore explore several methods of chaos and nonlinear dynamics to extract features from time-series data of the GM spatial structure of the brain on MRI. Such methods include approximate entropy (ApEn) [37], sample entropy (SampEn) [38], regularity dimension (RD) [39], recurrence plots (RP) [40], and the largest Lyapunov exponent (LLE) [41,42]. Because dynamic-time warping (DTW) is known as a popular approach for pattern matching in terms of time series [43,44], it is also applied in this study to establish the distance matrix for the tree reconstruction for the proposed brain age grouping, which can be compared with those obtained from the methods of chaos and nonlinear dynamics.

The rest of this paper is organized as follows. Section 2 presents the methods of chaos, nonlinear dynamics, dynamic-time warping, and phylogenetic tree reconstruction, which are implemented in this study. Section 3 describes the database and the preprocessing of the MRI of the brain. Section 4 consists of experimental tests for brain age grouping obtained from several methods, discussion and comparison of the results. Finally, Section 5 is the conclusion of the research, including the summary of new findings, limitation, and issues suggested for future work.

2. Methods

2.1. Approximate Entropy and Sample Entropy

The formulation of the approximate entropy (ApEn) [37] has its roots in the correlation dimension [45] and Kolmogorov-Sinai (K-S) entropy [46]. However, the correlation dimension and K-S entropy provide procedures for quantifying chaos, while ApEn does not. ApEn only measures how complex or predictable a dynamical system can be.

To assess the dynamical behavior of a time series $\mathbf{u} = (u_1, u_2, \dots, u_N)$, its dynamics is first reconstructed as a set of vectors of an embedding dimension m : $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N-m+1})$, where $\mathbf{x}_i = (u_i, u_{i+1}, \dots, u_{i+m-1})$, $i = 1, 2, \dots, N - m + 1$. Next, let r be a positive constant that indicates a tolerance for determining the similarity between a pair of the reconstructed vectors. The probability of reconstructed vector \mathbf{x}_i being similar to \mathbf{x}_j , $j = N - m + 1$, is expressed as

$$C_i(m, r) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} \theta(d(\mathbf{x}_i, \mathbf{x}_j)), \quad (1)$$

where $\theta(d(\mathbf{x}_i, \mathbf{x}_j))$ is the Heaviside step function defined as

$$\theta(d(\mathbf{x}_i, \mathbf{x}_j)) = \begin{cases} 1 & : d(\mathbf{x}_i, \mathbf{x}_j) \leq r \\ 0 & : d(\mathbf{x}_i, \mathbf{x}_j) > r \end{cases} \quad (2)$$

The distance measure $d(\mathbf{x}_i, \mathbf{x}_j)$ can be defined as

$$d(\mathbf{x}_i, \mathbf{x}_j) = \max_k (|u_{i+k-1} - u_{j+k-1}|), k = 1, 2, \dots, m. \quad (3)$$

Then, the probability of all the reconstructed vectors being similar to one another can be computed from $C_i(m, r)$ by:

$$C(m, r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log(C_i(m, r)). \quad (4)$$

Finally, with fixed m and r , ApEn is defined as a family of formulas as follows:

$$ApEn(m, r) = \lim_{N \rightarrow \infty} (C(m, r) - C(m + 1, r)). \quad (5)$$

Given m, r and N , ApEn is defined as a family of statistics:

$$ApEn(m, r, N) = C(m, r) - C(m + 1, r). \quad (6)$$

From the above mathematical expressions, $C(m + 1, r)$ will always have a value smaller or equal to $C(m, r)$. Therefore, $ApEn(m, r, N) \geq 0$. It can be interpreted that ApEn is the logarithmic likelihood of the reconstructed vectors that are close, and still remain close on the next incremental comparisons. A smaller value of ApEn indicates more self-similarity in data set. The study in [37] suggested that for $m = 2$ and $N = 1000$, values of r ranging from 0.1σ to 0.2σ , where σ is the standard deviation of the time series \mathbf{u} , would produce reasonable statistical validity of $ApEn(m, r, N)$.

Modifying ApEn, sample entropy (SampEn) [38] does not include self-similar patterns as ApEn does, introducing bias by including self-matching, particularly for a finite value of N . Thus, for given m, r and N , SampEn computes the probability of a reconstructed vector being similar to other reconstructed vectors as

$$A_i(m, r) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m} \theta(d(\mathbf{x}_i, \mathbf{x}_j)), i \neq j. \quad (7)$$

Then the probability that two time series match for a dimension m within a tolerance r is calculated by considering only the first $N - m$ reconstructed vectors of dimension m :

$$A(m, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i(m, r). \quad (8)$$

A family of formula for *SampEn* is defined as

$$SampEn(m, r) = \lim_{N \rightarrow \infty} \left(-\log \left(\frac{A(m + 1, r)}{A(m, r)} \right) \right), \quad (9)$$

in which the ratio $A(m + 1, r) / A(m, r)$ is interpreted as the conditional probability that two time series within a tolerance r for dimension m remain within r of each other for the next larger dimension [38].

A family of statistics for *SampEn* is given by

$$SampEn(m, r, N) = -\log \left(\frac{A(m + 1, r)}{A(m, r)} \right). \quad (10)$$

2.2. Regularity Dimension

The regularity dimension (RD) [39] has been introduced to overcome the difficulty in selecting the two critical parameters m and r of either ApEn or SampEn. Because SampEn is an improved version of ApEn, only the former is discussed in this context. The RD is formulated based on the definition of the power law and SampEn using the general rule of the exponent dimension defined as [47]

$$\Delta \propto s^{-D} \tag{11}$$

where Δ , α , s , and D stand for the number of increments, proportional to, scale size, and exponent dimension, respectively.

Rearrangement of Equation (11) to obtain the exponent dimension, we have

$$D \propto \frac{\log(\Delta)}{\log(1/s)} \tag{12}$$

Information, which can be the measure of complexity provided by SampEn, increases with decreasing size of the lag space [42]. The lag can be equivalent to the length m defined in SampEn. In other words, the information is approximately proportional to the log of $1/m$ (from now on information I is used to refer to SampEn). A straight line of the semi-log axes is therefore expected in a plot of $1/m$ versus I , and the mathematical relation is a logarithmic equation. Because the relation is a straight line, it has the following form:

$$I_m = a + D_m \log(1/m) \tag{13}$$

where I_m is SampEn denoting the information subject to m , a is a constant which is the intercept, and D_m is the slope of the straight line, which is also the regularity dimension.

Rearrangement of Equation (13) to solve for D_m yields

$$D_m = \frac{I_m}{\log(1/m)} - \frac{a}{\log(1/m)} \tag{14}$$

Setting a limit term on m , which indicates that the relation does not hold for very large m , gives

$$D_m = \lim_{m \rightarrow 0} \frac{I_m}{\log(1/m)} - \frac{a}{\log(1/m)} \tag{15}$$

When m gets smaller and smaller, $1/m$ and its log becomes larger and larger. Since a is a constant, the term $\frac{a}{\log(1/m)}$ approaches zero in the limit of m approaching zero, and therefore this term becomes negligibly small. Thus, Equation (15) can be simplified as

$$D_m = \lim_{m \rightarrow 0} \frac{I_m}{\log(1/m)} \tag{16}$$

It can be noted from Equation (16) that the regularity dimension D_m measures the rate of change of signal regularity/predictability with respect to $\log(1/m)$, it is the rate at which the entropy of a dynamical system is gained with increasing resolution (decreasing length m).

Alternatively, increasing the value of r , which is the tolerance of similarity, decreases the information in the sense that no information is gained when most reconstructed vectors are considered to be similar (signal is highly predictable or has low complexity). If r is large enough, then both $A(m, r)$ and $A(m + 1, r)$ become 1, then the value of ApEn is 0 ($\log(1/1) = 0$). Being analogous to the relation of a straight line developed for D_m , the regularity dimension D_r can be obtained as

$$I_r = a + D_r \log(1/r) \tag{17}$$

where I_r is ApEn, denoting the information subject to r , a is a constant which is the intercept, and D_r is the slope of the straight line, which is the regularity dimension.

Rearrange Equation (17), we can define D_r as

$$D_r = \frac{I_r}{\log(1/r)} - \frac{a}{\log(1/r)} \tag{18}$$

To indicate the relation that does not apply to very large r , a limit term is inserted on r , giving

$$D_r = \lim_{r \rightarrow 0} \frac{I_r}{\log(1/r)} - \frac{a}{\log(1/r)} \tag{19}$$

Removing the negligible term $\frac{a}{\log(1/r)}$ in the limit of r approaching zero gives

$$D_r = \lim_{r \rightarrow 0} \frac{I_r}{\log(1/r)} \tag{20}$$

As from now, it can be seen that either D_m expressed in Equation (16) or D_r expressed in Equation (20) can be obtained as the slope of the plot of I_m versus $\log(1/m)$ or I_r versus $\log(1/r)$ on the curve which is approximately linear or by least squares fitting of a straight line to the curve, respectively.

2.3. Recurrence Plots

A recurrence plot (RP) is used as visual information of an attractor constructed by associated phase-space vectors to study nonlinear dynamics and chaos [40]. A plot of numbers of nearest neighbors of a trajectory \mathbf{x}_i in an embedding dimension versus a set of indices is called an RP, which was introduced in [40] and studied in detail in [48]. An RP can be expressed in form of a matrix as [49]:

$$\mathbf{R}_{ij} = H(\epsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), i, j = 1, \dots, N, \tag{21}$$

where $\|\cdot\|$ is a norm, ϵ is a threshold distance, and $H(\cdot)$ is the Heaviside step function defined as

$$H(x) = \begin{cases} 1 & : x \geq 0 \\ 0 & : x < 0. \end{cases} \tag{22}$$

Thus, the recurrence plot is the visualization of a square recurrence matrix of distance elements within a cutoff threshold. If the distance between two points on the attractor is less than the given threshold, a point is plotted on the matrix of the RP: if $\mathbf{R}_{ij} = 1$ then a black dot is plotted at the location (i, j) of the matrix, which is otherwise plotted with a white dot. Since any trajectory is self-recurrent, that is $\mathbf{R}_{ii} = 1$; therefore an RP is always represented with a black main diagonal that is called the line of identity. An RP is characterized with various visual patterns, which can be helpful for gaining insights into the dynamics underlying the system under study. Its patterns are classified according to topological structures such as homogeneity, periodicity, drift, and transience [46].

These components of an RP describe an individual attractor and can quantify features according to a histogram of the diagonal line in the RP. A diagonal line of length l , where the path of a segment of the trajectory is almost parallel to another segment, is defined as

$$(1 - \mathbf{R}_{(i-1)(j-1)})(1 - \mathbf{R}_{(i+l)(j+l)}) \prod_{k=0}^{l-1} \mathbf{R}_{(i+k)(j+k)} = 1 \tag{23}$$

A vertical line of length v is defined as

$$(1 - \mathbf{R}_{(i)(j-1)})(1 - \mathbf{R}_{(i)(j+v)}) \prod_{k=0}^{v-1} \mathbf{R}_{(i)(j+k)} = 1 \tag{24}$$

The histogram of the number of length l in the RP is expressed by:

$$P(l) = \sum_{i,j=1}^N (1 - \mathbf{R}_{(i-1)(j-1)})(1 - \mathbf{R}_{(i+l)(j+l)}) \prod_{k=0}^{l-1} \mathbf{R}_{(i+k)(j+k)} \quad (25)$$

The diagonal line, vertical line, and histogram are the basis of the RP for deriving features to quantify the complexity of a dynamical system. This RP-based feature analysis is known as recurrence quantification (RQ). Three commonly used features of the RQ are the recurrence rate (RR), determinism (DET), and recurrence entropy (RE) [50]. The RR is a measure of the density of recurrence points in the RP. The DET is the ratio of recurrence points that configure the diagonal lines of at least length l_{\min} to all recurrence points in the matrix. The RE is formulated in terms of the Shannon entropy to measure the uncertainty of the RP with respect to the diagonal lines. These three common features are mathematically defined as follows [49].

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{ij} = \frac{1}{N^2} \sum_{l=1}^N lP(l) \quad (26)$$

$$DET = \frac{\sum_{l=l_{\min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (27)$$

$$RE = - \sum_{l=l_{\min}}^N p(l) \log p(l) \quad (28)$$

where

$$p(l) = \frac{P(l)}{\sum_{l=l_{\min}}^N P(l)} \quad (29)$$

2.4. Largest Lyapunov Exponent

One of the most well-known method for quantitative measures of chaos is the largest Lyapunov exponent (LLE) [42,51]. The value of the LLE can be negative, zero, or positive. A negative LLE indicates the convergence of two trajectories. The LLE of zero means on the average rate the two trajectories keep at the same distance of each other. A positive LLE implies the divergence of the trajectories, which is an indicator of chaos. In other words, a positive LLE expresses sensitive dependence on initial conditions by presenting the average rate over the whole attractor, at which two nearby trajectories become exponentially separate with evolution. A practical numerical technique for calculating the LLE is the method developed by Rosenstein *et al.* [41], which works well with small datasets; and is robust to changes in the embedding dimension, reconstruction delay, and noise level. Therefore, this method was adopted in this study to calculate the LLE, and briefly described as follows [52,53].

Given a time series or sequence of length N , the first step is to reconstruct the phase space of the dynamical system using the time-delay method [54]. Let m and L be the embedding dimension and time delay. The reconstructed phase space can be expressed in matrix form as

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M)^T \quad (30)$$

where \mathbf{X} is matrix of size $M \times m$, $M = N - (m - 1)L$, and $\mathbf{X}_i = (x_i, \dots, x_{i+(m-1)L})$ which is the state of the system at discrete time i .

$$d_j(0) = \min_{\mathbf{X}_{j^*}} \|\mathbf{X}_j - \mathbf{X}_{j^*}\| \quad (31)$$

where $|j - j^*| > MP$ where MP is the mean period which is the reciprocal of the mean frequency of the power spectrum.

The basic idea is that the LLE (λ_1) for a dynamical system can be defined as [41]

$$d(t) = c e^{\lambda_1 t} \tag{32}$$

where $d(t)$ is the average divergence of two randomly chosen initial conditions at time t , and c is a constant that normalizes the initial separation between neighboring points.

By the definition given in Equation (32), the j pair of nearest neighbors can be assumed to diverge at a rate measured by λ_1 as follows:

$$d_j(i) \approx c_j e^{\lambda_1(i\Delta t)} \tag{33}$$

where $d_j(i)$ is the distance between the j pair of nearest neighbors after i discrete-time steps which is $i\Delta t$, Δt is the sampling period of the time series, and c_j is the initial separation between two neighboring points.

Taking the logarithm of both sides of Equation (33), giving

$$\ln[d_j(i)] \approx \lambda_1(i\Delta t) + \ln(c_j) \tag{34}$$

where $d_j(i) = \|\mathbf{X}_j(i) - \mathbf{X}_{j^*}(i)\|$. Equation (34) gives a set of approximately parallel curves, one for each j ($j = 1, \dots, M$). If these curves are approximately linear, their slopes represent the LLE (λ_1). The LLE can be computed as the slope of a straight-line fit to the average logarithmic divergence curve defined by

$$s(i) = \frac{1}{i\Delta t} \langle \ln[d_j(i)] \rangle_j \tag{35}$$

where $\langle \cdot \rangle_j$ denotes the average over all values of j .

2.5. Dynamic-Time Warping

Dynamic-time warping (DTW) has been well known as a good method for comparing the similarity between two feature-based signals or sequences [44]. For DTW, the pairwise comparison does not require that two sequences are of the same length that is a condition for the Euclidean metric. DTW can nonlinearly warps by stretching or compressing the sequences to determine their best match based on certain optimal criteria.

DTW can be generally described as follows [53]. Let \mathcal{F} be a feature space, $\mathbf{r} = (r_1, r_2, \dots, r_n, \dots, r_N)$, and $\mathbf{t} = (t_1, t_2, \dots, t_m, \dots, t_M)$ be the real MRI-feature-based vectors of reference and test patterns, respectively. DTW aims to construct a warping path of sequential discrete points: $\mathbf{w} = (w_1, w_2, \dots, w_k, \dots, w_K)$, where $w_k = (n_k, m_k) \in [1, N] \times [1, M]$, a grid point along axes \mathbf{r} and \mathbf{t} . A valid warping path has to satisfy the following conditions:

1. Boundary constraint by imposing $w_1 = (1, 1)$, and $w_K = (N, M)$. This condition is also known as endpoint constraints.
2. Monotonicity property such that $n_1 \leq n_2 \dots \leq n_K$, and $m_1 \leq m_2 \dots \leq m_K$. This means a valid path must follow a monotonic order with respect to time.
3. Continuity condition by setting $n_{k+1} - n_k \leq 1$ and $m_{k+1} - m_k \leq 1$, $k \in [1, K - 1]$. This is also known as step-size constraints.

Given the above conditions imposed for the working of the DTW, an optimal quantification of pattern similarity between two feature vectors \mathbf{r} and \mathbf{t} is carried out on the basis of a defined local cost function by means of some local distance measure $\delta(r_n, t_m)$, such that $\delta : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{R} \geq 0$. Smaller value of $\delta(r_n, t_m)$ indicates feature elements r_n and t_m are more similar to each other. Finally, the optimal warping path is determined as the one whose total induced cost over all possible warping paths is minimum, that is

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} [DTW(\mathbf{r}, \mathbf{t})] = \arg \min_{\mathbf{w}} [\delta_{\mathbf{w}}(\mathbf{r}, \mathbf{t})] \tag{36}$$

where \mathbf{w}^* denotes the optimal warping path.

As DTW is based on the principle of dynamic programming (DP) that can provide elegant solutions to optimization problems, DP involves in making sequential decisions that must be made at various stages. Therefore, DP is formulated using the framework of a *stage* variable, *state* variable, and *decision* variables [55]. In order to prevent a DP algorithm from searching a large space of stages and states, which can be very computationally expensive, warping paths are usually limited to a search space in terms of a warping window and slope constraint, which are described as follows.

1. Warping window: By setting $|n_k - m_k| \leq \omega$, where ω is a positive integer representing the window bandwidth. This restriction means that only features within a warping-path window are considered.
2. Slope constraint: By introducing a slope-weighting vector (ϕ_H, ϕ_V, ϕ_D) , where ϕ_H , ϕ_V , and ϕ_D are the weights for the horizontal, vertical, and diagonal directions, respectively. The purpose of this slope constraint is to avoid having a warping path that is either too steep or too shallow, and prevent matching very short segments with very long ones.

Based on the DP formulation, the cumulative distance for each point of a warping path, denoted as $\gamma(n_k, m_k)$, is computed using the following backward recursion:

$$\gamma(n_k, m_k) = \delta(n_k, m_k) + \min[\gamma(n_k - 1, m_k - 1), \gamma(n_k - 1, m_k), \gamma(n_k, m_k - 1)] \tag{37}$$

Similarly, the optimal warping path \mathbf{w}^* can be found by back-tracking the index order, selecting the point with the lowest accumulative distance. Starting with the endpoint $w_K^* = (N, M)$ that has the smallest accumulative distance, the back-tracking algorithm for finding the optimal warping path is described as follows.

$$w_{k-1}^* = \begin{cases} (1, m_k - 1), & \text{if } n_k = 1 \\ (n_k - 1, 1), & \text{if } m_k = 1 \\ \arg \min [DTW(n_k - 1, m_k - 1), DTW(n_k - 1, m_k), DTW(n_k, m_k - 1)], & \text{otherwise} \end{cases} \tag{38}$$

2.6. Phylogenetic Tree Reconstruction

Phylogenetic tree reconstruction is a visualization method for studying dissimilarity between genetic data based on the assumption that genomes change or diverge at a slow rate due to accumulated mutations of nucleotides. Thus, the degree of similarity between two genomes can be assessed by quantifying the amount of their base differences: two genomes sharing a recent ancestor would be more similar than those of more distant descendants. The topology of the reconstructed phylogenetic tree can provide insights into relationships between different genomes. In general, a phylogenetic tree shows relationships among species, organisms, or objects for taxonomic purposes of their evolutions. The leaf nodes of the tree structure represent descendants, whereas the internal nodes and root of the tree stand for the common ancestors. Two descendants that belong to the same node or clade are called sister groups, which are in contrast to the outgroup. The tree branches represent the distances between the objects. Thus, by plotting the tree using the feature distances of the MRI-based brain age groups, their relationships can be easily visualized and validated. The notion of phylogenetic tree reconstruction has been described in [9,56] for studying age-related diseases.

Phylogenetic tree can be constructed with the choice of various clustering models. A popular procedure is the use of a similarity matrix based approach, such as the unweighted pair-group

method using arithmetic averaging (UPGMA) [57]. UPGMA is a hierarchical cluster analysis that builds nested hard clusters in a dataset by merging two clusters at each step based on the minimization of a similarity measure. The UPGMA algorithm can be described with the following procedure [58]:

1. Given dataset $\mathbf{X} \in \mathcal{R}^q$, $n = |\mathbf{X}|$, $\mathbf{U}_n = [\delta_{ij}]_{n \times n}$, where δ_{ij} is the Kronecker delta: 0 if $i \neq j$ and 1 if $i = j$ (each $\mathbf{x}_k \in \mathbf{X}$ is a singleton cluster at the number of clusters $c = n$, $\mathbf{V}(n) = 0$, where $\mathbf{V}(n)$ is the c -partition of \mathbf{X}).
2. At step k , $k = 1, \dots, n - 1$, $c = n - k + 1$, using \mathbf{U}_c to directly solve the measure of hard-cluster similarity (hard clustering means that each data point is a member of one and only one cluster) by minimizing the following function to identify the minimum distance as the similarity between any two data points in \mathbf{X} :

$$\underset{j,k}{\text{minimize}} \left[\mathbf{J}(\mathbf{u}_j, \mathbf{u}_k) = \frac{\sum_{i=t+1}^n \sum_{t=1}^{n-1} u_{ji} u_{kt} d(\mathbf{x}_i, \mathbf{x}_t)}{(\sum_{i=1}^n u_{ji})(\sum_{t=1}^n u_{kt})} \right], \quad (39)$$

where $\mathbf{u}_j, \mathbf{u}_k$ denote the j -th and k -th rows of \mathbf{U}_c , $u_{ji} \in [0, 1]$, and $d : \mathbf{X} \times \mathbf{X} \rightarrow \mathcal{R}^+$ is any measure of dissimilarity on \mathbf{X} , and d was used as a spectral-distortion measure in this study.

3. Let $(\mathbf{u}_r, \mathbf{u}_s)_c$ solve Equation (39). Merge u_r and u_s , thus constructing from \mathbf{U}_c the updated partition \mathbf{U}_{c-1} , record $\mathbf{V}(c - 1) = \mathbf{J}[(\mathbf{u}_r, \mathbf{u}_s)_c]$.
4. If $k < n - 1$, go to Step 2; if $k = n - 1$, $c = 2$. Merge the two remaining clusters, set $\mathbf{U}_1 = [1, \dots, 1]$, compute $\mathbf{V}(c - 1) = \mathbf{J}(\mathbf{u}_1, \mathbf{u}_2)$, and stop.

3. MRI Data and Preprocessing

The IXI (Information eXtraction from Images) dataset [59] was used in this study. This MRI database consists of 564 MR images from normal, healthy subjects, with ages ranging from 20 to 86 years old. The MRI acquisition protocol for each subject includes T1, T2 and PD-weighted images, collected on three different scanners. For the comparison of the brain structures of groups of younger and older subjects, most studies have reported age-related decreases in grey matter (GM) densities in specific brain sites measured by MRI [60]. Therefore, this study focuses on the extraction of the GM morphology for brain-age grouping, using T1-weighted images. The preprocessing of the T1-weighted images was performed using the SPM8 package [61]. All T1-weighted images were corrected for bias-field inhomogeneity, normalized and segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) [62].

The brain images are divided into 7 groups within 10 years of age as follows: 20–29 years old (100 subjects), 30–39 years old (99 subjects), 40–49 years old (89 subjects), 50–59 years old (99 subjects), 60–69 years old (120 subjects), 70–79 years old (49 subjects), and 80–86 years old (8 subjects).

To generate the time series of the preprocessed brain MRI data represented by the GM, using the radial scanning method [27] for each MRI slice, a corresponding time series was obtained by measuring the distances from subsequent outer boundary points to the GM center of mass, being described in [63]. By computing the distances between the boundary points and the GM center of each cortical sheet, these time series can capture the morphology of the cortical surface including its folding patterns and shape changes. Figure 1 shows examples of the MRI data and the corresponding generated time series of chronological brain ages of 20, 40, 60, and 80 years.

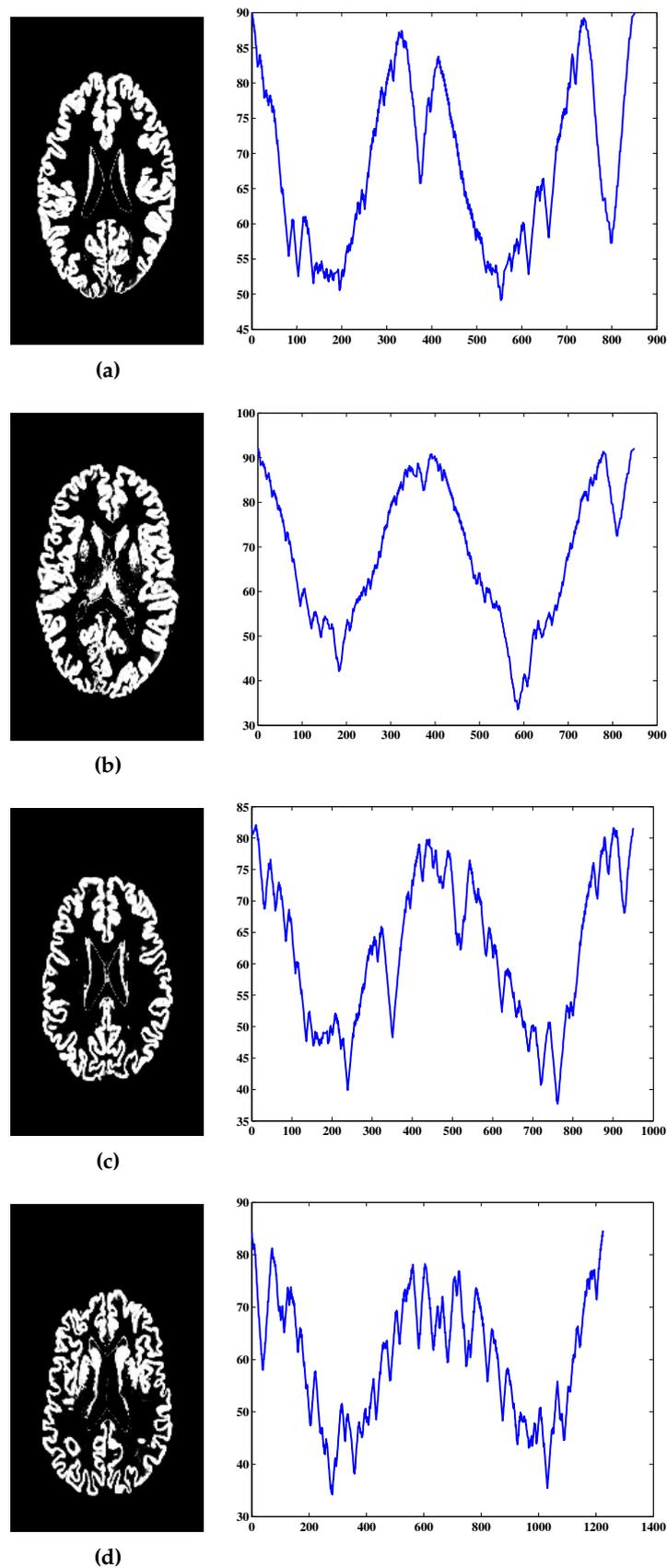


Figure 1. Examples of magnetic resonance imaging (MRI) scans (left) and time series of associated GM morphology (right) of a 20 year-old (a), 40 year-old (b), 60 year-old (c), and 80 year-old (d) healthy brain subjects; where each plot of the time series shows pixel-based distances measured from the gray-matter center of the brain (y -axis) to sequential outer boundary points (x -axis).

4. Results and Discussion

The sample entropy (SampEn), regularity dimension (RD), recurrence plots (RP), largest Lyapunov exponent (LLE), and dynamic-time warping (DTW) methods were used to obtain featured distances between the brain-age groups from the brain time-series data to construct the corresponding phylogenetic trees.

Using $m = 1, 2$ and 3 , and the tolerance $r = 0.2$ for computing SampEn values; three corresponding trees were constructed and shown in Figure 2. Table 1 shows the means and variances of SampEn values for $m = 1, 2$ and 3 . For $m = 1$, the tree topology provides a desirable result by partitioning the brain ages into two distinct clades of 20–59 and 60–86 (a clade represents a single branch on the tree to indicate a population having a common ancestor), in which 20–29 and 30–39, 40–49 and 50–59, 60–69 and 70–79 are sister groups. For $m = 2$, the relationships of the brain ages are well divided into three groups: 20–29 and 30–39, 40–49 and 50–59, and 60–69 and 70–79 and 80–86, where the latter two groups are closer to each other. As for $m = 3$, 20–29 and 30–39, and 40–49 are closer to 50–59, 60–69 and 70–79, while 80–86 is considered as an outgroup. In general, the three tree topologies shown in Figure 2 display reasonable relationships of the brain ages.

Utilizing SampEn values, with $m = 1, 2$ and 3 , D_r values were obtained to establish the trees of relationships among the brain-age groups, which are shown in Figure 3. For each fixed value of m , r was decreased from 0.5 to 0.05 with an interval of 0.05 [16]. The slopes of SampEn versus $\log(1/r)$ were taken at the first five points, which form a straight line, to calculate D_r . Table 2 shows the means and variances of D_r values for $m = 1, 2$ and 3 . For $m = 2$, the RD-based trees appropriately assign the relationships of the brain-age groups by separating the clade of 20–29 and 30–29 from the other brain-age groups, and the clade of 40–49 and 50–59 from the clade of 60–69, 70–79, and a single branch for 80–86. The RD-based tree, with $m = 1$, shows the grouping of the chronological brain ages into well separated clades, except the groups of 70–79 and 40–49 belong to the same clade. For $m = 3$, the RD-based tree locates the groups of 80–86 and 40–49 to the same clade, while other groups are appropriately partitioned.

Table 1. (Mean, variance) of sample entropy (SampEn) values of brain-age groups with $r = 0.2$, and $m = 1, 2$, and 3 .

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(4.5079, 0.0173)	(1.5254, 0.0032)	(1.2375, 0.0031)
30–39	(4.5115, 0.0139)	(1.5424, 0.0028)	(1.2400, 0.0026)
40–49	(4.5540, 0.0145)	(1.5653, 0.0024)	(1.2504, 0.0023)
50–59	(4.5838, 0.0143)	(1.5897, 0.0013)	(1.2673, 0.0015)
60–69	(4.6531, 0.0148)	(1.6067, 0.0008)	(1.2773, 0.0015)
70–79	(4.7003, 0.0089)	(1.6197, 0.0006)	(1.2812, 0.0010)
80–86	(4.7765, 0.0018)	(1.6158, 0.0006)	(1.3029, 0.0006)

Table 2. (Mean, variance) of regularity dimension (RD) values of brain-age groups with $m = 1, 2$, and 3 .

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(1.0482, 0.0002)	(0.7409, 0.0006)	(0.6907, 0.0020)
30–39	(1.0475, 0.0001)	(0.7448, 0.0005)	(0.6666, 0.0024)
40–49	(1.0519, 0.0001)	(0.7572, 0.0003)	(0.6441, 0.0024)
50–59	(1.0528, 0.0001)	(0.7664, 0.0002)	(0.6094, 0.0027)
60–69	(1.0524, 0.0001)	(0.7733, 0.0002)	(0.5997, 0.0024)
70–79	(1.0521, 0.0001)	(0.7753, 0.0002)	(0.5786, 0.0036)
80–86	(1.0516, 0.0001)	(0.7783, 0.0002)	(0.6282, 0.0008)

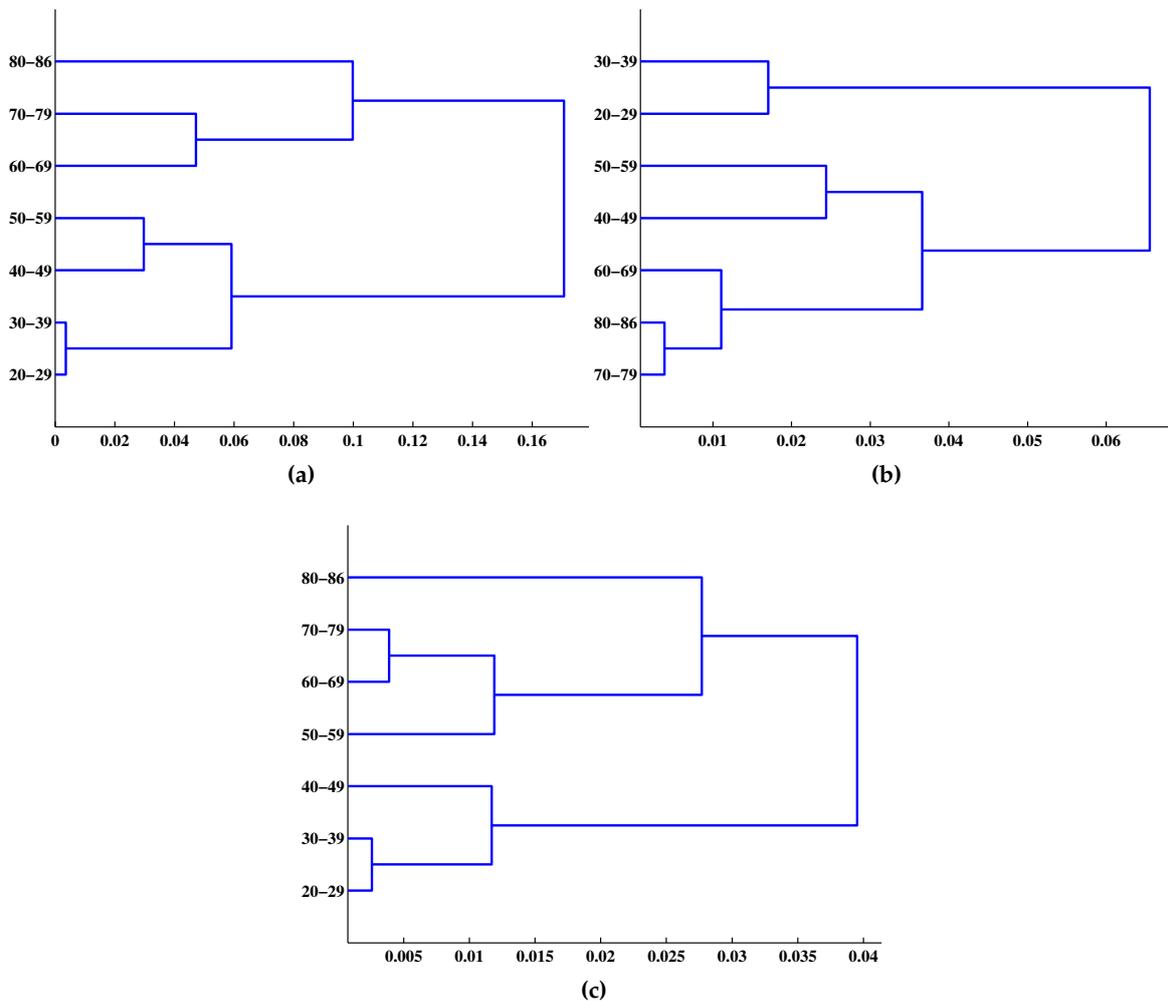


Figure 2. SampEn-based constructed trees of relationships of MRI-based healthy brain age groups, with $m = 1$ (a), $m = 2$ (b), and $m = 3$ (c); where values on the x -axis of each plot are the corresponding feature-based distances.

The trees provided by the RP, with $\epsilon = 5$, using its recurrence rate (RR) feature and $m = 1, 2$ and 3 , are shown in Figure 4a–c, respectively. Table 3 shows the means and variances of RR values for $m = 1, 2$ and 3 . While the group of over 80 years old can be well separated from other groups in all the trees; the groups of 20–29 and 40–49 and 50–59 belong to the same clade for $m = 1$; the groups of 20–29 and 60–69 are located in the same clade, also the groups of 30–39 and 50–59 belong to the same clade in the trees for $m = 2$ and 3 .

Table 3. (Mean, variance) of recurrence rate (RR) values of brain-age groups with $m = 1, 2$, and 3 .

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(0.0187, 0.0000)	(0.0161, 0.0000)	(0.0154, 0.0000)
30–39	(0.0192, 0.0000)	(0.0167, 0.0000)	(0.0159, 0.0000)
40–49	(0.0189, 0.0000)	(0.0165, 0.0000)	(0.0158, 0.0000)
50–59	(0.0189, 0.0000)	(0.0166, 0.0000)	(0.0159, 0.0000)
60–69	(0.0181, 0.0000)	(0.0158, 0.0000)	(0.0152, 0.0000)
70–79	(0.0176, 0.0000)	(0.0154, 0.0000)	(0.0147, 0.0000)
80–86	(0.0164, 0.0000)	(0.0142, 0.0000)	(0.0136, 0.0000)

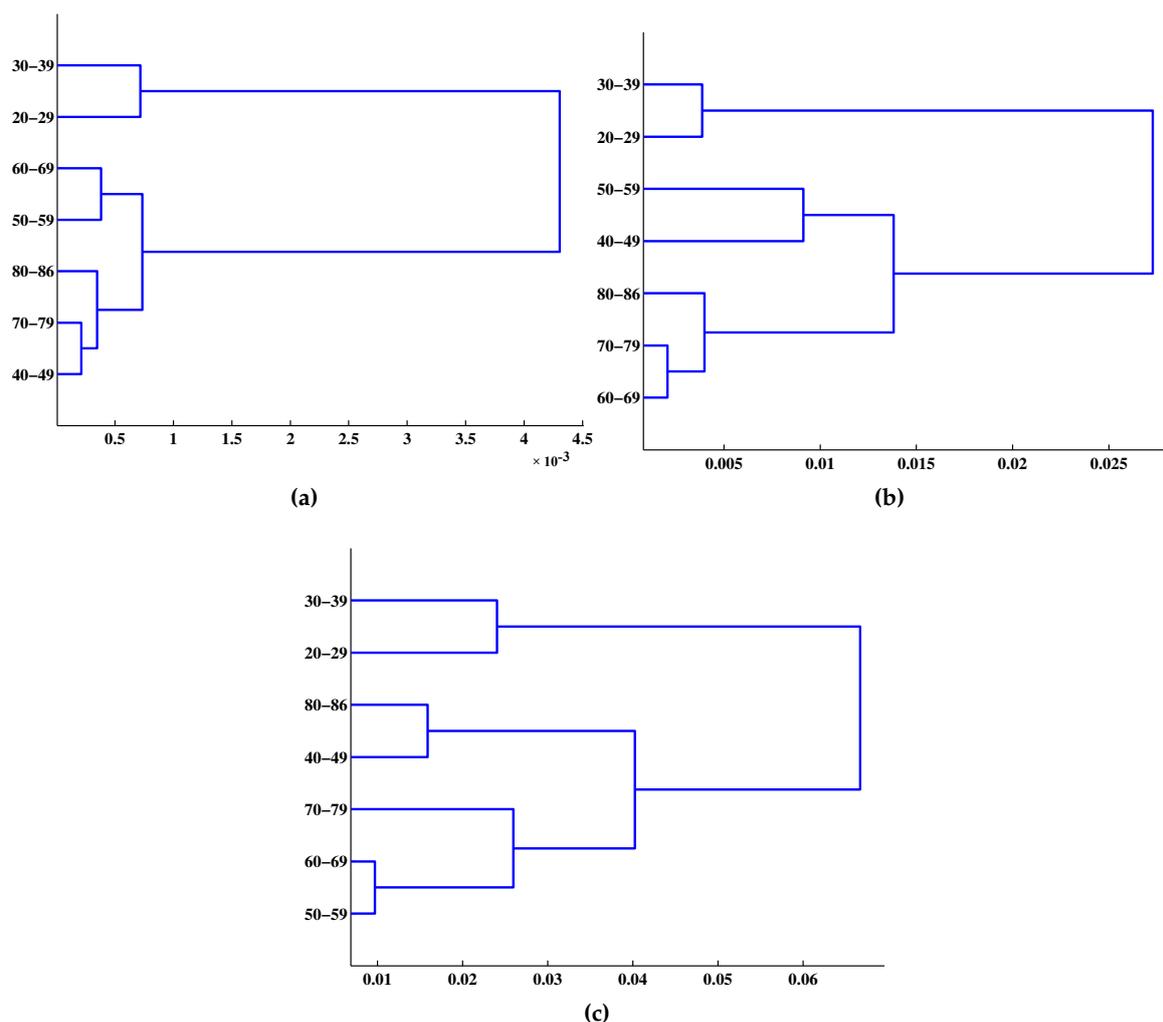


Figure 3. RD-based constructed trees of relationships of MRI-based healthy brain age groups, with $m = 1$ (a), $m = 2$ (b), and $m = 3$ (c); where values on the x -axis of each plot are the corresponding feature-based distances.

Using the recurrence entropy (RE) feature of the RP with $m = 1$, and $l_{\min} = 2$ and 6 , the relationships of the brain-age groups are shown in Figure 4d,e, respectively. The groups of 60–69 and 80–86 are assigned to the same clade in both sub-plots. Using the recurrence-entropy feature of the RP with $m = 2$ and $l_{\min} = 2$ and 6 , the relationships of the brain-age groups are shown in Figure 4f,g, respectively. For $l_{\min} = 2$, the groups of 60–69 and 80–86 are in the same clade. A reasonable result of the tree construction is obtained using $l_{\min} = 6$. Increasing m to 3, the corresponding Figure 4h,i show that the RP-based assignment of the age group of 80–86 is confused with the younger age groups (closer to 60–69 for $l_{\min} = 2$, and closer to 40–49 for $l_{\min} = 6$). Tables 4 and 5 show the means and variances of RE values for $l_{\min} = 2$ and 6 , with $m = 1, 2$ and 3 ; respectively.

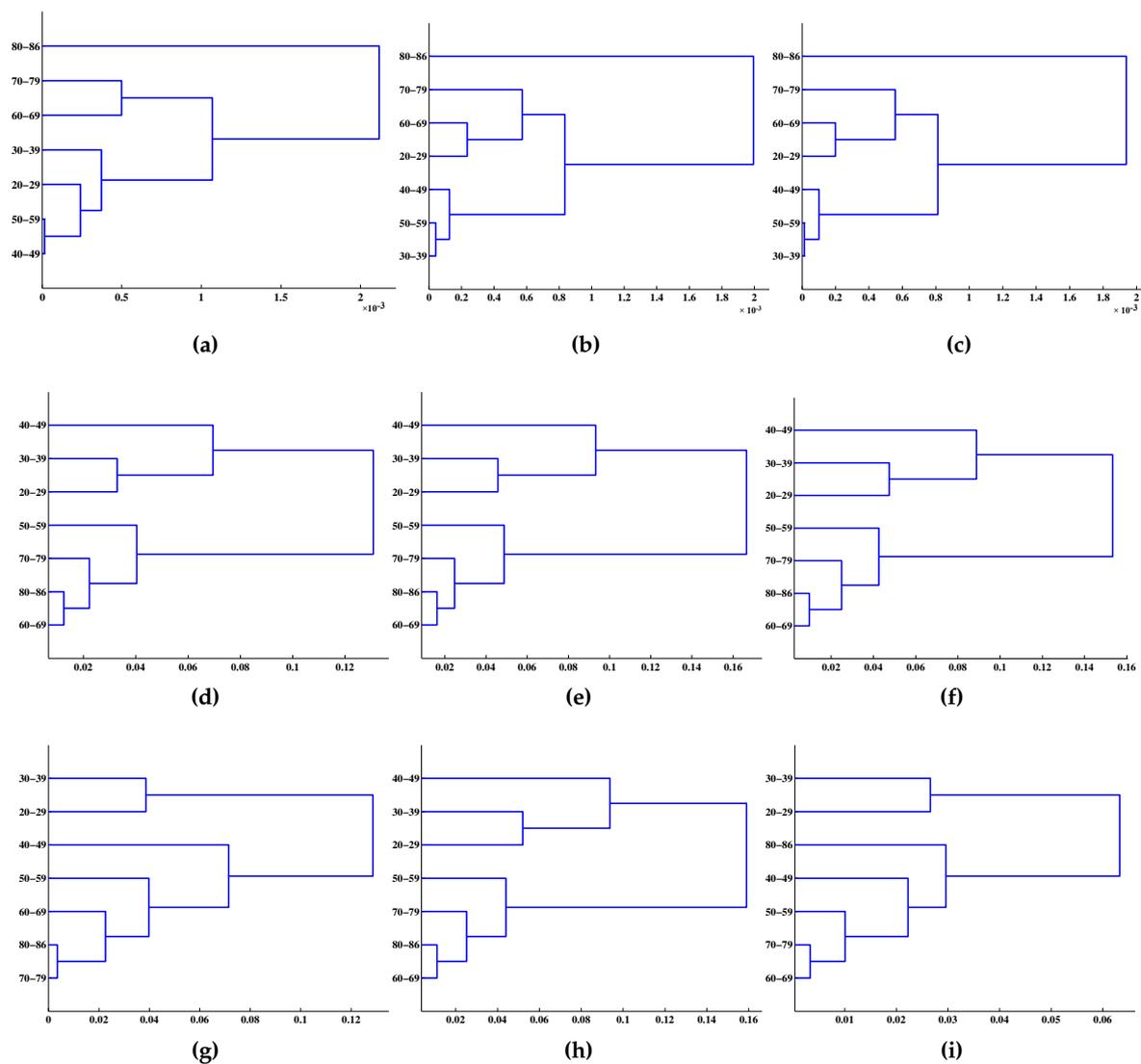


Figure 4. Recurrence plot (RP)-based constructed trees of relationships of MRI-based healthy brain age groups, where recurrence rate (RR) used as feature with $m = 1$ (a), $m = 2$ (b), and $m = 3$ (c); recurrence entropy (RE) used as feature with $m = 1$ and $l_{\min} = 2$ (d), $m = 1$ and $l_{\min} = 6$ (e), and $m = 2$ and $l_{\min} = 2$ (f); recurrence entropy (RE) used as feature with $m = 2$ and $l_{\min} = 6$ (g), $m = 3$ and $l_{\min} = 2$ (h), and $m = 3$ and $l_{\min} = 6$ (i). Values on the x-axis of each plot are the corresponding feature-based distances.

Table 4. (Mean, variance) of recurrence entropy (RE) values of brain-age groups with $l_{\min} = 2$, and $m = 1, 2$, and 3.

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(5.1295, 0.0134)	(4.6164, 0.0202)	(4.3049, 0.0224)
30–39	(5.0966, 0.0123)	(4.5690, 0.0170)	(4.2528, 0.0184)
40–49	(5.0435, 0.0090)	(4.5039, 0.0118)	(4.1852, 0.0125)
50–59	(4.9894, 0.0066)	(4.4417, 0.0087)	(4.1217, 0.0091)
60–69	(4.9627, 0.0056)	(4.4123, 0.0070)	(4.0917, 0.0071)
70–79	(4.9341, 0.0041)	(4.3826, 0.0053)	(4.0609, 0.0054)
80–86	(4.9502, 0.0027)	(4.4027, 0.0031)	(4.0806, 0.0033)

Table 5. (Mean, variance) of recurrence entropy (RE) values of brain-age groups with $l_{\min} = 6$, and $m = 1, 2$, and 3.

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(4.9974, 0.0243)	(4.4715, 0.0157)	(4.2885, 0.0068)
30–39	(4.9517, 0.0210)	(4.4329, 0.0114)	(4.2620, 0.0042)
40–49	(4.8813, 0.0142)	(4.3807, 0.0075)	(4.2346, 0.0027)
50–59	(4.8136, 0.0098)	(4.3391, 0.0043)	(4.2190, 0.0014)
60–69	(4.7811, 0.0081)	(4.3144, 0.0031)	(4.2106, 0.0008)
70–79	(4.7483, 0.0060)	(4.2935, 0.0016)	(4.2073, 0.0004)
80–86	(4.7649, 0.0038)	(4.2899, 0.0020)	(4.1883, 0.0002)

On the tree reconstruction using the LLE feature, Figure 5 shows the relationships of the chronological brain-age groups for $m = 1, 2$ and 3. The time delay L was selected to be 1, and the first five points of the divergence curve were used to calculate the slope of a straight line as the value of the LLE. Table 6 shows the means and variances of LLE values for $m = 1, 2$ and 3. It can be seen from all the three figures that the LLE-based trees tend to distinguish the clade of the age groups of 40–49, 50–59, 60–69, 70–79, and 80–86 from the clade of 20–29 and 30–39. In general, the LLE-based tree topologies show the groups of similar ages are more closely related to each other than they are to the outgroup, except for $m = 1$, the group of 80–86 is closer to the clade of 40–49 and 50–59 than the clade of 60–69 and 70–79, and for $m = 3$, 80–86 is slightly closer to 50–59 than the clade of 60–69 and 70–79.

On the use of the DTW for the phylogenetic tree reconstruction of the brain-age groups, the matching costs were calculated by adopting the Itakura and Sakoe-Chiba local constraints [64]. Figure 6a–c show the DTW-based trees using the Itakura cost constraints, using the three longest (first, second, and third) time series of the GM boundaries of each age group for pattern matching, respectively. Figure 6d–f show the DTW-based trees using the Sakoe-Chiba cost constraints, using the three longest (first, second, and third) time series of the GM boundaries of each age group for pattern matching, respectively. Figure 6a puts the group of 80–86 closer the clade of 20–29 and 40–49, and assigns 30–39 and 50–59 to the same clade. Figure 6b can separate the groups of 60–69, 70–79, and 80–86 from those of 20–29, 30–39, 40–49, 50–59 by assigning these two groups into two distinct clades; however, 20–29 is closer to the clade consisting of 40–49 and 50–59 than 30–39. Figure 6c shows that 30–39 is closer to the clade consisting of 50–59 and 60–69, and 20–29 is closer to 40–49. On the adoption of the Sakoe-Chiba local constraints, Figure 6d assigns 20–29 and 40–49 to the same clade. The topology of the tree shown in Figure 6e represents reasonable relationships between the age groups, and distinguishes the age groups of 20 to 59 from those of 60 to 86. The age group of 20–29 is the outgroup to the others for the DTW-based tree using the Sakoe-Chiba local constraints as can be seen from Figure 6f.

Table 6. (Mean, variance) of largest Lyapunov exponent (LLE) values of brain-age groups with $m = 1, 2$, and 3.

Age Group	$m = 1$	$m = 2$	$m = 3$
20–29	(0.3230, 0.0002)	(0.3546, 0.0001)	(0.3516, 0.0001)
30–39	(0.3267, 0.0001)	(0.3599, 0.0001)	(0.3566, 0.0001)
40–49	(0.3322, 0.0001)	(0.3642, 0.0001)	(0.3601, 0.0001)
50–59	(0.3328, 0.0001)	(0.3671, 0.0001)	(0.3624, 0.0000)
60–69	(0.3357, 0.0001)	(0.3677, 0.0000)	(0.3626, 0.0000)
70–79	(0.3361, 0.0000)	(0.3685, 0.0000)	(0.3628, 0.0000)
80–86	(0.3408, 0.0000)	(0.3722, 0.0000)	(0.3652, 0.0000)

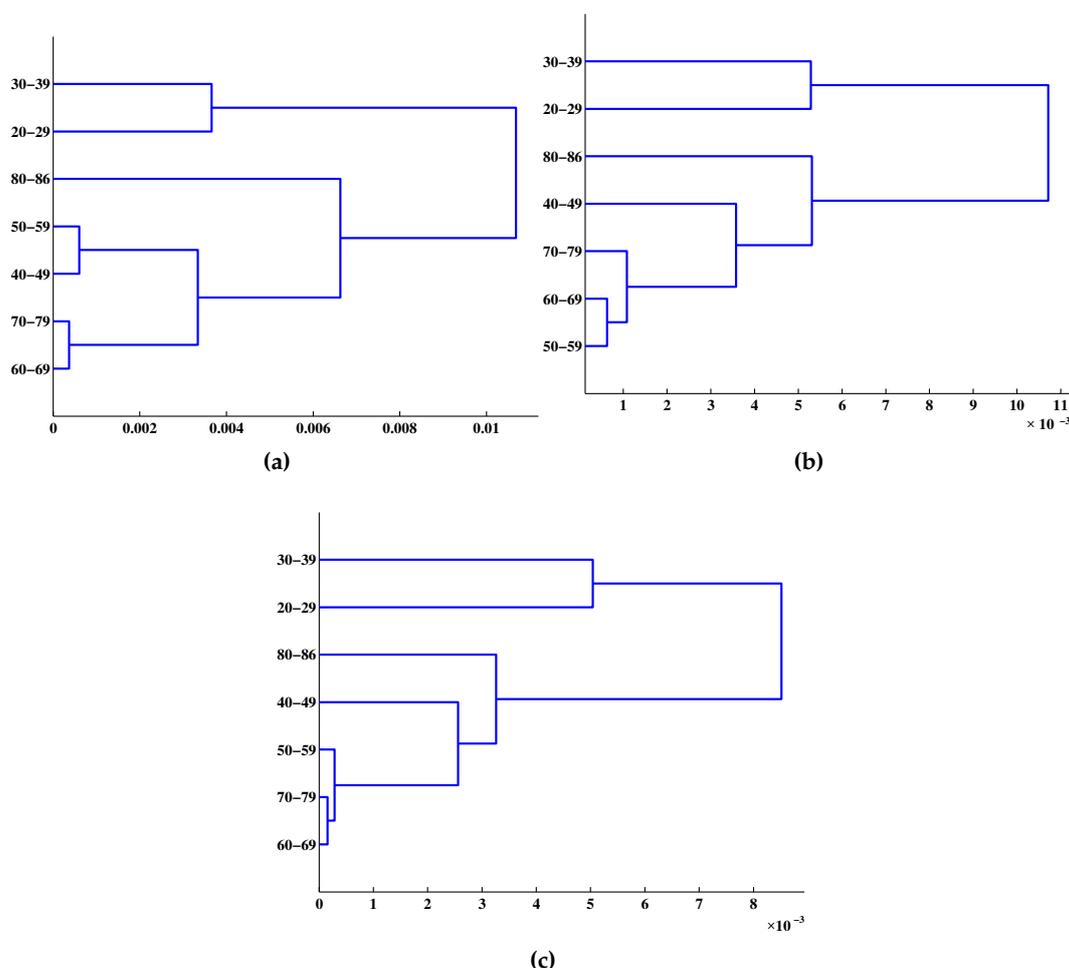


Figure 5. LLE-based constructed trees of relationships of MRI-based healthy brain age groups, with $m = 1$ (a), $m = 2$ (b), and $m = 3$ (c); where values on the x -axis of each plot are the corresponding feature-based distances.

Regarding the values given in Tables 1–6, the SampEn means tend to increase with the older-age groups, and become smaller with larger values for m . For $m = 1$, the means of RD get larger with the older-age groups; but for $m = 3$, the RD values are smaller for the older ages, except for the cohort of 80–86 years of age; and there is no clear trend of increasing or decreasing of the RD with the brain ages. The RD values are smaller for larger values of m . All RR and RE values are smaller for larger m . Except for 80–86 year-old cohort, the RE decreases with increasing ages of the brain. Except for the 20–29 year-old cohort for $m = 1$, and 20–29 and 50–59 year-old cohorts for $m = 2$ and 3, the RR tends to decrease with older age groups. All the means and variances presented in Tables 1–6 are truncated to four decimal digits, which make the RR variances equivalent to zero. Such small values of the RR variances can be due to both the insufficient samples of MRI data and the computational method.

In summary, with the given selection of parameters for each method, the SampEn-based trees, RD-based tree with $m = 2$, RP-based tree using the RE with $m = 2$ and $l_{\min} = 6$, and DTW-based tree using the Sakoe-Chiba constraints on the second largest GM-generated time series yield reasonable results to represent the chronological relationships of the healthy brain MRI groups. Among these mentioned reasonable tree topologies, those obtained from SampEn seem all valid for expressing the relationships of the brain age groups, and therefore the SampEn is the most preferred method with the currently available MRI data.

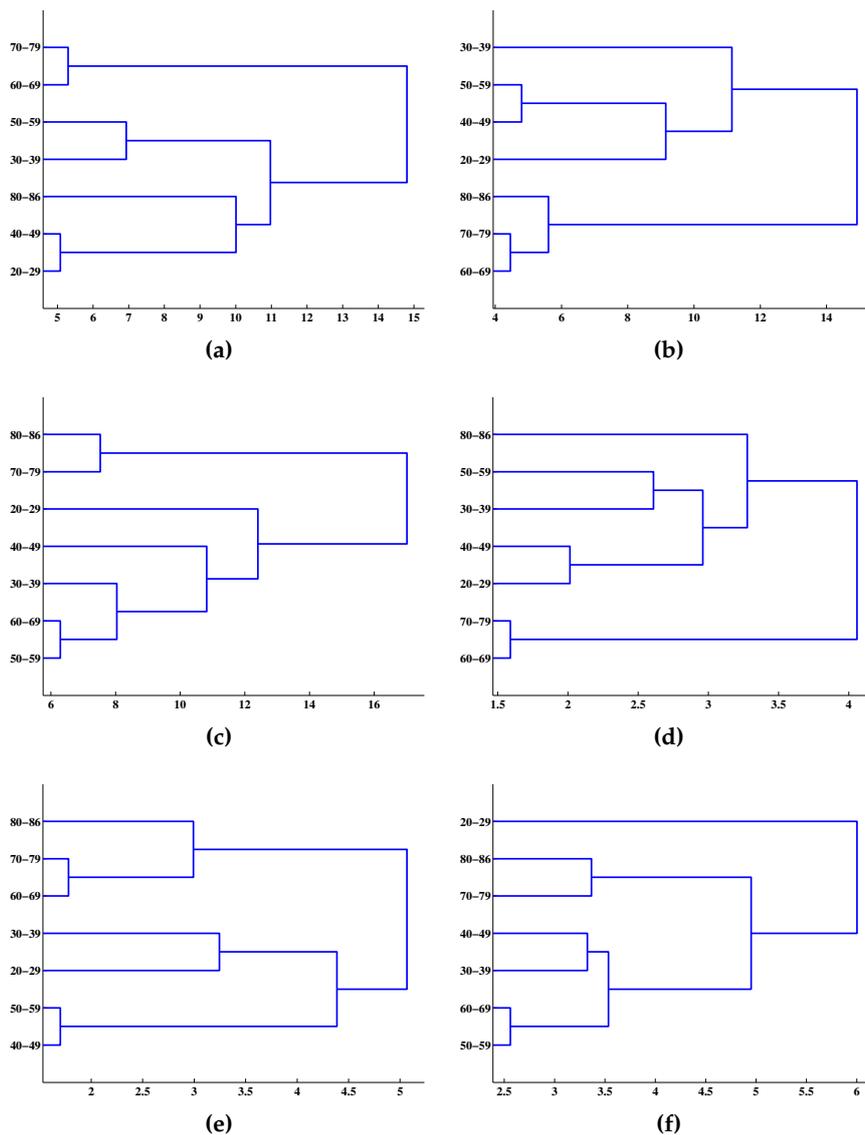


Figure 6. Dynamic-time warping (DTW)-based constructed tree of relationships of MRI-based healthy brain age groups using Itakura local constraints based on: largest (a), second largest (b), and third largest (c) gray matter (GM)-generated time series of brain MRI; Sakoe-Chiba local constraints based on: largest (d), second largest (e), and third largest (f) GM-generated time series of brain MRI. Values on the x -axis of each plot are the corresponding feature-based distances.

A common problem encountered with the computational methods studied here in assigning the group of 80–86 years of age to an appropriate branch of the tree is likely due to the limited data of the age group, which consists of only eight subjects. Other issues are due to data requirements and parameter selections imposed by different methods, which need to be further investigated to ensure their optimal implementations. For example, the LLE is a method for quantifying chaos (a positive LLE indicates chaos), the concept of the LLE is to keep track of two nearby orbits to determine the average logarithmic rate of separation of the orbits. If the two orbits depart too far from each other, one of them has to be adjusted back to the vicinity of the other along the line of separation, and this adjustment is known to be the most difficult and error-prone step for calculating the LLE [65]. When underlying equations for chaos are available, the numerical calculation of the LLE is relatively simple; but for experimental data, such a calculation is very difficult [65]. In this study, MRI-based

time-series data were analyzed using the Rosenstein method [41], which is particularly developed to compute the LLE from small recorded time series and robust to the choice of embedding dimension, reconstruction delay, and noise. In this study, the first five points of the divergence curves were used to estimate the LLEs using the least-squares fit. A practical reason for the poorer presentation of the brain-age relationships via the LLE-based tree topologies than other feature-based tree topologies is due to the inconsistency of the numbers of divergence points representing a straight line for the LLE calculation. Therefore, some method that can adaptively determine suitable numbers of points on the divergence curves for different MRI-based time series would be needed to improve the diagnostic performance of the Rosenstein method.

Finally, the idea is that if a valid tree structure that represents the chronological relationships within a group of brain ages can be constructed, the physiological age of a brain on MRI can be reliably predicted by assigning its physiological age to the age group whose feature-based distance is smallest among others. Furthermore, the constructed brain-age tree can be always updated and extended when more MRI data representing various age groups become available.

5. Conclusions

Aging is thought as the combined result of physical, biological, chemical, and psychological changes. Evidence from computed tomography imaging studies suggested that the cerebral ventricles dilate as a function of age, which is known as ventriculomegaly [66]. More recently, an MRI study found that regional areas of the brain of age-related disorder decreases in cerebral volume [67]. However, regional volume reduction is not uniform. Some brain regions shrink, whereas others remain relatively stable until the end of the life-span [14]. The structure and function of the brain are known to be very complex. Here, several methods of nonlinear dynamics and chaos, which are well recognized for the analysis of complexity, have been carried out to test their feasibility for an objective measure of the similarity and relationships of the brain-age groups using their MRI data.

Because of limited MRI data, the brain-age gap adopted in this study is 10 years to test the feasibility of the presented methods. Such an age gap can be large for applications in clinical settings. When new samples become available, the tree of brain-age groups can be reconstructed with smaller age gaps, or included in the existing tree model, whose new topology will be reassessed for validity of the relationships before putting the model to a practical use. Furthermore, not only the physiological MRI-based brain age can be predicted among the brain-age groups, inference about the brain under study from the established trees can also be made to discover a possible pattern of the new data. Such a proposed conceptual framework departs itself from other related studies that attempt to predict the brain age based on training data and classifiers.

Being different from the perspective of pattern classification, the purpose of the proposed methodology for grouping brain ages is to secure the reliability of the brain age prediction by examining relationships of the brain age groups, regardless of the age gaps. Validation of the proposed approach can be obtained by the visual inspection of the tree topology under study. Since there can be several valid tree topologies provided by the same or different methods, a question of interest is to determine which topology of relationships is the best. An answer to this question can be problem-dependent, may rely on the knowledge of the anatomy of the brain and clinical trials.

In general, chaos and nonlinear dynamics methods and their potential combination can be useful tools to extract discriminative features related to the variability of the complexity and dynamics of the brain structure on MRI for grouping brain ages, in which the physiological brain age of a subject with neurodegenerative disease can be detected if it is found older than the chronological age based on the matching of features between the constructed brain-age groups and the individual. There is another issue worth considering. Since the brain has many different areas and types of tissue (GM and WM), the vulnerability of different functions of the GW and WM of the brain to age-induced changes may not be consistent because GM consists of cell bodies in the cortex and subcortical nuclei, whereas WM consists of tightly packed myelinated axons connecting the neurons of the cerebral cortex to each other

and with the periphery [68]. Only the structural information of the GW is taken into account in this study. The inclusion of both types of the brain matter and extraction of features out of regional areas of the brain would be expected to gain further insight into the relationships of chronological brain ages captured on MRI.

Acknowledgments: Part of this work was carried out while the first author (TDP) was with the Medical Image Processing Lab at the University of Aizu. Hung T. Le, who is a graduate student of the University of Aizu, assisted TDP in part of the pre-processing of the MRI data.

Author Contributions: T.D.P. conceived, designed, and supervised the study, and wrote the paper. T.A. performed the computer programming and experiments. T.D.P., R.O. and Y.-F.C. supervised the work on DTW. T.D.P., T.A., R.O. and Y.-F.C. performed data analysis and interpretation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Teverovskiy, L.A.; Becker, J.T.; Lopez, O.L.; Liu, Y. Quantified brain asymmetry for age estimation of normal and AD/MCI subjects. In Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2008), Paris, France, 14–17 May 2008; pp. 1509–1512.
2. Franke, K.; Ziegler, G.; Kloppel, S.; Gaser, C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage* **2010**, *50*, 883–892.
3. Wang, B.; Pham, T.D. MRI-based age prediction using hidden Markov models. *J. Neurosci. Methods* **2011**, *199*, 140–145.
4. Dukart, J.; Schroeter, M.L.; Mueller, K. Age correction in dementia—matching to a healthy brain. *PLoS ONE* **2011**, *6*, e22193; doi:10.1371/journal.pone.0022193.
5. Kandel, B.M.; Wolk, D.A.; Gee, J.C.; Avants, B. Predicting cognitive data from medical images using sparse linear regression. In *Information Processing in Medical Imaging*; Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., ZÁúllei, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7917, pp. 86–97.
6. Irimia, A.; Torgerson, C.M.; Goh, S.Y.; van Horn, J.D. Statistical estimation of physiological brain age as a descriptor of senescence rate during adulthood. *Brain Imaging Behav.* **2015**, *9*, 678–689, doi:10.1007/s11682-014-9321-0.
7. Cole, J.H.; Leech, R.; Sharp, D.J. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann. Neurol.* **2015**, *77*, 571–581.
8. Spulber, G.; Niskanen, E.; MacDonald, S.; Smilovici, O.; Chen, K.; Reimanet E.M.; Jauhiainen, A.M.; Hallikainen, M.; Tervo, S.; Wahlund, L.-O.; et al. Whole brain atrophy rate predicts progression from MCI to Alzheimer’s disease. *Neurobiol. Aging* **2010**, *31*, 1601–1605.
9. Pham, T.D.; Salvetti, F.; Wang, B.; Diani, M.; Heindel, W.; Knecht, S.; Wersching, H.; Baune, B.T.; Berger, K. The hidden-Markov brain: Comparison and inference of white matter hyperintensities on magnetic resonance imaging (MRI). *J. Neural Eng.* **2011**, *8*, 016004; doi:10.1088/1741-2560/8/1/016004.
10. Su, L.; Wang, L.; Hu, D. Predicting the age of healthy adults from structural MRI by sparse representation. In *Intelligent Science and Intelligent Data Engineering*; Yang, J., Fang, F., Sun, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2013, Volume 7751, pp. 271–279.
11. Gaser, C.; Franke, K.; Kloppel, S.; Koutsouleris, N.; Sauer, H. BrainAGE in mild cognitive impaired patients: Predicting the conversion to Alzheimer’s disease. *PLoS ONE* **2013**, *8*, e67346, doi:10.1371/journal.pone.0067346.
12. Bigler, E.D. Traumatic brain injury, neuroimaging, and neurodegeneration. *Front. Hum. Neurosci.* **2013**, *7*, doi:10.3389/fnhum.2013.00395.
13. Sowell, E.R.; Peterson, B.S.; Thompson, P.M. Mapping cortical change across the human life span. *Nat. Neurosci.* **2003**, *6*, 309–315.
14. Raz, N.; Rodrigue, K.M. Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neurosci. Biobehav. Rev.* **2006**, *30*, 730–748.
15. Wang, B.; Pham, T.D. HMM-based brain age interpolation using kriging estimator. In Proceedings of the IEEE International Symposium on Image and Signal Processing and Analysis, Dubrovnik, Croatia, 4–6 September 2011; pp. 704–708.

16. Chen, Y.; Pham, T.D. Entropy and regularity dimension in complexity analysis of cortical surface structure in early Alzheimer's disease and aging. *J. Neurosci. Methods* **2013**, *215*, 210–217.
17. What is Alzheimer's? Available online: http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp (accessed on 16 August 2015).
18. Neeb, H.; Zilles, K.; Shah, N.J. Fully-automated detection of cerebral water content changes: Study of age- and gender-related H₂O patterns with quantitative MRI. *NeuroImage* **2006**, *29*, 910–922.
19. Ashburner, J. A fast diffeomorphic image registration algorithm. *NeuroImage* **2007**, *38*, 95–113.
20. Brown, T.A. *Genomics*, 2nd ed.; Wiley: New York, NY, USA, 2002.
21. Radford, A.; Atkinson, M.; Britain, D.; Clahsen, H.; Spencer, A. *Linguistics: An Introduction*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1999.
22. Douaud, G.; Refsum, H.; de Jager, C.A.; Jacoby, R.; Nichols, T.E.; Smith, S.M.; Smith, A.D. Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 9523–9528.
23. Gao, Y.; Riklin-Raviv, T.; Bouix, S. Shape analysis, a field in need of careful validation. *Hum. Brain Mapp.* **2014**, *35*, 4965–4978.
24. The Brain Geek. Available online: <http://thebraingeek.blogspot.jp/2012/04/folds-of-brain.html> (accessed on 7 September 2015).
25. Geschwind, D.H.; Rakic, P. Cortical evolution: Judge the brain by its cover. *Neuron* **2013**, *80*, 633–647.
26. Sun, T.; Hevner, R.F. Growth and folding of the mammalian cerebral cortex: From molecules to malformations. *Nat. Rev. Neurosci.* **2014**, *15*, 217–232.
27. Keogh, E.; Wei, L.; Xi, X.; Lee, S.H.; Vlachos, M. LB_Keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures. In Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, 12–15 September 2006; pp. 882–893.
28. Tak, Y.S.; Hwang, E. A leaf image retrieval scheme based on partial dynamic time warping and two-level filtering. In Proceedings of the 7th IEEE International Conference on Computer and Information Technology, Fukushima, Japan, 16–19 October 2007; pp. 633–638.
29. Bartolini, I.; Ciaccia, P.; Patella, M. WARP: Accurate retrieval of shapes using phase of Fourier descriptors and time warping distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 142–147.
30. Skarda, C.A.; Freeman, W.J. Chaos and the new science of the brain. *Concepts Neurosci.* **1990**, *1*, 275–285.
31. Liebovitch, L.S. *Fractals and Chaos Simplified for the Life Science*; Oxford University Press: New York, NY, USA, 1998.
32. Van Straaten, E.C.W.; Stam, C.J. Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI. *Eur. Neuropsychopharmacol.* **2013**, *23*, 7–18.
33. Strogatz, S.H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2nd ed.; Westview: Cambridge, MA, USA, 2014.
34. Pham, T.D. Classification of complex biological aging images using fuzzy Kolmogorov-Sinai entropy. *J. Phys. D Appl. Phys.* **2014**, *47*, doi:10.1088/0022-3727/47/48/485402.
35. Gutierrez-Tobal, G.C.; Alvarez, D.; Gomez-Pilar, J.; del Campo, F.; Hornero, R. Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women. *Entropy* **2015**, *17*, 123–141.
36. Pan, W.Y.; Su, M.C.; Wu, H.T.; Lin, M.C.; Tsai, I.T.; Sun, C.K. Multiscale entropy analysis of heart rate variability for assessing the severity of sleep disordered breathing. *Entropy* **2015**, *17*, 231–243.
37. Pincus, S.M. Approximate entropy (ApEn) as a complexity measure. *Chaos* **1995**, *5*, 110–117.
38. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049.
39. Pham, T.D. Regularity dimension of sequences and its application to phylogenetic tree reconstruction. *Chaos Soliton. Fract.* **2012**, *45*, 879–887.
40. Eckmann, J.P.; Kamphorst, S.O.; Ruelle, D. Recurrence plots of dynamical systems. *EPL Europhys. Lett.* **1987**, *4*, 973–977.
41. Rosenstein, M.T.; Collins, J.J.; DeLuca, C.J. A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D Nonlinear Phenom.* **1993**, *65*, 117–134.
42. Williams, G.P. *Chaos Theory Tamed*; Joseph Henry Press: Washington, DC, USA, 1997.

43. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process* **1978**, *26*, 43–49.
44. Rabiner, L.R.; Juang, B. *Fundamentals of Speech Recognition*; Prentice-Hall: Upper Saddle River, NJ, USA, 1993.
45. Grassberger, P.; Procaccia, L. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593.
46. Eckmann, J.P.; Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **1985**, *57*, 617–656.
47. Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*; W.H. Freeman: New York, NY, USA, 1991.
48. Casdagli, M.C. Recurrence plots revisited. *Phys. D Nonlinear Phenom.* **1997**, *108*, 12–44.
49. Marwan, N.; Romano, M.C.; Thiel, M.; Kurths, J. Recurrence plots for the analysis of complex systems. *Phys. Rep.* **2007**, *438*, 237–329.
50. Facchini, A.; Mocenni, C.; Vicino, A. Generalized recurrence plots for the analysis of images from spatially distributed systems. *Phys. D Nonlinear Phenom.* **2009**, *238*, 162–169.
51. Dingwell, J.B. Lyapunov exponents. In *Wiley Encyclopedia of Biomedical Engineering*; Metin, A., Ed.; John Wiley & Sons: New York, NY, USA, 2006.
52. Pham, T.D. The butterfly effect in ER dynamics and ER-mitochondrial contacts. *Chaos Soliton. Fract.* **2014**, *65*, 5–19.
53. Pham, T.D. Validation of computer models for evaluating the efficacy of cognitive stimulation therapy. *Wirel. Pers. Commun.* **2015**, doi:10.1007/s11277-015-3017-7.
54. Takens, F. Detecting strange attractors in turbulence. *Lect. Notes Math.* **1981**, *898*, 366–381.
55. Ecker, J.G.; Kupferschmid, M. *Introduction to Operations Research*; John Wiley & Sons: New York, NY, USA, 1988.
56. Pham, T.D.; Oyama-Higa, M.; Truong, C.T.; Okamoto, K.; Futaba, F.; Kanemoto, S.; Sugiyama, M.; Lampe, L. Computerized assessment of communication for cognitive stimulation for people with cognitive decline using spectral-distortion measures and phylogenetic inference. *PLoS ONE* **2015**, *10*, e0118739, doi:10.1371/journal.pone.0118739.
57. Michener, C.D.; Sokal, R.R. A quantitative approach to a problem in classification. *Evolution* **1957**, *11*, 130–162.
58. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum: New York, NY, USA, 1981.
59. IXI (Information eXtraction from Images) Dataset. Available online: <http://www.brain-development.org> (accessed on 4 December 2015).
60. Giorgio, A.; Santelli, L.; Tomassini, V.; Bosnell, R.; Smith, S.; de Stefano, N.; Johansen-Berg, H. Age-related changes in grey and white matter structure throughout adulthood. *Neuroimage* **2010**, *51*, 943–951.
61. SPM: Statistical Parametric Mapping. Available online: <http://www.fil.ion.ucl.ac.uk/spm> (accessed on 11 March 2015).
62. Ashburner, J.; Friston, K.J. Unified segmentation. *Neuroimage* **2005**, *26*, 839–851.
63. Chen, Y.; Pham, T.D. Development of a brain MRI-based hidden Markov model for dementia recognition. *BioMed. Eng. Online* **2013**, *12*, S2, doi:10.1186/1475-925X-12-S1-S2.
64. Theodoridis, S.; Piskakis, A.; Koutroumbas, K.; Cavouras, D. *Introduction to Pattern Recognition: A Matlab Approach*; Academic Press: New York, NY, USA, 2010.
65. Sprott, J.C. *Chaos and Time-Series Analysis*; Oxford University Press: Oxford, UK, 2003.
66. Cardoza, J.D.; Goldstein, R.B.; Filly, R.A. Exclusion of fetal ventriculomegaly with a single measurement: The width of the lateral ventricular atrium. *Radiology* **1988**, *169*, 711–714.
67. Raz, N.; Lindenberger, U.; Rodrigue, K.M.; Kennedy, K.M.; Head, D.; Williamson, A.; Dahle, C.; Gerstorf, D.; Acker, J.D. Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cereb. Cortex* **2005**, *15*, 1676–1689.
68. Craik, F.I.M.; Salthouse, T.A. *The Handbook of Aging and Cognition*, 3rd ed.; Psychology Press: New York, NY, USA, 2008.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).