

## RESEARCH ARTICLE

# Quantile Autoencoder With Abnormality Accumulation for Anomaly Detection of Multivariate Sensor Data

SEUNGHYOUNG RYU<sup>1</sup>, JIYEON YIM<sup>1,2</sup>, JUNGHOO SEO<sup>3</sup>, YONGGYUN YU<sup>1</sup>, AND HOGEON SEO<sup>1</sup>

<sup>1</sup>Artificial Intelligence Application & Strategy Team, Korea Atomic Energy Research Institute, Daejeon 34507, South Korea

<sup>2</sup>Department of Nuclear Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea

<sup>3</sup>SI Analytics Company, Daejeon 34047, South Korea

Corresponding author: Hogeon Seo (hogeony@kaeri.re.kr)

This work was supported in part by the Korea Atomic Energy Research Institute (KAERI) Research and Development Program under Grant KAERI-524450-22, and in part by the National Research Foundation of Korea (NRF) funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant NRF-2021R1F1A1051290.

**ABSTRACT** Anomaly detection (AD) is a crucial task in various industrial sectors where large amounts of data are generated from multiple sensors. Deep learning-based methods have made significant progress in AD, owing to big data and deep neural networks (DNN). Most methods for deep anomaly detection (DAD) utilize reconstruction error (i.e., the difference between the original and reconstructed values) as a measure of abnormality. However, AD performance can be improved by diversifying the source of anomaly score. To support this, we introduce the concept of anomaly source diversification and provide mathematical proofs to support this idea. In this regard, we propose a quantile autoencoder (QAE) with abnormality accumulation (AA) as a novel DAD approach that leverages data uncertainty and iteratively obtains reconstruction errors as additional sources. The anomaly score with QAE is derived from both the reconstruction error and the uncertainty term which is the range between the two quantiles. In addition, AA aggregates the errors obtained from the recursive reconstruction of the input, after which calculates the anomaly score based on the Mahalanobis distance. This process induces the score distributions of both the normal and abnormal samples farther apart by narrowing the width of the distributions, which contributes to the improvement of AD performance. The performance of the proposed QAE-AA was verified through the experiments on multi-variate sensor datasets in various domains; QAE-AA achieves 4-23% higher AUROC score on average compared to the other AD methodologies.

**INDEX TERMS** Deep learning, anomaly detection, autoencoder, quantile regression, anomaly score.

## I. INTRODUCTION

Anomaly detection (AD), which is also known as novelty detection or outlier detection, is a task that involves identifying abnormal cases in a pool of collected data or a data stream. Since the 1960s, it has been widely studied [1] and utilized in broad range of applications, such as, fraud detection [2]–[4], network security [5]–[7], video surveillance [8]–[10], medical diagnosis [11]–[13] and multiple sensor data [14]–[18]. In particular, as IoT and big data technologies

have become common, acquiring meaningful features for performing AD from massive sensor data becomes more challenging. Under these circumstances, recent advancements in neural networks and deep learning have significantly influenced the field of AD. Deep anomaly detection (DAD) methods have demonstrated improved performance in many complicated AD tasks [19].

DAD framework can be classified into supervised, unsupervised, and semi-supervised AD based on the problem and data formulation. Supervised AD can be achieved when both normal and abnormal samples are sufficient and labeled [20]. However, in general, abnormal samples are neither sufficient

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Hao Chen.

nor labeled in many real-world applications [19]. This is the reason why semi-supervised and unsupervised AD methods have been applied widely in DAD. Unlike supervised AD, which directly determines whether the input is normal or not, both the semi-supervised and unsupervised AD methods learn the features of normality, after which they calculate the anomaly score to measure a degree of abnormality.

A recent review of [21] further categorized DAD according to the role of deep neural network (DNN) in AD process. Two major branches among them are *generic normality feature learning* and *anomaly measure-dependent feature learning*. The former is based on general feature extraction methods including autoencoders and generative adversarial networks (GAN). In doing this, networks learn generic representations to well reconstruct/generate/predict normal data. Typically, these methods utilize reconstruction error as an anomaly score. Feature extraction of the latter approach is more dependent on the anomaly scoring function. By designing a loss function for specific anomaly measures, deep learning model could learn score-dependent latent features, e.g., nearest neighbor distance, one-class classification (DSVDD [22]), and clustering-based scores (DAGMM [23]).

We focus on the former approach, in particular, autoencoder-based methods with semi-supervised AD settings which are easy to implement and have straightforward intuitions in detecting anomalies [21]. Deep autoencoders perform dimensionality reduction, such as principal component analysis [24]–[26] and random projection [27]–[29]. By training an autoencoder with normal samples, the network shows lower reconstruction error on normal data, but poor on abnormal data that are not exposed during the training. This difference allows us to separate two classes by reconstruction error. However, AD performance can be further improved because the objective function of minimizing reconstruction error is not identical to maximizing anomaly detection performance. In this regard, one approach to improve DAD performance is leveraging additional sources for anomaly scoring (i.e., anomaly source diversification), and here we exploit *aleatoric uncertainty* and *recursive reconstruction errors*.

There are two types of uncertainty in deep learning depending on its causes: *epistemic* uncertainty and *aleatoric* uncertainty [30]. Epistemic uncertainty, also called model uncertainty, originates from the difference between training results (difference between the models). Epistemic uncertainty can be reduced by acquiring more data. On the other hand, aleatoric uncertainty, which is also called data uncertainty, is attributed to the data itself. Therefore, it is inherent and irreducible. Recently, epistemic uncertainty has been considered in DAD, because the reconstruction for the abnormal sample has a significant variance depending on the model [31]–[35]. On the other hand, aleatoric uncertainty has received less attention in the field of DAD and is used as a threshold for the classification of normal and abnormal [36].

In this study, we propose a novel DAD framework that considers aleatoric uncertainty by introducing a quantile

autoencoder (QAE). We leverage aleatoric uncertainty under the assumption of channel-wise consistency in normal conditions; that is, the inherent deviation of normal data would be less than abnormal data. Aleatoric uncertainty in terms of the range between two quantiles is used in the proposed framework with reconstruction errors. In addition, we propose the abnormality accumulation (AA) technique, which aggregates and calculates the anomaly score based on the errors of recursive reconstructions. This makes the difference between normal and abnormal distributions more evident. We verified the proposed framework with multi-variate sensor datasets in different domains. Each of the two methods contributes to anomaly source diversification. We further provide theoretical grounds that support the idea of anomaly source diversification under the assumption of Gaussian error distribution.

The main contributions of this study are as follows:

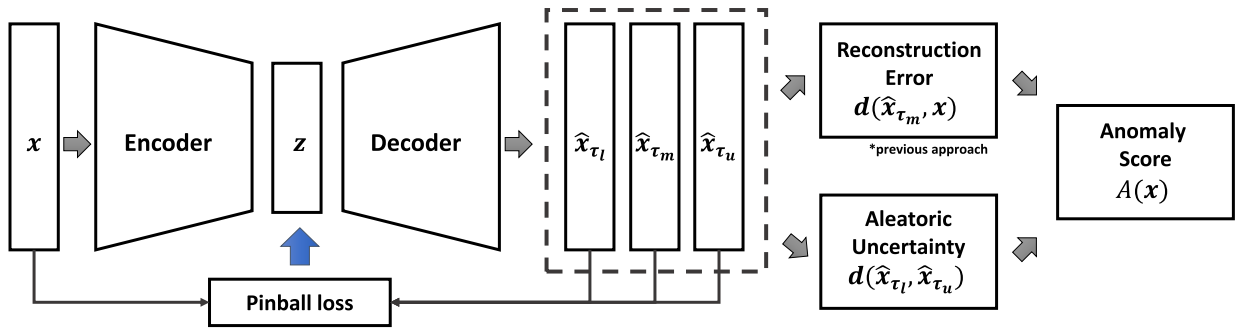
- We propose QAE for uncertainty-based DAD. Aleatoric uncertainty term, which is the range between the two quantiles, is additionally considered in anomaly scoring. To the best of our knowledge, it is the first time to utilize QAE and quantile range in AD as the source of anomaly score.
- We propose AA, in which recursive reconstruction error is additionally considered in anomaly scoring. AA decreases the overlapping region between anomaly score distributions on normal and abnormal data. Therefore, the two score distributions become more distinguishable, which facilitates the determination of the normal and anomalies. The performance of the proposed QAE-AA is tested using various multi-variate sensor datasets, and it demonstrates a significant improvement in AD performance in terms of AUROC.
- We introduce the concept of anomaly source diversification, which means it becomes easier to distinguish normals and abnormal as we diversify and gather more error sources in calculating anomaly scores. We provide mathematical proofs of why anomaly source diversification is helpful in reconstruction error-based DAD methods under the assumption of Gaussian error distribution. This explains the utilization of QAE, AA, and Mahalanobis distance in the proposed framework, and we further show that the empirical error distributions can be modeled by a mixture of Gaussians.

This paper consists of five sections, including the introduction. In Section II, we introduce the related DAD methods. Section III describes the proposed framework, QAE-AA, and concept of anomaly source diversification. The experiments' results and conclusions are presented in Sections IV and V, respectively.

## II. RELATED WORKS

### A. RECONSTRUCTION ERROR BASED METHODS

In DAD by generic normality feature learning, the difference between an input and the reconstructed output (e.g., mean



**FIGURE 1.** Structure of the QAE for AD. The QAE predicts the median value and two quantiles. The reconstruction error and the aleatoric uncertainty are used for anomaly scoring.

**TABLE 1.** Summary of literature on the recent relevant anomaly detection models.

Ref.	Year	Abbr	Backbone	Features
[37]	2015	VAE	VAE	First VAE-based anomaly detection.
[11]	2017	AnoGAN	GAN	First GAN framework for anomaly detection.
[38]	2018	GANomaly	GAN	Additional latent error is used with GAN AD score.
[22]	2018	DSVDD	NN	Finding minimum hypersphere in latent space of neural networks.
[23]	2018	DAGMM	NN	Fitting Gaussian mixture model in latent space of neural networks.
[39]	2019	RAPP	AE	Latent error with Mahalanobis distance.
[40]	2020	XAE	AAE	RAPP with extended autoencoder structure.
[41]	2020	DROCC	NN	Robust one class classification via neural network.
[42]	2020	DN2	NN	k-nearest neighbor approach in latent space from neural network.
[43]	2021	enGAN	GAN	Unified framework of GAN based models with ensemble.
[44]	2022	LUNAR	GNN	Local neighborhoods approach via graph neural network and negative samples.

squared error) is typically used as a measure of abnormality. In general, the autoencoder (AE) [45]–[47] and the variational autoencoder (VAE) [37], [48], [49] have been utilized because of their superior capability to learn latent representations. Once a neural network is trained using normal samples to minimize its reconstruction error, the normal samples are reconstructed effectively from lower-dimensional latent features, but the abnormal samples are not. Therefore, anomalies produce larger reconstruction errors and can be distinguished by properly specifying a threshold for the normal cases. In a recent work of [39], a novel DAD methodology namely reconstruction along projection pathways (RAPP) was introduced which leverages reconstruction errors in latent spaces. It first collects the latent features in each layer of the encoder from the first forward path. By inputting the output into the encoder network again, the difference in the latent features of the second forward path is used as an anomaly score.

Another branch is based on generative adversarial networks (GANs) [11], [38], [43], [50], [51]. By utilizing GAN architecture, a neural network can model the distribution of normal samples. Anomaly scoring in GAN approaches is also based on the reconstruction error, and discriminator loss is additionally used with the reconstruction error [11], [43], [50], [51]. GANomaly [38] performs anomaly scoring based on the reconstruction error of bottleneck features. Thus, the score is defined in the latent space. The unified framework of GAN-based DAD is well summarized in [43], and it shows

that the ensemble of anomaly scores from GAN variants further improves detection performance.

Previous studies show the utilization of the latent reconstruction errors, discriminator loss, and ensemble of reconstruction errors has improved the AD performance. This can be explained by the concept of anomaly source diversification. Similarly, uncertainty can also be an additional source, and contribute to the improved performance of DAD methods.

### B. ANOMALY DETECTION WITH UNCERTAINTY

There are various methods for dealing with uncertainty within a deep learning framework, such as Bayesian deep learning, the Monte-Carlo (MC) dropout technique, and deep ensembles [52]. In particular, the MC dropout technique is adopted in uncertainty-based DAD applications [31]–[35]. MC dropout utilizes dropout in both the training and inference stages. Therefore a single neural network can generate different outputs based on the probabilistic connections between neurons. The statistical information of the outputs can be obtained through MC sampling, where the dropout is activated during the inference stage. This approach mainly aims to exploit epistemic uncertainty. The following examples use the MC dropout technique for epistemic uncertainty quantification: [31] conducted a study on a deep learning-based time-series prediction with a confidence interval for the Uber dataset in which AD was performed by triggering

an alarm when the observed value fell outside the 95% predictive interval. Uncertainty in terms of the variance of the reconstruction errors was utilized as a weighting factor for the anomaly score [32]. In the field of medical imaging, uncertainty was adopted to diagnose diabetic retinopathy from fundus images in [33]. In the study conducted by [34], pixel-wise variations in retinal optical coherence tomography images were derived and utilized for the segmentation of abnormal areas. Similarly, the uncertainty of abnormal images in the MVTEC-AD dataset was derived, after which the area under the receiver operating characteristic (AUROC) scores between the residual-based and uncertainty-based detection results was compared [35].

In this study, we consider aleatoric uncertainty via QAE and use the uncertainty term to diversify the sources for anomaly scoring. Our approach is different from the previous methods in terms of the method for measuring uncertainty (aleatoric uncertainty with multiple quantile regression) and the way in which uncertainty is used for anomaly scoring (Mahalanobis distance-based anomaly score).

### III. PROPOSED METHODOLOGY

In this section, we describe the proposed QAE model and AA technique in detail.

#### A. QUANTILE AUTOENCODER FOR ANOMALY DETECTION

The aleatoric uncertainty is utilized for AD with QAE. Its basic concept is that the reconstruction of normal samples produces stable outputs within a certain range of variations for each input channel. In other words, *normal* cases reconstructed using *normal-fitted latent features* are likely to have low variability compared to the abnormal cases (due to the loss function that induces minimizing the variance of output). The consistency of normal samples is independently valid for each channel. To leverage aleatoric uncertainty, we propose a QAE which predicts multiple quantiles with a single neural network. Then, the range between two quantiles (upper and lower) is used as a degree of uncertainty. Then, the anomaly score is derived by *Mahalanobis distance* from both the reconstruction error and the uncertainty term, as illustrated in Fig. 1.

##### 1) QUANTILE AUTOENCODER

The basic AE performs a single-value reconstruction, which is the mean of Gaussian distribution by minimizing the mean squared error (MSE). The proposed QAE is a variation of the AE that predicts the different quantiles of the output distribution by minimizing the sum of pinball losses. Thus, the QAE performs multiple quantile regressions through a single neural network, which can be regarded as multi-task learning.

Let  $X$  be a random variable with a cumulative distribution function  $F_X(x)$ . Next, the  $\tau$ -th quantile  $x_\tau$  is  $F_X^{-1}(\tau)$ , where  $\tau \in (0, 1)$ . The proposed QAE  $Q$  comprises an encoder  $Q_{enc}$

and a quantile decoder  $Q_{dec}$  as follows:

$$\begin{aligned} Q_{enc}(x) &= z, \\ Q_{dec}(z) &= [\hat{x}_{\tau_l}, \hat{x}_{\tau_m}, \hat{x}_{\tau_u}] = \hat{\mathbf{x}}_\tau, \end{aligned} \quad (1)$$

where  $\tau_l < 0.5$ ,  $\tau_u > 0.5$ , and  $\tau_m = 0.5$ , that represent the lower, upper quantiles and median.

Quantile regression can be performed by minimizing the pinball loss [53]. For the  $\tau$ -quantile, the pinball loss is defined as follows:

$$L_\tau(x, \hat{x}_\tau) = \begin{cases} \tau(x - \hat{x}_\tau) & \text{if } x \geq \hat{x}_\tau \\ (1 - \tau)(\hat{x}_\tau - x) & \text{if } x < \hat{x}_\tau \end{cases} \quad (2)$$

where  $x$  represents an input and  $\hat{x}_\tau$  represents a prediction of  $\tau$ -quantile. The pinball loss function is a tilted absolute error function weighted by the target  $\tau$ . Intuitively, the pinball loss results in a higher penalty of overestimation ( $\hat{x}_\tau \geq x$ ) for low quantiles ( $\tau < 0.5$ ). Therefore, the network is trained to underpredict. Similarly, the network overpredicts for high quantiles  $\tau > 0.5$ . Finally, the loss for the proposed QAE,  $L_Q$ , is a summation of the pinball losses with multiple  $\tau$ s:

$$L_Q(x, \hat{\mathbf{x}}_\tau) = L_{\tau_l}(x, \hat{x}_{\tau_l}) + L_{\tau_m}(x, \hat{x}_{\tau_m}) + L_{\tau_u}(x, \hat{x}_{\tau_u}) \quad (3)$$

By training multiple quantiles at once, QAE could alleviate *crossing quantiles* problem [54].

##### 2) QUANTILE-BASED ANOMALY SCORING

In the proposed approach, anomaly scores are derived from the reconstruction error and the aleatoric uncertainty term which is the range between the predicted upper and lower quantiles. From the output of the QAE  $\hat{\mathbf{x}}_\tau$ , the reconstruction error  $\epsilon_{rec}$  and the uncertainty term  $\epsilon_{unc}$  in the form of row vector are defined as follows:

$$\begin{aligned} \epsilon_{rec} &= x - \hat{x}_{\tau_m}, \\ \epsilon_{unc} &= \hat{x}_{\tau_u} - \hat{x}_{\tau_l}. \end{aligned} \quad (4)$$

Then, typical MSE-based anomaly score  $A_r(x)$  and a simple version of the quantile-based anomaly score  $A_q(x)$  with  $\epsilon = [\epsilon_{rec}, \epsilon_{unc}]$  is expressed as follows:

$$A_r(x) = \|\epsilon_{rec}\|^2/d_{\epsilon_{rec}}, \quad (5)$$

$$A_q(x) = \|\epsilon\|^2/d_\epsilon \quad (6)$$

where  $d$  is the dimension of corresponding vector. Because  $\epsilon_{rec}$  and  $\epsilon_{unc}$  exist in different domains, the difference between their magnitude ranges has a significant impact on anomaly scoring. This problem can be alleviated through normalization. Previous work of [39] reported a similar problem when utilizing reconstruction errors in both the original space and the latent spaces. To match the range of the error terms, a normalized distance with orthogonalization and scaling was proposed which is identical to the Mahalanobis distance. We also calculate the Mahalanobis distance-based normalized anomaly score  $A_{nq}(x)$  as follows:

$$A_{nq}(x) = (\epsilon - \mu)S^{-1}(\epsilon - \mu)^T, \quad (7)$$

where  $\mu$  and  $S$  represent the channel-wise mean and covariance matrix obtained from  $\epsilon$  of the training set, respectively. This approach is also adopted in the AA technique described below.

**B. ABNORMALITY ACCUMULATION**

In addition to the QAE, we propose an anomaly scoring technique, namely abnormality accumulation (AA), to further improve AD performance. The basic concept of AA is calculating the anomaly score based on the aggregated errors using a single AE. Let the superscript  $i$  represent the number of iterations for AA. Then,  $Q^i$  is a composite function  $Q \circ Q \cdots \circ Q$  that  $Q$  is repeated  $i$  times; thus,  $Q^1(x)$  is equal to  $Q(x)$ . After the first iteration,  $\hat{x}_{\tau_m}^i$  is used as the input for the next iteration. Consequently,  $\epsilon^i$  contains the reconstruction error and uncertainty term obtained from the result of  $Q^i(x)$ . With a predefined number of iterations  $N$ , we aggregate  $\epsilon^i$ s into  $[\epsilon^i]_{i=1}^N$  and calculate normalized anomaly score with AA as follows:

$$A_{nqa}(x) = ([\epsilon^i]_{i=1}^N - \mu_A)S_A^{-1}([\epsilon^i]_{i=1}^N - \mu_A)^T, \quad (8)$$

where  $\mu_A$  and  $S_A$  represent the channel-wise mean and covariance matrix obtained from  $[\epsilon^i]_{i=1}^N$  of the training set. The illustration and algorithm of AA are presented in Fig. 2 and Algorithm 1.

**Algorithm 1** Abnormality Accumulation

- 1: **Input:** input sample  $x$ , QAE  $Q$ , number of iteration  $N$ .
- 2: set initial value  $\hat{x}_{\tau_m}^0 = x$ ;
- 3: **for**  $i = 1, 2, 3, \dots, N$  **do**
- 4:   obtain  $[\hat{x}_{\tau_l}^i, \hat{x}_{\tau_m}^i, \hat{x}_{\tau_u}^i] = Q(\hat{x}_{\tau_m}^{i-1})$ ;
- 5:   compute  $\epsilon_{rec}^i = x - \hat{x}_{\tau_m}^i$  and  $\epsilon_{unc}^i = \hat{x}_{\tau_u}^i - \hat{x}_{\tau_l}^i$ ;
- 6:   concatenate  $\epsilon_{rec}^i$  and  $\epsilon_{unc}^i$  into  $\epsilon^i = [\epsilon_{rec}^i, \epsilon_{unc}^i]$ ;
- 7: **end for**
- 8: aggregate  $\epsilon^i$ s to create a concatenated vector
- 9:  $[\epsilon^i]_{i=1}^N = [\epsilon^1, \epsilon^2, \epsilon^3, \dots, \epsilon^N]$ ;
- 10: compute the channel-wise mean  $\mu_A$  and the covariance matrix  $S_A$  from  $[\epsilon^i]_{i=1}^N$ ;
- 11: **Output:** anomaly score with AA,
- 12:  $A_{nqa}(x) = ([\epsilon^i]_{i=1}^N - \mu_A)S_A^{-1}([\epsilon^i]_{i=1}^N - \mu_A)^T$ .

**C. ANOMALY SOURCE DIVERSIFICATION**

As previously stated, the use of QAE and AA with Mahalanobis distance in the proposed DAD methods can be supported by the concept of anomaly source diversification which is based on the following proposition.

*Proposition 1: For  $k$  i.i.d. zero-mean Gaussian reconstruction error where the variance of normal is smaller than that of abnormal, MSE-based anomaly score distributions on normal and abnormal samples become more distinguishable as  $k$  increases.*

*Proof:* Let  $X_n$  and  $X_a$  be random variables of normal and anomaly that follow the distributions of  $P_{X_n}$  and  $P_{X_a}$ , respectively, where  $P_{X_n} \approx P_{X_a}$ . For AD, an AE  $Q$  with

its weight  $\theta$  is trained to minimize MSE between input and reconstructed output for normal data.

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{X_n} [||x - Q_{\theta}(x)||^2]. \quad (9)$$

We further assume that the well-trained AE  $Q_{\theta}$  has a zero-mean Gaussian distribution on its reconstruction errors,  $E_n = X_n - Q_{\theta}(X_n) \sim N(0, \sigma_n^2)$  and  $E_a = X_a - Q_{\theta}(X_a) \sim N(0, \sigma_a^2)$  with inequality of  $\sigma_n \leq \sigma_a$  from minimization of Eq. 9. Then, the squared errors,  $E_n^2$  and  $E_a^2$ , follow the gamma distributions of  $\Gamma(0.5, 2\sigma_n^2)$  and  $\Gamma(0.5, 2\sigma_a^2)$ , respectively.

If we consider multivariate case of  $k$  i.i.d. error distributions  $E_1, \dots, E_k$  in addition to the subscripts ( $n$  and  $a$ ),  $\bar{E}^2 = \frac{1}{k} \sum_i E_i^2$  follows the gamma distribution  $\Gamma(k/2, 2\sigma^2/k)$ , and its mean and variance can be obtained by the parameters of the gamma distribution, i.e.,  $\sigma^2$ , and  $2\sigma^4/k$ . Then,  $\bar{E}_n^2$  and  $\bar{E}_a^2$  have fixed means which are  $\sigma_n^2$  and  $\sigma_a^2$ . On the other hand, variances of the two distribution decreases as  $k$  increases because  $Var(\bar{E}_n^2) = 2\sigma_n^4/k$  and  $Var(\bar{E}_a^2) = 2\sigma_a^4/k$ . This makes the overlapping area between anomaly score distributions  $\bar{E}_n^2$  and  $\bar{E}_a^2$  decreases; thus, normal and abnormal become more distinguishable.  $\square$

Fig. 3 shows the changes in the anomaly score distribution according to different  $k$ . As the number of  $k$  increases, the separation between both distributions becomes evident.

The above proposition emphasizes the importance of obtaining as many i.i.d Gaussian errors as possible in AD. Anomaly source diversification aims to obtain different error sources to achieve the effect of increasing  $k$ . Typically, AD deals with multivariate data  $x \in \mathbb{R}^d$ , and the proposed AA technique further expands  $k = d$  to  $k = N \cdot d$  using a single neural network, where  $N$  indicates the number of iterations. Also, Mahalanobis distance corresponds to  $\bar{E}^2$  because orthogonalization and scaling make rotated axes to be independent. Therefore the AD performance can be improved by the proposed methodology.

Note that Proposition 1 assumes the ideal case of zero-mean Gaussian error distributions on both normal and abnormal data. Although normally distributed empirical errors can be found in the literature [55], the assumption of zero-mean Gaussianity on untrained abnormal data can be questioned because empirical errors are more likely to be skewed and biased. In this regard, we analyze the empirical error distributions of the real-world datasets used in this study. Fig. 4 shows histograms of reconstruction errors on abnormal data with Gaussian mixture model fitting results. As can be seen, the Gaussian mixture model well fits empirical error distribution, and we further generalize Proposition 1 to Gaussian distribution with arbitrary mean and variance as follows.

*Proposition 2: For  $k$  i.i.d. non-zero-mean Gaussian reconstruction errors, MSE-based anomaly score distributions on normal and abnormal samples become separable as  $k$  increases.*

*Proof:* For random variable  $X \sim N(\mu_x, 1)$ ,  $Z = \sum_{i=1}^k X_i^2$  follows non-central chi-squared distribution with

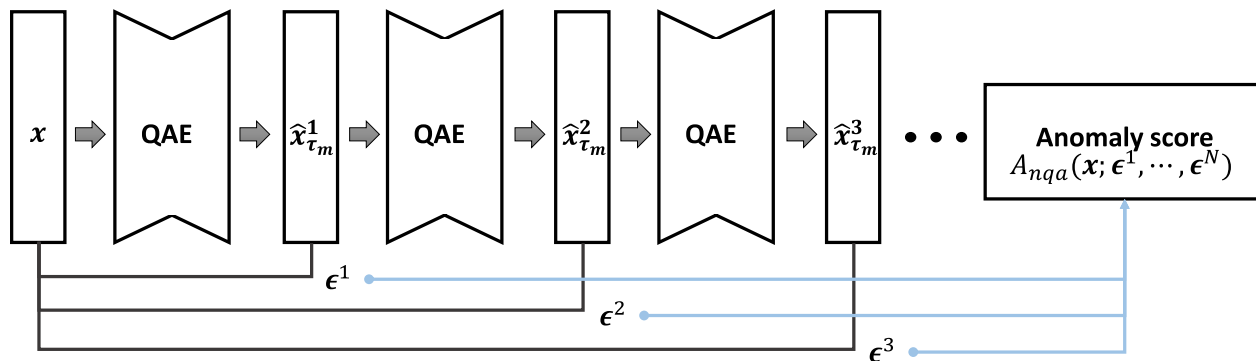


FIGURE 2. Illustration of AA. Abnormality becomes accumulated by aggregating  $\epsilon^i$ .

TABLE 2. Description of the benchmark datasets.

Dataset	$N$	$R_A$	$d_x$	$d_z$	Class	Domain	Anomaly Target	Description	Ref.
MI-F	25k	8.2%	58	23	2	CNC milling	Machine not completed	Real-valued sensor data (e.g., position, velocity, currents, etc.)	[56]
MI-V	23k	17.2%	58	23	2	CNC milling	Workpiece out-of-spec		
EOPT	10k	11.3%	20	6	2	Storage system	System failures	Binary & real-valued features (e.g., power, voltage, distance.)	[57]
RARM	20k	23.2%	6	3	2	Robotics	Malfunctions	Positions of robot arms	[58]
SNSR	58k	-	48	17	11	Electric Currents	Defective conditions	Electric current drive signals.	[59]

\* $N$ : number of data,  $R_A$ : ratio of anomaly in dataset,  $d_x$ : dimension of data,  $d_z$ : dimension of bottleneck feature.

\*\*This table is modified from [39].

$k$  degree of freedom and non-centrality parameter  $\lambda = \sum_{i=1}^k \mu_x^2$ . The mean and variance of  $Z$  is  $k + \lambda$  and  $2(k + 2\lambda)$ , respectively. If we denote reconstruction error with arbitrary mean and variance as  $Y = \sigma_y X \sim N(\mu_y, \sigma_y^2) = N(\mu_x \sigma_y, \sigma_y^2)$ , then squared mean of  $k$  i.i.d  $Y$ s can be represented by  $Z$  as follows.

$$\frac{1}{k} \sum_{i=1}^k Y_i^2 = \frac{\sigma_y^2}{k} \sum_{i=1}^k X_i^2 = \frac{\sigma_y^2}{k} Z \quad (10)$$

Then, the mean and variance of  $\frac{1}{k} \sum_{i=1}^k Y_i^2$  are  $\sigma_y^2 + \mu_y^2$  and  $\frac{2\sigma_y^2}{k}(\sigma_y^2 + 2\mu_y^2)$ , respectively. Likewise to the Proposition 1, variance of anomaly score distribution decreases as  $k$  increases but their mean is further apart by the bias of reconstruction error  $\mu_y^2$ ; mean and variance of mean square of  $k$  i.i.d zero-mean Gaussian  $N(0, \sigma^2)$  are  $\sigma^2$  and  $2\sigma^4/k$ .  $\square$

Thus, increasing  $k$  is still helpful for Gaussian with arbitrary mean and variance. In this perspective, the assumption of zero-mean can be considered the most difficult case in AD, because it is impossible to classify the normal and abnormal when  $\sigma_a$  is equal to or less than  $\sigma_n$ .

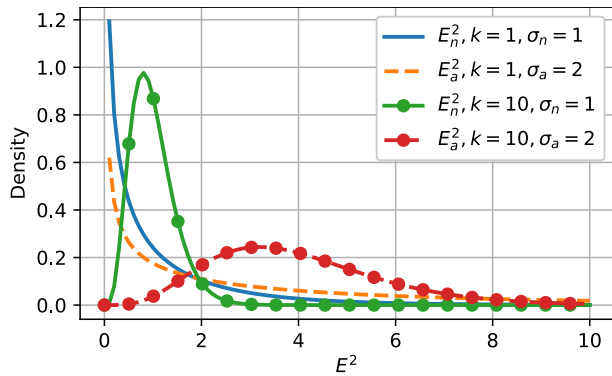
#### IV. EXPERIMENTS

To verify the effectiveness of the proposed methodology, we compare the AUROC score which is a well-known evaluation metric for classification models. In binary classification,

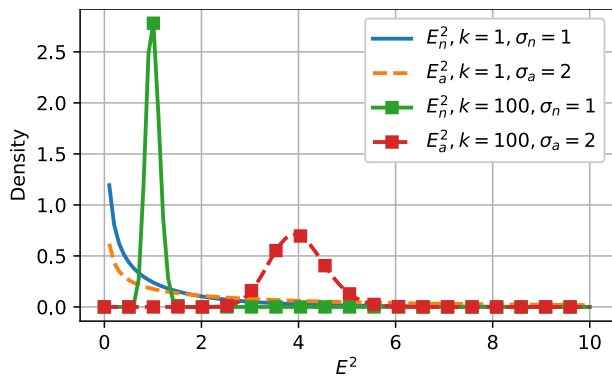
the perfect classifier has AUROC of 1, and the uniformly random classifier has AUROC of 0.5. During the experiment, we refer to the verification framework and datasets presented in the work of [39].

#### A. DATASETS AND PROBLEM SETTINGS

We compare the results of five different datasets; MI-F, MI-V, EOPT, RARM (binary class), and SNSR (multi class). Description of datasets is given in Table 2. For experiments, we set two normality conditions; unimodal normality and multimodal normality. Unimodal normality indicates a single normal class, and multimodal normality means a normal dataset is composed of multiple normal classes. The binary class datasets have labeled normal class. Therefore, the experiment is performed with unimodal normality. On the other hand, a multi-class dataset has no explicit labels for normal. Therefore, we set a target class that is considered normal in unimodal normality and abnormal in multimodal normality (thus remaining classes become normal, and in the following tables, MM is used for the abbreviation of multimodal normality.) We report averaged results on a different target class. For the training-test split, randomly selected 60% of the data in the normal class is used for training, and each half of the remaining data is used for the validation and normal test sets, respectively. All input features are normalized with z-score normalization.



(a)  $k = 1$  and  $k = 10$



(b)  $k = 1$  and  $k = 100$

**FIGURE 3.** Difference between gamma distributions of normal and abnormal data based on the increase in  $k$ : (a)  $k = 1$  and  $k = 10$ , (b)  $k = 1$  and  $k = 100$ . The normal and abnormal distributions of  $k = 100$  are more distinguishable than those of  $k = 10$ .

### B. NETWORK STRUCTURE AND EXPERIMENTAL SETUP

We build QAE with the same backbone network structure in [39] except for the final layer and loss function to provide the prediction of multiple quantiles  $\hat{x}_\tau$ . The target quantiles were set to  $\tau_l = 0.3$ ,  $\tau_m = 0.5$  and  $\tau_u = 0.7$ . Each  $Q_{enc}$  and  $Q_{dec}$  have 10 dense layers with the Leaky-ReLU activation function. The dimensions of  $z = Q_{enc}(x)$  listed in Table 2 are derived from the PCA results. The networks were trained using the Adam optimizer [60] and built using Pytorch [61].

We compared the results of the proposed methodology by following two steps for the ablation study. First, we compared the QAE with different anomaly score settings:  $A_r$ ,  $A_q$ ,  $A_{nq}$ . In doing so, MSE-based anomaly score from a basic AE trained to minimize MSE was used as the benchmark result. Second, we compared the results of  $A_{nqa}$  according to different numbers of iterations. Finally, we compared the AUROC score achieved via the proposed QAE with AA to other AD methodologies, including machine learning-based and recent deep learning-based methodologies reported in the works of [39], [40] (OCSVM to XAE); DN2 from [44]. In addition, we also compared the result of DSVDD [22] and DAGMM [23] models which are non-reconstruction

**TABLE 3.** AUROC (%) of QAE.

Dataset	AE		QAE	
	$A_r$	$A_r$	$A_q$	$A_{nq}$
MI-F	60.7	63.4 ± 1.0	47.4 ± 4.4	<b>74.6</b> ± 6.1
MI-V	89.7	73.0 ± 2.8	69.7 ± 3.6	<b>92.0</b> ± 3.4
EOPT	61.0	<b>76.2</b> ± 7.4	70.6 ± 5.7	61.6 ± 1.1
RARM	68.7	78.9 ± 0.9	78.1 ± 1.4	<b>80.4</b> ± 2.6
SNSR	97.9	96.5 ± 0.1	97.4 ± 0.9	<b>99.5</b> ± 0.0
SNSR <sub>MM</sub>	61.3	59.4 ± 0.6	62.4 ± 3.2	<b>67.3</b> ± 0.8

**TABLE 4.** AUROC (%) of QAE-AA.

Dataset	QAE-AA ( $A_{nqa}$ )					
	$N = 1$	2	3	4	5	6
MI-F	74.6 ± 6.1	79.4 ± 7.9	80.7 ± 8.0	<b>81.5</b> ± 7.2	81.2 ± 8.5	80.8 ± 9.1
MI-V	92.0 ± 3.4	<b>92.3</b> ± 4.1	92.0 ± 4.5	91.7 ± 4.9	91.3 ± 5.5	91.0 ± 5.7
EOPT	61.6 ± 1.1	62.0 ± 0.9	62.7 ± 0.5	62.9 ± 1.0	62.7 ± 1.2	<b>63.2</b> ± 1.1
RARM	80.4 ± 2.6	80.4 ± 3.0	<b>80.7</b> ± 2.9	<b>80.7</b> ± 2.8	80.6 ± 2.9	80.5 ± 2.7
SNSR	<b>99.5</b> ± 0.0	<b>99.5</b> ± 0.0	99.4 ± 0.1	99.3 ± 0.1	99.2 ± 0.1	99.1 ± 0.1
SNSR <sub>MM</sub>	67.3 ± 0.8	74.2 ± 1.0	<b>74.4</b> ± 0.9	74.1 ± 0.9	73.5 ± 0.8	72.8 ± 0.8

Bold : Best score,  $N = 1$  is identical to  $A_{nq}$  result.

error-based DAD methods. The above experimental setups are identical to those described in the work of [39], and the two models (DSVDD, DAGMM) are implemented with the same backbone networks as the QAE. Note that the experimental results described below are a summary of ten trials per model setting from different random seeds.

### C. EXPERIMENTAL RESULTS

Table 3 compares the mean and standard deviation of the AUROC results from the QAE with different scoring functions. Utilizing uncertainty terms without normalization ( $A_q$ ) induced a performance decrease compared to the  $A_r$  scores. But after the normalization, the AUROC of  $A_{nq}$  shows the best performance for all cases except EOPT. Compared to the  $A_r$  of basic AE, AD performance of  $A_{nq}$  is increased by 9% on average. Specifically, AUROC score is increased by 23% and 17% for MI-F and RARM, respectively.

Table 4 summarizes the AUROC of  $A_{nqa}$  (mean and standard deviation) by changing the number of iterations  $N$  from one to six ( $N = 1$  is equal to  $A_{nq}$ ). Although the increment is different for each dataset, the result shows that the AD performance can be improved by using additional reconstruction errors through iteration. Compared to the result without AA ( $A_{nq}$ ), MI-F and SNSR<sub>MM</sub> achieved 9% and 11% higher AUROC scores. Unlike the ideal case in Proposition 1, the

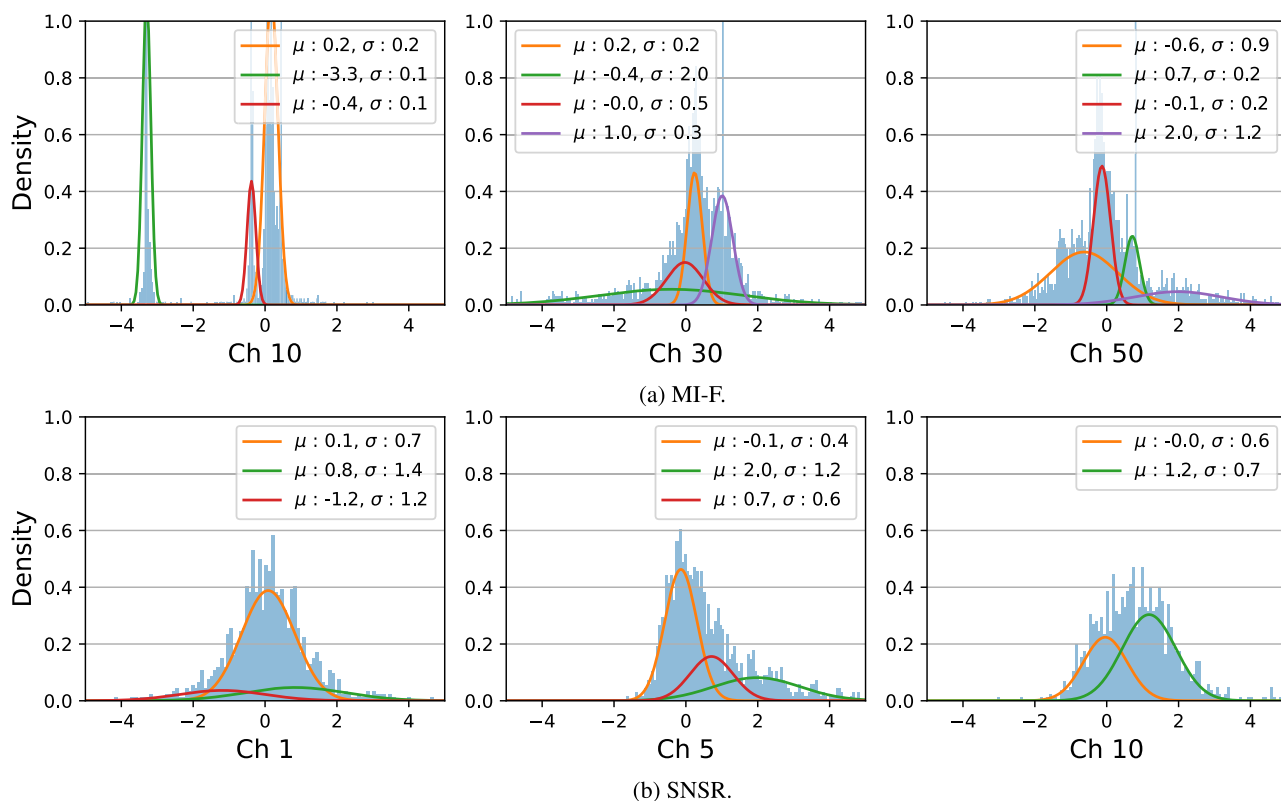


FIGURE 4. Error distribution with Gaussian mixture model fitting results: (a) MI-F, (b) SNSR.

TABLE 5. Comparison of AUROCs (%) to other benchmark results.

Dataset	Machine Learning				Deep Learning								
	OCSVM	ISOF	PCA	kPCA	AE	VAE	AAE	RAPP	XAE	DSVDD	DAGMM	DN2	QAE-AA
MI-F	77.7	<u>82.6</u>	54.6	<b>83.8</b>	60.7	59.1	63.2	70.5	65.3	73.4	70.8	77.3	80.7 ± 8.0
MI-V	84	83.9	87.4	88.8	89.7	84.5	89.5	<u>90.7</u>	89.6	66	74.6	62.5	<b>92.0</b> ± 4.5
EOPT	59.7	59.5	55.2	<b>69.3</b>	61	<u>67.5</u>	60.6	63.4	63.1	50.8	54.7	60.2	62.7 ± 0.5
RARM	74.9	75.2	58.7	73.4	68.7	58.7	66.4	68.6	67.3	75.6	<u>78</u>	-	<b>80.7</b> ± 2.9
SNSR	95.9	89.3	89.4	97	97.9	97.5	97.9	<u>99.1</u>	99	95.3	93.9	-	<b>99.4</b> ± 0.1
SNSR <sub>MM</sub>	52.4	52.6	53.6	58	61.3	60.8	61.2	<u>65.8</u>	-	50.6	51.7	-	<b>74.4</b> ± 0.9
Average	74.1	73.9	66.5	<u>78.4</u>	73.2	71.4	73.1	76.4	76.9	68.6	70.6	66.7	<b>81.7</b> ± 2.8

Bold : Best, Underline : Second best, Benchmark results are based on [39] (OCSVM~RAPP), [40] (XAE), and [44] (DN2)

iteration at which the performance is maximized differs for each data. But in most cases, significant performance changes occurred in the initial 2 ~ 3 iteration, and we compared the result of  $A_{nqa}$  with fixed iteration  $N = 3$  in Table 5.

Finally, Table 5 shows the performance comparison between the proposed methods and other AD methods. The characteristics of anomalies are different for each dataset, thus it is difficult to find a unique model that overwhelms the others for all cases. However, the proposed QAE-AA shows the best AUROC score in four out of six datasets; MI-V, RARM, SNSR, SNSR<sub>MM</sub>. For these datasets, AUROC increased 4.6% on average compared to the second-best result. In addition, in terms of the average score for all

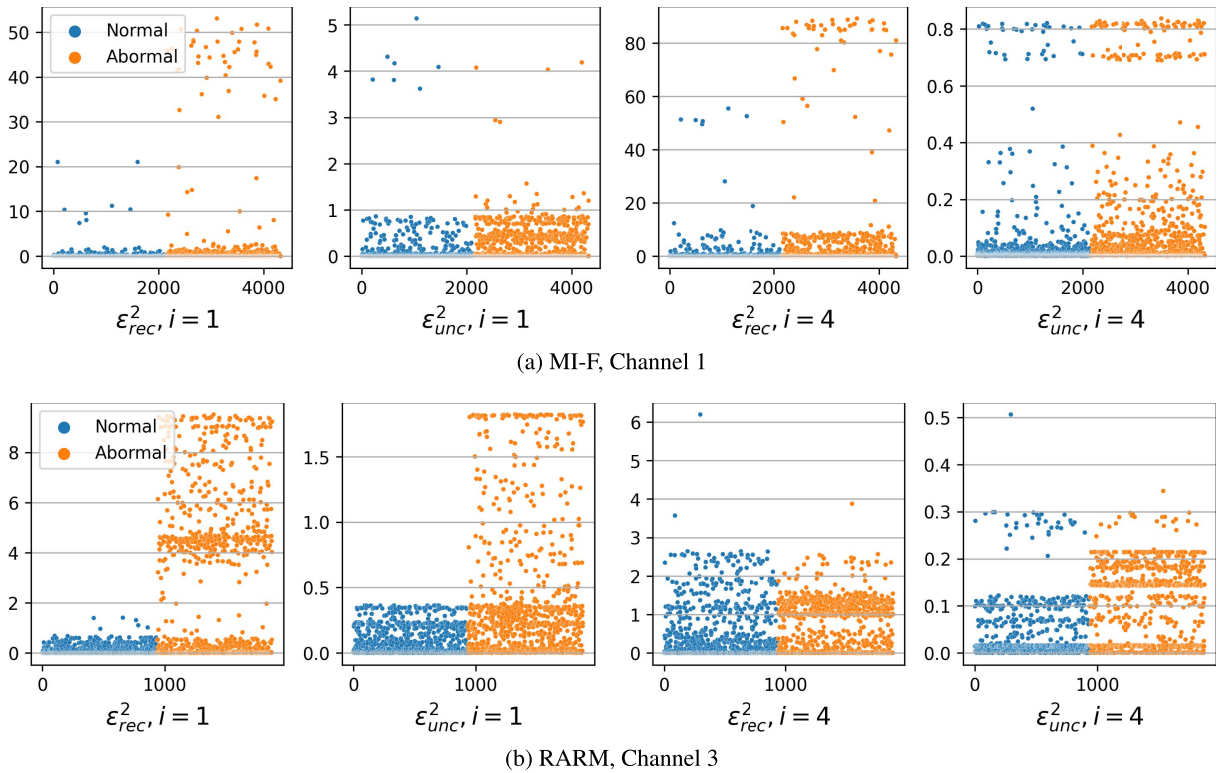
six datasets described at the bottom of the table, QAE-AA shows 4-23% higher AUROC; AUROC is increased 12% in overall.

This comparison result to various AD methodologies verifies the effectiveness of the proposed methodology, and shows that the concept of anomaly source diversification could be embodied by utilizing aleatoric uncertainty and iterative reconstruction errors.

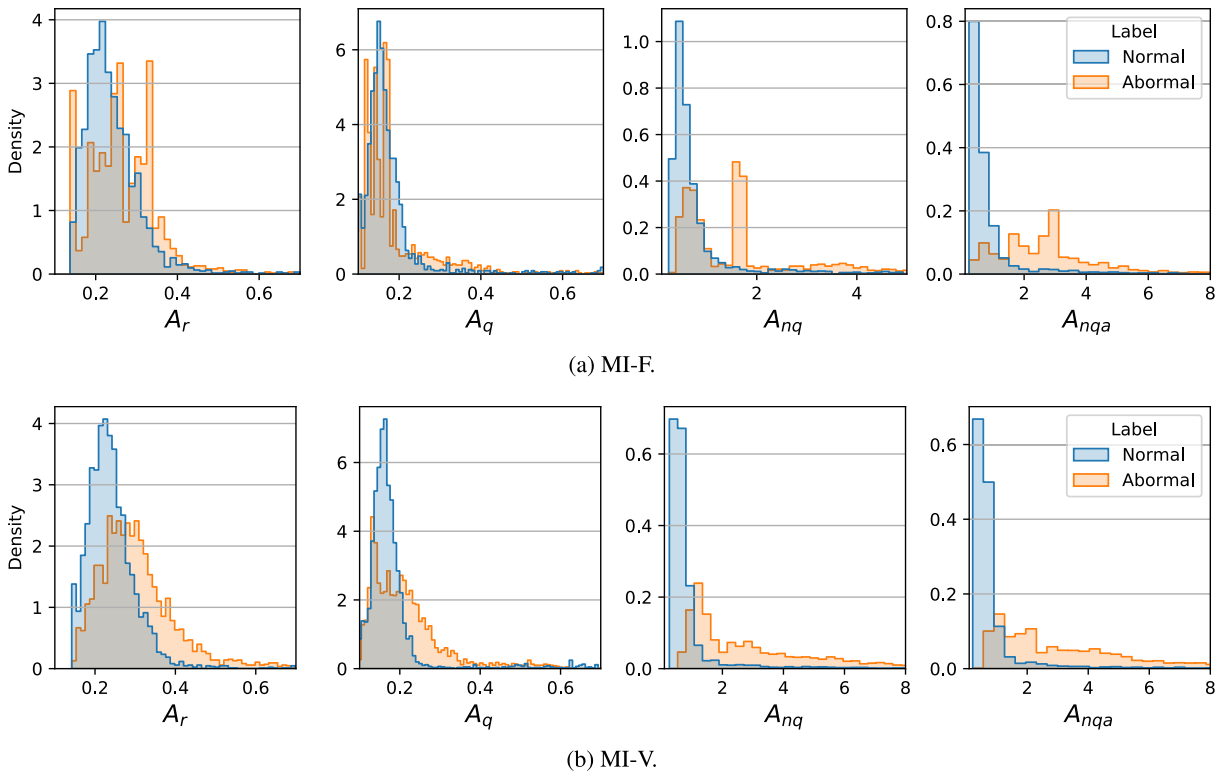
#### D. VISUALIZATION ON ANOMALY SCORE SOURCES

Fig. 5 shows examples of the QAE output (i.e., square of anomaly sources  $\epsilon_{rec}$  and  $\epsilon_{unc}$ ) on different target channels. The first two columns show the results of a single iteration





**FIGURE 5.** Examples of squared anomaly sources  $\epsilon_{rec}^2$  and  $\epsilon_{unc}^2$  on different target channel. x-axis indicates index of data and y-axis indicates value of corresponding anomaly source: (a) MI-F, Channel 1, (b) RARM, Channel 3.



**FIGURE 6.** Examples of anomaly score distribution of normal and abnormal data: (a) MI-F, (b) MI-V.

of QAE, whereas the right two columns show the results from the fourth iteration. Examples in Fig. 5 indicate that

both aleatoric uncertainty term and reconstruction error after iterations could be good sources for anomaly scoring. As can

be seen, abnormal samples are distributed in a higher error range. Thus, they can be classified according to the anomaly threshold.

### E. ANALYSIS ON SCORE DISTRIBUTION

Fig. 6 shows the changes in anomaly score distributions for MI datasets. Anomaly score distributions of normal and abnormal samples are illustrated in blue and orange histograms, respectively. Although the shapes do not perfectly match the ideal cases presented in Fig. 3, it is observed that the overlapping region between normal and abnormal distributions is reduced when applying QAE and AA. Therefore higher AUROC score can be achieved by the proposed methodology.

### V. CONCLUSION

In this research, we investigate the concept of anomaly score diversification, and propose QAE network and AA technique. The effectiveness of the proposed framework is verified with experiments on real-world datasets. Anomaly source diversification is inspired by the idea that diversifying error sources in the calculation of anomaly scores induces performance improvement in anomaly detection. We provide a theoretical background for this by showing that the distributions of mean square error anomaly scores on normal and abnormal become farther apart as the number of error sources under the assumption of Gaussian error increases.

In doing this, we propose a QAE that produces not only median but also quantiles to leverage aleatoric uncertainty as an additional error source for anomaly scoring. The outputs reconstructed from the abnormal samples are likely to have larger channel-wise uncertainty than that of normal samples likewise to the reconstruction errors. In addition, we introduce the AA technique that aggregates the errors via recursive reconstructions and then calculates anomaly score by using Mahalanobis distance. As the dimension of the errors is increased by the recursion, the difference between the anomaly score distributions of normal and abnormal samples becomes more apparent. The effectiveness of the proposed QAE-AA is verified with various datasets in the real world. QAE-AA obtained the highest AUROC score in four out of six datasets and achieved an average 4% to 23% higher AUROC score. These experimental results show that the proposed methodology can improve AD performance.

Recent works [44], [62] reported notable AD performance in some benchmark datasets by utilizing adversarial examples and time series AD setting, which can be additionally applied to the proposed QAE-AA framework. In this regard, our future research is moving toward overcoming the limitation and further improving AD performance. For example, AD performance on image data can be enhanced by additionally considering epistemic uncertainty or latent space errors.

### ACKNOWLEDGMENT

An earlier version of this paper was presented in part at the Workshop on AI for Design and Manufacturing (ADAM)

during the 36th AAAI Conference on Artificial Intelligence (AAAI-22), in February 2022 [63].

### REFERENCES

- [1] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.
- [2] A. V. Pawar, P. P. Kalavadekar, and S. N. Tambe, "A survey on outlier detection techniques for credit card fraud detection," *IOSR J. Comput. Eng.*, vol. 16, no. 2, pp. 44–48, 2014.
- [3] U. Porwal and S. Mukund, "Credit card fraud detection in e-commerce: An outlier detection approach," 2018, *arXiv:1811.02196*.
- [4] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *Int. J. Syst. Assurance Eng. Manage.*, vol. 8, no. 2, pp. 937–953, 2017.
- [5] D. Lee, "Anomaly detection in multivariate non-stationary time series for automatic DBMS diagnosis," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 412–419.
- [6] W. Aoudi, M. Iturbe, and M. Almgren, "Truth will out: Departure-based process-level detection of stealthy attacks on control systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2018, pp. 817–831.
- [7] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Comput.*, vol. 22, pp. 949–961, Sep. 2017.
- [8] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1577–1581.
- [9] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning deep representations of appearance and motion for anomalous event detection," 2015, *arXiv:1510.01553*.
- [10] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," *J. Imag.*, vol. 4, no. 2, p. 36, 2018.
- [11] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag. Cham, Switzerland: Springer*, 2017, pp. 146–157.
- [12] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 161–169.
- [13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [14] T. Kuzin and T. Borovicka, "Early failure detection for predictive maintenance of sensor parts," in *Proc. ITAT*. Princeton, NJ, USA: Citeseer, 2016, pp. 123–130.
- [15] P. Zhao, M. Kurihara, J. Tanaka, T. Noda, S. Chikuma, and T. Suzuki, "Advanced correlation-based anomaly detection method for predictive maintenance," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2017, pp. 78–83.
- [16] A. Beghi, L. Cecchinato, C. Corazzolo, M. Rampazzo, F. Simmini, and G. A. Susto, "A one-class SVM based tool for machine learning novelty detection in HVAC chiller systems," *IFAC Proc. Volumes*, vol. 47, no. 3, pp. 1953–1958, 2014.
- [17] J. E. Ball, D. T. Anderson, and C. S. Chan, Sr., "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, Sep. 2017, Art. no. 042609.
- [18] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2923–2960, 4th Quart., 2018.
- [19] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv:1901.03407*.
- [20] X. Gu, L. Akoglu, and A. Rinaldo, "Statistical analysis of nearest neighbor methods for anomaly detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32. Curran Associates, 2019, pp. 10923–10933. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/file/805163a0f0f128e473726ccda5f91bac-Paper.pdf>
- [21] G. Pang, C. Shen, L. Cao, and A. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2020.

- [22] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
- [23] 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 *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–19.
- [24] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.
- [25] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.
- [26] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 1997, pp. 583–588.
- [27] G. Pang, L. Cao, L. Chen, and H. Liu, "Learning representations of ultrahigh-dimensional data for random distance-based outlier detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2041–2050.
- [28] T. Pevný, "Loda: Lightweight on-line detector of anomalies," *Mach. Learn.*, vol. 102, no. 2, pp. 275–304, Feb. 2016.
- [29] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2006, pp. 287–296.
- [30] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5580–5590.
- [31] L. Zhu and N. Laptev, "Deep and confident prediction for time series at uber," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 103–110.
- [32] A. Legrand, H. Trannois, and A. Courmier, "Use of uncertainty with autoencoder neural networks for anomaly detection," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Jun. 2019, pp. 32–35.
- [33] C. Leibig, V. Alken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, Dec. 2017.
- [34] P. Seeböck, J. I. Orlando, T. Schlegl, S. M. Waldstein, H. Bogunović, S. Klimescha, G. Langs, and U. Schmidt-Erfurth, "Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 87–98, Jan. 2019.
- [35] A.-S. Collin and C. D. Vleeschouwer, "Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise," 2020, *arXiv:2008.12977*.
- [36] C. Xu and H. Chen, "A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data," *Energy Buildings*, vol. 215, May 2020, Art. no. 109864.
- [37] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture IE*, vol. 2, pp. 1–18, Dec. 2015.
- [38] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 622–637.
- [39] K. H. Kim, S. Shim, Y. Lim, J. Jeon, J. Choi, B. Kim, and A. S. Yoon, "RaPP: Novelty detection with reconstruction along projection pathway," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–14.
- [40] S. Y. Shin and H.-J. Kim, "Extended autoencoder for novelty detection with reconstruction along projection pathway," *Appl. Sci.*, vol. 10, no. 13, p. 4497, Jun. 2020.
- [41] S. Goyal, A. Raghunathan, M. Jain, H. V. Simhadri, and P. Jain, "DROCC: Deep robust one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 3711–3721.
- [42] L. Bergman, N. Cohen, and Y. Hoshen, "Deep nearest neighbor anomaly detection," 2020, *arXiv:2002.10445*.
- [43] X. Han, X. Chen, and L.-P. Liu, "GAN ensemble for anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 5, 2021, pp. 4090–4097.
- [44] A. Goodge, B. Hooi, S. K. Ng, and W. S. Ng, "LUNAR: Unifying local outlier detection methods via graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 1–9.
- [45] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. MLSDA 2nd Workshop Mach. Learn. Sensory Data Anal. (MLSDA)*, 2014, pp. 4–11.
- [46] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 665–674.
- [47] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Appl. Energy*, vol. 211, pp. 1123–1135, Feb. 2018.
- [48] H. Xu, Y. Feng, J. Chen, Z. Wang, H. Qiao, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, and D. Pei, "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications," in *Proc. World Wide Web Conf. World Wide Web (WWW)*, 2018, pp. 187–196.
- [49] J. Sun, X. Wang, N. Xiong, and J. Shao, "Learning sparse representation with variational auto-encoder for anomaly detection," *IEEE Access*, vol. 6, pp. 33353–33361, 2018.
- [50] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," 2018, *arXiv:1809.04758*.
- [51] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante, and K. Veeramachaneni, "TadGAN: Time series anomaly detection using generative adversarial networks," 2020, *arXiv:2009.07769*.
- [52] M. Abdar, F. Pourpanah, S. Hussain, D. Rezagadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," 2020, *arXiv:2011.06225*.
- [53] I. Steinwart and A. Christmann, "Estimating conditional quantiles with the help of the pinball loss," *Bernoulli*, vol. 17, no. 1, pp. 211–225, 2011.
- [54] N. Tagasovska and D. Lopez-Paz, "Single-model uncertainties for deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 6417–6428.
- [55] S. Ryu, S. Bae, J. Lee, and H. Kim, "Gaussian residual bidding based coalition for two-settlement renewable energy market," *IEEE Access*, vol. 6, pp. 43029–43038, 2018.
- [56] *CNC Milling Dataset*. Accessed: Aug. 17, 2021. [Online]. Available: <https://www.kaggle.com/shasun/tool-wear-detection-in-cnc-mill>
- [57] *High Storage System Data for Energy Optimization*. Accessed: Aug. 17, 2021. [Online]. Available: <https://www.kaggle.com/init-owl/high-storage-system-data-for-energy-optimization>
- [58] *ROS Anomaly Detector Package*. Accessed: May 27, 2021. [Online]. Available: [https://github.com/narayave/mh5\\_anomaly\\_detector](https://github.com/narayave/mh5_anomaly_detector)
- [59] *Dataset for Sensorless Drive Diagnosis Data Set*. Accessed: May 27, 2021. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/dataset+for+sensorless+drive+diagnosis>
- [60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [61] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035. [Online]. Available: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [62] T. Choi, D. Lee, Y. Jung, and H.-J. Choi, "Multivariate time-series anomaly detection using SeqVAE-CNN hybrid model," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2022, pp. 250–253.
- [63] H. Seo, S. Ryu, J. Yim, J. Seo, and Y. Yu, "Quantile autoencoder for anomaly detection," in *Proc. AAAI Workshop AI Design Manuf. (ADAM)*, 2021, pp. 1–6.



**SEUNGHYOUNG RYU** received the B.S., M.S., and Ph.D. degrees in electronic engineering from Sogang University, in 2014, 2016, and 2020, respectively. He is currently a Senior Researcher of the Artificial Intelligence Application & Strategy Team, Korea Atomic Energy Research Institute, South Korea. His research interests include applied artificial intelligence and energy ICT, specifically focused on deep learning for anomaly detection and multi-variate time-series forecasting.



intelligence and computer vision, specifically focused on deep learning application to object.

**JIYEON YIM** received the B.S. degree from the Department of Biomedical Engineering, Chonnam National University, in 2020. She is currently pursuing the M.S. degree with the Department of Nuclear Engineering, Ulsan National Institute of Science and Technology, South Korea. She is also a Research Student of the Artificial Intelligence Application & Strategy Team, Korea Atomic Energy Research Institute, South Korea. Her research interests include applied artificial



mization of musical instruments and nuclear reactor components. His current research interest includes the industrial application of deep learning.

**YONGGYUN YU** received the B.S. and Ph.D. degrees from the Department of Mechanical Engineering, KAIST, in 2001 and 2010, respectively. Then, he worked with the Mobile Harbor Center, KAIST, as a Postdoctoral Fellow and a Research Assistant Professor. He is currently a Principal Researcher and the Research Director of the Artificial Intelligence Application & Strategy Team, Korea Atomic Energy Research Institute, Daejeon, South Korea. His expertise covers design opti-



**JUNGHOON SEO** received the B.S. degree in electrical engineering and computer science from the Gwangju Institute of Science and Technology, in 2021. He is currently a Technical Lead of the Artificial Intelligence Research Institute, SI Analytics Inc, South Korea. His research interests include machine learning and computer vision, specifically focused on remote sensing and surveillance systems.



**HOGEOON SEO** received the B.S. degree in mechanical engineering and the Ph.D. degree in convergence mechanical engineering from Hanyang University, in 2013 and 2018, respectively. He is currently a Senior Researcher with the Artificial Intelligence Application & Strategy Team, Korea Atomic Energy Research Institute. His research interests include artificial intelligence for prognosis, nondestructive evaluation, and sensor fusion.

...