

# Improved Perfusion Pattern Score Association with Type 2 Diabetes Severity Using Machine Learning Pipeline: Pilot Study

Yuheng Chen, BS,<sup>1</sup> Wenna Duan, MS,<sup>1</sup> Parshant Sehrawat, MS,<sup>1</sup> Vaibhav Chauhan, BS,<sup>1</sup> Freddy J Alfaro, MD,<sup>2</sup> Anna Gavrieli, PhD,<sup>2</sup> Xingye Qiao, PhD,<sup>3</sup> Vera Novak, MD, PhD,<sup>2</sup> and Weiyong Dai, PhD<sup>1\*</sup>

**Background:** Type 2 diabetes mellitus (T2DM) is associated with alterations in the blood–brain barrier, neuronal damage, and arterial stiffness, thus affecting cerebral metabolism and perfusion. There is a need to implement machine-learning methodologies to identify a T2DM-related perfusion pattern and possible relationship between the pattern and cognitive performance/disease severity.

**Purpose:** To develop a machine-learning pipeline to investigate the method's discriminative value between T2DM patients and normal controls, the T2DM-related network pattern, and association of the pattern with cognitive performance/disease severity.

**Study Type:** A cross-sectional study and prospective longitudinal study with a 2-year time interval.

**Population:** Seventy-three subjects (41 T2DM patients and 32 controls) aged 50–85 years old at baseline, and 42 subjects (19 T2DM and 23 controls) aged 53–88 years old at 2-year follow-up.

**Field Strength/Sequence:** 3T pseudocontinuous arterial spin-labeling MRI.

**Assessment:** Machine-learning-based pipeline (principal component analysis, feature selection, and logistic regression classifier) to generate the T2DM-related network pattern and the individual scores associated with the pattern.

**Statistical Tests:** Linear regression analysis with gray matter volume and education years as covariates.

**Results:** The machine-learning-based method is superior to the widely used univariate group comparison method with increased test accuracy, test area under the curve, test positive predictive value, adjusted McFadden's R square of 4%, 12%, 7%, and 24%, respectively. The pattern-related individual scores are associated with diabetes severity variables, mobility, and cognitive performance at baseline ( $P < 0.05$ ,  $|r| > 0.3$ ). More important, the longitudinal change of individual pattern scores is associated with the longitudinal change of HbA1c ( $P = 0.0053$ ,  $r = 0.64$ ), and baseline cholesterol ( $P = 0.037$ ,  $r = 0.51$ ).

**Data Conclusion:** The individual perfusion diabetes pattern score is a highly promising perfusion imaging biomarker for tracing the disease progression of individual T2DM patients. Further validation is needed from a larger study.

**Level of Evidence:** 1

**Technical Efficacy:** Stage 1

J. MAGN. RESON. IMAGING 2019;49:834–844.

Type 2 diabetes mellitus (T2DM) is a metabolic disorder that increases the risk of cognitive impairment.<sup>1,2</sup> T2DM has been associated with alterations in the blood–brain barrier,<sup>3,4</sup> neuronal damage,<sup>5</sup> and arterial stiffness,<sup>6</sup> thus affecting cerebral metabolism and perfusion.<sup>7</sup> Impaired cerebral hemodynamics is considered a potential underlying cause

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com). DOI: 10.1002/jmri.26256

Received Jan 1, 2018, Accepted for publication Jun 26, 2018.

\*Address reprint requests to: W.D., N-14, Engineering Building, Department of Computer Science, Binghamton University, Binghamton, NY, 13902.  
E-mail: [w dai@binghamton.edu](mailto:w dai@binghamton.edu)

From the <sup>1</sup>Department of Computer Science, State University of New York at Binghamton, Binghamton, New York, USA; <sup>2</sup>Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA; and <sup>3</sup>Department of Mathematical Sciences, State University of New York at Binghamton, Binghamton, New York, USA

of cognitive decline.<sup>8,9</sup> Therefore, studies linking brain perfusion with cognitive impairment in T2DM are crucial to reveal the mechanism of cognitive deficits.

Studies using a variety of techniques that include phase-contrast (PC) magnetic resonance angiograph (MRA), single-photon emission computed tomography (SPECT), positron emission tomography (PET), pulse arterial spin labeling (PASL) magnetic resonance imaging (MRI), continuous arterial spin labeling (CASL) MRI, and pseudocontinuous arterial spin labeling (PCASL) MRI have investigated the effects of T2DM on brain perfusion and made attempts to correlate perfusion with cognitive performance.<sup>10–17</sup> However, the results from different studies have not been consistent. Most studies rely on the comparison of the whole-brain perfusion, gray matter (GM) perfusion, and several large regions of the cerebrum<sup>10–15</sup> for the discriminative value of brain perfusion images. Recently, a couple of groups applied voxel-by-voxel between-group comparisons by univariate statistical analysis.<sup>16,17</sup> The voxel-based techniques provided good insights in perfusion deficits of different brain regions, but lacked separation of different covariance sources (eg, signal variation source caused by cardiac pulsation), and hence may potentially have poor discriminative value to distinguish T2DM patients from controls at the individual level.

T2DM affects brain perfusion on a large volume of brain regions instead of just isolated regions.<sup>11,15,16,18</sup> Different brain regions work in coordination even during its resting state (without any explicit functional task) and the signals in the coordinated regions fluctuate in synchrony.<sup>19</sup> Separating the brain images into different noise variation sources may reflect the brain's underlying coordinated functional activities, and therefore holds great potential to reveal the underlying pathology. Machine-learning methods have been applied to study brain structural changes (eg, brain atrophy) for diagnosis, and to investigate task-related responses using functional MRI (fMRI).<sup>20–22</sup> These methods have been rarely used to investigate the resting brain functionality for disease effects.<sup>23</sup> Here, we aimed to identify T2DM-related brain covariance patterns with an improved discriminative value at the individual level by developing a method using pattern recognition and machine-learning methods.

## Materials and Methods

### Subjects

In all, 131 subjects, 50–85 years old, were enrolled in this 2-year study. All subjects signed an informed consent form (ICF) as approved by the Institutional Review Board (IRB) at Beth Israel Deaconess Medical Center. Of the 131 subjects, 73 subjects: 41 T2DM and 32 nondiabetic controls, were eligible and included in the baseline analysis, according to the inclusion criteria of the study. Of those, 42 subjects, 19 T2DM and 23 nondiabetic controls, who completed the 2-year follow-up, were included in the follow-up analyses. The inclusion and exclusion criteria for the

baseline study and rationale for sample size are described in the Appendix.

### Experimental Protocol and Data Acquisition

The experimental protocol, including the screening protocol of subjects, measurements of vital signs, walking test/gait assessment, a fasting blood draw to measure hematocrit, glucose, insulin, and glycated hemoglobin A1c (HbA1c) are described in the Appendix.

**COGNITIVE ASSESSMENT.** The cognitive assessment battery is a standard battery of cognitive tests that evaluates specific domains of cognition and daily living activities. It consists of measures of learning and memory (Hopkins Verbal Learning Test-Revised [HVLTR] <sup>24</sup> and Mini-Mental State Examination [MMSE]), measures of executive function (Verbal Fluency [VF],<sup>25</sup> Trail Making [TM], Clock Drawing [CD]), and measures of attention (Digit Span [DS]). HVLTR includes a Total Recall (HVLTR: Total Recall, total number of list items learned across trials), Delayed Recall (HVLTR: Delayed Recall, total number of list items recalled after the delay), and Retention (HVLTR: Retention, percentage of items from Total Recall that are subsequently recalled on Delayed Recall). MMSE assesses cognitive mental states. VF semantic fluency task requires the subject to generate items of a given semantic category (eg, animals) for 1 minute. The dependent variable for the fluency measure is the number of items generated for the semantic task (eg, animals), which will be referred to as VF: animals. Cognitive assessments were adjusted for education years.

**ASSESSMENT OF INSULIN RESISTANCE.** The homeostatic model assessment of insulin resistance (HOMA-IR) was calculated as the product of fasting glucose times insulin levels divided by 405.<sup>26</sup> The model was also used to estimate steady-state beta cell function (%B) and insulin sensitivity (%S), as percentages of a normal reference population.

**MRI ACQUISITION.** All 73 subjects (41 T2DM patients and 32 controls) at baseline and 42 subjects (19 T2DM and 23 controls) at follow-up were scanned at the same 3T, GE HDxt scanner (GE Healthcare, Milwaukee, WI) using a receive-only 8-channel head array coil and a body transmit coil. Both baseline and follow-up studies follow the same MRI acquisition protocols. Brain 3D perfusion images were obtained using the PCASL<sup>27</sup> with a 1.5-second labeling and 1.5-second postlabeling delay. Additional reference images were obtained for absolute perfusion quantification. Both perfusion images and reference images were acquired with a 3D stack of spirals RARE imaging sequence (resolution:  $4 \times 4 \times 4 \text{ mm}^3$ , matrix size:  $64 \times 64 \times 40$ , TR: 5 sec, acquisition time: 6 min, bandwidth: 62.5 kHz, three averages per label-control pair). T<sub>1</sub> anatomical images were acquired with a 3D magnetization prepared rapid acquisition gradient echo (MP-RAGE) sequence (resolution:  $0.94 \times 0.94 \times 3 \text{ mm}^3$ , matrix size:  $256 \times 256 \times 52$ ).

### Image Preprocessing

The quantitative perfusion image was calculated for each subject as previously described.<sup>28</sup> Perfusion images were normalized to the a priori gray matter template using the SPM8 software package (<http://www.fil.ion.ucl.ac.uk/spm/>). T<sub>1</sub> anatomical images served as intermediate images for perfusion image normalization, as they allow better alignment with the template. They were segmented by the "new segment" algorithm,<sup>29</sup> which output gray matter images as well as other images in the original image space. Subtraction images (between label and control) from PCASL acquisition were coregistered to the gray matter images from the segmentation and then the gray matter images were normalized to the gray matter template. The combined warping parameters from the coregistration and normalization were used to warp the quantitative perfusion maps from each subject to the template space. Quantitative perfusion images were smoothed using a Gaussian kernel with full-width at half-maximum (FWHM) of 8 mm. Image method development and analysis was performed by Y. Chen, who had no access to the document with clinical characteristics, cognitive performance, and mobility performance scores until the individual disease scores at baseline and follow-up were calculated.

### Logistic Regression (LR) Classification Pipeline

To avoid overfitting an LR model, we performed L2-regularized LR with a grid search to tune the regularization term,  $\lambda$ .  $\lambda$  was searched with the range [0 1] with step size 0.1. The L2-regularized LR model was used to distinguish T2DM patients from controls. Three measures: accuracy, area under the curve (AUC), and positive predictive value (PPV) were used as the performance measures for the LR model. Accuracy is evaluated as the ratio of correct numbers of prediction to the total number of predictions. AUC is calculated as the area under receiver operating characteristic curve. PPV is evaluated as the ratio of correct numbers of positive prediction to the total numbers of positive prediction.

The details of the classification pipeline were as follows. We used stratified k-fold (k = 10) cross-validation to evaluate the performance of the LR model. All subjects were divided into k folds, in which each fold has similar distribution of class labels (4 or 5 T2DM patients and 3 or 4 controls). For the *i*th partition, the *i*th fold served as the test set and the remaining k-1 folds served as the training set. First, each 3D image from a single training subject was rearranged into a one-dimensional vector (dimension: number of voxels), and all the training subjects form a two-dimensional matrix (dimension: number of training subjects  $\times$  number of voxels) that serves as the training dataset. The training data were demeaned for each voxel. Then principal components analysis (PCA) was performed on the training set to reduce the dimension and separate the covariant sources of data. From all the demeaned training data, PCA extracted the principal components (ie, the covariant noise sources). PCA also outputs the scores of each subject, which are the corresponding weights of the principal components. To remove the noisy components, 90% of total variance was used to select the number of principal components for further analysis. The full feature set is the set of potential candidate features for the LR classification model, including the PCA scores from the selected PCA components and a few basic variables (age, gender, hematocrit, and hypertension). The basic variables were selected based on the literature-reported factors

affecting perfusion. Second, feature scaling was applied to allow for faster convergence of the classification algorithm. For feature scaling of each feature, the mean of the feature was subtracted from each subject's feature value and then divided by the standard deviation of the feature. Third, the diabetes-related features were extracted from the full feature set using a backward stepwise search. Leave-one-out cross-validation accuracy served as the model selection criteria within the search. Fourth, an LR classification model was built using the selected features. Fifth, performance of the built LR model was evaluated using the test set. The test dataset was formed from the subjects in the test fold, also as a 2D matrix (dimension: number of test subjects  $\times$  number of voxels). The test set was projected to the selected principal components to derive the corresponding PCA scores by following the procedure the same as performed in the training dataset. Feature scaling on the test set was performed using the mean and standard deviations of the training set. Note that PCA and LR classification model building were performed on only the training set to keep the test set independent of the training set. The major processing steps of the classification pipeline, including PCA, feature selection, and LR classification model, is illustrated in Fig. 1.

Each 10-fold stratified cross-validation corresponds to a random 10-partition of the entire dataset. For each partition, we derived three performance measures (accuracy rate, AUC, and PPV). Consequently, one 10-fold cross-validation generates 10 values for each performance measure. We ran 100 iterations of the 10-fold cross-validation procedure and therefore output 1000 performance values for each measure.

### Pattern Generation and Pattern Score

**T2DM-RELATED COVARIANCE PATTERN.** We followed the LR classification pipeline to obtain the selected features and built an LR model using the selected features. The only difference from the LR classification pipeline was that all subjects were included to generate the covariance pattern. The T2DM-related pattern is  $Pattern = w_{PC} \cdot PC$ , where  $PC$  is the selected principal components, and  $w_{PC}$  is the weight vector associated with the selected principal components from the model. To visualize the pattern, we reshaped the vector back to the 3D image space.

To identify the reliable T2DM-related covariance pattern, a bootstrap estimation procedure was used to approximate the standard deviation of the population. We first generated the T2DM-related covariance pattern using all the T2DM patients and controls. Using the bootstrapping resampling method with replacement, we produced 500 surrogate datasets based on the original dataset while maintaining the same number of T2DM patients and controls. For each dataset, one T2DM-related covariance pattern was generated. From all 500 datasets, we calculated the standard deviation image using 500 bootstrapped covariance patterns. Z-scores were calculated using the ratio of the T2DM-related network pattern generated from the original dataset to the standard deviation image. A threshold of  $|z| > 3.27$  corresponding to a two-tailed  $P < 0.001$  was used to identify the voxels that significantly contribute to the T2DM-related pattern. For the

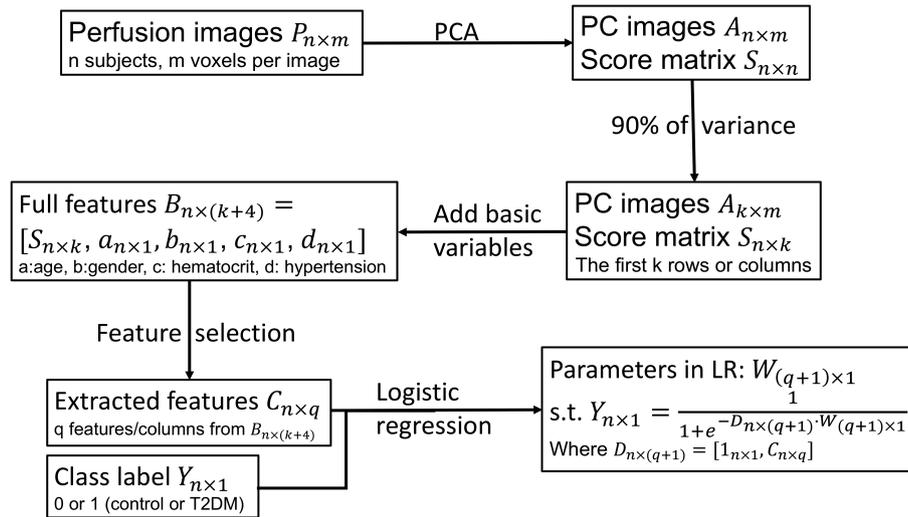


FIGURE 1: Schematic of the logistic regression classification pipeline.

clusters formed by connected voxels, a threshold for cluster size was calculated to correct for multiple comparisons. All clusters with size smaller than the threshold were removed.

**T2DM-RELATED PATTERN SCORES AT BASELINE.** Subject-specific T2DM-related pattern scores at baseline were obtained using the LR model with the selected features. Let  $\mathbf{x}$  be the vector with all the selected features from a subject and  $\mathbf{w}$  be the weight vector derived from the LR model. The subject's T2DM-related pattern score was defined as  $\text{Score} = \mathbf{w} \cdot \mathbf{x}$ .

**T2DM-RELATED PATTERN SCORES AT THE 2-YEAR FOLLOW-UP.** We calculated the T2DM-related pattern scores of the 42 subjects who were scanned at the 2-year follow-up. The pattern scores of the subjects at the follow-up were calculated as if the follow-up scans were images from the test dataset for the baseline training dataset.

### Simple LR Classification Model

In the proposed classification pipeline, we performed PCA and feature selection techniques to select the useful features out and built the classification model using the selected features. It is worth investigating whether the PCA and feature selection steps could improve the performance of the LR model. For comparison purposes, we generated a simple LR classification model using the imaging features from all the voxels in the brain and the same basic variables (age, gender, hematocrit, and hypertension). Feature scaling was also performed before building the LR model.

### Univariate Analysis

Univariate analysis is still the most frequently used image analysis to derive the disease-related pattern and discriminative value of an imaging contrast (eg, perfusion in the current project). For the univariate analysis, perfusion images were modeled as a multiple linear regression on a voxel-by-voxel basis using SPM8. Age, hematocrit, and hypertension were included as covariates. Gender was not included as a covariate because gender was significantly correlated with hematocrit, and hematocrit but not gender was correlated with

global perfusion in our dataset. The voxel-level significance threshold was set for  $P < 0.01$ , while the cluster-level threshold was set for  $P < 0.05$  in order to minimize any false-positive findings because of the multiple comparisons. The cluster with the most significant  $P$  value (ie, the smallest corrected cluster-level  $P$  value) from univariate analysis was used as the target region. The average perfusion on the target region was one of the features used to distinguish between T2DM and control group.

### Comparison of the Proposed LR Pipeline, Simple LR Classification Model, and Univariate Analysis

To compare the performance of our proposed LR pipeline against the performance of the simple LR classification method and univariate analysis, we calculated the performance measures (accuracy rate, AUC, and PPV) of the simple classification model and univariate analysis for the same 100 10-folds (1000 partitions). For univariate analysis, the regional perfusion value on the target region was corrected for the effects of age, hematocrit, and hypertension. A cutoff perfusion value that maximizes the sum of training accuracy and training PPV was derived. For the test set, the regional perfusion value was also corrected for the covariates, and the cutoff perfusion value was used to calculate performance measures.

The effect sizes for the proposed LR pipeline and univariate analysis were compared. Because of the difference in numbers of independent variables between the two models, the adjusted McFadden's R squares<sup>30,31</sup> were calculated to evaluate the overall effect size. McFadden's R square was chosen as a measurement of the effect size because it can serve as a uniform measurement for both logistic regression and multiple linear regression. Specifically, the adjusted McFadden's R square is defined as follows:

$$R_{McFadden}^2 = 1 - \frac{\log(L_c) - k}{\log(L_{null})} \quad (1)$$

where  $k$  is the number of predictors,  $L_c$  denotes the (maximized) likelihood value from the fitted model, and  $L_{null}$  denotes the corresponding value for the null model, ie, the model with only an intercept and no covariates.

**Correlation Analysis**

**ASSOCIATION OF T2DM-RELATED PATTERN SCORES AND DISEASE VARIABLES.** After obtaining the individual T2DM-related pattern scores, we calculated the sample Pearson correlation coefficient,  $r$ , between the pattern scores and each of the clinical variables (including variables for disease severity, cognitive functions, and mobility functions). The difference in GM volume between T2DM and control groups could potentially account for the perfusion differences, and hence pattern score differences between the two groups, GM volume was further included as a covariate in the partial correlation analysis.

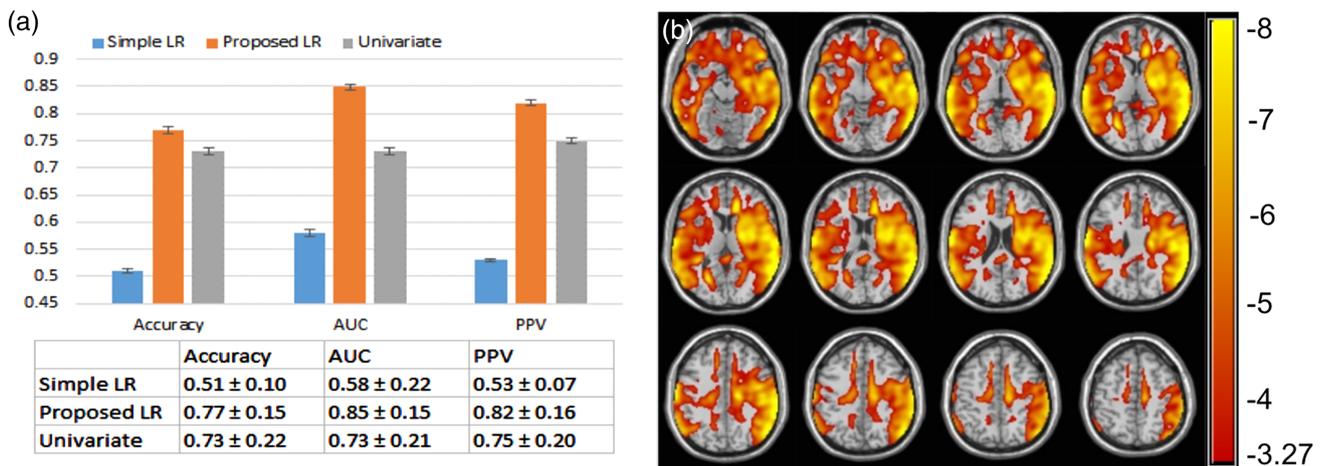
**ASSOCIATION OF LONGITUDINAL T2DM-RELATED PATTERN SCORE CHANGE AND CHANGE OF DISEASE VARIABLES.** In order to investigate whether the longitudinal T2DM-related pattern score change is related to the longitudinal change of cognitive and mobility performance, post-hoc correlation analysis was performed. Partial correlation coefficients were calculated between pattern score change and each of the baseline variables (including cognitive, mobility, and disease severity variables) in order to investigate which baseline variables can predict the longitudinal pattern score change. GM volume and education years were used as covariates. Education was included as an additional covariate because it was significantly different between the T2DM and control groups at 2-year follow-up ( $P = 0.012$ ).

**Results**

Table A1 (see Appendix) summarizes subjects' demographic and clinical characteristics, gait results, and cognitive scores at baseline and at the 2-year follow-up. At baseline, no differences between age, gender, education, and hematocrit values

were found between the two groups. The T2DM group had a higher prevalence of hypertension ( $P \leq 0.001$ ), body mass index (BMI) ( $P = 0.007$ ), fasting glucose ( $P \leq 0.001$ ), HbA1c ( $P \leq 0.001$ ), insulin level ( $P = 0.003$ ), and HOMA-IR ( $P \leq 0.001$ ) compared to controls. At the 2-year follow-up, the significant differences between the T2DM and control groups remained similar in demographics and clinical characteristics, while the years of education became significant.

For the proposed LR classification pipeline, the grid search of the regularization term showed that  $\lambda = 0.5$  gave the best performance for our dataset. Using this regularization term, the test accuracy rate, test AUC, and test PPV for the proposed LR classification pipeline were  $0.77 \pm 0.15$ ,  $0.85 \pm 0.15$ , and  $0.82 \pm 0.16$ , respectively, across 1000 different partitions. In contrast, the test accuracy rate, test AUC, and test PPV for the simple LR classification model were  $0.51 \pm 0.10$ ,  $0.58 \pm 0.22$ , and  $0.53 \pm 0.07$  and for the univariate analysis were  $0.73 \pm 0.22$ ,  $0.73 \pm 0.21$ , and  $0.75 \pm 0.20$ , respectively. The comparison of the performance measures between the proposed LR classification pipeline, simple LR classification method, and univariate analysis across the same 1000 different partitions are shown in Fig. 2a. We further compared the mean and standard deviation of each performance measure between the proposed LR classification pipeline, simple LR classification model, and univariate analysis using nonparametric Wilcoxon signed-rank tests. For all three performance measures: the proposed LR classification pipeline produced significantly larger means than the univariate analysis and simple LR classification model; the univariate analysis produced significantly larger means than the simple LR classification model; and the proposed LR classification pipeline produced significantly smaller standard deviation than the univariate analysis (see chart in Fig. 2a). The statistical



**FIGURE 2:** (a) (top) Average test accuracy, AUC, and PPV over 100 iterations of 10-fold cross-validation for simple logistic regression (LR), proposed LR, and univariate analysis. The error bars stand for standard errors. For test accuracy, AUC, and PPV, logistic regression produces significantly higher values than univariate analysis; (bottom) mean and standard deviation of test accuracy, AUC, and PPV for logistic regression and univariate analysis. (b) T2DM-related perfusion network pattern (shown as red-yellow color overlaid on brain axial slices) identified using machine-learning methods (principal component analysis, logistic regression classification, and backward stepwise feature selection).

difference of the three performances between the three methods can also be observed in Fig. 2a. Specifically, the proposed LR pipeline outperformed the univariate analysis in accuracy, AUC, and PPV measures by 4%, 12%, and 7%, respectively, and the univariate analysis outperformed the simple LR method in accuracy, AUC, and PPV measures by 22%, 15%, and 22%, respectively.

Adjusted McFadden's R square for the proposed LR classification pipeline was 0.315, which was an increase of 24% from the R square of 0.254 for the univariate analysis. The proposed LR classification pipeline can predict a much higher proportion of variability than the univariate analysis. This indicates that the features from the proposed LR pipeline can better explain the difference between T2DM patients and controls.

A T2DM-related covariance pattern was derived using bootstrap estimation procedure to show the effect of T2DM on perfusion. Figure 2b displays the T2DM perfusion pattern overlaid on brain axial slices. The T2DM-related perfusion covariance pattern appears in regions including basal ganglia, insula, limbic, temporal lobes, and regions of prefrontal cortex (Table 2). We found that these individual covariance pattern scores were significantly correlated with demographic and clinical characteristics, including insulin, HbA1c, fasting glucose, HOMA-IR, BMI, and systolic blood pressure (SBP) (Table 1). In addition, the individual pattern scores were also significantly correlated with mobility and cognitive functions: gait speed, HVLT: total recall, HVLT: delayed recall, and VF: animals. Even after removing outliers from our correlation calculations, the results remained significant ( $P < 0.05$ ,  $|r| > 0.3$ ) (Fig. 3). The significance and correlation coefficients are listed in Table 1. Only subjects whose demographic and clinical characteristic, mobility, and cognitive variables occurred less frequently (at large values or small values) were considered outliers. Subjects with abnormal pattern scores (eg, one subject had abnormally high pattern scores) were not considered as outliers initially. However, the association remained significant after removing the subject with the abnormal high pattern score. For the group of T2DM only, the covariance pattern scores were not associated with any of the above-mentioned variables.

For all subjects, the longitudinal covariance pattern score change was marginally associated with change of HbA1c between baseline visit and 2-year follow-up ( $P = 0.067$ ,  $r = 0.30$ ), and not associated with baseline cholesterol ( $P = 0.19$ ,  $r = 0.22$ ). In T2DM subjects only, the longitudinal covariance pattern score change was correlated with the change of HbA1c between baseline and 2-year follow-up ( $P = 0.0053$ ,  $r = 0.64$ ) (Fig. 4a). The association remained significant after excluding the subject with an abnormally high pattern score ( $P = 0.011$ ,  $r = 0.62$ ) and after excluding the subject with extremely large change (almost 4%) of HbA1c value ( $P = 0.023$ ,  $r = 0.56$ ). In T2DM subjects only, we also found an association between the longitudinal covariance pattern score change and baseline cholesterol level ( $P = 0.037$ ,  $r = 0.51$ ) (Fig. 4b). This result remained significant after excluding the subject with an abnormal high pattern score ( $P = 0.049$ ,  $r = 0.50$ ).

## Discussion

The individual perfusion pattern score is a highly promising perfusion imaging biomarker for tracing the disease progression of subjects with T2DM. We have shown that the proposed LR discrimination pipeline increased test accuracy, test AUC, and test PPV by 4%, 12%, and 7%, compared to the traditional univariate method. From the proposed method, we have derived a T2DM-related perfusion covariance pattern, including basal ganglia, insula, limbic, temporal lobes and regions of prefrontal cortex, and perfusion covariance pattern score for each individual subject. The perfusion pattern scores were associated with disease severity, mobility, and cognitive functions in the entire cohort at baseline. The change of the pattern scores at the follow-up was associated with the HbA1c change and the baseline cholesterol level within T2DM patients.

The proposed LR classification pipeline improved the performance measures dramatically compared to the simple LR method. Specifically, the accuracy rate of the simple LR method was only around random chance ( $0.51 \pm 0.10$ ), whereas the accuracy rate of the proposed LR classification pipeline was  $0.77 \pm 0.15$ , indicating that PCA feature reduction and LR feature selection are crucial steps for improved performance. Compared to the performance measures from

**TABLE 1. Significance and Correlation Coefficients for Association of Individual T2DM-Related Network Pattern Scores With Basic Disease Variables, Mobility Function, and Cognitive Functions**

	Insulin	HbA1c	Fasting glucose	HOMA-IR	BMI	SBP	Gait speed	HVLT: total recall	HVLT: delayed recall	VF: animal
<i>P</i> value	0.0015	0.00077	0.0053	0.00026	0.00021	0.0055	0.0031	0.00099	0.0051	0.0000076
<i>r</i> value	0.38	0.39	0.33	0.43	0.43	0.35	-0.36	-0.38	-0.33	-0.50

**TABLE 2. Anatomical Regions of T2DM-Related Perfusion Network Pattern**

Anatomical locations <sup>a</sup>	%Cluster	%Region
Basal Ganglia		
Caudate_R	0.54	53.12
Putamen_R	1.10	100.00
Putamen_L	0.89	85.63
Pallidum_R	0.29	100.00
Pallidum_L	0.26	87.37
Frontal Lobe		
Frontal_Inf_Orb_R	1.42	80.67
Frontal_Inf_Orb_L	1.22	70.12
Frontal_Inf_Oper_R	0.96	66.69
Frontal_Inf_Oper_L	0.41	38.15
Frontal_Inf_Tri_R	1.37	61.83
Frontal_Inf_Tri_L	1.31	50.34
Frontal_Med_Orb_R	0.50	56.78
Frontal_Mid_Orb_R	0.58	55.76
Frontal_Mid_Orb_L	0.50	54.28
Frontal_Sup_Orb_R	0.48	47.14
Frontal_Sup_Orb_L	0.56	56.70
Precentral_R	1.72	49.30
Rectus_R	0.33	43.36
Rectus_L	0.39	44.13
Rolandic_Oper_R	1.29	94.21
Rolandic_Oper_L	0.88	86.46
Insula		
Insula_R	1.71	93.67
Insula_L	1.02	53.44
Limbic		
Amygdala_R	0.25	95.97
Amygdala_L	0.12	52.27
Cingulum_Ant_R	0.66	48.82
Cingulum_Ant_L	0.47	32.79
Hippocampus_R	0.83	84.78
Hippocampus_L	0.51	53.22
Cingulum_Mid_R	1.01	44.62
ParaHippocampus_R	0.84	71.73
Cingulum_Post_R	0.14	41.49
Cingulum_Post_L	0.17	36.29
Temporal Lobe		
Temporal_Inf_R	2.69	73.24
Temporal_Inf_L	1.78	53.84

**TABLE 2. Continued**

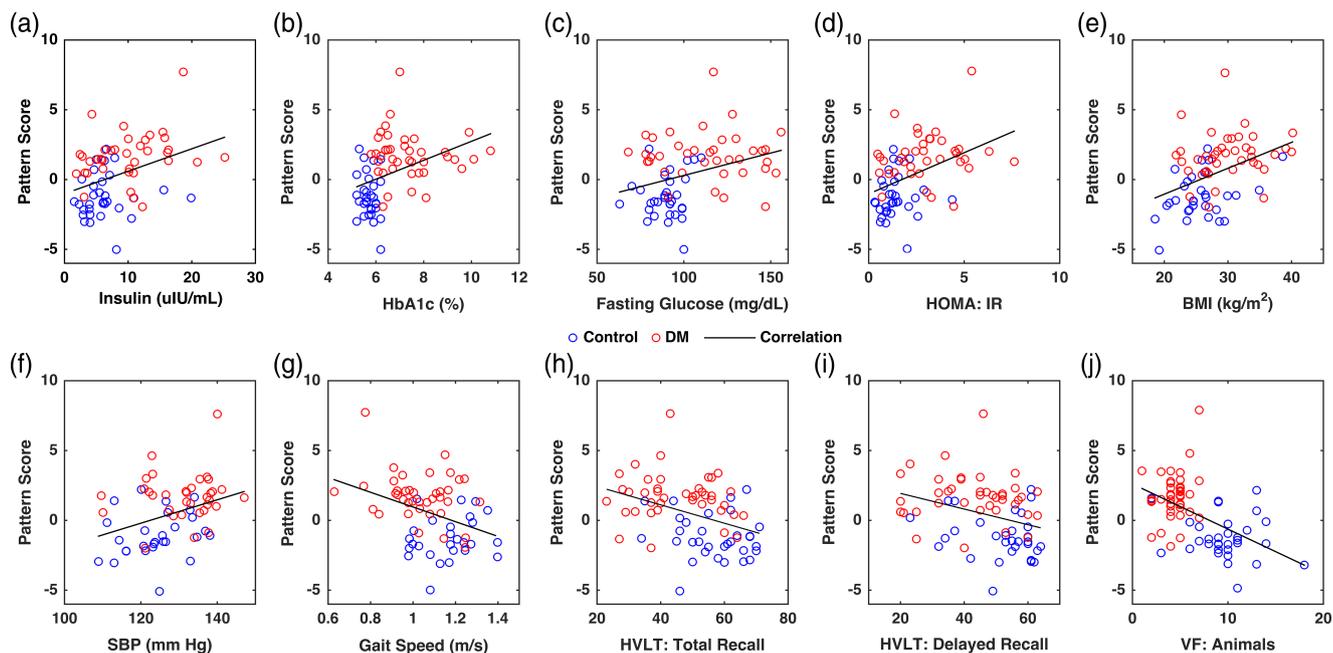
Anatomical locations <sup>a</sup>	%Cluster	%Region
Temporal_Mid_R	4.47	98.39
Temporal_Mid_L	4.72	92.65
Temporal_pole_Mid_R	0.61	50.13
Temporal_pole_Mid_L	0.36	45.96
Temporal_Sup_R	3.19	98.56
Temporal_Sup_L	1.99	84.02
Temporal_Pole_Sup_R	0.81	59.04
Temporal_Pole_Sup_L	0.43	32.61

<sup>a</sup>Labeling of the anatomical regions are based on the Automated Anatomical Labeling (AAL) atlas. %Cluster indicates the percentage of each cluster that falls within the defined region, %Region indicates the percentage of each defined region that falls within the cluster. The listed anatomical regions are either “%Cluster” > 1% or “%Region” > 30%.

the univariate analysis, the performance measures from the proposed LR classification pipeline were not only significantly improved, but also were more consistent even with different training samples (shown as significantly reduced standard deviation). This suggests that the proposed LR classification pipeline is more generalizable and captures more disease-related information.

The association between T2DM and changes in perfusion pattern in the insula region is consistent with a previous <sup>15</sup>O PET longitudinal study that reported perfusion changes in the same area.<sup>32</sup> Additionally, T2DM has been associated with changes in perfusion patterns in the basal ganglia,<sup>17</sup> and the changes of perfusion in limbic, temporal lobes, and prefrontal regions have been linked to the changes of cerebral glucose metabolism in prediabetes and T2DM with the use of <sup>18</sup>F fluorodeoxyglucose PET.<sup>33</sup> The decreased perfusion in temporal lobes, medial frontal, and inferior frontal regions are also in accordance with reduced functional connectivity or fluctuation amplitude using resting-state fMRI.<sup>34,35</sup>

The change of the pattern scores between baseline and follow-up was significantly associated with the HbA1c change even for the T2DM subjects only. It is worth noting that the follow-up subjects were completely independent samples, which was not involved in the LR discrimination model-building process. Therefore, in some extent, the longitudinal study can validate the performance of the pattern scores that were derived from the LR model using only the subjects at baseline. The association result between HbA1c change and the change of pattern scores indicates that T2DM affects the brain perfusion with a fixed spatial pattern. Resting-state fluctuation amplitude in the middle temporal gyrus has been reported to inversely correlate with HbA1c.<sup>35</sup> Our longitudinal study adds to

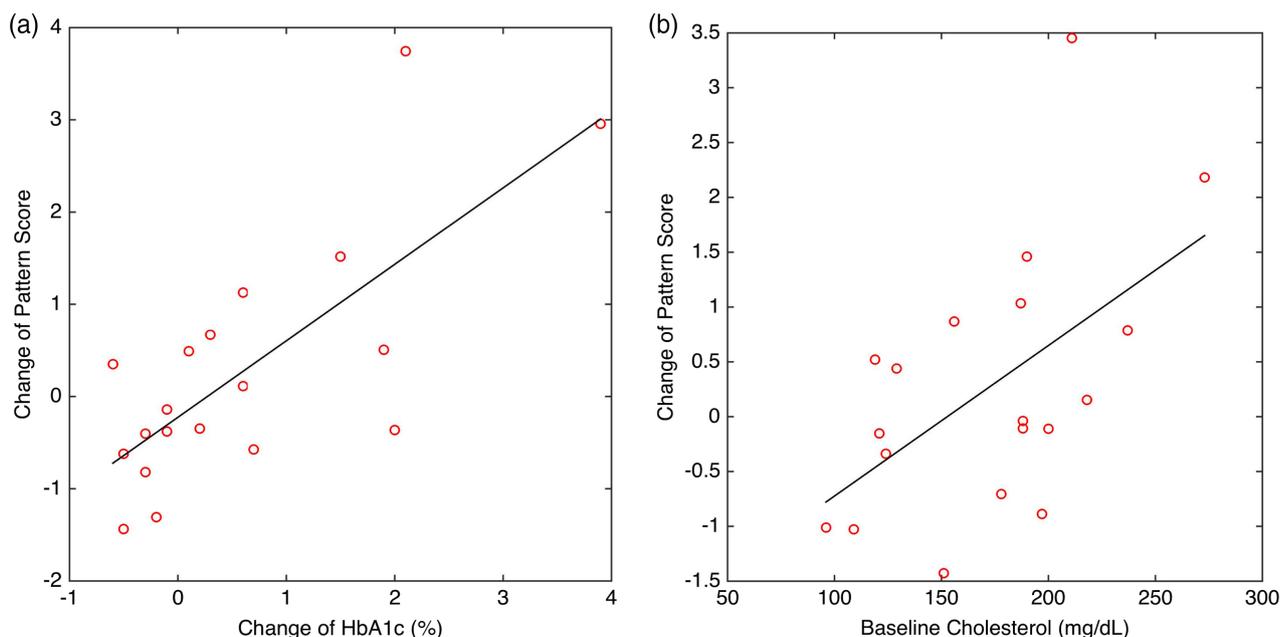


**FIGURE 3:** Significant association ( $P < 0.05$ ,  $|r| > 0.3$ ) of perfusion covariance pattern scores with (a) insulin level, (b) HbA1c, (c) fasting glucose, (d) HOMA-IR, (e) BMI, and (f) SBP, (g) gait speed, (h) HVLT: total recall, (i) HVLT: delayed recall, (j) VF: animals, even after removing the outliers.

the literature that the change of longitudinal perfusion pattern score closely follows the change of HbA1c in T2DM subjects. This finding also reinforces the importance of HbA1c as a significant predictor of brain function, which extends to the literature that glycemic control may potentially prevent a decline of brain function.

We found that the baseline cholesterol level was a significant predictor of the longitudinal change of perfusion pattern scores. Our study confirms that a very high cholesterol level is detrimental to brain function, which lends a potential

intervention point for T2DM. Previous cross-sectional studies have supported high cholesterol level as a risk factor to T2DM. An earlier structural study has linked T2DM to reduced hippocampal and prefrontal volumes and established a negative association between obesity and hippocampal volume.<sup>36</sup> BMI has been associated with T2DM.<sup>37</sup> Considering the positive correlation between BMI, obesity, and cholesterol level,<sup>38,39</sup> the literature is joining together to support the benefit of maintaining a right level of cholesterol for reduced risk for T2DM. We did not observe the association between longitudinal perfusion



**FIGURE 4:** Greater longitudinal increase in T2DM-related pattern scores at the 2-year follow-up is associated with (a) more increase in HbA1c and (b) higher baseline cholesterol. T2DM = type 2 diabetes mellitus; HbA1c = glycated hemoglobin A1c.

change and change of HbA1c, and between longitudinal perfusion change and baseline variables using the traditional univariate method.<sup>40</sup> Our results highlight that the advanced machine-learning-based methods on brain perfusion images can improve sensitivity for characterization of disease status and detection of risk factors for T2DM.

The study is not without limitations. The machine-learning-based pipeline was applied to a not-so-large sample size. However, we strived to reduce the potential overfitting problems in the dataset with k-fold cross-validation and feature selection. We applied the derived model to a completely independent dataset to assess the ability of our model for detecting the association of longitudinal change and change of disease-related variables. We expect that the model will lose the capability to track the longitudinal change if the model overfits the baseline data to a large extent. However, we found significant correlation between longitudinal change of perfusion scores and disease severity with an even smaller sample size at the follow-up. These longitudinal results indicate that the individual covariance pattern scores may serve as a T2DM biomarker, and that maintenance of blood sugar and reduction of cholesterol may mitigate cognitive decline and mobility impairment of T2DM. The covariance pattern scores hold great potential to monitor the longitudinal disease progress of T2DM. Our T2DM subjects were randomly drawn from those diagnosed with T2DM and treated with oral agents and/or insulin for more than 5 years and had similar MMSE, HVLT: Retention, TM compared to controls. The derived model should only reflect the variability of the T2DM subjects within the range of cognitive states and may not represent the global patterns of T2DM in a wide range of cognitive states. A study with larger sample size is warranted to validate the generalizability of the derived LR model and clinical significance of the promising biomarker.

In conclusion, we developed a machine-learning-based method for discriminating T2DM from controls and identifying a T2DM-related pattern using perfusion images measured with the ASL technique. The developed discrimination method increased test accuracy, test AUC, and test PPV by 4%, 12%, and 7%, compared to the traditional univariate method. We also derived a T2DM-related perfusion pattern. More important, the T2DM-related individual perfusion pattern scores were significantly associated with disease severity, as well as with mobility and cognitive function. The T2DM-related pattern scores hold great promise to become a biomarker to trace the progression of T2DM at the individual level.

## Acknowledgments

This work was supported by NIH-NIDDK 1R01DK103902-01A1 and NIH-NIA 1R01-AG0287601A2 to Dr. Vera

Novak. The project described was supported by Grant Number UL1 RR025758- Harvard Clinical and Translational Science Center and M01-RR-01032, from the National Center for Research Resources

## APPENDIX

### **Subject Inclusion and Exclusion Criteria at Baseline**

Subjects with T2DM were treated for diabetes for more than 5 years. Nondiabetic controls were age- and sex-matched with normal fasting glucose and HbA1c. Inclusion criteria were: age 50–85, diagnosis of T2DM and treated for more than 5 years, nondiabetic controls, hypertensive (BP > 140/90 mmHg and/or treated for hypertension) and normotensive (BP < 140/90 mmHg and no medical history of hypertension). Exclusion criteria were type 1 diabetes, heart disease, major surgery in the previous 6 months, stroke, carotid artery stenosis, liver or renal insufficiency, severe hypertension (SBP > 200 mm Hg or diastolic blood pressure [DBP] > 10 mm Hg or taking three or more antihypertensive medications), seizures, malignant tumors, recreational drug or alcohol abuse, BMI >40 kg/m<sup>2</sup>, dementia, or subthreshold MMSE score ( $\leq 24$ ). MRI exclusion criteria included incompatible metal implants, pacemakers, and claustrophobia. T2DM and control subjects were consecutively recruited from advertisement in the community. We used frequency quota sampling to match age ( $\pm 5$  years) and sex distribution between the groups.

Subjects were excluded from the baseline analyses based on the following reasons: consent withdrawal ( $n = 11$ ), lost to follow-up ( $n = 10$ ), MMSE  $\leq 24$  ( $n = 3$ ), stroke/TIA ( $n = 2$ ), arrhythmia ( $n = 4$ ), active cancer ( $n = 2$ ), smoking ( $n = 1$ ), heart failure ( $n = 1$ ), MRI exclusion ( $n = 1$ ), renal insufficiency ( $n = 1$ ), T2DM < 5 years ( $n = 3$ ), uncontrolled hypertension ( $n = 3$ ), unidentified neurological disorders ( $n = 2$ ), poorly controlled glycemia ( $n = 4$ ), adverse events ( $n = 1$ ), and incomplete datasets ( $n = 9$ ). For the 2-year follow-up analyses, subjects were excluded for: consent withdrawal ( $n = 5$ ), lost to follow-up ( $n = 25$ ), and dementia ( $n = 1$ ).

### **Rationale for Sample Size**

Sample size was calculated based on our preliminary data that indicated that with 60 subjects (30 T2DM patients and 30 controls) we will have 87% power to detect differences of 10.8 mL/100 mg/min in global perfusion measured by PCASL between T2DM patients ( $41.8 \pm 12.8$  mL/100 mg/min, HbA1c > 7%) and controls ( $52.6 \pm 14.2$  mL/100 mg/min, HbA1c < 7%) based on two-sided *t*-test and  $\alpha = 0.05$ .

### **Experimental Protocol**

Subjects were screened using a medical history questionnaire, autonomic function questionnaires, ECG, and laboratory measures (blood, glucose, and renal panels). After enrollment, subjects came for their 2-day inpatient visit at the Beth Israel Deaconess Medical Center (BIDMC) clinical research center

(CRC). On Day 1 patients had vital signs measured including SBP and DBP, and anthropometric measurements including height, weight, BMI, and a cognitive assessment battery testing. On Day 2, subjects had a fasting blood draw for hematocrit, glucose, insulin, and HbA1c, cognitive assessment, walking test, and MRI scans. Two years later, subjects came for a follow-up visit including the exact same measurements and tests. Cognitive and functional assessments were done by the research fellows trained in these procedures and by the study Principal Investigator (V.N.). This was an observational study and there was no blinding to the study procedures.

### Walking Test/Gait Assessment

Subjects completed two 6-min walking tests on a 75-m course of an 80 x 4 m indoor hallway. Subjects were instructed to walk at their natural and comfortable pace. The time taken to complete each 75-m-length course as well as the total distance walked were recorded. No assistant devices were used for ambulation. Gait speed was calculated by dividing distance (m) by time (s).

### References

1. Biessels GH, Deary IJ, Ryan CM. Cognition and diabetes: a lifespan perspective. *Lancet Neurol* 2008;7:184–190.
2. Cukierman T, Gerstein HC, Williamson JD. Cognitive decline and dementia in diabetes—systematic overview of prospective observational studies. *Diabetologia* 2005;48:2460–2469.
3. Starr JM, Wardlaw J, Ferguson K, et al. Increased blood-brain barrier permeability in type II diabetes demonstrated by gadolinium magnetic resonance imaging. *J Neurol Neurosurg Psychiatry* 2003;74:70–76.
4. Mogi M, Horiuchi M. Neurovascular coupling in cognitive impairment associated with diabetes mellitus. *Circ J* 2011;75:1042–1048.
5. Umegaki H. Type 2 diabetes as a risk factor for cognitive impairment: current insights. *Clin Interv Aging* 2014;9:1011–1019.
6. Zhou H, Zhang X, Lu J. Progress on diabetic cerebrovascular diseases. *Bosn J Basic Med Sci* 2014;14:185–190.
7. Roberts RO, Knopman DS, Cha RH, et al. Diabetes and elevated hemoglobin A1c levels are associated with brain hypometabolism but not amyloid accumulation. *J Nucl Med* 2014;55:759–764.
8. Brundel M, van den Berg E, Reijmer YD, et al. Utrecht Diabetic Encephalopathy Study, cerebral haemodynamics, cognition and brain volumes in patients with type 2 diabetes. *J Diabetes Complic* 2012;26:205–209.
9. Xia W, Wang S, Rao H, et al. Disrupted resting-state attentional networks in T2DM patients. *Sci Rep* 2015;5:11148.

**TABLE A1. Demographic, Clinical Ratings, Postural Control, and Cognitive Scores at Baseline and Follow-Up**

	Baseline			Two-year follow-up		
	T2DM (n = 41)	Controls (n = 32)	P-value	T2DM (n = 19)	Controls (n = 23)	P-value
Age (years)	65.51 ± 8.30	67.28 ± 10.08	NS	66.94 ± 8.18	67.09 ± 9.29	NS
Gender (females, N, %)	22 (54%)	16 (50%)	NS	12 (52%)	12 (63%)	NS
Education (years)	15.35 ± 3.78	16.05 ± 2.98	NS	14.03 ± 2.93	16.35 ± 2.81	0.012
Hematocrit (%)	38.83 ± 3.70	39.54 ± 3.89	NS	38.58 ± 3.21	38.71 ± 3.56	NS
Hypertension (N, %)	32 (78%)	7 (22%)	≤ 0.001	15 (79%)	5 (22%)	≤ 0.001
Diabetes Duration (years)	9.93 ± 7.91	—	—	9.63 ± 6.91	—	—
BMI (kg/m <sup>2</sup> )	29.12 ± 6.77	25.17 ± 6.68	0.007	29.46 ± 5.43	24.05 ± 3.08	≤ 0.001
Insulin (uIU/ml)	13.61 ± 13.30	6.26 ± 3.77	0.003	15.19 ± 17.55	6.06 ± 4.30	0.020
Fasting Glucose (mg/dl)	119.70 ± 36.78	89.94 ± 10.22	≤ 0.001	112.69 ± 31.81	91.59 ± 8.92	0.005
HOMA-IR*	3.87 ± 3.42	1.40 ± 0.83	≤ 0.001	4.35 ± 4.67	1.33 ± 0.92	≤ 0.001
HbA1c (%)	7.34 ± 1.25	5.72 ± 0.30	≤ 0.001	7.82 ± 1.79	5.64 ± 0.33	≤ 0.001
Gait Speed (m/s)	1.03 ± 0.15	1.16 ± 0.13	≤ 0.001	1.04 ± 0.14	1.18 ± 0.13	0.004
MMSE	28.59 ± 1.52	28.94 ± 1.56	NS	28.76 ± 1.35	28.83 ± 1.70	NS
HVLT: Total Recall	12.61 ± 13.04	30.40 ± 5.12	≤ 0.001	24.26 ± 6.07	27.96 ± 5.50	0.045
HVLT: Delayed Recall	43.03 ± 12.98	52.16 ± 10.86	0.002	42.47 ± 15.42	53.48 ± 11.70	0.012
HVLT: Retention	81.87 ± 16.79	87.20 ± 16.66	NS	79.34 ± 17.95	84.63 ± 15.90	NS
VF: Animal	4.39 ± 1.38	9.72 ± 3.14	≤ 0.001	8.37 ± 3.30	11.48 ± 3.22	0.004

10. Tiehuis AM, Vincken KL, van den Berg E, et al. Cerebral perfusion in relation to cognitive function and type 2 diabetes. *Diabetologia* 2008;51:1321–1326.
11. Nagamachi S, Nishikawa T, Ono S, et al. Regional cerebral blood flow in diabetic patients: evaluation by N-isopropyl-123I-IMP with SPECT. *Nucl Med Commun* 1994;15:455–460.
12. Sabri O, Hellwig D, Schreckenberger M, et al. Influence of diabetes mellitus on regional cerebral glucose metabolism and regional cerebral blood flow. *Nucl Med Commun* 2000;21:19–29.
13. Rusinek H, Ha J, Yau PL, et al. Cerebral perfusion in insulin resistance and type 2 diabetes. *J Cereb Blood Flow Metab* 2015;35:95–102.
14. Novak V, Zhao P, Manor B, et al. Adhesion molecules, altered vasoreactivity, and brain atrophy in type 2 diabetes. *Diabetes Care* 2011;34:2438–2441.
15. Last D, Alsop DC, Abduljalil AM, et al. Global and regional effects of type 2 diabetes on brain tissue volumes and cerebral vasoreactivity. *Diabetes Care* 2007;30:1193–1199.
16. Xia W, Rao H, Spaeth AM, et al. Blood pressure is associated with cerebral blood flow alterations in patients with T2DM as revealed by perfusion functional MRI. *Medicine* 2015;94:e2231.
17. Jansen JF, van Bussel FC, van de Haar HJ, et al. Cerebral blood flow, blood supply, and cognition in Type 2 Diabetes Mellitus. *Sci Rep* 2016; 6:10.
18. Novak V, Last D, Alsop DC, et al. Cerebral blood flow velocity and periventricular white matter hyperintensities in type 2 diabetes. *Diabetes Care* 2006;29:1529–1534.
19. Fox MD, Snyder AZ, Vincent JL, et al. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* 2005;102:9673–9678.
20. 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;21:46–57.
21. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 2009;45(1 Suppl):S199–209.
22. Casanova R, Hsu FC, Sink KM, et al. Alzheimer's disease neuroimaging, Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One* 2013;8:e77949.
23. Bron EE, Steketee RM, Houston GC, et al. Alzheimer's disease neuroimaging, diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. *Hum Brain Mapp* 2014;35:4916–4931.
24. Shapiro AM, Benedict RH, Schretlen D, Brandt J. Construct and concurrent validity of the Hopkins Verbal Learning Test-revised. *Clin Neuropsychol* 1999;13:348–358.
25. Benton AL, Hamsher K. *Multilingual aphasia examination*. Manual of instructions (2nd ed). 1989, Iowa City, IA: AJA Associates.
26. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;28:412–419.
27. Dai W, Garcia D, de Bazelaire C, Alsop DC. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn Reson Med* 2008;60:1488–1497.
28. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magn Reson Med* 1998;40:383–396.
29. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage* 2005;26:839–851.
30. Long JS, Freese J. *Regression models for categorical dependent variables using Stata*. 2006, College Station, TX: Stata Press.
31. Long JS. *Regression models for categorical and limited dependent variables*. 1997, Thousand Oaks, CA: Sage Publications.
32. Beason-Held LL, Thambisetty M, Deib G, et al. Baseline cardiovascular risk predicts subsequent changes in resting brain function. *Stroke* 2012; 43:1542–1547.
33. Baker LD, Cross DJ, Minoshima S, Belongia D, Watson GS, Craft S. Insulin resistance and Alzheimer-like reductions in regional cerebral glucose metabolism for cognitively normal adults with prediabetes or early type 2 diabetes. *Arch Neurol* 2011;68:51–57.
34. Musen G, Jacobson AM, Bolo NR, et al. Resting-state brain functional connectivity is altered in type 2 diabetes. *Diabetes* 2012;61:2375–2379.
35. Xia W, Wang S, Sun Z, et al. Altered baseline brain activity in type 2 diabetes: a resting-state fMRI study. *Psychoneuroendocrinology* 2013;38:2493–2501.
36. Bruehl H, Wolf OT, Sweat V, Tirsli A, Richardson S, Convit A. Modifiers of cognitive function and brain structure in middle-aged and elderly individuals with type 2 diabetes mellitus. *Brain Res* 2009;1280:186–194.
37. Zhao Q, Laukkanen JA, Li Q, Li G. Body mass index is associated with type 2 diabetes mellitus in Chinese elderly. *Clin Interv Aging* 2017;12:745–752.
38. Faheem M, Qureshi S, Ali J, et al. Does BMI affect cholesterol, sugar, and blood pressure in general population? *J Ayub Med Coll Abbottabad* 2010;22:74–77.
39. Nwaiwu O, Ibe BC. Relationship between serum cholesterol and body mass index in Nigeria schoolchildren aged 2-15 years. *J Trop Pediatr* 2015;61:126–130.
40. Dai W, Duan W, Alfaro FJ, Gavrieli A, Kourtellis F, Novak V. The resting perfusion pattern associates with functional decline in type 2 diabetes. *Neurobiol Aging* 2017;60:192–202.