Isometric Feature Embedding for Content-based Image Retrieval

Hayato Muraki[1], Kei Nishimaki[1], Shuya Tobari[1], Kenichi Oishi[2], Hitoshi Iyatomi[1]
for the Alzheimer's Disease Neuroimaging Initiative*.
[1]*Department of Applied Informatics, Graduate School of Science and Engineering, Hosei University, Tokyo, Japan*
[2]*Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, USA*
{hayato.muraki.5m, kei.nishimaki.1106, shuya.tobari.7b}@gmail.com, koishi2@jhmi.edu, iyatomi@hosei.ac.jp

***Abstract*—Content-based image retrieval (CBIR) technology for brain MRI is needed for diagnostic support and research. To realize practical CBIR, it is necessary to obtain a low-dimensional representation that simultaneously achieves (i) data integrity, (ii) high disease retrieval capability, and (iii) interpretability. However, conventional methods based on machine learning techniques such as variational autoencoders (VAE) cannot acquire representations that satisfy these requirements, and an ad-hoc classification model must be prepared for disease retrieval. In this paper, we propose Isometric Feature Embedding for CBIR (IE-CBIR), a low-dimensional representation acquisition framework that simultaneously satisfies the above requirements. In the evaluation experiment using the ADNI2 dataset of t1-weighted 3D brain MRIs from 573 subjects (3,557 cases in total), the low-dimensional representation acquired by IE-CBIR (1/4,096 of the number of elements compared to the original) achieved a classification performance of 0.888 in F1 score and 91.5% in accuracy for Alzheimer's disease and normal cognitive subjects without the need for ad hoc models, while achieving a high preservation of the original data. This diagnostic performance outperformed machine learning methods such as CNNs (76%-91% accuracy), which specialize in classification without considering the acquisition of low-dimensional representations and their interpretability.**

***Index Terms*—ADNI, CBIR, dimensional reduction, 3D brain MRI**

I. Introduction

Magnetic resonance imaging (MRI) is widely used for diagnosing neurological disorders since it can capture the anatomical features of the brain without radiation exposure [1]. Picture archiving and communication system (PACS) [2] store MRIs with the associated clinical information and centralize management of scanned images. In PACS, MRIs are generally retrieved and registered using specialized keywords that indicate the anatomical and clinical features of the brain [3]. However, selecting the specific keywords requires a high level of expertise and experience. In addition, the human cost of associating images with specific keywords is extremely expensive. Therefore, content-based image retrieval (CBIR), which retrieves MRIs of similar cases from only input images without using specific keywords, is desired [4]. The brain MRIs are high-dimensional data, and it is not possible to directly calculate similarity using machine learning methods. Therefore, for practical CBIR realization, it is necessary to obtain a low-dimensional representation that:

(i) preserves the original information (i.e., data integrity),
(ii) has high disease retrieval capability,
(iii) has high interpretability.

As a first study aimed at realizing this objective, Arai et al. compressed approximately six million dimensions of brain MRI to 150 dimensions using 3D convolutional autoencoders (3D-CAE) [5]. Despite such a large dimensional compression, their model suppressed the reconstruction error to 8.4% of the range that the pixel could take. Indicating that it could obtain a low-dimensional representation that retained detailed structural information of the brain. However, this method focused only on (i) data integrity. Onga et al. proposed a dimensional reduction method called disease-oriented data concentration with metric learning (DDCML) [6], in which subjects with the same diagnostic information are embedded closely on a low-dimensional representation. They added constraints of metric learning, which has been successful in fields such as general object recognition [7], to the loss function of the 3D-CAE. As a result, their method made it possible to evaluate similarity calculations focusing on diagnostic information without an ad-hoc discrimination model. Nishimaki et al [8]. proposed localized variational autoencoder (Loc-VAE), a structurally interpretable low-dimensional representation acquisition method based on VAE [9]. In general, the VAE is trained so that each dimension of the low-dimensional representation follows a normal distribution by introducing KL divergence. Therefore, the independence and continuity of each dimension are strongly guaranteed compared to CAE. In addition, Loc-VAE succeeded in obtaining a more structurally interpretable representation by introducing a new constraint that each dimension of the low-dimensional representation retains information about a local region of the input image. These methods have achieved some success in obtaining low-dimensional representations for realizing CBIR for brain MRIs. However, they have not yet achieved all three requirements listed above, and further research is needed.

On the other hand, quantitative analysis and theoretical elu-

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

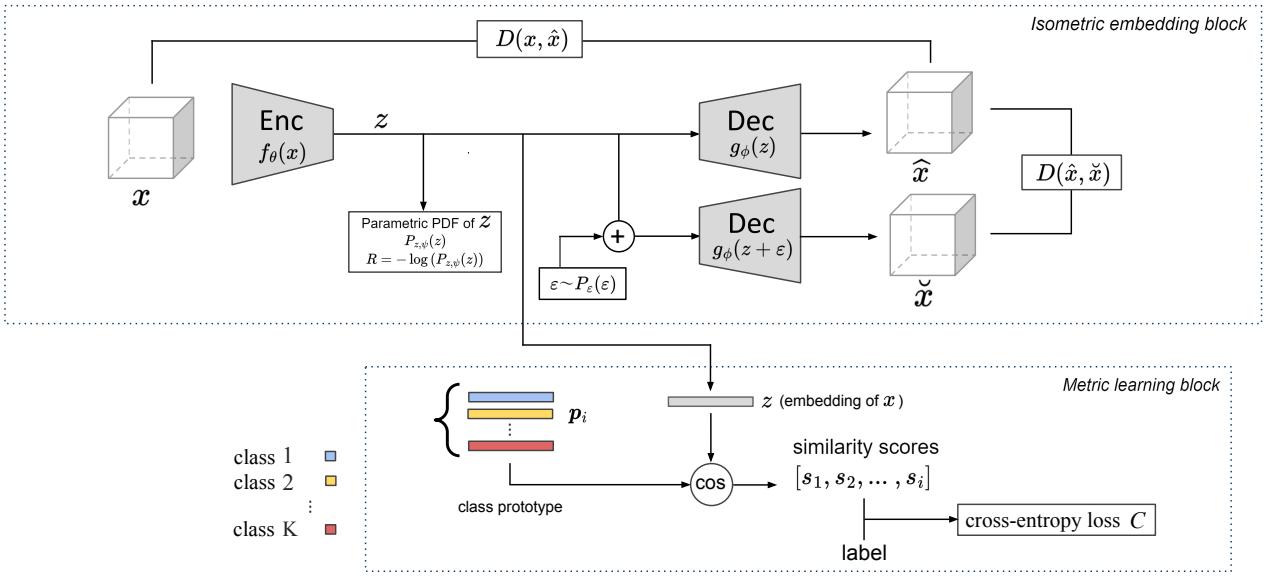


Fig. 1. The overview of isometric feature embedding for content-based image retrieval (IE-CBIR).

cidation of machine learning models proceeded actively with the development of large generative models. The generative models generally train the probability distribution of the input dataset, but it is difficult to directly obtain the probability density function (PDF) $P_{\boldsymbol{z}}(\boldsymbol{z})$ $(\boldsymbol{z} \in Z)$. Therefore, the VAE and its advanced model (Gaussian mixture model) [10] have calculated the probability density function of the latent space Z. However, the latent space obtained by VAE has no isometric relationship with the real space [11]–[13]. In other words, the distance between data points in the real space does not correspond to the distance in the latent space. In such a situation, a simple estimation of $P(\boldsymbol{z})$ can not replace the estimation of the probability density function $P_{\boldsymbol{x}}(\boldsymbol{x})$ $(\boldsymbol{x} \in X)$ in the real space X. To make interpretable CBIR, it is necessary to embed into a low-dimensional representation that maintains isometric with the real space [14]. Kato et al. proposed a rate-distortion optimization guided autoencoder (RaDOGAGA) that maintains isometry between the real space and the latent space and demonstrated its excellent performancel [15].

In addition, for the realization of the clinical CBIR, it is important to obtain a low-dimensional representation that can be clustered by disease features while maintaining isometricity with real space. Metric learning [16] (or recently called contrastive learning), in which data belonging to the same class are learned to have similar representations, while data belonging to different classes are learned to have distant representations, has received particular attention in recent years and is used in a variety of research areas, from basic to applied. Chen et al. proposed Baseline++, a general-purpose and effective machine learning concept that replaces the last linear classifier or similar small networks used for fine-tuning of arbitrary classifiers, etc., with distance-based classifiers [17]. Baseline++ allows each discriminative class to be corresponding class prototype [18] and updates it with training data to obtain a low-dimensional representation representative of the class. In addition to achieving significant accuracy gains in a variety of classification tasks, this method also provides a high degree of model explainability, since classification is only possible by the distance to the obtained low-dimensional representation. By training distance-based classifiers, clusters are formed and, variation within classes is explicitly reduced, and excellent performance is achieved in many recognition tasks [19], [20].

In this study, we propose isometric feature embedding for content-based image retrieval (IE-CBIR), which is a new low-dimensional representation acquisition framework. Our IE-CBIR generates a low-dimensional representation that simultaneously satisfies the above three important requirements for realizing CBIR for brain MRIs. The IE-CBIR is based on the ideas of RaDOAGA [15], a method of latent space acquisition that guarantees isometry between real and latent space, and Baseline++ [17], which allows classification on the distance of the low-dimensional representation itself. This achieves both high preservation of information on changes in brain shape related to diseases and the explanatory power of measuring similarity directly on the low-dimensional representations. Our proposed IE-CBIR simultaneously achieves conservation, explainability that allows direct similarity measurement in a low-dimensional representation, and continuity.

II. IE-CBIR

Our IE-CBIR is an effective and generic method for obtaining low-dimentional representation of the clinical CBIR. Fig. 1 shows an overview of the IE-CBIR. The IE-CBIR consists of the Isometric embedding block (upper part in Fig. 1) based on

the RaDOGAGA and the Metric learning block (lower part in Fig. 1) based on the Baseline++. By integrating two blocks, the IE-CBIR acquires a low-dimensional representation $\boldsymbol{z} \in \mathbb{R}^D$ that guarantees the isotonicity of the original data with the real space and allows data exploration using itself. In other words, we can obtain a low-dimensional representation $\boldsymbol{z}$ that directly calculates the distance between diseases in the latent space.

A. Isometric embedding block

The Isometric embedding block is based on the RaDOGAGA, a machine learning method that aims to efficiently compress input information and achieve low distortion. The RaDOGAGA is an encoder-decoder type unsupervised learning model similar to the VAE. High bit rate is realized by constraints that shorten the average coding length of $\boldsymbol{z}$ in the generated representation, i.e., the entropy of the distribution of $\boldsymbol{z}$ is reduced, and low data distortion is a model realized by making each dimension real space and latent space isometric, which was not considered in the VAE. Let $f_\theta(\boldsymbol{x})$, $g_\phi(\boldsymbol{z})$, and $P_{\boldsymbol{z},\psi}(\boldsymbol{z})$ be the parametric encoder, decoder, and the PDF of the latent variable with parameters θ, ϕ, and ψ. $D(\cdot, \cdot)$ is the general reconstruction loss, and in this paper we used the root mean squared error (RMSE). The loss function for this block consists of three components as in the original paper [15].

First, $R = -\log(P_{\boldsymbol{z},\psi}(\boldsymbol{z}))$ is the loss for higher bitrates in the model defined by the RaDOGAGA and is defined assuming a code length of $\boldsymbol{z}$. The probability model for $\boldsymbol{z}$ here can be chosen arbitrarily, and we assumed the same normal distribution as in the original paper. Second, $D(\boldsymbol{x}, \hat{\boldsymbol{x}})$ is the reconstruction error. Third, $D(\hat{\boldsymbol{x}}, \breve{\boldsymbol{x}})$ is the difference between the reconstruction from $\boldsymbol{z}$ and the reconstruction from $\boldsymbol{z} + \boldsymbol{\epsilon}$ $(\in \mathbb{R}^D)$. $\boldsymbol{\epsilon}$ is a uniform noise with mean 0 and variance σ^2.

B. Metric learning Block

The Metric learning block cooperates with the previous block to update the low-dimensional representation $\boldsymbol{z}$ to a representation in which CBIR can only be realized by performing a neighborhood search on it. Specifically, based on the Baseline++ approach, the prototype $\boldsymbol{p}_i \in \mathbb{R}^D$ (colored vectors in Fig.1) corresponding to each disease i that has the same dimensionality as $\boldsymbol{z}$ is updated by the metric learning based on cosine similarity to obtain more appropriate representation. The loss in this block is obtained by the cross-entropy loss C with the similarity score obtained by the cosine similarity between the representation $\boldsymbol{z}$ of each case and the prototype $\boldsymbol{p}_i$. The inference is based on the distance between the low-dimensional representation and the prototype of each class. This framework, which allows class estimation based on neighborhood search alone, is very important in the CBIR and is the main reason why we focused on this method.

However, the Baseline++ is designed to replace the fully connected layer at the back end of the model. That is, it assumes that its low-dimensional representation $\boldsymbol{z}$ is acquired based on supervised learning and contains strong class information inside the representation (i.e., typically a representation of the penultimate layer of the model). In addition, the low-dimensional representation $\boldsymbol{z}$ obtained in our previous block is a representation obtained by unsupervised learning, so the expected results cannot be obtained by using Baseline++ as is. Therefore, unlike the original Baseline++, IE-CBIR updates not only the prototype $\boldsymbol{p}_i$, but also the low-dimensional representation $\boldsymbol{z}$, by simultaneously learning the encoder of the isometric embedding block using the gradient information of the cross-entropy loss C.

TABLE I
NUMBER OF CASES PER DATASET

	# cases		# patients	
	CN	AD	CN	AD
Train	1,844	1,003	265	192
Test	523	187	72	44
Total	2,367	1,190	337	236

C. Entire objective of IE-CBIR

The coordination of these two blocks allows the IE-CBIR to efficiently compress the original information while achieving a low-distortion and highly explainable low-dimensional representation that can directly measure similarity in a low-dimensional representation. The overall loss L for IE-CBIR to obtain a low-dimensional representation $\boldsymbol{z}$ is

$$L = R + \lambda_1 D(\boldsymbol{x}, \hat{\boldsymbol{x}}) + \lambda_2 D(\hat{\boldsymbol{x}}, \breve{\boldsymbol{x}}) + \lambda_3 C. \tag{1}$$

Here, λ_1, λ_2, and λ_3 are loss coefficients.

III. EXPERIMENT

A. Datasets and preproccessing

The performance of our IE-CBIR was evaluated using data from the publicly available ADNI2[1] dataset labeled Alzheimer's disease (AD) and Cognitive normal (CN). Table I shows a breakdown of the data used in this experiment. The images of the same subjects were not included in model training and evaluation. This was to prevent bias caused by mixing similar data into training and validation fold.

The N4 bias field correction [21] was applied to all images for correcting intensity non-uniformity. In addition, all images were skull-stripped using a deep learning-based method [22]. To fix the brain position, each brain was transformed in MNI space (standard reference space for brain MRI) using only rotation and translation (i.e., the size of the brain remains the same). The matrix size was then unified to $80 \times 112 \times 80$ and the resolution to $2\,mm \times 2\,mm \times 2\,mm$ by zero padding and bi-linear interpolation. The MRIs often have different intensity distributions and contrasts depending on the protocol or operator used during the scan. Therefore, the intensity normalization was performed. Specifically, the standard deviation σ of the intensity was calculated for each case, and negative values and values greater than 4σ were replaced by 0 and 4σ, respectively. Then, we scaled the data to the range of [0, 1].

[1] https://adni.loni.usc.edu

TABLE II
DETAILED ARCHITECTURE OF THE IE-CBIR †

Encoder	Act.	Output shape
Input image	-	$1 \times 80 \times 112 \times 80$
Conv block	$\begin{bmatrix} 3 \times 3 \times 3, & 32 \\ 3 \times 3 \times 3, & 32 \end{bmatrix}$	$32 \times 80 \times 112 \times 80$
Avg pool	-	$32 \times 40 \times 56 \times 40$
Conv block	$\begin{bmatrix} 3 \times 3 \times 3, & 32 \\ 3 \times 3 \times 3, & 64 \end{bmatrix}$	$64 \times 40 \times 56 \times 40$
Avg pool	-	$64 \times 20 \times 28 \times 20$
Conv block	$\begin{bmatrix} 3 \times 3 \times 3, & 64 \\ 3 \times 3 \times 3, & 128 \end{bmatrix}$	$128 \times 20 \times 28 \times 20$
Avg pool	-	$128 \times 10 \times 14 \times 20$
Conv block	$3 \times 3 \times 3, 128$	$128 \times 10 \times 14 \times 10$
Res block	$3 \times 3 \times 3, 128$	$128 \times 10 \times 14 \times 10$
Conv	$3 \times 3 \times 3, 128$	$1 \times 10 \times 14 \times 10$
latent $\boldsymbol{z}$	-	$1,400 \times 1 \times 1 \times 1$

Decoder	Act.	Output shape
latent $\boldsymbol{z}$	-	$1 \times 10 \times 14 \times 10$
Conv block	$3 \times 3 \times 3, 128$	$128 \times 10 \times 14 \times 10$
Res block	$3 \times 3 \times 3, 128$	$128 \times 10 \times 14 \times 10$
Conv block	$3 \times 3 \times 3, 128$	$128 \times 10 \times 14 \times 10$
Upsample	-	$128 \times 20 \times 28 \times 20$
Conv block	$\begin{bmatrix} 3 \times 3 \times 3, & 128 \\ 3 \times 3 \times 3, & 64 \end{bmatrix}$	$64 \times 20 \times 28 \times 20$
Upsample	-	$64 \times 40 \times 56 \times 40$
Conv block	$\begin{bmatrix} 3 \times 3 \times 3, & 64 \\ 3 \times 3 \times 3, & 32 \end{bmatrix}$	$64 \times 40 \times 56 \times 40$
Upsample	-	$32 \times 80 \times 112 \times 80$
Conv block	$3 \times 3 \times 3, 32$	$32 \times 80 \times 112 \times 80$
Conv	$3 \times 3 \times 3, 1$	$1 \times 80 \times 112 \times 80$
ReLU	-	$1 \times 80 \times 112 \times 80$

† All the models used in this experiment for comparison (CAE, VAE, RaDOGAGA, DDCML) have this configuration.

B. Configuration of IE-CBIR

The IE-CBIR is composed of an encoder and a decoder, each of which consists of a 10-layer CNN, similar to the conventional CAE, VAE, and RaDOGAGA. The specific configuration of the IE-CBIR is shown in Table II. In this experiment, the size of the input and output of these models (i.e., reconstruction) are $80 \times 112 \times 80$ as described above and the dimension of the low-dimensional representation $\boldsymbol{z}$, was set to 1,400 (i.e., $D = 1,400$). The Conv block comprises convolution, batch normalization, and rectified linear unit (ReLU) activation. The Residual block has a configuration published in the literature [23], combining multiple convolutions with skip connections to stabilize training as the number of layers increases.

C. Evaluation

The effectiveness of the low-dimensional representation obtained by the proposed IE-CBIR was evaluated in (1) the reconstruction capability of the original brain, and (2) the data retrieval capability and its interpretability. The first evaluation corresponds to the requirement (i) for low-dimensional representation described above, and the latter corresponds to its (ii) and (iii). The performance of the proposed (E) IE-CBIR has been compared with (A) 3D-CAE, (B) 3D-VAE, (C) RaDOGAGA [15] and (D) DDCML [6]. The network architecture of all models, including the dimensionality of $\boldsymbol{z}$, is the same as that of the proposed IE-CBIR, as shown in Table II.

1) Data integrity (data reconstruction capability): How well the low-dimensional representation $\boldsymbol{z}$ preserves the information of the original 3D brain image is evaluated. This evaluation corresponds to the requirement (i). Root mean square error (RMSE) and structural similarity (SSIM) were obtained from the input image $\boldsymbol{x}$ and the reconstructed image $\hat{\boldsymbol{x}}$.

2) Data retrieval capability and its interpretability: How accurately the disease can be estimated from the obtained low-dimensional representation $\boldsymbol{z}$ is evaluated in terms of the ability to classify CN and AD classes. We then discuss whether the estimation process is easy to understand for users. This evaluation corresponds to the (ii) and (iii) of the requirements. Since the low-dimensional expressions obtained by the 3D-CAE, 3D-VAE and RaDOGAGA are obtained by unsupervised manner, a meaningful search cannot be expected if clustering is performed as is. Therefore, a logistic regression with L1 regularization model (LR) was separately trained for class estimation of each model. The DDCML and our IE-CBIR do not require such an additional learning model because they can estimate the disease class by the distance to each disease using the low-dimensional representation itself. For the DDCML, classes were estimated by K-means algorithm according to the original paper [6]. For our IE-CBIR, classes were estimated by nearest neighbor estimation to the class prototype $\boldsymbol{p}_i$. The macro F1 score was used for this evaluation.

IV. RESULTS

Table III summarizes the performance of each method, including our proposed IE-CBIR. The proposed IE-CBIR achieved an extremely high classification performance of 0.888 in macro F1 score and 91.5% in accuracy only by neighborhood search from the generated low-dimensional representation. Fig. 2 visualizes the distribution of each method's low-dimensional representation $\boldsymbol{z}$ with t-SNE [24].

V. DISCUSSION

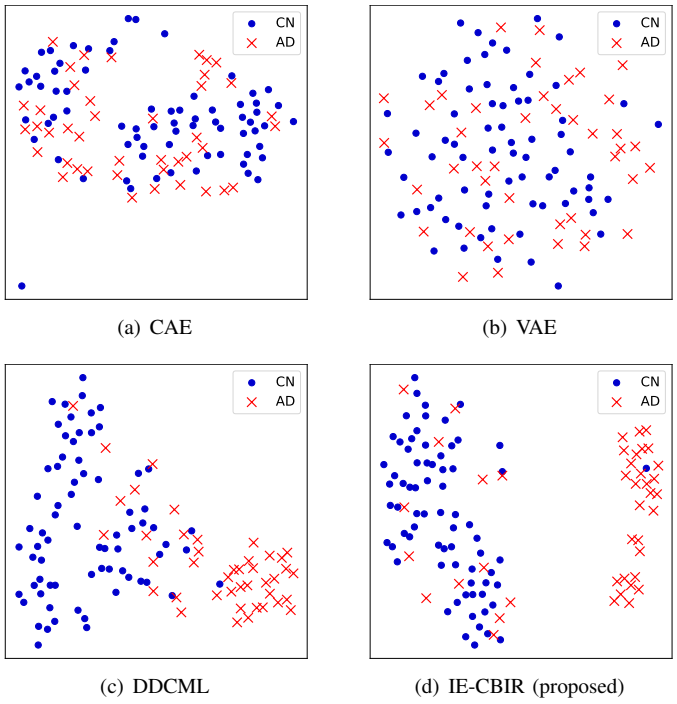
A. Data integrity: data reconstruction capability

For the reconstruction of brain structures or the preservation of original data, the best results were obtained for the representation produced by the 3D-CAE, where the model is trained with a reconstruction-only loss function. This is because the model is specialized for reconstruction. Due to the sophisticated learning algorithm of the RaDOGAGA, which provides a mapping of the input and output spaces to a space of guaranteed equidistance, the method RaDOGAGA and the proposed method IE-CBIR, which applies it, have

TABLE III
THE OVERALL PERFORMANCE OF EACH MODEL

		(1) Data reconstruction		(2) Data retrieval capability		
		RMSE	SSIM	Recall	Precision	macro F1
A)	3D-CAE	0.0649	0.922	0.690	0.658	0.777
B)	3D-VAE	0.0786	0.880	0.573	0.485	0.607
C)	RaDOGAGA [15]	0.0775	0.882	0.599	0.577	0.749
D)	DDCML [6]	0.0902	0.836	0.786	0.891	0.874
E)	IE-CBIR	0.0758	0.887	0.802	0.867	**0.888** ‡

‡ 91.5% in accuracy



(a) CAE (b) VAE

(c) DDCML (d) IE-CBIR (proposed)

Fig. 2. Low-dimentional representation visiualized by t-SNE

been able to achieve 3D-CAE-like performance in terms of data preservation.

Although the DDCML has the great advantage of being able to perform data exploration from the low-dimensional representation itself, as described below, the reconstruction capability were relatively low because metric learning is applied directly to the low-dimensional representation obtained by 3D-CAE, which inhibits data reconstruction to some extent.

While preserving the structural information of the original data is an important aspect of acquiring low-dimensional representations, it is supposed to note that in the present evaluation, the high degree of preservation is not directly related to the disease classification ability. This is because brain structures contain a large amount of information derived from individual differences (e.g., wrinkle regions of the brain) that do not contribute to disease classification. However, this metric will become significant in the future as the number of diseases covered increases and the model deals with unlearned diseases.

B. Data retrieval capability and its interpretability

The IE-CBIR and DDCML have excellent data retrieval capability (ii) and interpretability (iii), which allows data retrieval from the obtained low-dimensional representation z itself using only nearest neighbor search without the need for ad hoc discriminators. This desirable feature for CBIR can also be confirmed by the scatter plot of t-SNE in Figure 2. Among them, the IE-CBIR in particular improved the classification performance of AD and CN by 11.1 points, even when compared to CAE, which best preserves brain structures. The final AD and CN classification performance of 0.888 in macro F1 and 91.5% in accuracy is comparable or better than the accuracies reported in the literature under proper evaluation (i.e., without data leakage; 76%-91% accuracy [25]) using machine learning models such as CNNs. It should be emphasized that they do not take into account the acquisition of low-dimensional representations and their interpretability. This may be due to the advantages of the RaDOGAGA, such as the isometric embedding in the input and output space and the independence of each dimension, as well as our original learning method based on Baseline++, where the low-dimensional representation $\boldsymbol{z}$ and each disease prototype $\boldsymbol{p}_i$ are made close to each other by metric learning.

VI. CONCLUSION

In this paper, we proposed a new framework IE-CBIR for realizing CBIR of brain MRIs. To realize practical CBIR, it is necessary to obtain a low-dimensional representation that has (i) data integrity, (ii) high disease retrieval capability, and (iii) its interpretability. The experiments have shown that the IE-CBIR is able to realize these requirements and has demonstrated superior performance to existing techniques. In particular, the IE-CBIR's advantage of being able to retrieve similar cases only by neighbor search is very useful for the realization of the CBIR. This not only eliminates the need for ad hoc machine learning models, but is also a very important advantage in terms of interpretability of low-dimensional representations. We plan to validate the IE-CBIR on a larger scale and with more diverse data in the near future.

ACKNOWLEDGMENT

This research was supported in part by the Ministry of Education, Science, Sports and Culture of Japan (JSPS KAK-

ENHI), Grant-in-Aid for Scientific Research (C), 21K12656, 2021-2024.

The MRI data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] P. G. Morris, "Nuclear Magnetic Resonance Imaging in Medicine and Biology," 1986.

[2] R. H. Choplin, J. Boehme 2nd, and C. Maynard, "Picture archiving and communication systems: an overview." *Radiographics*, vol. 12, no. 1, pp. 127–129, 1992.

[3] S. Murala and Q. J. Wu, "Local Mesh Patterns Versus Local Binary patterns: Biomedical Image Indexing and Retrieval," *IEEE journal of biomedical and health informatics*, vol. 18, no. 3, pp. 929–938, 2013.

[4] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data," *Journal of digital imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.

[5] H. Arai, Y. Chayama, H. Iyatomi, and K. Oishi, "Significant Dimension Reduction of 3D Brain MRI using 3D Convolutional Autoencoders," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5162–5165.

[6] Y. Onga, S. Fujiyama, H. Arai, Y. Chayama, H. Iyatomi, and K. Oishi, "Efficient feature embedding of 3D brain MRI images for content-based image retrieval with deep metric learning," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 3764–3769.

[7] H. Oh Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep Metric Learning via Lifted Structured Feature Embedding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4004–4012.

[8] K. Nishimaki, K. Ikuta, Y. Onga, H. Iyatomi, and K. Oishi, "Loc-VAE: Learning Structurally Localized Representation from 3D Brain MR Images for Content-Based Image Retrieval," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2022, pp. 2433–2438.

[9] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," in *International conference on learning representations*, 2016.

[10] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection," in *International conference on learning representations*, 2018.

[11] N. Chen, A. Klushyn, R. Kurle, X. Jiang, J. Bayer, and P. Smagt, "Metrics for Deep Generative Models," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2018, pp. 1540–1550.

[12] H. Shao, A. Kumar, and P. Thomas Fletcher, "The Riemannian Geometry of Deep Generative Models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 315–323.

[13] C. Geng, J. Wang, L. Chen, W. Bao, C. Chu, and Z. Gao, "Uniform Interpolation Constrained Geodesic Learning on Data Manifold," *arXiv preprint arXiv:2002.04829*, 2020.

[14] J. McQueen, M. Meila, and D. Joncas, "Nearly Isometric Embedding by Relaxation," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[15] K. Kato, J. Zhou, T. Sasaki, and A. Nakagawa, "Rate-distortion optimization guided autoencoder for isometric embedding in Euclidean latent space," in *International Conference on Machine Learning*. PMLR, 2020, pp. 5166–5176.

[16] J. Hu, J. Lu, and Y.-P. Tan, "Deep Transfer Metric Learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 325–333.

[17] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang, "A Closer Look at Few-shot Classification," *arXiv preprint arXiv:1904.04232*, 2019.

[18] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-shot Learning," *Advances in neural information processing systems*, vol. 30, 2017.

[19] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Metric Learning for Large Scale Image Classification: Generalizing to New Classes at Near-Zero Cost," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12*. Springer, 2012, pp. 488–501.

[20] H. Qi, M. Brown, and D. G. Lowe, "Low-shot learning with imprinted weights," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5822–5830.

[21] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 Bias Correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[22] K. Nishimaki, K. Ikuta, S. Fujiyama, K. Oishi, and H. Iyatomi, "PCSS: Skull Stripping with Posture Correction from 3D Brain MRI for Diverse Imaging Environment," *IEEE Access*, 2023.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[24] L. Van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[25] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Medical image analysis*, vol. 63, p. 101694, 2020.