CrossMark

A new image-based immersive tool for dementia diagnosis using pairwise ranking and learning

Wei Huang[1] · Shuru Zeng[1] · Jing Li[1] · Guang Chen[2]

Abstract Dementia disease is globally acknowledged as one of the most severe non-communicable diseases nowadays. Identifying different stages of dementia disease is significant in its later treatment for delaying the onset and progression of the disease. Among diverse types of tools utilized in dementia disease diagnosis, brain scanning is generally accepted as an effective and affordable way at present. There are several kinds of medical images incorporated in contemporary dementia studies, and magnetic resonance images receives vast popularity. In this study, arterial spin labeling, an emerging perfusion functional-magnetic resonance imaging technique, is adopted in a newly proposed image-based immersive tool for dementia disease diagnosis. Novel pairwise ranking and learning techniques based on a new continuous and differentiable surrogated Kendall-Tau rank correlation coefficient is proposed to realize the immersive tool. Extensive experiments based on a database composed of images acquired from 350 demented patients are carried out with several popular pattern recognition diagnosis tools being compared. Their results undergo rigorous and comprehensive statistical analysis, and the superiority of the newly proposed image-based immersive tool in dementia disease diagnosis has been demonstrated.

✉ Wei Huang
huangwei@ncu.edu.cn

Shuru Zeng
floraru@163.com

Jing Li
jing.li.2003@gmail.com

Guang Chen
chenguang322@gmail.com

[1] School of Information Engineering, Nanchang University, Nanchang, China

[2] Xian Communications Institute, Xi'an, China

1 Introduction

Dementia disease is widely acknowledged as a broad category of brain functions degeneration, which may result in gradual decrease of normal capabilities of thinking, memory, language, motivation, etc, for ordinary people in a long term. There are various forms of dementia disease, including vascular dementia, lewy body dementia, frontotemporal dementia, etc [29]. Among them, Alzheimer's Disease (AD) is generally regarded as the most common form. According to statistics provided by the World Health Organization, AD is often diagnosed worldwide in patients over 60 years old, and is now considered as one of the five most severe non-communicable diseases in the whole world (i.e. others include cardiovascular disease, cancer, diabetes and chronic lung disease) [29]. According to another population study conducted by the United Nations, there are more than 26.6 million AD patients diagnosed globally [28], and 1 in 85 worldwide people is predicted to be suffering from AD by the year 2050 [3]. Thus, dementia disease becomes an actual threat in many countries of aging societies nowadays. Accurate diagnosis and timely treatment is essential to delay the onset and progression of dementia disease [3].

Identifying the progression of dementia disease into various stages accurately is often of great importance to understand mechanisms of the disease, making correct treatments to corresponding symptoms of the disease possible at a later stage [3]. Thus, in order to perceive the progression of dementia disease accurately in clinical diagnosis, a variety of methods have been proposed and utilized to date. Popular diagnosis methods include pathography analysis, cognitive examination, brain scanning, etc. Pathography is helpful to predict curable symptoms of demented patients who may usually suffer from other forms of diseases (e.g., stroke, heart disease, renal failure, etc) at the same time [20]. Cognitive examination evaluates the progression of demented patients through a series of cognition tests based on diverse cognitive capabilities of patients, including short-memory, long-memory, logic analysis, etc [8, 22]. Popular cognitive examinations include Mini-Mental State Examination (MMSE) [8] and Addenbrooke's Cognitive Examination (ACE) [22]. Although these cognitive exams require few trainings for clinicians and are relatively easy to be carried out by them, outcomes of those exams could be highly biased by patients specialities. For example, patients of high-level education suffering from dementia disease are more likely to outperform ordinary patients of low-level education without dementia disease in those cognitive exams. For brain scanning, it is accepted as an effective and affordable way in dementia diagnosis nowadays. There are several imaging tools incorporated for dementia diagnosis, including Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), etc. Among them, MRI receives vast popularity, because of its prominent capability in both generating high-resolution images of brain tissues and free of ionizing radiation exposure, compared with other scanning tools such as CT and PET. Most contemporary MRI scanning techniques can be categorized into structural MRI (sMRI) and functional MRI (fMRI), and both of them have already been adopted in various dementia studies nowadays [21, 23].

In this study, a novel fMRI-based immersive tool based on new raking techniques is introduced for dementia diagnosis for the first time. Generally speaking, the purpose of ranking is to sort a list of items according to a system of rating or a record of performance, and most ranking approaches can be categorized into pointwise ranking approaches and pairwise ranking approaches. For pointwise ranking, provided an image list

$d = \{d_1, d_2, \ldots, d_n\}$, pointwise ranking aims to assign each image a discrete category: $\{(d_1, c_1), (d_2, c_2), \ldots, (d_n, c_n)\}$, in which $\{c_1, c_2, \ldots, c_n\} \in C$; $C = \{c_1 \succ c_2 \succ \ldots c_m\}$ is a set of m (where $m \geq n$) ordered categories, in which $\succ$ denotes an order between various categories. Since elements in C are ordered discrete values, pointwise ranking is also known as ordinal regression, which is between regression (outputs: real values that can be ordered) and classification (outputs: non-ordered discrete values) [5, 13]. Representative pointwise ranking approaches include constrained ordinal regression [5], Pranking [6], OAP-BPM [12], ranking with large margin principals [26], etc. Although pointwise approaches are convenient to implement due to their close resemblance to both regression and classification, their drawbacks are obvious: they can only deal with judgements in the form of absolute values. Non-absolute preference, such as pairwise preference and partial/full list orders, cannot be handled by those pointwise ranking approaches. For pairwise ranking, however, it focuses on data "pairs", instead. Take images for illustration purposes, in the learning stage of pairwise ranking approaches, image pairs (d_α, d_β) $(\alpha, \beta \in \{1, 2, \ldots, n\}, \alpha \neq \beta)$ are collected from an image list $d = \{d_1, d_2, \ldots, d_n\}$. For each pair, a label $r \in \{+1, -1\}$ is assigned indicating the order of two images (d_α, d_β): $+1$ shows d_α should be ranked before d_β, while -1 to the contrary. The main idea of pairwise ranking approaches is similar towards that of the well-known binary-class classification, and pairwise ranking approaches often formulate the ranking task as a classification problem accordingly [4, 9, 15]. Conventional classification methods, such as boosting, Support Vector Machine (SVM), and Artificial Neural Network (ANN), have been incorporated leading to corresponding pairwise ranking methods, such as RankBoost [9], RankingSVM [15], and RankNet [4], respectively. There are several advantages with pairwise ranking approaches compared with pointwise ranking approaches. First, existing classification methodologies can be conveniently adopted in pairwise ranking approaches [4, 9, 15]. Second, pairwise preference, rather than absolute relevance, is relatively easy to obtain under certain circumstances [15], making pairwise approaches more adaptive to be utilized in practical applications.

In this paper, a novel pairwise ranking technique and it associated learning method will be investigated in the newly proposed fMRI-based immersive tool for dementia diagnosis. Main contributions of this paper can be revealed as follows: 1) Technically, new pairwise ranking and learning techniques are introduced in this paper for the first time; 2) It is also the first attempt to perform dementia disease diagnosis from the new perspective of pairwise ranking, rather than conventional classification or clustering perspectives. The organization of this paper is described as follows. First, the conventional position-based ranking evaluation measure, Kendall-Tau rank correlation coefficient, is introduced in Section 2. The Kendall-Tau coefficient forms the theoretical basis of the new pairwise ranking technique introduced in this study. The learning of ranking functions in the pairwise ranking procedure is to be realized via an optimization process based on the Kendall-Tau coefficient, which is, however, neither continuous nor differentiable with respect to its discrete pair-counting terms. Hence, a new continuous and differentiable surrogate Kendall-Tau coefficient is proposed and its corresponding ranking functions learning method based on the new surrogate coefficient is introduced therein. All the above novel techniques and essential derivations are explicitly elaborated in Section 2. In Section 3, extensive experiments are conducted to evaluate the performance of the newly introduced image-based immersive tool for dementia diagnosis. fMRIs acquired from 350 real patients of different dementia disease progressions are utilized to construct a database for performance evaluation. The newly proposed tool has been compared with several other tools based on well-established pattern recognition techniques, and all experimental results of all compared methods are evaluated from the statistical point of view. Finally, the conclusion of this study is drawn in Section 4.

2 Methodology

2.1 Kendall-Tau rank correlation coefficient

The Kendall-Tau rank correlation coefficient is a conventional ranking performance evaluation measure named after the British statistician Sir Maurice Kendall [18]. Take images for illustration purposes, the definition of Kendall-Tau rank correlation coefficient (i.e., KT) is described in (1).

$$\text{KT} = \frac{N}{N_n} = \frac{P - Q}{N_n} = \frac{P - Q}{\frac{1}{2}n(n - 1)} \tag{1}$$

where, P and Q denote pairwise comparisons in a ranked images list. To be specific, P and Q are counting numbers of concordant image pairs and discordant image pairs in a ranked images list, respectively. N_n is a normalization term denoted by the number of image pairs in a ranked list consisting of n images (i.e., Thus N_n is equivalent to the number of 2-combinations from n images: $C_n^2 = \frac{n!}{(n-2)!\cdot 2!} = \frac{1}{2}n(n - 1)$).

The idea of concordant/discordant pairs can be described in Fig. 1. Provided an image pair (x, y) where images x and y both record dementia disease in this study, x contains more serious disease severity than y (i.e., $x > y$ in Fig. 1). Suppose x and y are included in the same images list together with other images which also record the dementia disease, and the listed images are ranked in a descending order of the disease severity. When x is ranked before y, it matches the fact that x contains more severe disease severity than y, and (x, y) forms a concordant pair thereafter (i.e., illustrated in Fig. 1). Otherwise, (x, y) forms a discordant pair.

Generally speaking, the range of KT is within $[-1, +1]$, and higher KT values indicate better ranking performance within the ranked list. In this study, KT is picked to evaluate the ranking performance, as the characteristics of data pairs is naturally inherited in this conventional position-based ranking evaluation measure, making it suitable for pairwise ranking. Based on KT, function learning can be carried out within an optimization process. However, the optimization cannot be executed directly on KT for ranking functions learning, because the original KT is neither continuous nor differentiable in terms of its discrete pair-counting terms P and Q. Hence, a new continuous and differentiable rank correlation measure is necessary for functions learning in the pairwise ranking.

2.2 Continuous and differentiable surrogate Kendall-Tau coefficient

In order to obtain a continuous and differentiable surrogated measure based on the original KT, terms P and Q in (1) are represented mathematically first. Given two images $x = (\Re_x, \ell_x)$ and $y = (\Re_y, \ell_y)$, where $\Re_{(x)}$ and $\ell_{(x)}$ denote extracted low-level visual features
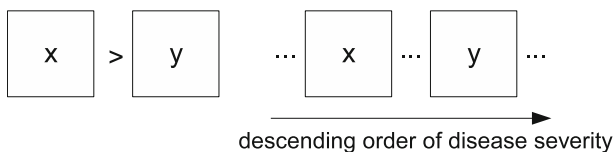

Fig. 1 An illustration of concordant pairs in the Kendall-Tau rank correlation coefficient (1) for dementia disease diagnosis

from image x and annotated disease severity of image x by clinicians, respectively. P and Q in (1) can be re-written via (2) and (3).

$$\text{Concordant pair } (P): \quad \text{sgn}(s_{(x,y)}) \cdot \text{sgn}(\ell_{(x,y)}) = 1 \tag{2}$$

$$\text{Discordant pair } (Q): \quad \text{sgn}(s_{(x,y)}) \cdot \text{sgn}(\ell_{(x,y)}) = -1 \tag{3}$$

where, $\text{sgn}(\cdot)$ is a sign (or signum) function, whose outcome is $+1$ when its variable is non-negative and -1 otherwise; $s_{(x,y)}$ is a pre-defined function measuring the disease severity difference between images x and y as the following form in this study: $s_{(x,y)} = < a, \Re_x - \Re_y >$, in which $<, >$ denotes an inner product between vector a and the feature vector difference between images x and y; $\ell_{(x,y)}$ is the direct difference of annotated disease severities by clinicians between x and y: $\ell_{(x,y)} = \ell_x - \ell_y$. Hence, the vector a actually performs a scaling on the feature space, and elements within the vector a are parameters to learn in this study.

Provided image x contains more severe dementia disease than image y, the disease severity in image x reflected by $s_{(x)} = < a, \Re_x >$ should be higher than that in image y (i.e. $s_{(x)} > s_{(y)}$). Meanwhile, the disease severity of image x annotated by clinicians should also be higher than that of image y (i.e. $\ell_{(x)} > \ell_{(y)}$). If the above conditions holds, (x, y) constitutes a concordant pair, P increases by 1 as indicated by (2) (i.e., $s_{(x,y)} = s_{(x)} - s_{(y)} > 0$ and $\ell_{(x,y)} = \ell_x - \ell_y > 0$). The above explanation is also valid when x contains less severe disease than image y and (x, y) constructs a concordant pair in a ranked images list of an ascending order of disease severities. Otherwise, (x, y) forms a discordant pair, (3) holds, and Q increases by 1.

After substituting terms P and Q into the original KT coefficient in (1), it can be further re-written as (4).

$$\text{KT} = \frac{1}{N_n} \sum_{x, y \in D, x \neq y} \big(\text{sgn}(s_{(x,y)}) \cdot \text{sgn}(\ell_{(x,y)}) \big) \tag{4}$$

in which, D denotes all images to be ranked. However, the above equation is still not ready for direct optimizations because of the well-known step transition characteristics of sign functions in (4). Therefore, the problem is further handled by approximating the original discrete sign function using a continuous hyperbolic tangent function. An illustration of the above approximation is shown in Fig. 2. The detailed approximation is explained in (5), where ξ denotes the variable of corresponding functions.

$$\text{sgn}(\xi) \simeq \tanh(\xi) = \frac{\sinh(\xi)}{\cosh(\xi)} = \frac{\frac{e^\xi - e^{-\xi}}{2}}{\frac{e^\xi + e^{-\xi}}{2}} = \frac{e^\xi - e^{-\xi}}{e^\xi + e^{-\xi}} = \frac{e^{2\xi} - 1}{e^{2\xi} + 1} \tag{5}$$

When incorporating (5) into (4), a novel continuous and differentiable surrogate Kendall-Tau coefficient, named surrogate Kendall-Tau coefficient (SKT), can be written as follows.

$$\text{SKT} = \frac{1}{N_n} \cdot \sum_{x, y \in D, x \neq y} \left(\frac{\exp\big(2(s_{(x,y)})\big) - 1}{\exp\big(2(s_{(x,y)})\big) + 1} \cdot \frac{\exp\big(2(\ell_{(x,y)})\big) - 1}{\exp\big(2(\ell_{(x,y)})\big) + 1} \right) \tag{6}$$

It can be easily observed that, the above SKT coefficient avoids discrete and non-differentiable problems inherited in the original KT coefficient, which makes direct optimizations feasible on SKT for the following functions learning.

2.3 Functions learning in pairwise ranking

A corresponding functions learning algorithm via direct optimization based on SKT via gradient ascent is elaborated in Table 1. The most critical step in this algorithm is to calculate
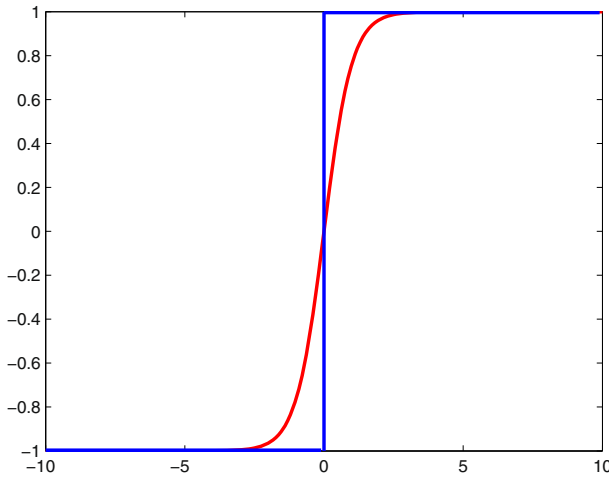
Fig. 2 An illustration of approximating a discrete sign function (*in blue*) via a continuous hyperbolic tangent function (*in red*)

the gradient of SKT with respect to the parameter to learn a (i.e. $\bigtriangledown$SKT(a)) in Steps T4 and T5. Detailed derivation is explained and demonstrated as follows.

In (6), the second term $\frac{\exp(2(\ell_{(x,y)}))-1}{\exp(2(\ell_{(x,y)}))+1}$ within the brackets of the Right Hand Side (RHS) of SKT can be treated as a coefficient in the derivation of the gradient, since this term is not related to parameters a to learn. Thus, the following derivation only focuses on the first

Table 1 An algorithm of functions learning based on SKT via pairwise ranking

Inputs	Images for training $\{x \in \chi\}$; Images for validation $\{x_v \in \chi_v\}$;
	Number of Iterations T; Learning rate η.
Training	
T1.	Initialize the parameter a in function $s_{(x)}$ as a_0
T2.	For t = 1 to T
T3.	Set $a = a_{t-1}$
T4.	Feed $\{x \in \chi\}$ to (8) to calculate the gradient
T5.	Update a via the gradient ascent: $a = a + \eta \cdot \bigtriangledown$SKT$(a)$
T6.	Set $a_t = a$
T7.	End for T2
Validation	T learned functions $s_{(x)}$ with T corresponding learned parameters a
V1.	For j = 1 to T
V2.	Feed j^{th} learned function $s_j(x)$ to $\{x_v \in \chi_v\}$ for ranking validation images
V3.	Calculate its corresponding SKT value using (6)
V4.	End for V1
V5.	Determine a_{opt} and its corresponding function $s_{opt}(x)$ with the highest SKT
Outputs	Learned function: $s_{opt}(x)$

term $\frac{\exp(2(s_{(x,y)}))-1}{\exp(2(s_{(x,y)}))+1}$ of RHS in SKT. For the ease of writing, the second term $\frac{\exp(2(\ell_{(x,y)}))-1}{\exp(2(\ell_{(x,y)}))+1}$ is denoted as term $coeff$. After applying the differentiation, the gradient of SKT(a) can be further derived as follows.

$$
\begin{aligned}
\bigtriangledown \text{SKT}(a) &= \frac{1}{N_n} \cdot \left(\sum_{x,y \in D, x \neq y} \frac{2 \cdot \exp\left(2s_{(x,y)}\right)\left(s_{(x,y)}\right)' \cdot \left(\exp\left(2s_{(x,y)}\right)+1\right)}{\left(\exp\left(2s_{(x,y)}\right)+1\right)^2} \right.\\
&\qquad\left. \frac{-2 \cdot \exp\left(2s_{(x,y)}\right)\left(s_{(x,y)}\right)' \cdot \left(\exp\left(2s_{(x,y)}\right)-1\right)}{\left(\exp\left(2s_{(x,y)}\right)+1\right)^2} \cdot coeff \right)\\
&= \frac{1}{N_n} \cdot \left(\sum_{x,y \in D, x \neq y} \frac{4 \cdot \exp\left(2s_{(x,y)}\right)\left(s_{(x,y)}\right)'}{\exp^2 2s_{(x,y)} + 2\exp\left(2s_{(x,y)}\right)+1} \cdot coeff \right)\\
&= \frac{1}{N_n} \cdot \left(\sum_{x,y \in D, x \neq y} \frac{4 \cdot \left(s_{(x,y)}\right)'}{\exp\left(2s_{(x,y)}\right)+\exp\left(2s_{(y,x)}\right)+2} \cdot coeff \right) \quad (7)
\end{aligned}
$$

Thus, after replacing the term $coeff$ with its original mathematical form, the gradient can be re-written as the explicit form in (8), where $s_{(x,y)} = s_{(x)} - s_{(y)} = <a, \Re_x - \Re_y>$.

$$
\bigtriangledown \text{SKT}(a) = \frac{1}{N_n} \cdot \left(\sum_{x,y \in D, x \neq y} \frac{4 \cdot \left(s_{(x,y)}\right)'}{\exp\left(2s_{(x,y)}\right)+\exp\left(2s_{(y,x)}\right)+2} \cdot \frac{\exp\left(2(\ell_{(x,y)})\right)-1}{\exp\left(2(\ell_{(x,y)})\right)+1} \right) \quad (8)
$$

In the above equation, a is the parameter to learn and there are several elements within it (i.e. the extracted feature vector $\Re_x$ is 8-dimensional in this study, hence there are also 8 elements within a to be determined). Therefore, learning a in this study is actually solving a multi-dimensional optimization problem. It is commonly acknowledged that, sometimes high-dimensional problems can be really tricky to tackle, as it is often hard to find the global optimum. In many studies of computer vision, researchers try to look for a local minimum which is good enough for specific applications, instead. In Table 1, T iterations are executed in the training step and there are T learned parameters a obtained (i.e., from Steps T2 to T7 in Table 1). After that, an optimal a_{opt} with the highest SKT value evaluated based on the validation data is sorted out (i.e., from Steps V1 to V5 in Table 1). Learned function $s_{opt}(x)$ with the optimal a_{opt} will be utilized in the next testing step for dementia disease diagnosis.

2.4 Disease severity diagnosis

An undiagnosed image x will be sorted together with other diagnosed images, their ordering in a ranked images list is to be determined by the learned function $s_{opt}(x)$, according to the ordering of computed score $s_{opt}(x) = <a_{opt}, \Re_x>$ of all images. When image x is located at position i of the list, its grade g_{x_i} can be interpolated using both scores of itself computed as $s_{opt}(x_i)$ and those of its neighboring images (i.e., $s_{opt}(x_{i-1})$ and $s_{opt}(x_{i+1})$), as well as their annotated grades (i.e., $g_{x_{i-1}}$ and $g_{x_{i+1}}$), which are known diagnosis results provided by clinicians. The grading strategy to predict the dementia disease severity of image x can

then be explicitly described as the following piecewise function. In this way, the dementia disease severity diagnosis task is accomplished.

$$
g_{x_i} = \begin{cases}
g_{x_{i+1}}, & \text{if } g_{x_{i+1}} = g_{x_{i-1}} \\
g_{x_{i+1}} + \dfrac{s_{opt}(x_i) - s_{opt}(x_{i+1})}{s_{opt}(x_{i-1}) - s_{opt}(x_{i+1})} \times (g_{x_{i-1}} - g_{x_{i+1}}), & \text{if } g_{x_{i-1}} > g_{x_{i+1}} \\
g_{x_{i-1}} + \dfrac{s_{opt}(x_{i-1}) - s_{opt}(x_i)}{s_{opt}(x_{i-1}) - s_{opt}(x_{i+1})} \times (g_{x_{i+1}} - g_{x_{i-1}}), & \text{if } g_{x_{i-1}} < g_{x_{i+1}}
\end{cases}
\tag{9}
$$

3 Experiments and analysis

3.1 Data description and pre-processings

In order to demonstrate the superiority of the new fMRI-based immersive tool for dementia diagnosis, clinical data obtained from 350 patients of different disease progressions, including 110 Alzheimer's Disease (AD) patients, 120 Mild Cognitive Impairment (MCI) patients and 120 Non-Cognitive Impairment (NCI) patients acquired in the affiliated hospital of Nanchang University, is utilized. Informed consent was obtained from all patients for research purpose. The averaged age of these patients is 70.56 ± 7.20 years old. Arterial Spin Labeling (ASL) images, which is an emerging perfusion fMRI technique and a new indicator in contemporary dementia studies [21], are acquired using a SIEMENS 3T TIM Trio MR scanner for each single patient. Acquisition parameters of the ASL scanning include: labeling duration = 1500 ms, post-labeling delay = 1500ms, TR/TE = 4000/9.1ms, ASL voxel size = $3 \times 3 \times 5 \ mm^3$.

When the acquisition of ASL images is accomplished, they need to be pre-processed before feeding into the fMRI-based immersive tool for dementia diagnosis. The reason is because that, ASL images often suffer from the Partial Volume Effects (PVE) problem, which is mainly caused by signal cross-contamination due to pixel heterogeneity and limited spatial resolution of ASL images [24]. Since PVE indicates the loss of apparent activity in small objects because of the limited resolution of an imaging system, ASL images with the PVE problem become more prone to under-estimate the measured perfusion, making the fMRI-based immersive tool inaccurate to reveal the actual brain atrophy of demented patients.

Therefore, PVE needs to be properly corrected in the pre-processing step of ASL images. In this study, the popular regression-based method proposed in [1] is incorporated for PVE correction. The main idea of this method can be described as follows. When correcting PVE on pixel i of an ASL image, its neighbors are necessary to be incorporated for adding up extra information to solve the PVE problem. For instance, given an adopted neighbor of size $n \times n$, a regression matrix P of the size $n^2 \times 3$ can be formulated using P_{GM}, P_{WM}, and P_{CSF}, which include fractional Gray Matter (GM), While Matter (WM), and Cerebro-Spinal Fluid (CSF) tissue volume of all n^2 neighbor pixels respectively as P's three columns. Unknowns to be solved in the PVE correction task on pixel i can be obtained using $(P^T P)^{-1} P^T \hat{M}$, where $\hat{M}$ depicts a matrix with magnetization on all n^2 neighbor pixels as its elements obtained in the ASL scanning process; T and -1 represent the transpose and the inverse of a matrix, respectively. For probability maps P_{GM}, P_{WM}, and P_{CSF}, they are generated using the SPM toolbox [27] based on High-resolution Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) T1-weighted MRI images [2] acquired with ASL images simultaneously, in the ASL scanning protocol of this study. The above

obtained maps are then co-registered towards their corresponding ASL images after motion correction for every patient using the FSL toolbox [7].

3.2 Experiments and analysis on dementia diagnosis

After all the above pre-processing steps are executed, ASL images can be utilized as fMRI in the new image-based immersive tool for dementia disease diagnosis. The pre-defined parameters in this novel immersive tool are set as: the number of iterations $T = 100$ and the learning rate $\eta = 0.01$ in Table 1, through trial-and-error for optimal performance. Mean ASL signal calculated from the segmented left & right hippocampus, the left & right parahippocampal gyrus, the left & right putamen, and the left & right thalamus (i.e. the above tissue segmentation is realized via the IBA-SPM toolbox [14]) from ASL images after PVE correction is utilized to construct a 8-dimensional feature vector $\Re_{(x)}$ for image x, following literatures in clinical dementia studies [10, 11, 19].

Other popular pattern recognition tools widely utilized in conventional disease diagnosis studies, including Linear Regression (denoted as "LR"), Support Vector Regression (denoted as "SVR" as a non-linear regression tool) and Ranking-Support Vector Machine (denoted as "RankingSVM" as a popular ranking tool) are implemented in this experiment based on all patients data for dementia diagnosis performance evaluation. Since learning is incorporated in the newly proposed immersive tool (denoted as "Our Method"), parameters of other methods are also determined via learning for fair comparisons. For SVR and RankingSVM, Gaussian Radial Basis Function (RBF) are adopted as kernels; Gaussian widths are learned via the popular radius/margin bound algorithm [17], and SVM-light

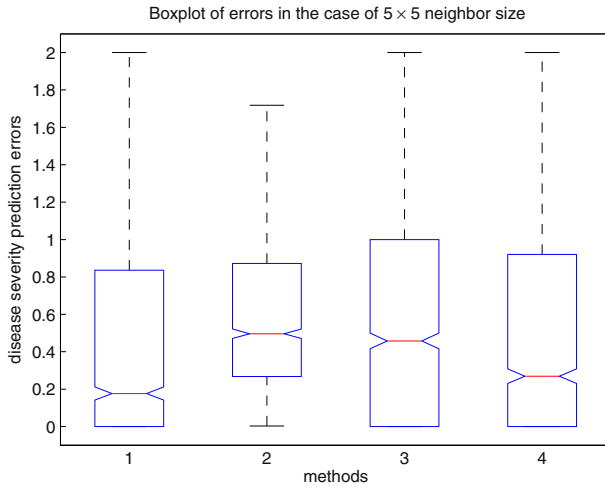| Table 2 The constitution of all 20 trials in 5-fold cross validation	Trial	Training	Validation	Testing
	I	1,2,3	4	5
	II	1,2,3	5	4
	III	1,2,4	3	5
	IV	1,2,4	5	3
	V	1,2,5	3	4
	VI	1,2,5	4	3
	VII	1,3,4	2	5
	VIII	1,3,4	5	2
	IX	1,3,5	2	4
	X	1,3,5	4	2
	XI	1,4,5	2	3
	XII	1,4,5	3	2
	XIII	2,3,4	1	5
	XIV	2,3,4	5	1
	XV	2,3,5	1	4
	XVI	2,3,5	4	1
	XVII	2,4,5	1	3
	XVIII	2,4,5	3	1
	XIX	3,4,5	1	2
	XX	3,4,5	2	1

Fig. 3 Boxplot of disease severity prediction errors among all compared methods in the case of 5×5 neighbor size (i.e., 1-Our Method; 2-LR; 3-SVR; 4-RankingSVM)

toolbox [16] is utilized for their implementations. For LR and Our Method, disease severities of different progressions of dementia, i.e., AD, MCI and NCI, are labeled as 3, 2 and 1 respectively. Regression coefficients in LR are determined via labels and regressors (i.e., the 8-dimensional feature vector) of the training data.

In our experiments, there are totally 3 different sizes of neighbors incorporated in the regression-based PVE correction method, in order to evaluate the diagnosis performance of all compared methods regarding different PVE correction situations when dealing with ASL images. To be specific, 5×5, 9×9 and 15×15 neighbors are adopted to represent small, medium, and large sizes of neighbors, respectively in PVE correction. The whole dataset
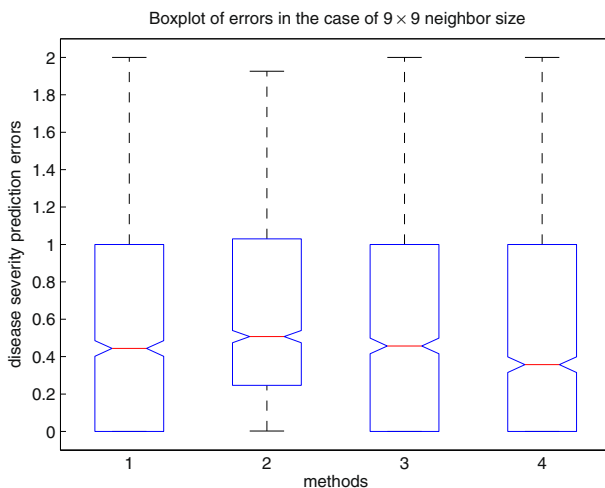

Fig. 4 Boxplot of disease severity prediction errors among all compared methods in the case of 9×9 neighbor size (i.e., 1-Our Method; 2-LR; 3-SVR; 4-RankingSVM)
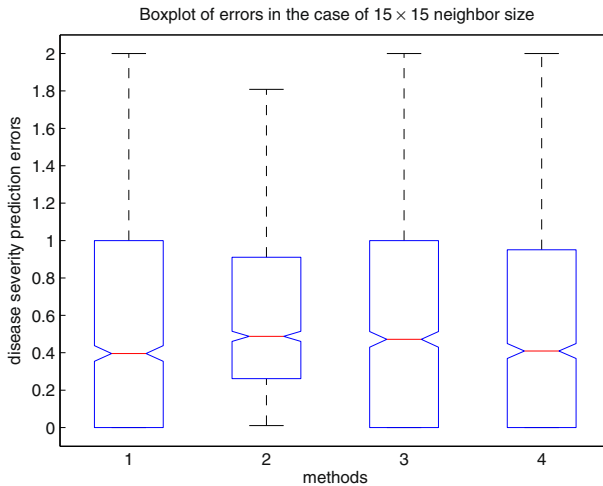
Fig. 5 Boxplot of disease severity prediction errors among all compared methods in the case of 15×15 neighbor size (i.e., 1-Our Method; 2-LR; 3-SVR; 4-RankingSVM)

of 350 patients is equally divided into 5 subsets to conduct a 5-fold cross validation for statistical evaluation. In each subset, patients with different dementia disease severities are roughly equivalent (i.e., 22 AD/ 24 MCI/ 24 NCI in each subset). Since there are training, validation and testing phases in the newly proposed immersive tool, and numbers of subsets utilized in them are 3, 1 and 1 individually in each trial of the 5-fold cross validation, the total number of trials in the whole 5-fold cross validation is $C_5^3 \cdot C_2^1 \cdot C_1^1 = 20$, where $C_{(*)}^{(\star)}$ denotes the number of combinations of $\star$ objects from a set of $*$ objects. Details of the constitution of all 20 trials in the 5-fold cross validation is elaborated in Table 2. For other compared methods without validation (i.e., "LR", "SVR" and "RankingSVM"), all non-testing subsets (i.e., training+validation subsets) are utilized for parameters learning in each trial.

The prediction error measuring the difference between the disease severity prediction generated by a diagnosis tool of one patient and her/his corresponding annotated disease severity by clinicians is utilized to evaluate the diagnosis performance among all compared methods. The smaller the prediction error, the superior the corresponding method becomes. Based on all disease severity prediction errors generated by all compared methods in the

Table 3 Multiple comparison test of all compared methods based on disease severity prediction errors in the case of 5×5

Method I	Method II	Prediction Error Difference (I-II)	A 95 % Confidence Interval
Our Method	LR	−0.1905	[−0.2353, −0.1457]
Our Method	SVR	−0.1028	[−0.1475, −0.0580]
Our Method	RankingSVM	−0.0196	[−0.0644, 0.0252]
LR	SVR	0.0877	[0.0429, 0.1325]
LR	RankingSVM	0.1709	[0.1261, 0.2157]
SVR	RankingSVM	0.0832	[0.0384, 0.1280]

Table 4 Multiple comparison test of all compared methods based on disease severity prediction errors in the case of 9 × 9

Method I	Method II	Prediction Error Difference (I-II)	A 95 % Confidence Interval
Our Method	LR	−0.1283	[−0.1773, −0.0794]
Our Method	SVR	−0.0085	[−0.0575, 0.0404]
Our Method	RankingSVM	−0.0024	[−0.0514, 0.0465]
LR	SVR	0.1198	[0.0709, 0.1687]
LR	RankingSVM	0.1259	[0.0770, 0.1748]
SVR	RankingSVM	0.0061	[−0.0428, 0.0550]

above three cases of PVE correction, three box-and-whisker plots are depicted in Figs. 3, 4 and 5, respectively. In each box, a red horizontal line is drawn across each box representing the median of prediction errors, while the upper and lower quartiles of prediction errors are depicted by blue lines above and below the median. A vertical dashed line is drawn up from the upper and down from the lower quartiles to their most extreme data points, which are within a 1.5 Inter-Quartile Range (IQR) [25]. It can be observed that boxes of Our method, SVR and RankingSVM are significantly lower than that of LR, which reveals that disease severity prediction errors generated by LR cannot be compared with others. However, which method among Our method, SVR and RankingSVM can produce the best diagnosis outcomes is still obscure based on box-and-whisker plots alone.

In order to clear the above doubt, a designated statistical analysis composed of one-way ANalysis Of VAriance (ANOVA) followed by a post-hoc multiple comparison test [25] is further conducted. ANOVA is a popular correction of models analyzing the difference between diverse group means and their associated variations in statistics [25]. To be specific, in one-way ANOVA, means of prediction errors from all methods are compared to test a hypothesis (H_0) that, all prediction error means of various methods could be equivalent, against the general alternative that at least one method is different. P-value is used here as an indicator to reveal whether H_0 exists or not. In this study, p-values for cases of 5 × 5, 9 × 9, and 15 × 15 are all nearly 0, which strongly suggests that H_0 is an invalid hypothesis for all cases. Hence, the next step is to do more detailed paired comparisons. The reason to conduct paired comparisons here is because that, the generative alternative against H_0 is too general to reveal which method is superior from the statistical point of view. Therefore, a post-hoc multiple comparison test is adopted to investigate it.

Table 5 Multiple comparison test of all compared methods based on disease severity prediction errors in the case of 15 × 15

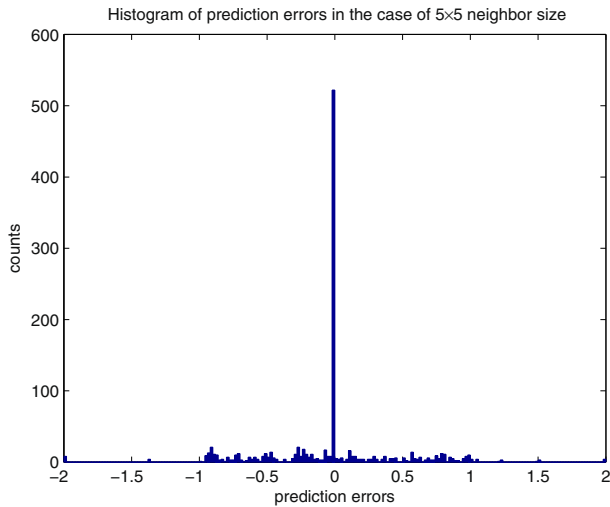| Method I	Method II	Prediction Error Difference (I-II)	A 95 % Confidence Interval
Our Method	LR	−0.1266	[−0.1733, −0.0800]
Our Method	SVR	−0.0462	[−0.0928, 0.0005]
Our Method	RankingSVM	−0.0133	[−0.0599, 0.0333]
LR	SVR	0.0805	[0.0338, 0.1271]
LR	RankingSVM	0.1133	[0.0667, 0.1600]
SVR	RankingSVM	0.0329	[−0.0138, 0.0795]

Fig. 6 Histogram of disease severity prediction errors provided by Our Method in the case of 5×5 neighbor size

Entries in Tables 3, 4 and 5 are results of multiple comparison tests based on prediction errors generated by all methods for cases of 5×5, 9×9, and 15×15, respectively. Each row indicates a paired comparison between two methods, and there are two types of estimations for each paired comparison: one is a single-value estimation, which estimates the difference of prediction errors within two compared methods by a single value; the other is an interval estimation conducted via a 95 % Confidence Interval (CI), which estimates a range that the prediction error difference is likely to be included. For instance, the first row of Table 3 is
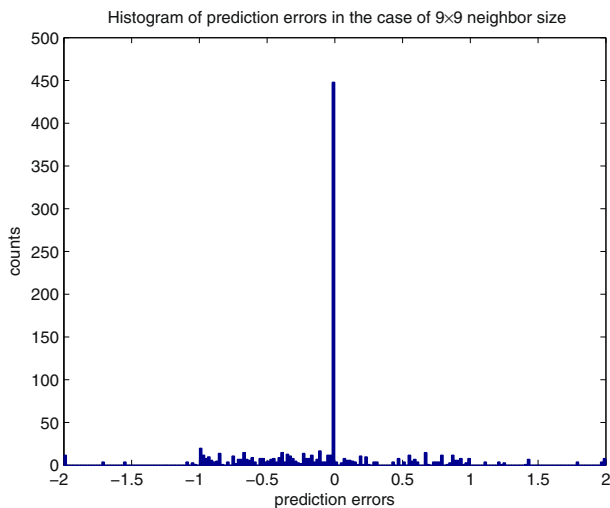

Fig. 7 Histogram of disease severity prediction errors provided by Our Method in the case of 9×9 neighbor size

about the paired comparison between Our Method and LR in the case of 5×5. The prediction error difference of the single-value estimation is -0.1905 (using Our Method minus LR), which suggests that Our Method can produce less disease severity prediction errors than LR, from the single-value estimation perspective. The prediction error difference is likely to fall within a 95 % CI $[-0.2353, -0.1457]$. Since both its upper and lower bounds are both negative, it gives a strong indication (> 95 %) that, the prediction error difference (using Our Method minus LR) is negative. Hence, Our Method is superior to LR in the case of 5×5 from both single-value and interval estimation perspectives. For paired comparisons between Our Method and others, the analysis is similar. The prediction errors of Our Method is 0.1028 and 0.0196 smaller than SVR and RankingSVM, respectively from the single-value estimation perspective in Table 3. The 95 % CIs for the two paired comparisons are $[-0.1475, -0.0580]$ and $[-0.0644, 0.0252]$. One thing to clarify here is that, the upper bound of the 95 % CI between Our Method and RankingSVM is positive. It suggests that, RankingSVM can be marginally better than Our Method in certain patients diagnosis (i.e., RankingSVM is superior in 28.13 % diagnosis in the case of 5×5, following a general assumption that the 95 % CI $[-0.0644, 0.0252]$ is uniformly distributed). However, Our Method still dominates in more than half of all patients diagnosis following the above analysis, if a positive upper bound exists in the 95 % CI. Similar conclusions can also be drawn in Tables 4 and 5 regarding comparisons between Our Method and others. To sum up, after conducting the one-way ANOVA followed by multiple comparison tests, Our method outperforms others from the statistical point of view.

In Fig. 6, a histogram depicting the distribution of disease severity prediction errors obtained by Our Method in the case of 5×5 neighbor size is illustrated based on all diagnosis results obtained from the 5-fold cross validation. The number of testing data in Fig. 6 is 1400 (which equals to 350×4, as each testing subset will be utilized $C_4^3 \times C_1^1 = 4$ times brought by different combinations of training and validation subsets in the 5-fold cross validation). It can be observed that, the prediction error of most cases is within the range $[-1, 1]$. In Figs. 7 and 8, similar observations exist for cases of 9×9 and 15×15 neighbor sizes.
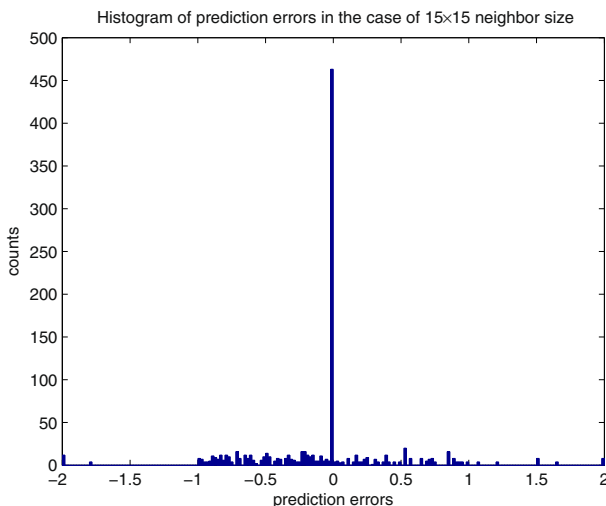

Fig. 8 Histogram of disease severity prediction errors provided by Our Method in the case of 15×15 neighbor size

Statistics of prediction errors for cases of 5×5, 9×9 and 15×15 are 0.5086 ± 0.4748, 0.4318 ± 0.4418, and 0.4843 ± 0.4625 (i.e., means $\pm$ standard deviation), respectively. The accuracies for the three cases are 95.43 %, 97.71 %, and 96.86 %, respectively when taking diagnosis outcomes with the absolute prediction error less than 1 as accurate diagnosis following our senior clinicians suggestions. Overall, the superiority of the newly proposed fMRI-based immersive tool in dementia disease diagnosis can be revealed by the above extensive experiments and comprehensive analysis from the statistical point of view.

3.3 Discussion

In this study, the linear ranking function $s_{(x)} = <a, \Re_x>$ is utilized for each image x, where $\Re_x$ denotes the extracted feature from image x and a represents the unknown parameter to learn. In this section, the choice of linear and non-linear ranking functions and their influence on dementia disease diagnosis performance is investigated and discussed. To be specific, an exponential form $s_{(x)} = \exp(<a, \Re_x>)$ is incorporated as the compared non-linear ranking function in this fMRI-based immersive tool for dementia disease diagnosis. To reveal its performance, the same 5-fold cross validation using the same combinations of training, validation and testing data is adopted, and the same statistical analysis is performed with this non-linear ranking function. It turns out that, prediction errors (i.e., means $\pm$ standard deviation) for cases of 5×5, 9×9 and 15×15 are 0.4331 ± 0.4425, 0.5089 ± 0.4758, and 0.4849 ± 0.4632, respectively when the non-linear ranking function is incorporated. The accuracies for the above three cases are 97.43 %, 95.29 %, and 96.86 %, respectively when taking diagnosis outcomes with the absolute prediction error less than 1 as accurate diagnosis results following the same strategy as that of the linear ranking function. After comparing the above outcomes with ones of the linear ranking function accordingly within the same immersive tool, it can be concluded that, the dementia disease diagnosis performance brought by both linear and non-linear ranking functions in this newly proposed fMRI-based immersive tool are comparable. Thus, both linear and non-linear ranking functions are applicable.

4 Conclusion

In this study, a novel image-based immersive tool for dementia disease diagnosis via new pairwise ranking and learning techniques is proposed for the first time. Arterial spin labeling is incorporated as images for diagnosis. Extensive experiments and their comprehensive statistical analysis demonstrate the superiority of the novel image-based immersive tool in dementia disease diagnosis. Main contributions of this study can be summarized as: 1) Technically, new pairwise ranking and learning techniques based on a new continuous and differentiable surrogated Kendall-Tau rank correlation coefficient are introduced in this paper; 2) It is also the first attempt to perform dementia disease diagnosis from the new perspective of pairwise ranking. Future efforts will be emphasized on investigating other types of image-based immersive tools with more sophisticated ranking models equipped for diverse disease diagnosis, based on different modalities of medical images.

References

1. Asllani I, Borogovac A, Brown T (2008) Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. Magn Reson Med 60:1362–1371
2. Brant-Zawadzki M, Gillan G, Nitz W (1992) MPRAGE: a three-dimensional, T1-weighted, qradient-echo sequence–initial experience in the brain. Radiology 182(3):769–775
3. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi M (2007) Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3(3):186–191
4. Burges C, Shaked T, Renshaw E, Lazier A, Deeds M, Hamilton N, Hullender G (2005) Learning to rank using gradient descent. In: International Conference on Machine Learning, pp 89–96
5. Chu W, Keerthi SS (2005) New approaches to support vector ordinal regression. In: International conference on machine learning, pp 145–152
6. Crammer K, Singer Y (2001) Pranking with ranking. Advances in neural information processing systems, pp 641–647
7. FMRIB Software Library (FSL) toolbox. http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
8. Folstein M, Folstein S, McHung P (1975) Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3):189–198
9. Freund Y, Iyer R, Schapire R, Singer Y (2003) An efficient boosting algorithm for combining preference. J Mach Learn Res 4:933–969
10. Galton C, Patterson K, Graham K, Lambon-Ralph M, Williams G, Antoun N, Sahakian B, Hodges J (2001) Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. Neurology 57(2):216–225
11. Gold G, Eniko K, Herrmann F, Canuto A, Hof P, Jean-Pierre M, Constantin B, Giannakopoulos P (2005) Cognitive consequences of thalamic, basal ganglia, and deep white matter lacunes in brain aging and dementia. Stroke 36(6):1184–1188
12. Harrington EF (2003) Online ranking/collaborative filtering using the perceptron algorithm. In: International conference on machine learning, pp 250–257
13. Herbirch R, Graepel T, Obermayer K (2000) Chapter 7 - Large margin rank boundaries for ordinal regression. Advances in large margin classifiers, pp 115–132
14. Individual Brain Atlases using Statistical Parametric Mapping (IBA-SPM) Software. http://www.thomaskoenig.ch/Lester/ibaspm.htm
15. Joachims T (2002) Optimizing search engines using clickthrough data. ACM special interest group on knowledge discovery and data mining, pp 133–142
16. Joachims T SVM light - An implementation of support vector machine in C. http://svmlight.joachims.org
17. Keerthi S (2002) Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. IEEE Trans Neural Netw 13(5):1225–1229
18. Kendall M (1938) A new measure of rank correlation. Biometrika 30:81–93
19. Laakso M, Partanen K, Riekkinen P, Lehtovirta M, Helkala E, Hallikainen M, Hanninen T, Vainio P, Soininen H (1996) Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia an MRI study. Neurology 46(3):678–681
20. Mahendra B (1987) A pathography of dementia. Dementia 1:189–202
21. Malpass K (2012) Alzheimer disease: arterial spin-labeled MRI for diagnosis and monitoring of AD. Nat Rev Neurol 8(3):847–849
22. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges J (2006) The Addenbrooke's cognitive examination-revised (ACE-R): a brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry 21(11):1078–1085
23. Musiek E, Chen Y, Korczykowski M, Saboury B, Martinez P, Reddin J, Alavi A, Kimberg D, Wolk D, Julin P, Newberg A, Arnold S, Detre J (2012) Direct comparison of fluorodeoxyglucose positron emission tomography and arterial spin labeling magnetic resonance imaging in Alzheimer's disease. Alzheimers Dement 8(1):51–59
24. Partial volume imaging. http://en.wikipedia.org/wiki/Partial_volume_(imaging)
25. Rice J (2007) Mathematical statistics and data dnalysis, 2nd edn. Duxbury Press
26. Shashua A, Levin A (2002) Ranking with large margin principle: two approaches. Advances in neural information processing systems, pp 937–944
27. Statistical Parametric Mapping (SPM) toolbox. http://www.fil.ion.ucl.ac.uk/spm/
28. United Nations. World population prospects. http://www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf
29. World Health Organization. The top 10 causes of death. http://www.who.int/mediacentre/factsheets/fs310/en/index2.html

Wei Huang obtained his B.Eng. and M.Eng. from Harbin Institute of Technology, China in 2004 and 2006, respectively. He obtained his Ph.D. from Nanyang Technological University, Singapore in 2011. Before joining Nanchang University as an Associate Professor, he worked in the University of California San Diego, USA, as well as Agency for Science Technology and Research, Singapore as Post-doctoral Research Fellows. Dr. Huang has published 40+ academic papers, and is now the principal investigator in 2 NSFC grants and several other national/provincial grants. Dr. Huang's research interests mainly include but not limited to medical image processing, pattern recognition, and computer vision.

Shuru Zeng obtained her B.Eng. degree from Jiangxi Agricultural University in 2014. She is now a M.Eng. candidate in the Department of Computer Science, Nanchang University under the supervision of Prof. Wei Huang. Her research interests mainly include image processing and computer vision.
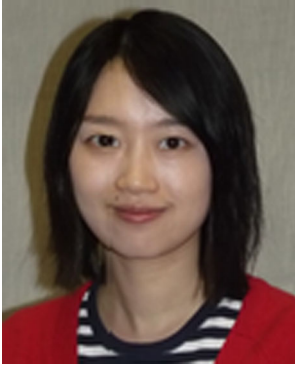
Jing Li received the Ph.D. degree from the Department of Electronic and Electrical Engineering at the University of Sheffield, U.K., in 2012. She is currently an Associate Professor with the School of Information Engineering at Nanchang University, China. Her research interests include content-based image retrieval, pattern recognition, and video analysis. She has authored or coauthored in various peer-reviewed journals, such as IEEE Transactions on Industrial Informatics, Information Sciences (Elsevier), etc.
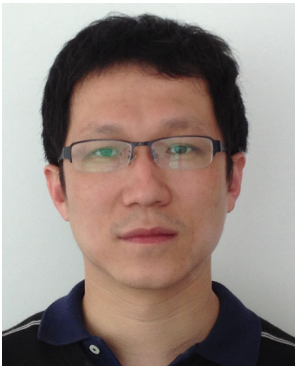
Guang Chen received his B.E. degree from the Nanjing Institute of Communication Engineering, China in 2001. He received his M.E.degree from the Xidian University, China in 2008. He is now a faculty in Xi'an Communications Institute. His current research interests include software engineering, information forensics and signal processing.