Journal of Neuroscience Methods 199 (2011) 140–145

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

MRI-based age prediction using hidden Markov models

Bing Wang, Tuan D. Pham*

Bioinformatics Research Group, School of Engineering and Information Technology, The University of New South Wales, Canberra ACT 2600, Australia

ARTICLE INFO

Article history:
Received 10 January 2011
Received in revised form 12 April 2011
Accepted 14 April 2011

Keywords:
MRI
Age prediction
Hidden Markov models
Wavelet transforms
Vector quantization
Kullback–Leibler divergence

ABSTRACT

Cortical thinning and intracortical gray matter volume losses are widely observed in normal ageing, while the decreasing rate of the volume loss in subjects with neurodegenerative disorders such as Alzheimer's disease is reported to be faster than the average speed. Therefore, neurodegenerative disease is considered as accelerated ageing. Accurate detection of accelerated ageing based on the magnetic resonance imaging (MRI) of the brain is a relatively new direction of research in computational neuroscience as it has the potential to offer positive clinical outcome through early intervention. In order to capture the faster structural alterations in the brain with ageing, we propose in this paper a computational approach for modelling the MRI-based structure of the brain using the framework of hidden Markov models, which can be utilized for age prediction. Experiments were carried out on healthy subjects to validate its accuracy and its robustness. The results have shown its ability of predicting the brain age with an average normalized age-gap error of two to three years, which is superior to several recently developed methods for brain age prediction.

1. Introduction

Age-related changes in brain morphology, including cortical thinning and gray matter (GM) atrophy, have been widely observed in ageing people. These brain changes can be firstly observed on the magnetic resonance imaging (MRI) in middle age (Salat et al., 2004; Chan et al., 2003). These alternations in the brain follow certain patterns. The morphological changes may be accelerated in some specific brain regions. An example includes the prefrontal cortex where age related changes can be greater than in other regions (Sowell et al., 2003). Meanwhile, neurodegenerative diseases, such as Alzheimer's disease (AD) and dementia accelerate brain tissue loss at a faster rate than the normal brain ageing process. Furthermore, recent reports (Vivek et al., 2006; Spulber et al., 2010) have revealed that brain tissue atrophy caused by AD is more regionally specific than normal ageing. These two factors exhibit properties that are distinguishable from the normal ageing brain morphology. Therefore, it is possible to estimate the stages of ageing according to the morphology observed on MRI data. The ability to predict deviations in brain morphology, from the normal ageing pattern, before the pathological onset has the potential of improving clinical diagnosis and treatment by early intervention (Christos et al., 2009; Spulber et al., 2010; Sluimer et al., 2009; Driscoll et al., 2009; Fotenos et al., 2008).

Several studies have been directed at detecting accelerated ageing in the brain: if the predicted age according to brain images is older than the subject's real age, then this could be evidence of fast ageing. It has been realised that in order to identify faster brain atrophy, the construction of a healthy MRI-based brain ageing model is required to validate the accuracy and robustness of the prediction (Franke et al., 2010).

There are four exploratory modelling methods for modelling neuronal ageing using MRI data. The first is support vector machines (SVM) (Lao et al., 2004) which assigns subjects into four age stages. The second is relevance vector machine for regression (RVR) based age prediction using principal component analysis for feature selection (Franke et al., 2010). The method was applied to 550 healthy subjects and a mean absolute error of 4.98 years was obtained. In (Ashburner, 2007), age prediction was estimated by using RVR, which yielded a root mean squared error of 6.5 years. The fourth example is quantitative brain water maps (BWM) which predicts brain age with a median absolute deviation of 6.3 years between real and predicted ages (Neeb et al., 2006).

All of these prediction methods are based on high dimensional morphological analysis (HDMA) (Fan et al., 2007). Although the HDMA approach has been widely applied, it has several technical challenges, including the need for large training data and effective feature selection. In order to overcome the limitation of the HDMA, and to detect subtle structural changes of the brain for identifying accelerated ageing, we propose to build a structural brain model for each subject by using the framework of hidden Markov models (HMMs). We then estimate the brain age of a target subject by com-

* Corresponding author. Tel.: +61 2 62688410.
E-mail addresses: bing.wang@student.adfa.edu.au (B. Wang), t.pham@adfa.edu.au (T.D. Pham).

0165-0270/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2011.04.022

puting the similarity between the constructed HMM of the subjects by using the Kullback–Leibler divergence.

The rest of the paper is organised as follows. In Section 2, we briefly present the mathematical components for building an HMM of the MRI-based brain structural model. Section 3 presents the concept of the Kullback–Leibler divergence as a basis for the HMM-based similarity analysis. Section 4 illustrates and discusses the performance of the proposed approach using a public MRI database. Finally, Section 5 is the summary of the findings, and highlights of some suggested issues for further study

2. HMM for MRI-based brain structural model

In order to construct an HMM for the brain model, we extract a useful feature of the MRI of the brain using the wavelet coefficients. The wavelet coefficients are then coded using a vector quantization technique for modelling the hidden states of the HMM. We also discuss other components of the HMM and finally provide a summary of the computational procedure for building the HMM for the MRI-based brain structural model.

2.1. Wavelet-feature extraction and symbol coding

The wavelet transform is a widely used signal processing tool which is often applied in multisolution analysis and image compression. It is capable of capturing data features based on different frequency bands. An example includes a 2D image: vertical, horizontal spatial frequency characteristics of the image. This procedure intuitively decomposes images into a number of scales, each of which represents a certain coarseness of the data under study. The discrete wavelet transform (DWT) is commonly used for signal analysis that decomposes original signals into approximation and oscillation (detail) coefficients. Approximation is a basic and smooth representation of the original signal, while oscillation provides the high frequency components within the original signal.

In brain MRI, DWT can be used to extract basic voxel intensity distribution patterns to characterise the distribution patterns of the brain tissue and CSF, while discarding the high frequency oscillations. The high frequency components may contain noise that is introduced by the registration and normalization steps, which human eyes cannot distinguish. The DWT can be applied on different scales of the brain MRI; for example, in each block of scattered MRI slices, on each slice of MRI or the whole volume. We are interested in exploring the local structure of the brain tissues and CSF. We also aim to capture the gradual structural changes within the brain, accompanied with ageing. Therefore, we implemented the DWT in each block of scattered MRI slices of every subject.

In the present work, only GM, WM and CSF regions of the brain MRI data were used for the wavelet feature extraction, as our purpose is to detect the structural and volume changes of different tissues rather than the intensity or density alterations. We divided each axial slice of a brain MR image into $N \times N$ blocks, rearranged the voxel values of each block into a vector to extract its wavelet approximation. The approximation of the DWT reflects how different tissues are spatially distributed in each block of the MR images. For example, when the atrophy of GM exists in one block, the position of different tissue indices would alter. Thus, the brain structural changes with ageing can be captured by this feature, and further reflected in the brain model we are going to build. In order to avoid the effect of noise and registration errors in the detail coefficients of the wavelet transform, we only preserved the wavelet approximation coefficients.

To implement the wavelet transforms, two sets of functions are involved in the procedure: scaling and wavelet functions representing low and high pass filters, respectively. The Daubechies wavelet transform was applied in our study. The high-pass and low-pass functions are given as follows.

$$y_{\text{high}}[k] = \sum_n x[n]g[2k-n] \tag{1}$$

$$y_{\text{low}}[k] = \sum_n x[n]h[2k-n] \tag{2}$$

where y_{high}, y_{low} are the outputs of the highpass g and lowpass h filters after resampling by two. The procedure can be repeatedly applied to the approximation of the last scale to produce approximation and detail on a coarser scale until a desired level is reached which forms a pyramidal structure (Mallat, 1989).

The wavelet coefficient vectors defined above are expected to be able to extract substantial information of the strcutures of the tissues and CSF. However, since we extract these vectors from each block where there is inevitable spatial relevance leading to redundancy. A data reduction step is required and discussed in the following procedure using a vector quantization (VQ) technique. Vector quantization is a data compression method, which utilizes codevectors to represent the source vectors in their proximity. VQ can reduce the amount of data, storage requirement and computational complexity. A set of codevectors which best represent a training dataset is called a codebook. Suppose we have a group of M source vectors. $\mathbf{T} = \{x_1, x_2, \ldots, x_M\}$, every vector is k-dimensional. $\mathbf{x_m} = (x_{m1}, x_{m2}, \ldots, x_{mk})$, $m = 1, 2, \ldots, M$. Let N be the number of codevectors and $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c_N}\}$, every $\mathbf{c_n}$ is k-dimensional as $\mathbf{x_m} : \mathbf{c_n} = (c_{n1}, c_{n2}, \ldots, c_{nk})$, $n = 1, 2, \ldots, N$. Let S_n represent the encoding space associated with code vector $\mathbf{C_n}$, $P = \{S_1, S_2, \ldots, S_N\}$ is the partition of the encoding space. If a source vector x_m is located in an encoding space S_n, then its approximation is C_n, denoted by $Q(\mathbf{x_m}) = c_n$. The average distortion is given by:

$$D_{\text{VQ}} = \frac{1}{MK} \sum_{t=1}^{M} (||\mathbf{x}_t - Q(\mathbf{x}_t)||)^2 \tag{3}$$

The VQ process can be shortly described as follows: given $\mathbf{T}$ and N, find $\mathbf{C}$ and P, such that D_{VQ} is minimized. However, $\mathbf{C}$ and P must follow the two criteria (Gersho and Gray, 1992). One is the nearest neighbour condition: the encoding region S_n consists of all vectors that are closer to $\mathbf{c_n}$ than any of the other codevectors. For example, if one vector is determined to belong to region S_n, then the distance between the vector and the center of S_n should be shorter than any distances between the vector and the centres of other regions. For vectors that are on the boundary of any region, a tie-breaking procedure can be applied to determine which region these vectors belong to. The other criterion is the centroid condition: codevector $\mathbf{c_n}$ should be the average of all training vectors that are in encoding region S_n. The most commonly used VQ method is the LBG algorithm (Gray, 1984). The LGB starts with an initial codebook, and then iteratively splits the training data into two codevectors until the desired number of the codevectors is reached. The data to be quantized in this study are the wavelet coefficient vectors which are extracted from the MRI data of each subject.

In order to compare the similarity between subjects, the code vectors should be consistent from subject to subject. However, the more the number of subjects increases, the more the computer memory requires. In order to avoid the problem of large computational load in the VQ design, we propose to build the same codebook for every two subjects for all the pairwise combinations. In other words, the codevectors of every two subjects are pooled together for the construction of the codebook, and then based on the feature (wavelet) vectors of each subject, the state transitions of each subject can be obtained from the pooled codebook. The number of code vectors was experimentally chosen to be 32, which is among

the range between 32 and 256 as it could provide the best trade-off between effective analysis and efficient computation.

2.2. HMM of MRI-based brain structure

Hidden Markov models has been well known for its ability of stochastic modelling in many pattern classification problems (Rabiner and Juang, 1989; Ibrahim et al., 2006; Ying et al., 2010). In our work, we use the framework of HMM to model the relationship of the distributions of the intracortical tissues, CSF, and GM in a dynamic fashion. The features of these brain components are expected to be useful for identifying the brain age of the subjects. Each HMM, denoted as λ, is characterised by a set of parameters: $\lambda = (A, B, \pi)$, where A denotes the transition matrix of the hidden states, B the matrix of the emission probabilities of the symbols within the hidden states, and π the probability distribution of the initial hidden states. An observation sequence usually carries observable features in the original system. In brain scans, cortical thinning and GM volume decreasing are the observable alterations with ageing which could be seen as shown in Fig. 1. Furthermore, the change of the brain tissues with normal ageing has been statistically validated based on large samples (Salat et al., 2004). Thus, in the context of age estimation, the GM portion in each block of the MRI slices can be used for constructing the observation sequences. The sequences of both observation and hidden states are extracted from the whole volume of every subject. Therefore, each subject can be represented by an HMM. The orders of the two sequences are the same: on each axial slice, the order is from left to right, row by row; and the slices were selected from superior to inferior.

In our study, the tissues and CSF distribution patterns are used as the hidden states that are represented by the VQ indices of the wavelet coefficient vectors. Using the VQ indices in each block of an image and the corresponding GM volume within the same block, we can obtain the initial transition probability distribution. The transition distribution of the HMM describes the changes in pattern of the brain tissues and CSF when the GM volume is observed in each image block according to the order mentioned above. This spatial transition is of the basic structure of the tissues and CSF in different locations of the human brain. The three parameters of an HMM are then re-estimated using the Baum–Welch algorithm which is also referred to as the forward–backward algorithm (Rabiner and Juang, 1989).

The definition of an HMM is as follows. Let N be the number of states, M the number of observation symbols, $V = \{v_1, v_2, \ldots, v_M\}$ the set of distinct M individual symbols, $O = (o_1, o_2, \ldots, o_T)$ an observation sequence, where T is the number of observations. We denote $A = \{a_{ij}\}$ to be the state-transition probability distribution, $B = \{b_j(k)\}$ the observation symbol distribution, and $\pi = \{\pi_i\}$ the initial state distribution. These three probabilistic measures are mathematically defined as follows.

$$a_{ij} = P(q_{t+1} = j | q_t = i), \quad 1 \le i, \quad j \le N$$

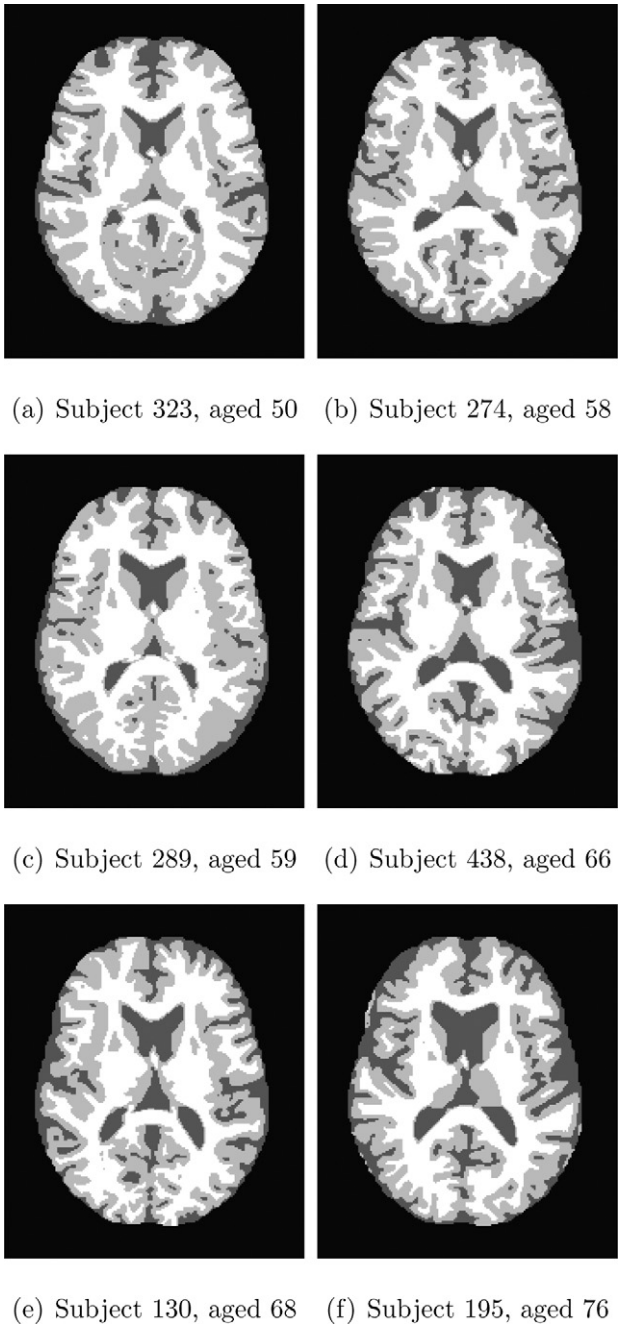
where q_t denotes the state at time t, and a_{ij} is the probability of state i going to state j.

$$b_j(k) = P(o_t = v_k | q_t = j), \quad 1 \le j \le N; \quad 1 \le k \le M$$

which is the probability of symbol v_k being in state j. The initial state probability is given as

$$\pi_i = P(q_1 = i), \quad 1 \le i \le N.$$

After construction of each HMM, we are interested in calculating the probability measure for an observation sequence O given the model λ, that then can be used for the similarity comparison between the two subjects.



(a) Subject 323, aged 50 (b) Subject 274, aged 58

(c) Subject 289, aged 59 (d) Subject 438, aged 66

(e) Subject 130, aged 68 (f) Subject 195, aged 76

Fig. 1. Gray matter (GM) alteration in ageing brains: GM volume of subject 438 (aged 66) has remarkable tissue loss, while subject 323 (aged 50) has relatively more GM.

3. Similarity measure of HMMs of brain structure

To compute the similarity between the HMMs of every two subjects, the Kullback–Leibler divergence (KLD) is applied herein. Suppose $\lambda_1 = (A_1, B_1, \pi_1)$, and $\lambda_2 = (A_2, B_2, \pi_2)$ are two HMMs of the two subjects. The KLD allows the calculation of the probabilistic distance between two probability distributions. This distance can represent the similarity between the two subjects, and is defined as

$$d(\lambda_1, \lambda_2) = 1 - \exp[D_s(\lambda_1, \lambda_2)] \tag{4}$$

where D_s is the symmetrised version of the approximate KLD of λ_1 and λ_2, which is expressed as

$$D_s(\lambda_1, \lambda_2) = \frac{D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1)}{2} \quad (5)$$

in which $D(\lambda_1, \lambda_2)$ is the empirical KLD between λ_1 and λ_2, which was introduced by Juang and Rabiner (1985). The models are assumed to be ergodic, having arbitrary observation densities, and the dissimilarity is defined as the mean divergence of the observation sample. This approximate KLD is given by (Juang and Rabiner, 1985)

$$D(\lambda_1, \lambda_2) = \frac{1}{T_2} \log \frac{P(O_{\lambda_2}|\lambda_1)}{P(O_{\lambda_2}|\lambda_2)} \quad (6)$$

where $O_{\lambda_2} = (o_1 o_2 \ldots o_{T_2})$ is a sequence of observations generated by model λ_2, and T_2 is the length of sequence O_{λ_2}.

The above equation represents how well model λ_1 describes the observation sequence that is used to construct model λ_2, relative to how well model λ_2 describes the observations used to construct itself. We can similarly define $D(\lambda_2, \lambda_1)$ as

$$D(\lambda_2, \lambda_1) = \frac{1}{T_1} \log \frac{P(O_{\lambda_1}|\lambda_2)}{P(O_{\lambda_1}|\lambda_1)} \quad (7)$$

where $O_{\lambda_1} = (o_1 o_2 \ldots o_{T_1})$ is a sequence of observed symbols generated by model λ_1.

Finally, the brain age can be predicted by selecting the known-age brain model which has the smallest distance to the unknown model. The procedure for MRI-based age prediction can be summarised as follows.

Procedure for MRI-based age prediction

1. Extract brain tissues and CSF distribution patterns.
2. Extract wavelet coefficients (approximate) of tissues and CSF.
3. Build VQ codebook of the wavelet coefficients.
4. Estimate the transition distribution of the hidden states using the VQ codebook.
5. Extract observation sequences based on the volumetric measurements of the GM.
6. Estimate the observation symbol distribution based on the hidden state distribution and obervation sequences.
7. Assume equal initial probability distribution of the hidden states.
8. Reestimate $\lambda = (A, B, \pi)$ using the Baum–Welch algorithm.
9. Estimate brain age using the KLD.

4. Experiment

To test the proposed approach, we used the brain MR images of healthy subjects from the publicly accessible ABSIS database (http://www.oasis-brains.org/), which contains 20 subjects considered as suitable in this study as they were cognitively normal, have complete MRI scans, clinical test records and their age gap is not very large. These subjects are of ages from 50 to 86. We used the leave-one-out method to predict the brain age of each subject by comparing the model of each subject with those of the other 19 subjects. To test if smaller datasets can give similar results, we divided the dataset into two different subsets. The first and second subsets contain 11 and 8 subjects, respectively. In each subset, we followed the same leave-one-out method to predict the brain age.

Human brains differ in shape and size; and their MRI intensity is sensitive to all acquisition conditions, including MR protocols, MR scanners, and MR adjustments (Collewet et al., 2004). Therefore, in order to ensure the analysis accuracy, we chose skull-eliminated, normalized, atlas-registered, 3-D T1 weighted images of the 19 subjects. Furthermore, to avoid the effect of intensity variation, we used the regions of interest in the images of the subjects, which consist of only three components in the image: GM, WM and CSF. Examples of the images of some subjects are shown in Fig. 1.

Table 1
Age prediction of 11 heathy subjects (first subset): normalized mean absolute error of age prediction = 2.47 years, normalized average age gap = 2.5 years.

Subject ID	Actual age	Predicted age
2	55	55
10	74	55
11	52	55
85	70	65
130	68	65
195	76	65
199	69	76
207	51	55
208	55	55
241	74	69
322	65	76

Each brain scan was partitioned into 16×16 blocks, which allows a sufficient smallest size for the extraction of the wavelet features while maximise the information for modelling the state and obervation transitions of the brain structural properties. In this study, the voxel values of the tissues and CSF within each image block were re-arranged into vectors, and then one-dimensional wavelet decomposition was applied on four levels to each of the vectors. Because there was a large amount of extracted wavelet coefficient vectors, VQ was employed to reduce the number of the wavelet-coefficient vectors. Having mentioned before, the number of the hidden states in our model was experimentally chosen to be 32.

The root mean squared error (RMSE) was used to evaluate the accuracy of the age prediction, and expressed as follows.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i (g_i' - g_i)^2}{n}} \quad (8)$$

where n is the number of samples, g_i the real age, and g_i' the predicted age. To be more meaningful in the interpretation of the measure of the prediction accuracy of the brain age (Franke et al., 2010), the mean absolute error is commonly used and defined as

$$\mathrm{MAE} = \frac{\sum_i |g_i' - g_i|}{n} \quad (9)$$

Both RMSE and MAE are reasonable measures of the average prediction errors, due to the use of large databases which cover continuous ages in the age range which allows age gap not being greater than one. In our case, as the complete age samples were not available, we therefore introduce an average age gap to normalize the age-gap variation. The average normalized age-gap error of prediction, denoted as ANGE, is defined as

$$\mathrm{ANGE} = \frac{\sum_i |g_i' - g_i|}{m \times n} \quad (10)$$

where $m > 0$ is the average age gap of all samples. When $m = 1$ (10) becomes identical to (9).

Table 1 shows the results of the age prediction of the first subset (11 subjects), in which the normalized age gap provided by the proposed method is 2.5 years. In the second subset (8 subjects), the normalized age gap is 3.71 years as showed in Table 2, and the value in all 20 subjects is 2.57 years showed in Table 3. It can be seen that the proposed approach in all the three experiments provided similar prediction accuracies. Fig. 2 shows the bar chart of the prediction accuracies of the proposed approach in the three experiments and the comparisons with other results obtained from other three studies. The first bar shows the ANGE (2.47 years) of the proposed method (HMM) in the first subset containing 11 subjects, the second bar shows the ANGE of 2.16 years obtained by the HMM using the second subset containing 8 subjects, the ANGE of 2.41

Table 2
Age prediction of 8 heathy subjects (second subset): normalized mean absolute error of age prediction = 2.16 years, normalized average age gap = 3.71 years.

Subject ID	Actual age	Predicted age
85	70	76
116	52	55
195	76	66
323	50	66
357	55	52
438	66	76
130	68	76
274	58	66

Table 3
Age prediction of 20 heathy subjects: normalized mean absolute error of age prediction = 2.41 years, normalized average age gap = 2.57 years.

ID	Actual age	Predicted age	ID	Actual age	Predicted age
2	55	66	241	74	71
10	74	86	274	58	66
11	52	52	280	78	71
85	70	65	317	86	68
116	52	52	322	65	76
130	68	65	323	50	51
195	76	65	356	68	66
199	69	76	357	55	52
207	51	65	398	71	78
208	55	55	438	66	65

years given by the HMM using the whole database is shown by the third bar, the fourth bar shows the MAE (4.98 years) of age prediction obtained by the RVR method applied in (Franke et al., 2010), the fifth bar shows the age prediction accuracy (6.3 years) using the BWM reported in (Neeb et al., 2006), and finally the last bar shows the RMSE (6.5 years) using the RVR method reported in (Ashburner, 2007). The results obtained from the three experiments have shown the superior performance of the proposed method to the HDMA-based algorithms.

The approach used in (Franke et al., 2010) and (Ashburner, 2007) is a relevance vector regression model for brain age prediction. Models of dependency of a real age on the brain morphology require sufficient training data to ensure the accurate age prediction for new subjects. Such required training samples are expected to cover most age stages, which are costly to acquire. The advantages of our proposed approach are two-fold. Firstly, its classification is based on a KLD measure which is much simpler for computer implementation than other existing methods for MRI-based brain age prediction. Secondly, our proposed approach has been able to effectively perform the prediction by using much smaller training samples: the construction of each HMM for the representation of a particular age requires only at least one MRI dataset for the training of the HMM parameters. Furthermore, in comparison with the study in (Neeb et al., 2006) using cerebral water content map for the age prediction, we extracted the gray matter atrophy information as the features of ageing, which has been proved to be a linear changing factor associating to ageing. The incorporation of this factor can help improve the indicator of ageing.

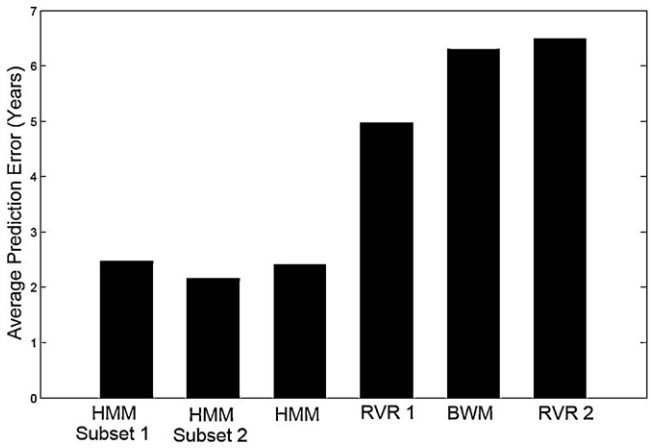


Fig. 2. Comparison of age prediction accuracy (units are in years of age): first bar (proposed method using 11 subjects), second bar (proposed method using 8 subjects), third bar (proposed method using 20 subjects), fourth bar (Franke et al., 2010), fifth bar (Neeb et al., 2006), and sixth bar (Ashburner, 2007).

5. Conclusion

We have presented an effective method for predicting ageing rate using MRI of the brain. The experimental result have shown the most favourable performance of the proposed method in comparison with other methods in that the former can model the brain structure with a small sample size (as small as one MRI for each age), and applies a much simpler classification rule for the prediction task. Another advantage of the proposed modelling is the selection of an effective feature of the brain. Although other tissues, including white matter and cerebral spinal fluid, change their morphological patterns with ageing (Pham et al., 2010; Zhang et al., 2006); GM atrophy, however, is known for playing the most leading role in the ageing brain (linearly progresses with ageing). We therefore used only gray matter as the selected feature of ageing in the HMM and have achieved superior results

The accuracy of the proposed method does not seem to suffer when reducing the training set from 20 subjects to 11; thus suggesting the method can work well without the requirement of a large training dataset, which is not easy to acquire in practice. However, neither 11 nor 20 subjects used in this study conform a large enough training dataset. In other words, a larger dataset could better represent the variability among subjects of the same age, and thus better accuracy could be expected.

In the present study, we used the average normalized age-gap error (ANGE) to evaluate the accuracy of brain age prediction. However, it is likely that the minimum (youngest) and maximum (oldest) ages of the subjects do not have enough samples for comparison. Neither the ANGE nor other existing measures of the age prediction accuracy could address this issue. Therefore, a more appropriate measure of the age prediction accuracy should be developed to handle such a problem. This is an extrapolation of the brain age prediction problem and quite a challenging techinical issue. As another issue, we only used the wavelet coefficients as the feature of the brain MRI in the HMM; thus the use of other effective image features may improve the age prediction accuracy. This work is currently underway in our on-going research in computational neuroscience.

Acknowledgement

The authors would like to express their gratitude to the two anonymous reviewers, whose valuable comments and suggestions help improved the quality of the revised manuscript.

References

Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage 2007;38(1):95–113.
Chan D, Janssen J, Whitwell J, Watt H, Jenkins R, Frost C, et al. Change in rates of cerebral atrophy over time in early-onset Alzheimer's disease: longitudinal MRI study. The Lancet 2003;362(9390):1121–2.
Christos D, Feng X, Yang A, Yong F, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain 2009;132(8):2026–35.

Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magnetic Resonance Imaging 2004;22(1):81–91.
Driscoll I, Davatzikos C, An Y, Wu X, Shen D, Kraut M, et al. Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology 2009;72(22):1906.
Fan Y, Shen D, Gur R, Gur R, Davatzikos C. Compare: classification of morphological patterns using adaptive regional elements. IEEE Transactions on Medical Imaging 2007;26(1):93–105.
Franke K, Ziegler G, Klöppel 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(3):883–92.
Fotenos AF, Mintun MA, Snyder AZ, Morris JC, Buckner RL. Brain volume decline in aging: evidence for a relation between socioeconomic status, preclinical Alzheimer disease, and reserve. Archives of Neurology 2008;65(1):113.
Gersho A, Gray RM. Vector quantization and signal compression. Netherlands: Springer; 1992.
Gray R. Vector quantization. IEEE ASSP Magazine 1984;1(6):4–29.
Ibrahim M, John N, Kabuka M, Younis A. Hidden Markov models-based 3D MRI brain segmentation. Image and Vision Computing 2006;24(10):1065–79.
Juang B, Rabiner L. A probabilistic distance measure for hidden Markov models. ATT Technical Journal 1985;64(2):391–408.
Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 2004;21(1):46–57.
Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989;11(7):674–93.
Neeb H, Zilles K, Shah NJ. Fully-automated detection of cerebral water content changes: study of age- and gender-related H_2O patterns with quantitative MRI. NeuroImage 2006;29(3):910–22.
Pham TD, Elfiqi HZ, Knecht S, Wersching H, Baune BT, Berger K. Structural simplexity of the brain. Journal of Neuroscience Methods 2010;188(1):113–26.
Rabiner L, Juang B. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 1989;77(2):257–86.
Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RDR, Busa E, et al. Thinning of the cerebral cortex in aging. Cerebral Cortex 2004;14(7):721–30.
Sluimer JD, Van Der Flier WM, Karas GB, van Schijndel R, Barnes J, Boyes RG, et al. Accelerating regional atrophy rates in the progression from normal aging to Alzheimer's disease. European Radiology 2009;19(12):2826–33.
Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. Nature Neuroscience 2003;6(3):309–15.
Spulber G, Niskanen E, MacDonald S, Smilovici O, Chen K, Reimanet EM, et al. Whole brain atrophy rate predicts progression from MCI to Alzheimer's disease. Neurobiology of Aging 2010;31(9):1601–5.
Vivek S, Howard C, Lerch JP, Evans AC, Dorr AE, Kabani NJ. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. Brain 2006;129(11):2885–93.
Ying W, Susan R, Christos D. Spatio-temporal analysis of brain MRI images using hidden Markov models. In: Jiang T, Navab N, Pluim J, Viergever M, editors. Medical image computing and computer-assisted intervention—MICCAI 2010. Lecture notes in computer science, vol. 6362; 2010. p. 160–8.
Zhang L, Liu JZ, Dean D, Sahgal V, Yue GH. A three-dimensional fractal analysis method for quantifying white matter structure in human brain. Journal of Neuroscience Methods 2006;150(2):242–53.