Received: 3 May 2021 | Revised: 14 September 2021 | Accepted: 14 September 2021

DOI: 10.1002/mp.15269

RESEARCH ARTICLE

MEDICAL PHYSICS

Unsupervised anomaly detection in MR images using multicontrast information

Byungjai Kim[1] | **Kinam Kwon**[2] | **Changheun Oh**[1] | **Hyunwook Park**[1]

[1] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea

[2] Samsung Electronics, Maetan-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do, Republic of Korea

Correspondence
HyunWook Park, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea.
Email: hwpark@kaist.ac.kr

Abstract

Purpose: Anomaly detection in magnetic resonance imaging (MRI) is to distinguish the relevant biomarkers of diseases from those of normal tissues. In this paper, an unsupervised algorithm is proposed for pixel-level anomaly detection in multicontrast MRI.

Methods: A deep neural network is developed, which uses only normal MR images as training data. The network has the two stages of feature generation and density estimation. For feature generation, relevant features are extracted from multicontrast MR images by performing contrast translation and dimension reduction. For density estimation, the distributions of the extracted features are estimated by using Gaussian mixture model (GMM). The two processes are trained to estimate normative distributions well presenting large normal datasets. In test phases, the proposed method can detect anomalies by measuring log-likelihood that a test sample belongs to the estimated normative distributions.

Results: The proposed method and its variants were applied to detect glioblastoma and ischemic stroke lesion. Comparison studies with six previous anomaly detection algorithms demonstrated that the proposed method achieved relevant improvements in quantitative and qualitative evaluations. Ablation studies by removing each module from the proposed framework validated the effectiveness of each proposed module.

Conclusion: The proposed deep learning framework is an effective tool to detect anomalies in multicontrast MRI. The unsupervised approaches would have great potentials in detecting various lesions where annotated lesion data collection is limited.

KEYWORDS
anomaly detection, magnetic resonance imaging, multicontrast images, singularity problem, unsupervised learning

1 | INTRODUCTION

Detection and segmentation of medical images are useful tools for finding biomarkers that distinguish normal tissue characteristics from abnormalities. Recently, deep learning approaches have been developed for detection tasks and their data-driven markers have provided convincing results.[1,2] Supervised learning has shown outstanding performances in specified disease detection of medical images.[3] However, the supervised learning requires a large amount of annotated images.

Unsupervised anomaly detection is gaining interest in medical imaging fields.[4] Instead of using annotated anomaly images for network training, the unsupervised approaches estimate the distributions of easily accessible normal images and distinguish exceptional data from large normal datasets.[5] The methods could be grouped into two representative categories: reconstruction-based[6–8] and clustering-based methods.[9] The reconstruction-based methods usually employ a deep autoencoder, assuming that the deep autoencoder trained by only normal images cannot properly

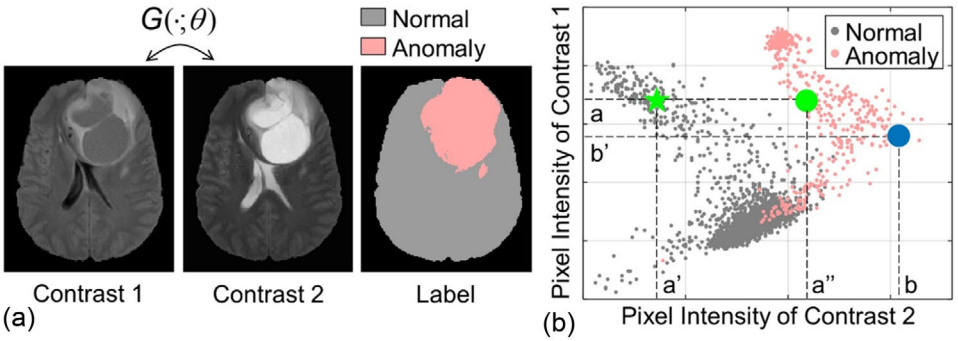

FIGURE 1 (a) Two magnetic resonance (MR) images with different contrasts and a label image of anomaly region. (b) Intensity distribution of the two images. The green and blue points in (b) depict the two cases of how anomaly can be detected in contrast-to-contrast translation

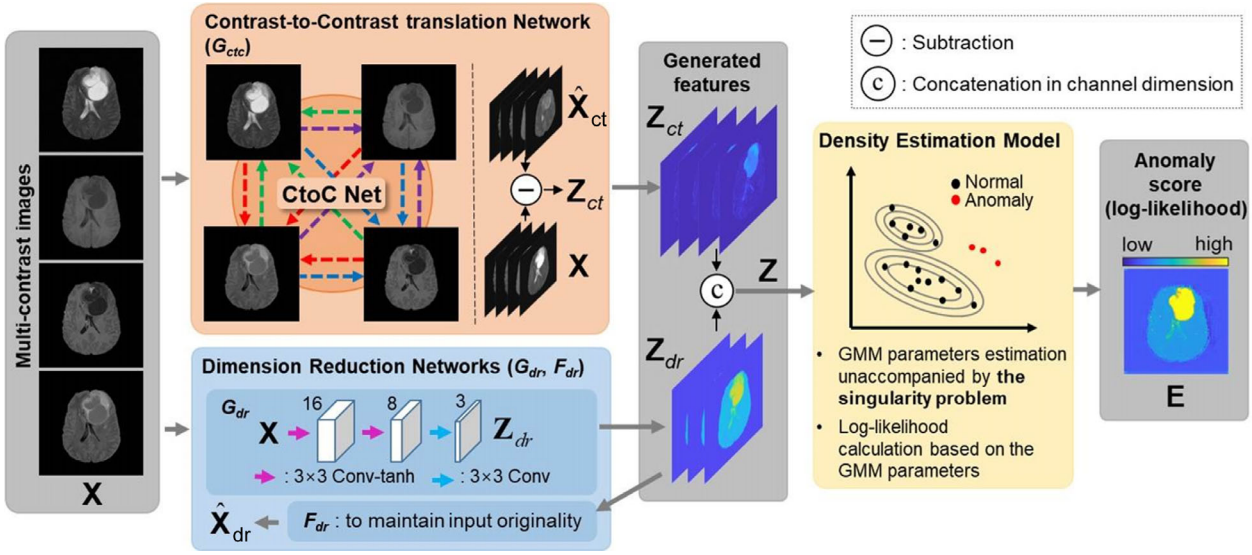

FIGURE 2 Overall diagram of the anomaly detection for magnetic resonance images (ADM) network in the test phase. Two different types of features are extracted from multicontrast magnetic resonance (MR) images, which are reconstruction based (the orange box) and clustering based (the blue box), respectively. The dotted arrows in the contrast-to-contrast translation (CtoC) network denote contrast translation between multicontrast images. In a density estimation model (the yellow box), normative distributions estimated in training phases are used to detect anomalies. The region considered anomalies has high intensity in **E**

reconstruct anomaly images. Meanwhile, the clustering-based methods use neural networks to extract latent features from high-dimensional images, and the clustering algorithms such as one-class support vector machine (OC-SVM)[10] classify the features. These methods have provided promising results for magnetic resonance imaging (MRI) and optical coherence tomography (OCT). However, the reconstruction-based methods have difficulty in selecting a degree of dimension reduction (DR).[6] The clustering-based methods have the separated pipelines of feature estimation and clustering. The approaches could easily lead to suboptimal performance because the first step of feature estimation is performed without information from the subsequent clustering models.[9]

A set of multicontrast images are usually obtained for accurate diagnosis in MRI. Figure 1 shows examples of the multicontrast MR images and their anomalous pixel intensity values. The pixel intensities of anomalies can be distinguished from normal intensities in the two-dimensional contrast space, while the normal and abnormal intensities are significantly overlapped in each dimension (Figure 1b). Actually, many clinical protocols for diagnosis consist of multicontrast images. The aspect is well reflected in several public datasets, such as The Cancer Imaging Archive (TCIA),[11] Ischemic Stroke Lesion Segmentation (ISLES),[12] and Brain Tumor Segmentation (BraTS).[13] Several anomaly detection approaches utilizing the inherent properties of multicontrast images have been developed for detecting epilepsy or multiple sclerosis lesions.[6,14–17]

In this work, a deep learning method is proposed for unsupervised abnormal tissue detection in multicontrast MR images. A deep learning framework is proposed,

which is called anomaly detection for MR images (ADM) network. The ADM network can effectively estimate the density of normal data in an unsupervised manner and can achieve significant improvement in anomaly detection, thanks to several contributions. First, two relevant features are extracted from multicontrast MRI. One is derived through learning of contrast-to-contrast translation (CtoC) between multicontrast MR images, and the other is the low-dimensional representation of multicontrast images. Experimental results demonstrate that the contrast translation is an effective way to capture the accurate distribution of normal data, which is distinct from that of anomaly data. Second, the two different types of features are collaboratively used in a following density estimation (DE) model for Gaussian mixture model (GMM). Also, the two steps of generating features and estimating density are jointly optimized so that the generated features become suitable for the anomaly detection tasks. Third, the proposed DE model effectively prevents the singularity problem that frequently occurs in estimating GMM parameters and critically interrupts the joint learning processes. The proposed method was compared with previous anomaly detection algorithms, in terms of how accurately anomaly pixels can be detected in multicontrast MR images. Qualitative and quantitative analyses demonstrate the superiority and the feasibility of the proposed method in pixel-wise anomaly detection.

1.1 | Related works

A typical approach for unsupervised anomaly detection is to use reconstruction errors, assuming that anomalies cannot be effectively reconstructed by models representing normal data distributions. Thus, the degree of errors is the criterion for determining whether the data are normal or not. Most of the methods assume that there are great losses in the DR of anomalies when the process is optimized only for normal data. In this respect, several works have employed principal component analysis (PCA) which is a typical unsupervised DR algorithm.[18,19] Recently, a deep autoencoder has shown remarkable performance.[20–22] The deep autoencoder compresses input data into low-dimensional latent variables in the encoder part. In training, the reduced latent variables represent the common features of training datasets, from which a decoder accurately reconstructs the original input data. So, the deep autoencoder trained only with normal datasets cannot properly reconstruct anomalies.[6–8]

However, the performances of the reconstruction-based works are limited by relying only on assumption of "data compressibility" that anomalies cannot be reconstructed with low-dimensional representations only for normal data. We can expect high reconstruction error when anomalies are constructed with low-dimensional representation of normal data. However, plenty of anomalies would still have low level of reconstruction error especially when the dimension increases. It is usually difficult to estimate the intrinsic dimensions of complex normal data.[23] That is why determining the degree and the manner of "data compression" to discriminate anomalies would be more challenging.

Another common method for anomaly detection is to use a clustering algorithm to estimate the probability distribution of normal data. Conventional clustering and classification algorithms including kernel DE, SVM, and GMM have been used in anomaly detection.[24–27] Several researches have adopted deep neural networks as a preprocessor for the conventional clustering algorithms, which transforms input data into low-dimensional latent representation.[9,14,28–30] The clustering algorithms employing latent representation from neural networks have shown higher performance especially in high-dimensional image data. Nevertheless, the performances of these methods were limited by the fact that the latent features were not suitable for following clustering models.

To overcome the issues, a research group proposed a clustering objective function for a neural network,[31] which was based on support vector data description.[32] The latent variables induced by the objective function were directly used for anomaly detection without clustering models. In another way, some groups have developed joint learning of DR and clustering models.[33–35] In the joint learning frameworks, neural networks estimated latent representations suitable for clustering models (K-means or GMM). Especially, the joint learning framework with GMM showed outstanding performance in the anomaly detection field.[36] However, estimating the parameters of GMM in the joint learning model may introduce the singularity problem. The singularity problem is that a variance of the Gaussian model becomes zero in a specific dimension so that its covariance matrix becomes singular (or noninvertible). The singular matrix critically interrupts the learning processes of estimating GMM parameters.[36,37]

Recently, powerful deep neural networks have been developed for general representation learning.[38,39] By using the neural networks, several anomaly detection techniques have shown promising results. First, a variational autoencoder (VAE) was one of popular neural networks adopted for unsupervised anomaly detection,[6,16,40] which used probabilistic latent spaces to describe normative distribution effectively.[41] Second, a generative adversarial network (GAN)[39] has been recently used for anomaly detection techniques.[42–44] Some of these networks combined an autoencoder with GAN for relevant latent representations.[7] In the deep neural network framework, the latent space of normal images was directly guided by GAN.[45] The VAE- and GAN-based methods improved anomaly detection performance. However, since their architectures are based

on autoencoders, the difficulty of choosing the reduction degree still remains.

2 | MATERIALS AND METHODS

A new anomaly detection algorithm is proposed for multicontrast MR images. The proposed framework, which is called the ADM network, has three modules of deep neural networks (Figure 2). The ADM method adopts two types of feature generations that are reconstruction based and clustering based, which complement each other. And then, a DE module is followed. The joint learning strategy of the three modules is adopted to improve feature quality and detection performance. The motivation of each module is explained in detail as follows.

2.1 | Motivations of the proposed method

A feature for the proposed anomaly detection (Z_{ct}) is derived from the contrast relationship between multicontrast images. Learning of CtoC translation is a way of generating a specific contrast image from the other contrast images. A network trained with only normal data can effectively capture the relationship of multicontrast normal images. In the test phase, a specific contrast image is generated from the other contrast images by the trained network. Then, the reconstruction error between the generated image and the true image is the basis for detecting anomalies.

Figure 1b shows an example of how the contrast translation works for anomaly detection. The green points of the star and circle in Figure 1b depict the case when normal and anomaly tissues have similar intensities (*a*) in contrast 1 but significantly different (*a*' and *a*", respectively) in contrast 2. A neural network trained with only normal data for contrast translation from 1 to 2 poorly reconstructs the intensity of contrast 2 of the green circle, and its reconstruction error would be the difference between the two green points of the star and circle in the dimension of contrast 2 (i.e., $|a' - a''|$). The blue point in Figure 1b describes another case where anomaly data have a completely different intensity (*b*) from normal data distribution in contrast 2. A neural network for contrast translation from 2 to 1 would have significant reconstruction errors on the unseen input data of blue circle. The reconstruction errors are derived using the relationship between multicontrast images. That is, the proposed method is not based on the data compressibility of anomaly, unlike previous reconstruction-based methods using autoencoders. The effectiveness of the proposed reconstruction-based method is validated in our experiments.

Another type of feature (Z_{dr}) is the low-dimensional representation of multicontrast images. Normal intensities are distributed differently from anomaly intensities in multicontrast images (Figure 1b). To extract relevant features from the multicontrast images, a neural network is adopted for DR. The trained network with only normal multicontrast images would represent a low-dimensional latent space for normal data so that anomalies could have different distributions in the latent space. Note that the joint learning scheme is used to estimate efficient features for clustering.[28,30] That is, the neural network for DR and the following model for clustering are jointly optimized in the proposed method. The effectiveness of the joint learning is demonstrated by experiments.

A DE model is adopted to properly describe normal feature distribution as GMM parameters. A log-likelihood function with the estimated GMM parameters is calculated for scoring the anomaly degree. However, estimating GMM parameters is vulnerable to the singularity problem when a covariance matrix becomes singular, and learning processes are interrupted due to the singularity problem. Even though several approaches have been developed to prevent this singularity problem,[36,37] these approaches would be effective only if the feature distributions do not change during DE. Therefore, we propose a DE model to prevent the singularity problem in the joint learning, by ensuring a positive definite covariance matrix.

2.2 | Architectures of ADM network

2.2.1 | Contrast-to-contrast translation network

The CtoC network (G_{ct}) generates a specific contrast image from the collaborative knowledge of other contrast images (Figure 3). Based on the architecture showing impressive results in style transfer,[46] the CtoC network is designed to perform contrast translation with the prior information of a target contrast. First, we assume that there is a set (X) of four contrast input images (x^a, x^b, x^c, and x^d) as follows:

$$X = \{x^a, x^b, x^c, x^d\} \tag{1}$$

Assuming that an image of *b* contrast is generated by G_{ct}, it can be described as follows:

$$\mathbf{x}^b = G_{ct}(\mathbf{X}^b, \theta_{ct}) \tag{2}$$

$$\mathbf{X}^b = \{\mathbf{x}^a, \mathbf{m}^b, \mathbf{x}^c, \mathbf{x}^d\}\ , \tag{3}$$

where θ_{ct} is the parameters of the network G_{ct}, and $\mathbf{x}^b$ and $\mathbf{X}^b$ are the output and the input of the CtoC network, respectively. Here, m^b is an image whose whole pixels are filled with the median value of $\mathbf{x}^b$. The median value as prior information serves to consider the

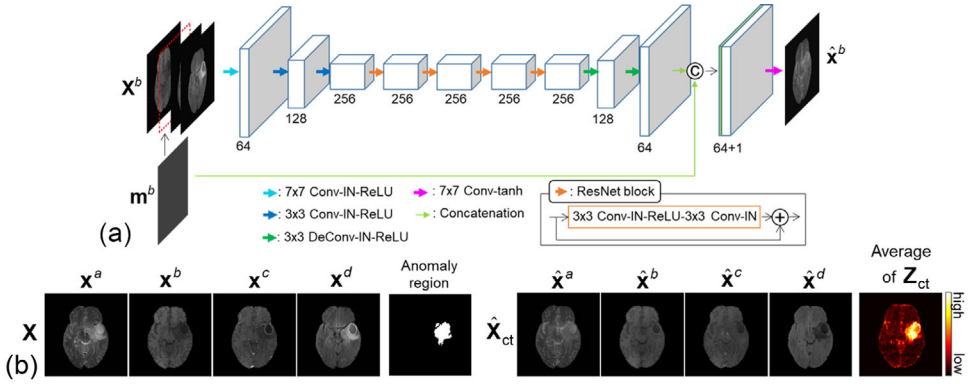

FIGURE 3 (a) The architecture of the contrast-to-contrast translation (CtoC) network. The $\mathbf{m}^b$ is fed into the first and the last layers in the CtoC network to provide an intensity bias in synthesizing a target contrast image. (b) Examples of the input **X** and output $\mathbf{X}_{ct}$ of the CtoC network and their anomaly region. For visualization, the reconstruction errors of multicontrast images are averaged

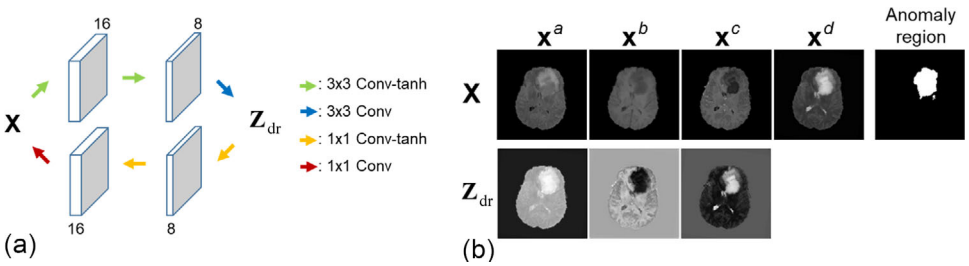

FIGURE 4 (a) Architecture of the dimension reduction (DR) network. (b) Examples of the input **X** and output Z_{dr} of the DR network and their anomaly region. The DR network reduces the contrast dimension of the input **X** and produces features Z_{dr} containing new contrast information which are used in the density estimation network

intensity diversity of open datasets. A deep neural network is trained for all cases of contrast translation. The generated image has the contrast corresponding to the channel location of **m**. For example, when m is the image of a channel *b*, the network generates a *b* contrast image. Reconstruction errors between the generated image and the corresponding true image are used as features ($\mathbf{Z}_{ct}$) for anomaly detection in a test phase. Examples of the generated images, **X**, and their Z_{ct} are shown in Figure 3b and represented as follows:

$$\begin{aligned}\mathbf{X}_{ct} &= \{\mathbf{x}^a, \mathbf{x}^b, \mathbf{x}^c, \mathbf{x}^d\} \\ &= \left\{G_{ct}(\mathbf{X}^a;\theta_{ct}), \ldots, G_{ct}(\mathbf{X}^d;\theta_{ct})\right\}\end{aligned} \tag{4}$$

$$\mathbf{Z}_{ct} = \text{abs}(\mathbf{X} - \mathbf{X}_{ct}) \tag{5}$$

where abs(X) is the absolute value of the elements of X.

2.2.2 | Dimension reduction network

The DR network (G_{dr}) generates features, $\mathbf{Z}_{dr}$, that correspond to the low-dimensional representations of input images. DR is performed along the channel dimension, whereas the spatial dimension of $\mathbf{Z}_{dr}$ is the same as that of input images. That is, each pixel intensity in $\mathbf{Z}_{dr}$ is a feature used in determining whether the pixel is anomaly or not. A reconstruction network (F_{dr}) is adopted to prevent $\mathbf{Z}_{dr}$ from losing key information of input **X**. The two networks are described in Figure 4 and formulated as follows:

$$\mathbf{Z}_{dr} = G_{dr}(\mathbf{X};\theta_{dr,1}) \tag{6}$$

$$\mathbf{X}_{dr} = F_{dr}(\mathbf{Z}_{dr};\theta_{dr,2}) \tag{7}$$

where $\theta_{dr,1}$ and $\theta_{dr,2}$ are the network parameters of G_{dr} and F_{dr}, respectively, and $\mathbf{X}_{dr}$ is the reconstructed images from Z_{dr}.

The convolutional layers in G_{dr} estimate the pixel-level features $\mathbf{Z}_{dr}$ by considering spatial information in a receptive field. In F_{dr}, 1×1 convolutional layers are used to offer a strong constraint that $\mathbf{Z}_{dr}$ maintains the spatial specificity of original input **X**. The four multicontrast images are converted to three feature images. In training, G_{dr} and F_{dr} are jointly optimized with a following DE model.

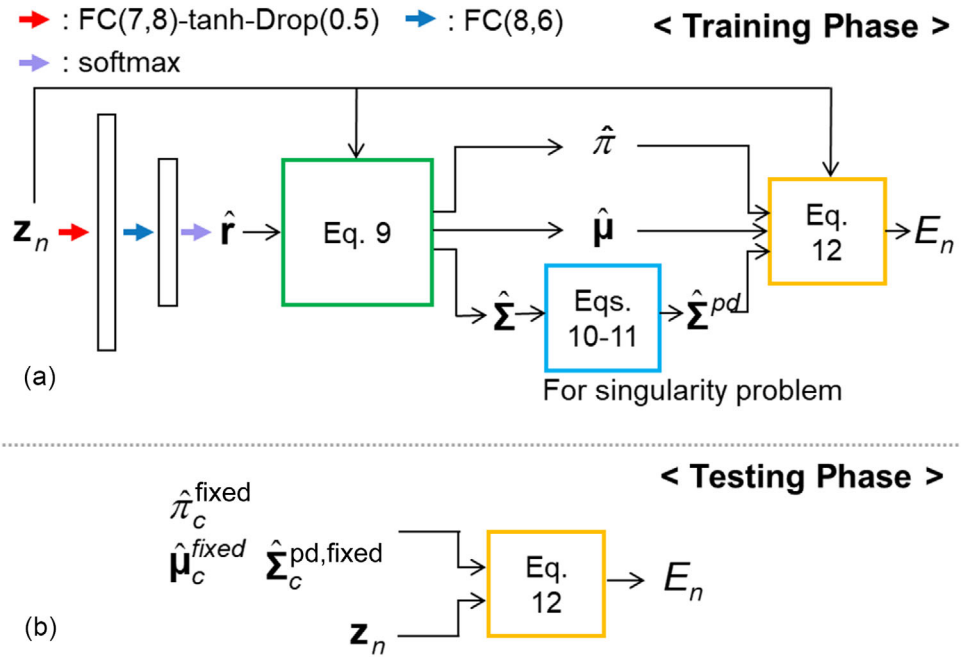

FIGURE 5 The architecture of the density estimation (DE) model. In the training phase (a), the DE model is trained to estimate the Gaussian mixture model (GMM) parameters of given features $\mathbf{z}_n$ by reducing the log-likelihood E_n. In testing phase (b), the DE calculates the E_n of $\mathbf{z}_n$ with $\hat{\pi}_c^{\text{fixed}}$, $\hat{\mu}_c^{\text{fixed}}$, and $\hat{\Sigma}_c^{\text{pd,fixed}}$, which are the GMM parameters estimated from the training data at the end of the training phase

2.2.3 | Density estimation model

The DE model estimates GMM parameters well representing the normal features of training data and then calculates anomaly scores for test features by using the estimated GMM parameters (Figure 5). The goodness-of-fit of the estimated GMM parameters is measured by a log-likelihood function. Log-likelihood is a loss function for training and also an anomaly score function for testing. Detailed configurations are as follows.

A vector along the channel dimension of the concatenated matrix of $\mathbf{Z}_{\text{ct}}$ and $\mathbf{Z}_{\text{dr}}$ is a feature for pixel-wise anomaly detection, which is $\mathbf{z}_n (1 \le n \le N)$. Note that N is the total number of $\mathbf{z}_n$ in spatial and batch dimensions. The proposed DE model is based on previous works,[36,47] where a neural network estimates the probability $\hat{\mathbf{r}}_n$ that the input feature $\mathbf{z}_n$ belongs to each Gaussian model. The number of Gaussian models is a hyperparameter C and $\hat{\mathbf{r}}_n$ has C elements of $\hat{r}_{n,c}$ $(1 \le c \le C)$. The GMM parameters of the fraction $(\hat{\pi}_c)$, mean $(\hat{\boldsymbol{\mu}}_c)$, and covariance $(\hat{\boldsymbol{\Sigma}}_c)$ are calculated as follows:

$$\hat{\mathbf{r}}_n = G_{\text{de}}(\mathbf{z}_n; \theta_{\text{de}}) \tag{8}$$

$$\hat{\pi}_c = \sum_n \frac{\hat{r}_{n,c}}{N}, \quad \hat{\boldsymbol{\mu}}_c = \frac{\sum_n \hat{r}_{n,c} \cdot \mathbf{z}_n}{\sum_n \hat{r}_{n,c}}$$

$$\hat{\boldsymbol{\Sigma}}_c = \frac{\sum_n \hat{r}_{n,c} (\mathbf{z}_n - \hat{\boldsymbol{\mu}}_c)(\mathbf{z}_n - \hat{\boldsymbol{\mu}}_c)^{\mathsf{T}}}{\sum_n \hat{r}_{n,c}} \tag{9}$$

where G_{de} and θ_{de} are the neural network and its parameters, respectively, and the subscript of c is an index of Gaussian models. Six Gaussian models are assumed in our experiments.

The proposed DE model effectively prevents the singularity problem by introducing a lower bound (ε) for eigenvalues of a covariance matrix. The procedure ensures a covariance matrix to be positive definite as follows:

$$\hat{\boldsymbol{\Lambda}}_c^{\text{p}} = \text{ReLU}\left(\hat{\boldsymbol{\Lambda}}_c - \varepsilon\right) + \varepsilon \tag{10}$$

$$\hat{\boldsymbol{\Sigma}}_c^{\text{pd}} = \hat{\mathbf{Q}}_c \cdot \hat{\boldsymbol{\Lambda}}_c^{\text{p}} \cdot \hat{\mathbf{Q}}_c^{\mathsf{T}} \tag{11}$$

where $\mathbb{Q}_c$ and $\mathbb{\Lambda}_c$ are the eigenvectors and the eigenvalues of $\mathbb{\Sigma}_c$, respectively, the superscript T is the matrix transpose, and $\hat{\boldsymbol{\Sigma}}_c^{\text{pd}}$ is a final covariance matrix. The value of ε was 10^{-6} in our experiments. Finally, the log-likelihood of $\mathbf{z}_n$ is calculated as follows.:

$$E_n = -\log\left(\sum_c \hat{\pi}_c \frac{\exp(-0.5(\mathbf{z}_n - \hat{\boldsymbol{\mu}}_c)^{\mathsf{T}} (\hat{\boldsymbol{\Sigma}}_c^{\text{pd}})^{-1} (\mathbf{z}_n - \hat{\boldsymbol{\mu}}_c))}{\sqrt{\det(2\pi \hat{\boldsymbol{\Sigma}}_c^{\text{pd}})}}\right) \tag{12}$$

where det(·) denotes the determinant of a matrix. In the training phase, the DE model estimates GMM parameters for each batch data and minimizes their

log-likelihood. That is, the DE model is optimized to be a density estimator. At the end of the training, final GMM parameters for normative distributions are estimated from the large amount of normal data. In the test phase, the estimated GMM parameters are used in calculating log-likelihoods for test data, which are anomaly scores.

Details of the parameters of the ADM network are explained in Section I of the Supporting Information.

2.3 | Objective function and joint learning strategy

The proposed framework is optimized in two steps by using the two objective functions as follows:

$$\min_{\theta_{ct}} \|\mathbf{X} - \mathbf{X}_{ct}\|_2^2 \tag{13}$$

$$\min_{\theta_{dr,1},\theta_{dr,2},\theta_{de}} \|\mathbf{X} - \mathbf{X}_{dr}\|_2^2 + \frac{\lambda}{N}\sum\nolimits_n E_n \tag{14}$$

where $\|\cdot\|_2$ is the L2 norm of a vector. First, the G_{ct} is optimized to minimize the objective function in Equation (13), then the feature $\mathbf{Z}_{ct}$ is generated from G_{ct}. With the estimated $\mathbf{Z}_{ct}$, the G_{dr} and the G_{de} are jointly optimized to minimize the objective function in Equation (14). The DE network finds the parameters of Gaussian mixtures, properly presenting the feature distributions by reducing E_n. As E_n decreases, the features of $\mathbf{Z}_{dr}$ are densely distributed and have cluster-friendly properties. In the test phase, the features of anomalies are effectively discriminated by their E_n values. The number of training epochs were 50 for BraTS dataset and 200 for ISLES dataset, respectively. The number of epochs was empirically determined based on the saturation of the loss values of Equations (13) and (14) (Figure S1). The λ is a hyperparameter for the ADM network. In practice, λ was set as 0.0005 and 0.005 for BraTS and ISLES datasets, respectively, with which desirable results of the area under curve (AUC) were obtained using the validation data.

2.4 | Datasets

Two open datasets of BraTS 2019[13] and ISLES 2015[12] were used in the experiments. Both datasets are three-dimensional whole brain MR images with pixel-wise annotations of anomalies. By using the annotations, the datasets could be divided into normal and abnormal data. Normal data are only used for training of unsupervised anomaly detection.

BraTS is a dataset for glioblastoma and lower grade glioma, and has the four contrast images of T2-FLAIR, T2-weighted, T1-weighted, and T1-weighted with contrast agent. First, 241 patients in BraTS were randomly divided into two parts of 221 and 20 patients. From the 221 patients, 7790 normal image sets and 14 185 anomaly image sets were used as training and validation data, respectively. The whole image sets from the 20 patients (2100 sets) were used as test data.

ISLES is a dataset for ischemic stroke lesions and has four contrast images such as diffusion weighted, T1-weighted, T2-FLAIR, and T2-weighted images. From 18 patients, 828 normal image sets and 900 anomaly image sets were selected and used as training and validation data, respectively. In additions, 576 image sets from other six patients were used as test data. Anomaly detection in ISLES was more challenging due to its small amount of data. In additional experiments, the training data were augmented with the images of the three common MR contrasts in BraTS and ISLES, which were FLAIR, T1-weighted, and T2-weighted images.

Images from the two datasets went through the three preprocessing steps of skull stripping, co-registration between multicontrast images, and interpolation with 1 mm^3 resolution. The intensity distribution of whole brain images was normalized to be a normal distribution with zero mean and unit variance. This normalization was performed for each contrast image. All images were resized to 128×128 by using bilinear interpolation.

Data augmentation methods with flip, translation, crop, and intensity scaling were performed for training. Similarly, to increase the total number of test datasets and the validation data of ISLES, we applied the augmentation methods so that 6300 and 6912 sets are used as test data for BraTS and ISLES, respectively, and 6300 sets are used as validation data for ISLES.

2.5 | Comparison models

The proposed method was compared with several anomaly detection algorithms, which have been developed for general applications as well as medical images. For the three groups described in the related works, the representative works of the competing models were selected. In particular, an anomaly detection algorithm with GMM was selected and compared with the proposed method in terms of the singularity problem of GMM. The selected works are as follows: (1) Gaussian mixture variational autoencoder (GMVAE) is a reconstruction-based technique to detect anomalies in MRI, which is based on a VAE and a maximum-a-posterior (MAP) reconstruction.[41] (2) Deep support vector data description (DSVDD)[31] is a clustering-based anomaly detection algorithm introducing a clustering objective function. (3) fanoGAN[45] uses a new neural network with GAN for anomaly detection. Adversarial learning with GAN enables to secure the relevant latent representations of normal data. (4) Deep autoencoding GMM (DAG)[36] is an autoencoder adopting GMM for anomaly detection. (5) A modified DAG (DAG_c) including

convolutional layers is more appropriate for pixel-level anomaly detection. The configuration of each method is described in Section III of the Supporting Information.

In addition, variants of the proposed framework were employed for the ablation study to demonstrate the importance of each component as follows. (6) ADM_{ct} and (7) ADM_{dr} are the proposed frameworks without the DR and the CtoC networks, respectively. (8) ADM_{woj} is the proposed framework without the joint optimization of feature generation and DE. That is, the two loss terms in Equation (14) were used separately to optimize the corresponding network.

2.6 | Metrics and hyperparameter selection

Several metrics were adopted to quantify the performance of anomaly detection. The AUC of the receiver operating characteristic (ROC) curve was measured to compute the general performance of anomaly detection. Precision, recall, and F_1 scores were reported to evaluate how accurately true positives were detected. For each method, a threshold value to calculate those metrics was selected as the value that reports the best F_1 score on validation data. Note that anomaly class was assumed as positive. Black background regions in brain images were not considered as normal class and excluded in calculating the metrics. The pixel-level annotations provided by each dataset were used as the ground truth.

Hyperparameters for the proposed method and other competing methods were carefully adjusted to report their best performance in the comparison study. We adopted holdout cross-validation by dividing clinical datasets into normal data for training and abnormal data for validation. The AUC values for the validation data were used to select hyperparameters. Detailed explanations for each method are given in Section III of the Supporting Information.

2.7 | Lower bound for the singularity problem

DE using GMM has been used in various clustering tasks including anomaly detection.[34,37,47] The singularity problem is a known issue with GMM and happens more easily in the joint learning of feature generation and DE. We performed an experiment to investigate how the singularity problem occurred during the joint learning and to demonstrate the effectiveness of the proposed lower bound method.

In this study, three types of the DE models were compared, which were the baseline model without any constraints, the proposed model using the lower bound for eigenvalues, and the comparison model with the loss penalizing small diagonal entries in a covariance matrix.[36] For the comparison model, the term of $L_p = \lambda_p \cdot \sum_i^d (1/(\hat{\Sigma}_c)_{ii})$ is added to the loss of Equation (14) with λ_p of 10^{-5}, where d is the number of diagonal entries and $(\hat{\Sigma}_c)_{ii}$ indicates ii-element of $\hat{\Sigma}_c$. This method was the generalized version of adding a small value to the diagonal entries. The joint learning model of ADM_{dr} was employed to visualize its three-dimensional feature space.

3 | RESULTS

3.1 | Quantitative analyses

Quantitative analysis results are denoted by AUC, precision, recall, and F_1 score for BraTS and ISLES in Tables 1–4. Each method was trained three times and then the mean and the standard deviation of each metric were measured. Examples of anomaly score maps for each method are depicted as qualitative results as shown in Figure 6. The results for validation data usually show lower performance than those for test data, since validation data consisting of only anomaly images have a higher anomaly ratio than test data. For subjective evaluation, the segmented anomalous regions for images in Figure 6 are shown in Figure S2.

Anomaly detection in ISLES is more challenging than in BraTS, because of the small amount of training data in ISLES. The quantitative values in ISLES are usually lower than those in BraTS. The precision values for ISLES dataset are very low, since insufficient normal data cannot represent the general properties of normal tissues. Nevertheless, the proposed ADM method still shows higher performance than the competing models.

3.1.1 | Comparison study

Tables 1 and 2 show the quantitative results for comparison study with the mean and the standard deviation of each metric. The F_1 score of the ADM is mostly higher than those of the competing models (BraTS validation/test = 0.575/0.671, ISLES validation/test = 0.396/0.276). In addition, the ADM also achieves the highest AUC value for BraTS validation/test data (0.883/0.934) and ISLES test data (0.870), while DSVDD provides a higher AUC value for ISLES validation data and the ADM is ranked second. For the precision values, the proposed method is usually ranked high, except for in the ISLES validation data, as follows: BraTS validation/test = first/second and ISLES validation/test = fifth/first. For recall values, the proposed method is ranked as validation/test = fourth/third for BraTS and validation/test = second/first for ISLES.

The AUC values show general performance in anomaly detection. For the AUC values of test data,

TABLE 1 Quantitative results in comparison studies for Brain Tumor Segmentation (BraTS) datasets

		GMVAE	DSVDD	fanoGAN	DAG	DAG_c	ADM
Validation	AUC	0.759 ± 0.020	0.826 ± 0.006	0.764 ± 0.000	0.836 ± 0.001	0.843 ± 0.027	**0.883 ± 0.002**
	Precision	0.255 ± 0.051	0.425 ± 0.023	0.426 ± 0.014	0.430 ± 0.029	0.449 ± 0.062	**0.578 ± 0.042**
	Recall	0.519 ± 0.101	0.597 ± 0.036	0.392 ± 0.011	0.587 ± 0.049	**0.618 ± 0.040**	0.574 ± 0.021
	F_1	0.334 ± 0.029	0.496 ± 0.003	0.408 ± 0.000	0.495 ± 0.002	0.519 ± 0.056	**0.575 ± 0.010**
Test	AUC	0.776 ± 0.034	0.870 ± 0.000	0.832 ± 0.000	0.904 ± 0.000	0.899 ± 0.003	**0.934 ± 0.011**
	Precision	0.332 ± 0.023	**0.725 ± 0.026**	0.642 ± 0.007	0.585 ± 0.021	0.583 ± 0.027	0.702 ± 0.012
	Recall	0.415 ± 0.084	0.569 ± 0.022	0.430 ± 0.005	**0.677 ± 0.025**	**0.677 ± 0.033**	0.643 ± 0.045
	F_1	0.366 ± 0.040	0.635 ± 0.005	0.517 ± 0.003	0.627 ± 0.003	0.625 ± 0.000	**0.671 ± 0.028**

Abbreviations: ADM, anomaly detection for magnetic resonance images; AUC, area under curve; DAG, deep autoencoding Gaussian mixture model; DAG_c, modified DAG including convolutional layers; DSVDD, deep support vector data description; fanoGAN, Fast unsupervised anomaly detection with generative adversarial networks; GMVAE, Gaussian mixture variational autoencoder.
The highest score for each metric (each row) are shown in bold values.

TABLE 2 Quantitative results in comparison studies for Ischemic Stroke Lesion Segmentation (ISLES) datasets

		GMVAE	DSVDD	fanoGAN	DAG	DAG_c	ADM
Validation	AUC	0.694 ± 0.016	**0.844 ± 0.003**	0.723 ± 0.001	0.743 ± 0.002	0.725 ± 0.004	0.814 ± 0.014
	Precision	0.240 ± 0.032	0.382 ± 0.006	**0.554 ± 0.001**	0.523 ± 0.006	0.419 ± 0.082	0.414 ± 0.013
	Recall	0.214 ± 0.007	**0.470 ± 0.007**	0.223 ± 0.000	0.248 ± 0.003	0.276 ± 0.017	0.381 ± 0.020
	F_1	0.226 ± 0.015	**0.421 ± 0.006**	0.318 ± 0.000	0.336 ± 0.003	0.329 ± 0.017	0.396 ± 0.016
Test	AUC	0.595 ± 0.025	0.609 ± 0.008	0.660 ± 0.000	0.650 ± 0.000	0.643 ± 0.003	**0.870 ± 0.010**
	Precision	0.047 ± 0.006	0.049 ± 0.002	0.176 ± 0.001	0.087 ± 0.002	0.059 ± 0.002	**0.203 ± 0.027**
	Recall	0.451 ± 0.075	0.408 ± 0.026	0.222 ± 0.001	0.279 ± 0.017	0.398 ± 0.021	**0.433 ± 0.005**
	F_1	0.085 ± 0.009	0.088 ± 0.004	0.196 ± 0.001	0.133 ± 0.001	0.103 ± 0.002	**0.276 ± 0.025**

Abbreviations: ADM, anomaly detection for magnetic resonance images; AUC, area under curve; DAG, deep autoencoding Gaussian mixture model; DAG_c, modified DAG including convolutional layers; DSVDD, deep support vector data description; fanoGAN, Fast unsupervised anomaly detection with generative adversarial networks; GMVAE, Gaussian mixture variational autoencoder.
The highest score for each metric (each row) are shown in bold values.

a paired *t*-test analysis was performed between ADM and competing methods. In the statistical analyses, *p* values were below 0.05 for every case. Consequently, the ADM method usually achieved higher performance in anomaly detection. These quantitative results are well reflected in anomaly score maps. The proposed method shows low anomaly scores in most normal regions (Figure 6) and it leads to low false positives and high performance. In contrast, some competing models provide high anomaly scores for normal tissues, especially around cerebrospinal fluid (CSF).

The GMVAE has lower precision than the ADM and it leads to low performance in AUC and F_1. The DSVDD usually shows high recall and low precision. For ISLES validation data, the DSVDD usually has higher performance than the ADM. The fanoGAN provides relatively high anomaly scores in normal regions, but sufficient differences between normal and anomaly regions can be observed in anomaly score maps (Figure 6). The DAG_c that is the convolutional variant of DAG usually outperforms DAG in most evaluations and is second only to the ADM for BraTS.

3.1.2 | Ablation study

Tables 3 and 4 show the quantitative results for ablation study with the mean of each metric. For BraTS, the AUC values of ADM_{ct} (validation/test = 0.861/0.930) are comparable to those of ADM_{dr} (validation/test = 0.865/0.922). However, for the small dataset of ISLES, the AUC values of ADM_{ct} (validation/test = 0.768/0.839) are higher than those of ADM_{dr} (validation/test = 0.734/0.662). The ADM_{woj} that is a combination of ADM_{ct} and ADM_{dr} outperforms each variant for AUC values. In addition, the joint learning in the proposed method provides further improvement, given that ADM usually shows better performance in the quantitative results than ADM_{woj}.

3.1.3 | Analyses with anomaly size

We analyzed the performance of the ADM in terms of the size of anomaly region as shown in Figure 7. The anomaly size was measured as the number of

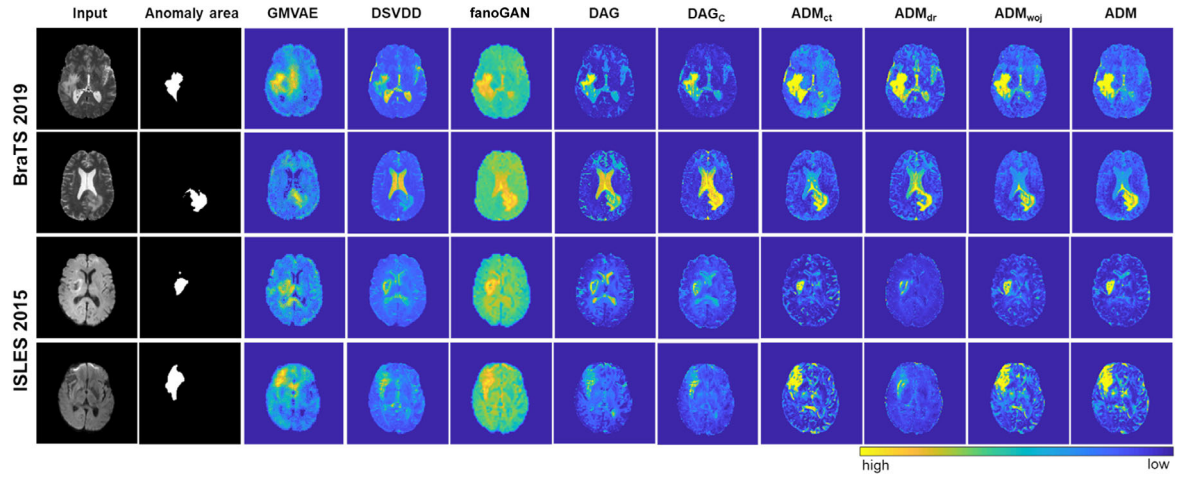

FIGURE 6 Resultant anomaly score maps (*E*) for test and validation data in Brain Tumor Segmentation (BraTS) 2019 and Ischemic Stroke Lesion Segmentation (ISLES) 2015. One of multicontrast images is shown as an input image. High intensity values mean the high probability of anomaly

TABLE 3 Quantitative results in ablation studies for Brain Tumor Segmentation (BraTS)

		ADM$_{ct}$	ADM$_{dr}$	ADM$_{woj}$	ADM
Validation	AUC	0.861	0.865	0.878	**0.883**
	Precision	0.527	**0.623**	0.576	0.578
	Recall	0.522	0.550	0.569	**0.574**
	F_1	0.521	**0.584**	0.572	0.575
Test	AUC	0.930	0.922	0.933	**0.934**
	Precision	0.724	**0.797**	0.725	0.702
	Recall	0.607	0.643	0.638	**0.653**
	F_1	0.658	**0.712**	0.678	0.671

Abbreviations: ADM, anomaly detection for magnetic resonance images; ADM$_{ct}$, proposed frameworks without the dimension reduction network; ADM$_{dr}$, proposed frameworks without the contrast-to-contrast translation network; ADM$_{woj}$, proposed framework without the joint optimization of feature generation and density estimation; AUC, area under curve.
The highest score for each metric (each row) are shown in bold values.

pixels in the annotated lesion and an AUC value for an image is calculated. To provide quantitative results, the whole range of anomaly size (0∼1600 pixels) was divided into eight ranges with the step size of 200 pixels. The mean and the standard deviation of AUC values for each range were calculated. In addition, the number of samples for each range was counted. The anomalies of ISLES data are mostly smaller than those of BraTS data.

3.2 | Lower bound for the singularity problem

Figure 8a,b describes the features **Z** of the ADM$_{dr}$ without any constraints (the baseline model), just before the singularity problem occurs, where the two figures describe the same feature space but have different view angles. The features were properly described by the Gaussian model but collapsed to certain dimension. That is, the features were placed on a two-dimensional plane (depicted as red line in Figure 8b) and the direction perpendicular to the red plane was the collapsed dimension. Consequently, the eigenvalue of a covariance matrix along the dimension was close to zero (three eigenvalues were $3.1 \times 10^{-7}, 0.0021$, and 0.0166) and the covariance matrix became a singular matrix (noninvertible). The reason why the singularity problem occurs is that the dimension of feature space is higher than the intrinsic dimension of features. However, it is difficult to find the intrinsic dimension of latent features of inputs.[23] So, the singularity problem would be inevitable.

We compared the three DE models by visualizing the log-likelihood loss of Equation (14) (Figure 8c). The ends of the curves of "baseline" and "Ref. 36" mean the interruption of training due to the singularity problem. The baseline DE model without any constraints

TABLE 4 Quantitative results in ablation studies for Ischemic Stroke Lesion Segmentation (ISLES)

		ADM_{ct}	ADM_{dr}	ADM_{woj}	ADM
Validation	AUC	0.768	0.734	0.787	**0.814**
	Precision	0.267	**0.613**	0.300	0.414
	Recall	0.351	0.265	0.367	**0.381**
	F_1	0.303	0.371	0.330	**0.396**
Test	AUC	0.839	0.662	0.856	**0.870**
	Precision	0.122	0.111	0.158	**0.203**
	Recall	**0.537**	0.211	0.455	0.433
	F_1	0.198	0.145	0.234	**0.276**

Abbreviations: ADM, anomaly detection for magnetic resonance images; ADM_{ct}, proposed frameworks without the dimension reduction network; ADM_{dr}, proposed frameworks without the contrast-to-contrast translation network; ADM_{woj}, proposed framework without the joint optimization of feature generation and density estimation; AUC, area under curve.
The highest score for each metric (each row) are shown in bold values.

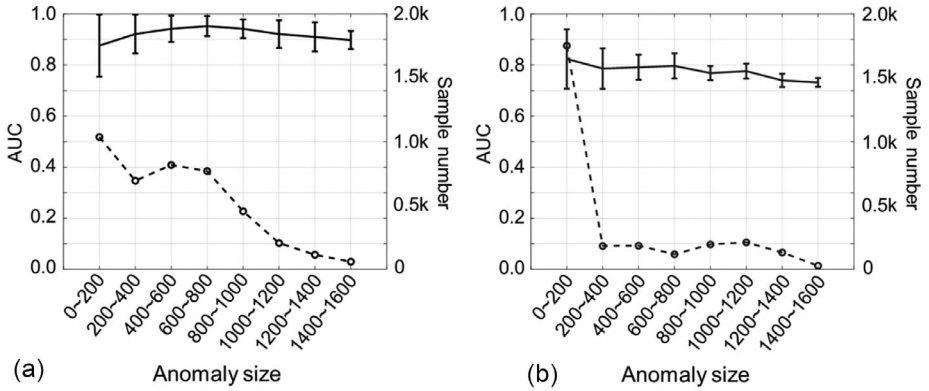

FIGURE 7 Analyses on area under curve (AUC) values with respect to the size of anomaly region for Brain Tumor Segmentation (BraTS) test data (a) and Ischemic Stroke Lesion Segmentation (ISLES) test data (b). The mean and the standard deviation of AUC values are depicted with solid lines. In additions, on the right side of each figure, the number of anomalies in each range is depicted with dashed lines

was easily interrupted (the blue line). Even the DE model with the loss in "Ref. 36" also falls in interruptions (the orange line). Furthermore, since the loss directly affected the log-likelihood value, the log-likelihood loss was slowly minimized compared to that of "baseline". In contrast, the proposed DE model effectively prevented the singularity problem while the intended minimization of log-likelihood was not disturbed (the yellow line).

4 | DISCUSSIONS

In this study, we developed an unsupervised algorithm for pixel-level anomaly detection (i.e., anomaly segmentation) in multicontrast MRI, based on combination of feature generation and DE with GMM. Two relevant features were extracted from multicontrast MR images and collaboratively used for unsupervised anomaly segmentation. The three components of the ADM network, which were CtoC, DR, and DE modules, were well trained to do their corresponding tasks correctly as shown in Figure S1. To evaluate the performance of the proposed method, six previous anomaly detection algorithms were compared in quantitative and qualitative comparison study. The proposed method demonstrated high performance in anomaly segmentation for glioblastoma and ischemic stroke lesion, compared to the previous approaches as shown in Tables 1 and 2. The CtoC network generated relevant reconstruction-based features by using the contrast relationship between multicontrast images and outperformed the autoencoder-based works. To validate the effectiveness of each proposed module, we performed the ablation study by removing each module from the proposed framework (three variants of the proposed method were used). The quantitative results in Tables 3 and 4 demonstrate the performance improvement after each module is added. Furthermore, the proposed method using the lower bound for the eigenvalues of covariance matrices effectively prevented the singularity problem in joint learning with GMM.

The proposed ADM model adopted a CtoC network to generate the reconstruction-based features. The effectiveness of the CtoC network was shown in the experimental results in Figure 6 and Tables 1 and 2.

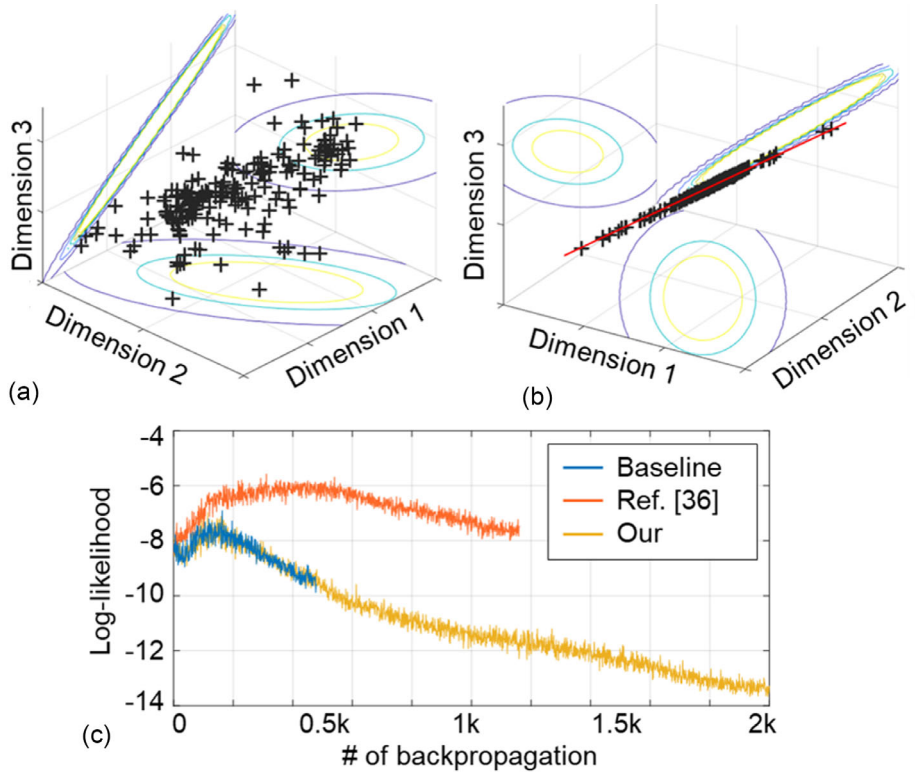

FIGURE 8 (a and b) Features and their latent spaces in the anomaly detection for magnetic resonance images (ADM_{dr}) are visualized with different view angles. (c) The evolution of log-likelihood loss for the three density estimation (DE) models

Especially for ISLES, the competing models had a difficulty in generalizing normal tissue characteristics due to limited training data and showed low performance for test data (AUC < 0.8 in Table 2). In contrast, ADM_{ct} showed higher performance for ISLES test, compared to competing models (AUC $= 0.839$) and led to the high performance of the ADM. Furthermore, the low-dimensional representation from the DR network could be another type of features for anomaly detection in multicontrast MRI and could be collaboratively used with the features from the CtoC network. The ablation study demonstrated the role of each module in quantitative ways.

Considering the high AUC and F_1 scores in the ADM model, the proposed method generally showed superior performance to the competing methods for segmenting glioblastoma and ischemic stroke lesions. Low anomaly scores in most of normal regions and small false positives resulted in high precision of the proposed method, although some methods provided higher recall values. For GMVAE, fanoGAN, DAG, and DAG_c employing autoencoders for feature generation, the fact that those methods only rely on the data compressibility would limit their performances. Although DSVDD was comparable with the ADM for ISLES validation, it had low performance for ISLES test. Test data would have new types of normal tissues, which are not included in train data. The DSVDD seems to have a difficulty in generalizing normative distributions for ISLES data. That is why some normal regions, especially CSF, have higher energy than anomaly regions (Figure 5).

We analyzed the quantitative results of the ADM network according to the lesion size as shown in Figure 7. The AUC value usually decreases as the size of a lesion decreases. The ADM network shows stable performance for lesions of various size in BraTS. On the other hand, in ISELS dataset which has insufficient training data, the AUC values are relatively low especially for the small size of lesions. Furthermore, although the overall range of the lesion size of ISLES is similar to that of BraTS, ISLES has a lot of small size lesions. That would be a reason why anomaly detection in ISLES was more challenging.

The singularity problem is inevitable in joint learning with GMM, as shown in Figure 8a,b. Reducing the likelihood of features in joint learning helps a neural network generate latent features to be suitable for Gaussian models. However, if the given dimension of the latent space is larger than the intrinsic feature dimension, the features collapse into the surplus latent dimension during training and the singularity problem occurs as shown in Figure 8b. It is difficult to find the intrinsic dimension of latent features of inputs.[23] Therefore, instead of challenging network design for an appropriate latent space, using a lower bound for the eigenvalue of covariance matrices fundamentally prevents the singularity problem. We demonstrated the efficacy of the proposed method in the comparison study.

The method is expected to be applicable to various applications with GMM.

The proposed method demonstrated the possibility of detecting anomalies in the two datasets of BraTS and ISLES. However, the quantitative performance in disease segmentation is still inferior to that from supervised segmentation techniques. With an advantage of unsupervised learning that requires only normal data for training, using a large amount of normal data in Refs.48,49 would be considered to further improve the performance of the proposed method and be applicable to actual clinical practices. The straightforward experiments with augmented normal data for ISLES (Section IV in the Supporting Information) demonstrated that the increase of normal data could improve the quantitative performance of anomaly detection.

The ADM network has shown reliable performance on open datasets of brain lesion, whose images were collected from various institutes. There have been problems with applicability to various clinical datasets, since the contrasts of MR images may vary according to clinical institutes even with the same MR protocols. To resolve the problems, we practically used two strategies. First, the CtoC network used the median value of a target contrast image (**m** in Equation (3)) as an input. Contrast translation by the CtoC network was not a deterministic mapping so that reconstruction errors can occur even with normal images. The median value provided an intensity bias in synthesizing the target contrast image. Second, data augmentation with random scaling of image intensity was also adopted to make the ADM network to be trained using datasets with the normative distributions of diverse contrasts. The data augmentation prevented the ADM network from overfitting to specific contrasts. As shown in Section V of the Supporting Information, further experiments were performed using the variants of the ADM network with or without both strategies. The AUC values for BraTS test were measured for evaluation (Table S3). The results show the effectiveness of the two strategies.

There are further works for the ADM network. First, there are some hyperparameters that affect the performance of the ADM network (Section II of the Supporting Information). To obtain better performance, AUC values for validation data were used for adjusting the values of λ (Equation (14)) of the ADM network. Therefore, the hyperparameter selection for the unsupervised models was performed by using label information in the validation data. A new hyperparameter selection method without label information should be developed for fully unsupervised learning and it would be a future work. Second, the usefulness of the ADM network was proven in its segmentation performance with BraTS and ISLES, where brain lesions were large and relatively evident. Evaluation to detect or segment small and subtle brain lesions, such as multiple sclerosis and vascular diseases, still remains.

5 | CONCLUSIONS

We proposed an unsupervised anomaly detection algorithm for MRI, which extracted two relevant features from multicontrast images and found anomaly regions by using a GMM. The proposed method has advantages in terms of pixel-wise detection performance and a learning strategy with GMM. The experimental results show that the proposed approach has higher performance than the state-of-the-art anomaly detection approaches. In particular, by forcing an estimated covariance matrix to be positive definite, the singularity problem can be prevented in GMM. The unsupervised approach has great potential in detecting various lesions in multicontrast MRI.

ACKNOWLEDGMENTS

This work was supported by the ICT R&D program of MSIT/IITP [2017-0-01779, A machine learning and statistical inference framework for explainable artificial intelligence].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
2. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85-117.
3. Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging*. 2016;35(5):1153-1159.
4. Taboada-Crispi A, Sahli H, Hernandez-Pacheco D, Falcon-Ruiz A. Anomaly detection in medical image analysis. In: *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. IGI Global; 2009:426-446.
5. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. *ACM Comput Surv (CSUR)*. 2009;41(3):1-58.
6. Baur C, Wiestler B, Albarqouni S, Navab N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. *International MICCAI Brainlesion Workshop*. Springer; 2018:161-169.
7. Chen X & Konukoglu E Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. arXiv:1806.04972; 2018.
8. Seeböck P, Orlando JI, Schlegl T, et al. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT. *IEEE Trans Med Imaging*. 2019;39(1):87-98.
9. Seeböck P, Waldstein SM, Klimscha S, et al. Unsupervised identification of disease marker candidates in retinal OCT imaging data. *IEEE Trans Med Imaging*. 2018;38(4):1037-1047.
10. Chen Y, Zhou XS, Huang TS. One-class SVM for learning in image retrieval. Proceedings 2001 International Conference on Image Processing. Vol 1. IEEE; 2001:34-37.

11. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045-1057.
12. ISLES Challenge 2015 Ischemic Stroke Lesion Segmentation.
13. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv:1811.02629; 2018.
14. Alaverdyan Z, Jung J, Bouet R, Lartizien C. Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: application to epilepsy lesion screening. *Med Image Anal*. 2020;60:101618.
15. Bowles C, Qin C, Guerrero R, et al. Brain lesion segmentation through image synthesis and outlier detection. *NeuroImage Clin*. 2017;16:643-658.
16. Uzunova H, Schultz S, Handels H, Ehrhardt J. Unsupervised pathology detection in medical images using conditional variational autoencoders. *Int J Comput Assist Radiol Surg*. 2019;14(3):451-461.
17. Nair T, Precup D, Arnold DL, Arbel T. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med Image Anal*. 2020;59:101557.
18. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? *JACM*. 2011;58(3):1-37.
19. Xu H, Caramanis C, Sanghavi S. Robust PCA via outlier pursuit. *IEEE Trans Inf Theory*. 2012;58(5):3047-3064.
20. Andrews JT, Morton EJ, Griffin LD. Detecting anomalous data using auto-encoders. *Int J Mach Learn Comput*. 2016;6(1):21.
21. Chen J, Sathe S, Aggarwal C, Turaga D. Outlier detection with autoencoder ensembles. Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM; 2017:90-98.
22. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313(5786):504-507.
23. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(8):1798-1828.
24. Kim J, Scott CD. Robust kernel density estimation. *J Mach Learn Res*. 2012;13(1):2529-2565.
25. Xiong L, Póczos B, Schneider J. Group anomaly detection using flexible genre models. *Advances in Neural Information Processing Systems*. ACM; 2011:1071-1079.
26. Zimek A, Schubert E, Kriegel HP. A survey on unsupervised outlier detection in high-dimensional numerical data. Stat Anal Data Min *ASA Data Sci J*. 2012;5(5):363-387.
27. Mourão-Miranda J, Hardoon DR, Hahn T, et al. Patient classification as an outlier detection problem: an application of the one-class support vector machine. *Neuroimage*. 2011;58(3):793-804.
28. Erfani SM, Rajasegarar S, Karunasekera S, Leckie C. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit*. 2016;58:121-134.
29. Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. ACM; 2014:4-11.
30. Xu D, Ricci E, Yan Y, Song J, Sebe N. Learning deep representations of appearance and motion for anomalous event detection. Proceedings of the British Machine Vision Conference (BMVC). 2015.
31. Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification. Proceedings of the 35th International Conference on Machine Learning. PMLR; 2018:4393-4402.
32. Tax DM, Duin RP. Support vector data description. *Mach Learn*. 2004;54(1):45-66.
33. Yang B, Fu X, Sidiropoulos ND, Hong M. Towards k-means-friendly spaces: simultaneous deep learning and clustering. Proceedings of the 34th International Conference on Machine Learning. PMLR; 2017:3861-3870.
34. Yang X, Huang K, Goulermas JY, Zhang R. Joint learning of unsupervised dimensionality reduction and Gaussian mixture model. *Neural Process Lett*. 2017;45(3):791-806.
35. Yang X, Huang K, Zhang R. Unsupervised dimensionality reduction for Gaussian mixture model. International Conference on Neural Information Processing. Springer; 2014:84-92.
36. Zong B, Song Q & Min MR et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. International Conference on Learning Representations. 2018.
37. Variani E, McDermott E, Heigold G. A Gaussian mixture model layer jointly optimized with discriminative features within a deep neural network architecture. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2015:4270-4274.
38. Kingma DP, Welling M. An introduction to variational autoencoders. arXiv:1906.02691; 2019.
39. Goodfellow IG, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D. Generative adversarial nets. *Advances in Neural Information Processing Systems*. ACM; 2014:2672-2680.
40. Zimmerer D, Isensee F, Petersen J, Kohl S, Maier-Hein K. Unsupervised anomaly localization using variational auto-encoders. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2019:289-297.
41. Chen X, You S, Tezcan KC, Konukoglu E. Unsupervised lesion detection via image restoration with a normative prior. *Med Image Anal*. 2020;64:101713.
42. Abati D, Porrello A, Calderara S, Cucchiara R. Latent space autoregression for novelty detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2019:481-490.
43. Sun L, Wang J, Huang Y, Ding X, Greenspan H, Paisley J. An adversarial learning approach to medical image synthesis for lesion detection. *IEEE J Biomed Health Inform*. 2020;24(8):2303-2314.
44. Xia T, Chartsias A, Tsaftaris SA. Pseudo-healthy synthesis with pathology disentanglement and adversarial learning. *Med Image Anal*. 2020;64:101719.
45. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal*. 2019;54:30-44.
46. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. European Conference on Computer Vision. Springer; 2016:694-711.
47. Zhang C, Woodland PC. Joint optimisation of tandem systems using Gaussian mixture density neural network discriminative sequence training. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2017:5015-5019.
48. Van Essen DC, Smith SM, Barch DM, et al. The WU-Minn human connectome project: an overview. *Neuroimage*. 2013;80:62-79.
49. Poldrack RA, Gorgolewski KJ. Making big data open: data sharing in neuroimaging. *Nat Neurosci*. 2014;17(11):1510-1517.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kim B, Kwon K, Oh C, Park H. Unsupervised anomaly detection in MR images using multicontrast information. *Med. Phys.* 2021;48:7346–7359. https://doi.org/10.1002/mp.15269