Semi-supervised Anomaly Detection Algorithm using Probabilistic Labeling (SAD-PL)

Kibae Lee¹, (Student Member, IEEE), Chong Hyun Lee¹, (Member, IEEE), Jongkil Lee²

¹Department of Ocean System Engineering, Jeju National University, Jeju 63243, South Korea
²Department of Mechanical Engineering Education, Andong National University, Andong 36729, South Korea

Corresponding author: Chong Hyun Lee (e-mail: chlonglee@jejun.ac.kr).

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development in Korea under the contract No. UD200011DD.

ABSTRACT To detect abnormal data via semi-supervised learning, unlabeled data are generally assumed to be normal data. This assumption, however, causes inevitable performance degradation when a small fraction of abnormal data is included in the unlabeled dataset. To overcome the degradation and to maintain stable detection performance, we propose a semi-supervised anomaly detection algorithm using probabilistic labeling (SAD-PL) for unlabeled data. The proposed SAD-PL is composed of two steps: (1) estimating local outlier factor (LOF) scores of latent vectors from both labeled and unlabeled data and (2) estimating labeling probability on the unlabeled data by using the prior missing probability of the labeled data via the Neyman-Pearson (NP) criterion. The SAD-PL runs iteratively by using the proposed complementary learning functions until the rate of label changes is lower than the predefined threshold. Experimental results reveal that the SAD-PL shows superior detection probability over the existing algorithms and stable performance regardless of the normal to abnormal data ratio in unlabeled data and the ratio of change variation of unlabeled data statistics to labeled data statistics.

INDEX TERMS Anomaly detection, Semi-supervised learning, Probabilistic labeling, Neyman-Pearson criterion and Local outlier factor

I. INTRODUCTION

Anomaly detection is used for detecting abnormal samples that deviate from the predefined normality [1]. It has various applications, such as in medicine, security, and manufacturing [1]. Further applications include intrusion detection in cybersecurity [2] – [4], industrial fault and damage detection in monitoring sensor data [5] – [7], and acoustic novelty detection for audio surveillance and underwater sonar systems [8] – [13]. Typical anomaly detection methods based on unsupervised learning assume that most of the samples are normal. The unsupervised approaches, primarily treated as a one class classification problem, learn features of normal samples [14] – [18]. Typical anomaly detection methods such as the one class support vector machine (OC-SVM) [14] and support vector data description (SVDD) [15] attempt to learn compact descriptions of normal samples. Recent deep learning approaches have shown outstanding performance by overcoming the problems with shallow learning on high-dimensional data [16] – [18]. Deep SVDD [16], a representative deep approach, trains a neural network, while minimizing the volume of a hypersphere that encloses normal samples in latent space. Most unsupervised approaches that are not trained on abnormalities have limited detection performance.

Some labeled data, as well as unlabeled data, may be utilized in real-world applications, and an especially small number of anomalous samples in labeled data can be used. Song et al. [19] and Akvay et al. [20] proposed semi-supervised anomaly detection models that use reliable normal samples in unlabeled data for training. However, since these models do not train the abnormalities like the unsupervised approaches, they have limited performance. Ruff et al. [21] proposed a deep semi-supervised anomaly detection (Deep SAD) that learns anomalous samples in labeled data. By assuming that most unlabeled samples are normal, the Deep SAD trains the normal data to be concentrated in the center of the latent space, and then trains labeled abnormal samples to move away from the center of the latent space. An unsupervised Outlier Exposure (OE) approach for learning OE data and unlabeled data has been proposed based on a similar assumption of semi-supervised anomaly detection [22]. The anomaly detection method based on unsupervised OE learning
trains a binary classifier on the OE data and the unlabeled normal and abnormal data. Hyper plane classification (HSC) [23] is a class classification algorithm based on unsupervised OE learning that uses the relative distance in latent space for training. The Deep SAD and HSC are advanced anomaly detection algorithms since they consider abnormal samples or OE data in the training process. However, their assumption of most unlabeled samples being normal inevitably causes performance degradation as the number of abnormal samples in the unlabeled dataset increases. To overcome the degradation, a new semi-supervised anomaly detection algorithm that can learn unlabeled normal and abnormal data efficiently is required.

In this paper, we propose an anomaly detection algorithm based on probabilistic normal or abnormal labeling for each sample in unlabeled data. The proposed algorithm, denoted as the semi-supervised anomaly detection algorithm using probabilistic labeling (SAD-PL), involves the two-step probabilistic labeling process: (1) computing the local outlier factor (LOF) score of latent vectors from both labeled and unlabeled data and (2) estimating the labeling probability on the unlabeled data by using the missing probability of labeled normal data via the Neyman-Pearson (NP) criterion. The SAD-PL runs until the labeling change rate becomes lower than the preset threshold.

Our paper is organized as follows: Section 2 describes the SAD-PL and evaluates the performance using toy data. In Section 3, the experimental results on the image dataset are described in comparison with existing algorithms. Conclusions are presented in Section 4.

II. SAD-PL

In this section, we introduce the SAD-PL based on semi-supervised learning with probabilistic labeling. Figure 1 shows the proposed algorithm along with the existing semi-supervised algorithms of the Deep SAD and HSC. The proposed SAD-PL uses feature representations \( \phi(x) \) through autoencoder pretraining and learns the encoded normal samples to close centroid \( c \) in latent space without the decoding network. Then, the SAD-PL is trained according to the proposed probabilistic labeling, which uses the LOF score \( s = \text{LOF}_k(\phi(x)) \). The LOF scoring the density based on the relative distance between \( k \) neighbor samples is known to show robust performance in the multimodal normality case [24]. To detect a small number of group anomaly samples, the SAD-PL sets \( k \) large enough to cover the relative distance between normal and abnormal samples. The probability used for labeling is computed by using the missing probability of the labeled normal data \( P_M \) as

\[
P_M = \int_\eta^\infty p(s|H_0)ds \tag{1}
\]

where \( p(s|H_0) \) denotes the probability density function of \( s = \text{LOF}_k(\phi(x)) \) obtained from labeled normal data \( (H_0) \). \( \eta \) denotes the threshold for the given \( P_M \). Note that \( \eta \) varies as the SAD-PL. These changes in \( \eta \) cause the probabilistic label \( y(t) \) to change each training epoch \( t \). However, labels \( y_i(t), i \leq n \) for \( n \) labeled samples \( x_i \) are fixed as follows:

\[
y_i(t) = \begin{cases} 1 & \text{if } y_i(t) > \eta \text{ and } y_i(t) \neq 1 \text{ for previous epochs} \\ 0 & \text{if } y_i(t) < \eta \text{ and } y_i(t) \neq 0 \text{ for previous epochs} \\ 1 \text{ or } 0 & \text{if } y_i(t) = \eta \text{ or } y_i(t) \text{ is unchanged} \\ -1 \text{ or } 1 & \text{if } y_i(t) \text{ is unknown} \end{cases}
\]

FIGURE 1. Overview of the different approaches to semi-supervised anomaly detection. Left: Semi-supervised anomaly detection models aim to separate normal and abnormal samples in latent space. Deep SAD learns normal and abnormal samples with asymmetric objectives. HSC is trained by complementary losses based on radial basis function. Deep SAD and HSC is trained assuming that the samples in unlabeled data are normal. SAD-PL learns both labeled and unlabeled data with complementary objectives depending on labeling probability.
\[ y_t(t) = \begin{cases} 1, & \forall t \text{ for normal} \\ 0, & \forall t \text{ for abnormal} \end{cases} \] (2)

For unlabeled data \( x_u \), the labels \( y_u(t) \), \( u > n \) are defined as follows:
\[
y_u(t) = \begin{cases} 1 - P_M, & s_u \leq \eta \\ P_M, & s_u > \eta \end{cases}
\] (3)

where \( s_u \) represents the LOF score of \( x_u \). According to \( P_M \) determined by the NP criterion, \( y_u(t) \) has a corresponding probability and consequently implies the probabilistic label for the unlabeled \( x_u \). For network \( \phi \) training, the SAD-PL uses datasets \( \{ x_i, y_i(t), l \leq n \} \) and \( \{ x_u, y_u(t), u > n \} \), which are composed of \( n \) labeled and \( (m - n) \) unlabeled data from the corresponding probabilistic labels estimated in (2) and (3). The resultant objective function of the SAD-PL can be described as follows:
\[
\min_w \frac{1}{m} \sum_{i=1}^{m} y_i(t) d(\phi(x_i)) + \left(1 - y_i(t)\right) \left(1 - d(\phi(x_i))\right) + \frac{1}{2} \sum_{j=1}^{J} \|W^j\|_F^2
\] (4)

where \( W^j \) is the weights of layer \( j \in \{1, \cdots, J\} \), \( \| \cdot \|_F \) denotes the Frobenius norm and \( d(\phi(x_i)) \) is the function of the distance from \( c \) using Geman-McClure loss and defined as follows:
\[
d(\phi(x_i)) = \frac{\|\phi(x_i) - c\|^2}{\|\phi(x_i) - c\|^2 + 1}
\] (5)

The proposed SAD-PL objective function consists of a weight decay regularizer on \( W^j \) multiplied by hyperparameter \( \lambda > 0 \) for preventing overfitting and two complementary learning functions of symmetrical losses for normal and abnormal samples, which are multiplied by complementary probabilistic labels \( y_i(t) \). The SAD-PL learns the normal samples closer to \( c \) and the abnormal samples away from \( c \) for the labeled data. For unlabeled data training, the network uses complementary losses with probability labeling \( y_u(t) \). To avoid a trivial solution of \( W = 0 \), the bias is not updated during the learning process. By adopting the Geman-McClure loss in (5), the SAD-PL can prevent divergence in the learning process with a limited loss of 0 to 1 and obtain robust stability for mislabeled data [25]. Note that the SAD-PL object function has both regression and classification properties since it learns by computing distance loss based on the estimated probabilistic labels \( y_u(t) \) for unlabeled data.

**Soft Labeling:** To find the optimum \( P_M \) for unlabeled data labeling, the SAD-PL adopts the ensemble networks \( [\phi_1, \cdots, \phi_Q] \), where \( \phi_q \) is the model trained with \( P_M(q) \). To estimate the optimum \( P_M \), the SAD-PL uses the fact that the LOF difference between normal and abnormal samples in well-trained networks becomes relatively larger than that in ill-trained networks. To determine the optimum \( P_M(q) \), we compute the LOF difference \( \Delta s(q) \) for \( \phi_q \) as follows:
\[
\Delta s(q) = |\tilde{s}(q)_0 - \tilde{s}(q)_1|
\] (6)

where \( \tilde{s}(q)_0 \) and \( \tilde{s}(q)_1 \) represent the average LOFs of the labeled normal and abnormal samples, respectively. Then, from \( Q \) differences \( [\Delta s(1), \cdots, \Delta s(Q)] \), the optimum \( P_M(q) \) is determined by finding the maximum \( \Delta s(q) \) at a given epoch \( T \). The SAD-PL learns the model by using the threshold \( \eta_s \).
obtained from $P_M(\omega)$ and the estimated probabilistic labeling procedure described above.

**Hard Labeling:** By setting $P_M(\omega) = 0$, the SAD-PL can reduce the computational cost of memory and time. This version of the SAD-PL is known as hard labeling and learns the network by using the threshold denoted as $\eta_H$ in Figure 1.

The proposed SAD-PL runs until the labeling change rate $\Delta y_u(t)$ at iteration $t$ and is lower than the preset threshold $\epsilon$. The $\Delta y_u(t)$ is computed as

$$\Delta y_u(t) = \frac{1}{m - n} \sum_{i=n+1}^{m} \left| \frac{y_u(t) - y_u(t-1)}{1 - 2P_M(\omega)} \right|$$

(7)

Algorithm 1 shows the overall learning procedure of the SAD-PL.

The proposed SAD-PL is related to the Deep SAD algorithm. Setting $y_i(t)$ as 1 for the normal samples, $y_i(t) = -1$ for the abnormal samples, $y_u(t) = 1 \forall t$, setting $d(\phi(x_i)) = ||\phi(x_i) - c||^2$ and $d(\phi(x_i))$ to $y_i(t)$ power in Equation (4), gives the Deep SAD objective function as follows [21]:

$$\min_{\mathbf{W}} \frac{1}{m} \sum_{i=1}^{m} (||\phi(x_i) - c||^2)^{y_i(t)} + \frac{1}{m} \sum_{u=n+1}^{m} ||\phi(x_u) - c||^2 + \frac{\lambda}{2} \sum_{j=1}^{J} ||W_j||^2_F$$

(8)

where the hyperparameter $\zeta$ controls the balance of learning between labeled and unlabeled terms.

Additionally, the proposed SAD-PL is related to the HSC model with the objective function of cross entropy for relative distance in latent space [23]. By taking the negative and logarithmic distance losses in Equation (4) and replacing $d(\phi(x_i))$ with the radial basis function $h(\phi(x_i)) = \exp[-(\sqrt{||\phi(x)||^2} + 1 - 1)]$ gives the HSC objective function as follows [23]:

$$\min_{\mathbf{W}} -\frac{1}{m} \sum_{i=1}^{m} y_i(t) \log h(\phi(x_i)) + (1 - y_i(t)) \log[1 - h(\phi(x_i))] + \frac{\lambda}{2} \sum_{j=1}^{J} ||W_j||^2_F$$

(9)

where $y_i(t) = 1 \forall t, i \in \{n + 1, \cdots, m\}$ for unlabeled data.

**A. COMPARISON OF ANOMALY DETECTION MODELS**

We compare the anomaly detection models described above with toy examples. The training data consist of a total of 10000 samples, of which 10% are labeled data and 90% are unlabeled data. Five percent of labeled data and 1% of unlabeled data consist of abnormal samples. We generate the normal samples in the training data as two-dimensional big moon and small moon patterns. We also add Gaussian noise of variance 0.2 and 0.25 to the labeled and unlabeled data, respectively. The abnormal samples in labeled data are located to be clearly distinguished from the normal samples. On the other hand, the abnormal samples in unlabeled data are located adjacent to the boundaries of the normal samples. The test data consist of 1000 normal and abnormal samples. The normal samples contain Gaussian noise with a variance of 0.3 in big moon and small moon patterns. The abnormal samples are generated with a uniform distribution.

The models for comparison use the same encoding network of the pretrained autoencoder. The encoding network consists of two hidden layers of 100 nodes, followed by ELU activations that represent a two-dimensional input in a two-dimensional latent space. For autoencoder pretraining, we employ the above architectures for the encoding networks and...
then construct decoding networks symmetrically. We set $\zeta = 1$ for the Deep SAD, and $k = 100$ and $\varepsilon = 0.0001$ for the SAD-PL. We use $P_M = 0.04$ obtained from soft labeling training. For all models, we set $\lambda = 10^{-6}$ and use the Adam optimizer with a learning rate of $10^{-5}$. In addition, we train all models except the SAD-PL in epoch 300. Figure 2 shows the decision boundaries with training data and the test AUC (area under the receiver operating characteristic curve) of anomaly detection models. The decision boundaries in Figure 2 are represented with an upper bound on 10% of the anomaly score normalized via min-max scaling. The anomaly score is computed $\|\phi(x) - c\|^2$ for the Deep SAD and the SAD-PL and $\|\phi(x)\|^2$ for HSC. In (a) and (c) of Figure 2, the Deep SAD and HSC create the decision boundaries along the perimeter of the large moon and small moon patterns because the unlabeled data are assumed to be normal. Therefore, the abnormal samples close to the normal samples in unlabeled data are located within the decision boundaries, and the decision boundaries are created wider in a region where the abnormal sample in the labeled data does not exist. However, the SAD-PL represents the tight decision boundaries with complementary learning for the labeling probability of the unlabeled data in (e) of Figure 2. In particular, the SAD-PL has higher anomaly scores than the existing methods in areas separated inside the large moon and small moon patterns. In (b) of Figure 2, the supervised deep SAD is trained with only normal samples in unlabeled data, since the labels are not used on the unlabeled term in the objective function. For this reason, the supervised Deep SAD represents compact decision boundaries compared to the semi-supervised Deep SAD. However, the supervised Deep SAD, which learns relatively few abnormal samples, creates wider decision boundaries than the SAD-PL. (d) and (f) in Figure 2 show the decision boundaries and test AUC for the supervised HSC and SAD-PL learning the unlabeled data with labels. The supervised HSC and SAD-PL have extremely tight decision boundaries and high AUCs of 99.25% and 99.28%, respectively. These results show improved AUCs of more than 4.02% compared to the semi-supervised HSC, but only up to a 0.62% difference in AUCs compared to the semi-supervised SAD-PL.

Figure 3 shows the training AUCs of the comparative models according to epoch and the $\Delta y_{\eta}(t)$ in the SAD-PL.

![Figure 3](image)

**FIGURE 3.** Training AUC of anomaly detection models according to epoch and change rate of probabilistic labels in SAD-PL model.

learning. In Figure 3, the HSC, which learns distance based on a radial basis function, achieves higher learning efficiency than the Deep SAD. However, the HSC and the Deep SAD, assuming the unlabeled data are normal, show slower learning caused by adversarial learning between the abnormal samples in labeled and unlabeled data. However, the SAD-PL achieves high learning efficiency with complementary learning based on probabilistic labeling of unlabeled data. We can also predict the completion of learning as $\Delta y_{\eta}(t)$ converges to zero as the epoch increases.

**B. CHARACTERISTICS OF PROBABILISTIC LABELING**

The SAD-PL sets the labeling probability on unlabeled data with $P_M$ of the labeled data via the NP criterion. It determines $\eta$ by Equation (1) and sets the labeling probability by (3). Typically, unlabeled data consist of a large number of samples with various statistics, such as variance, rather than labeled data. Therefore, the NP condition may not be met on unlabeled data for the $\eta$ set in labeled data. The proposed algorithm repeats the probabilistic labeling in an iteration of learning to change normal samples in labeled and unlabeled data into a similar probability distribution of LOF scores. In this learning process, the missing probability for the normal unlabeled samples converges to the predefined $P_M$. Figure 4 shows histograms of LOF scores for the normal samples in an iteration of training. In (a) and (b) of Figure 4, the probability distributions of LOF scores for normal samples in labeled and unlabeled data become similar with each iteration of training. As a result, the missing probability for the normal unlabeled samples varies from 20.25% to 4.02%, close to the predefined $P_M = 0.04$. Figure 5 shows that the labeling accuracy for the normal unlabeled data, which is mostly composed of normal samples, reaches approximately 96%.

To verify the soft probabilistic labeling, we compute the difference in LOF scores $\Delta s(q)$ according to epoch. Figure 6

![Figure 4](image)

**FIGURE 4.** Histogram for LOF score according to epochs (a) Normal samples of labeled data (b) Normal samples of unlabeled data.
filters, followed by a final dense layer of 128 units. For the pretraining autoencoder, we employ identical encoding networks and then construct the decoding networks symmetrically, where we replace max-pooling with simple upsampling and convolutions with deconvolutions. We use a batch size of 200 and set $\lambda = 10^{-6}$. We also use the Adam optimizer with a learning rate of $10^{-5}$. For experiments on the Deep SVDD, we use the one-class Deep SVDD model [16]. We run all experiments for $\xi \in \{0.01, 0.1, 1, 0, 100\}$ of the Deep SAD and show the best results. The SAD-PL is evaluated via two separate hard and soft labels. We set $k = 200$ and complete learning when $\Delta W_i(t)$ is less than $\epsilon = 0.001$. In soft labeling, we set $Q = 101$ by setting $P_M = 0$ to $P_M = 0.2$ with an $\delta = 0.002$ interval and select the $P_M(a)$ in which $\Delta s$ is maximum in $T = 10$. We train the deep models for 300 epochs. The image data used in the experiments are normalized through min-max scaling.

We use a typical one vs. rest evaluation method on the MNIST and CIFAR10 datasets [29]. On MNIST and CIFAR 10, we set the ten classes to be normal classes and let the remaining nine classes represent anomalies. We use the original training and test data. In the training data, we constitute most of the data as the normal class and replace a small amount with the data from the abnormal class according to the experimental scenario. We also divide the training data into labeled and unlabeled data, while organizing the abnormal samples in labeled data from a single anomalous class. However, the abnormal samples in unlabeled data equally contain samples of all anomalous classes. The class for abnormal samples in labeled data is randomly determined in each experiment. This gives training set sizes of approximately 6000 for MNIST and 5000 for CIFAR10. Both test sets have 10000 samples, including samples from the nine anomalous classes for each setup. On MNIST-C, we set original images to be normal and corrupted images to be abnormal. We use pre-configured training and test data. We also make the training data into most normal samples and a few abnormal samples according to the scenario. The training data are divided into labeled and unlabeled data. We organize the abnormal samples in labeled data using corrupted images of the same type. The abnormal samples in unlabeled data include all kinds of corrupted images equally. The type of corrupted images for abnormal samples in labeled data is randomly determined in each experiment. This configuration gives a training set size of approximately 60000 and a test set size of 160000 for MNIST-C.

We use 5% as labeled data and 95% as unlabeled data in the training data. We constituted 98% normal samples and 2% abnormal samples in the labeled data, along with 99% normal samples and 1% abnormal samples in the unlabeled data. We present the evaluation results for the models with an average AUC of 30 times. Table 1 shows the evaluation results on MNIST dataset. In Table 1, the SAD-PL (Hard) and the SAD-PL (Soft) indicate the evaluation results for hard and soft labeling, respectively. The SAD-PL achieves a better
performance than existing unsupervised and semi-supervised anomaly detection models with an average AUC of at least 93.65%. The SAD-PL also represents an average AUC that is at least 2.69% higher than the supervised deep SAD by training abnormal samples in unlabeled data. The semi-supervised SAD-PL has an improved average AUC of 1.97% via soft labeling, which is only 3.86% lower than that of the supervised SAD-PL. Additionally, in the experiment in which digit ‘0’ is a normal class on MNIST, the SAD-PL is trained

with \( P_M = 0.04 \) determined by soft labeling. Figure 7 shows the labeling accuracy and \( \Delta y_M(t) \) in the iteration of the SAD-PL training with soft labeling. As shown in Figure 7, the labeling accuracy reaches approximately 96%, corresponding to \( P_M = 0.04 \), and \( \Delta y_M(t) \) converges to 0.1% or less as the epoch increases.

Tables 2 and 3 show the evaluation results for anomaly detection models on CIFAR10 and MNIST-C, respectively. In Table 2, the SAD-PL proposed achieves a better detection
performance than existing unsupervised and semi-supervised methods with an average AUC of at least 64.37%. The SAD-PL with hard labeling also represents an average AUC that is 1.69% higher than the supervised deep SAD. We can see that the SAD-PL with soft labeling has a 5.22% improvement in the average AUC compared to the SAD-PL with hard labeling. This result represents an average AUC that is only 2.43% lower than the supervised SAD-PL. In Table 3, the SAD-PL similarly achieves a higher performance than the existing unsupervised and semi-supervised models, with an average AUC of at least 90.05% in the experiment using the MNIST-C dataset. The SAD-PL also represents an average AUC that is at least 5.22% higher than the supervised deep SAD. The proposed method with soft labeling has a 2.38% improvement in the average AUC compared to the SAD-PL with hard labeling.

Next, we investigate the effect of including labeled data during training on the MNIST, CIFAR10 and MNIST-C datasets by increasing the ratio of labeled training data from 0% to 10% and presenting the averaged AUCs of 30 times. We compute the average AUCs according to the one vs. rest evaluation method on the MNIST and CIFAR10 datasets. The ratio of abnormal samples in labeled and unlabeled data is maintained at 5% and 1%, as in the previous experiments. Figure 8 shows variations of performance in the average AUCs for the semi-supervised models according to the ratio of labeled data during training. The HSC, a classification model, shows high improvement in the average AUC as the labeled data increase. In particular, HSC presents a higher average AUC than the Deep SAD when the ratio of the labeled data is 5% or more. The Deep SAD, a regression model that is trained with distance loss, shows robust performance, even with a small amount of labeled data. However, the SAD-PL, having both properties of regression and classification by learning with complementary objectives, represents a high average AUC compared to the existing HSC and the Deep SAD. The SAD-PL with hard labeling has a similar average AUC to the HSC average as the ratio of the labeled data increases. Note that the SAD-PL with soft labeling represents a remarkably high average AUC, even with a high proportion of labeled data during training.

Similar to the experiment above, we investigate the effect
of including abnormal samples in the unlabeled data during the training with the MNIST, CIFAR10 and MNIST-C datasets. To do this, we increase the anomaly ratio, which is the proportion of abnormal samples in unlabeled training data, from 0% to 10% and represent the average AUCs of 30 times. We also use the one vs. rest evaluation method on the MNIST and CIFAR10 datasets. For the experiment, we keep the labeled data at 95% of the training data and include 98% and 2% of normal and abnormal samples, respectively. Figure 9 shows the variations of performance for the average AUCs of the unsupervised and semi-supervised models according to the anomaly ratio in unlabeled data during training. In Figure 9, the unsupervised and semi-supervised approaches assume that the unlabeled data consisting of only normal samples represent performance degradation in the average AUCs as the anomaly ratio in unlabeled data increases. However, the semi-supervised models have a high performance in the average AUC compared to the unsupervised models by learning abnormal samples in labeled data. However, the SAD-PL represents the improvement in performance for the average AUCs as the anomaly ratio in unlabeled data increases via probabilistic labeling. This performance variation appears in both hard and soft labeling.

VI. CONCLUSION

Unlike existing semi-supervised anomaly detection algorithms, which are trained by assuming that most of the samples in unlabeled data are normal, we propose the SAD-PL, which can be applied when abnormal samples are included in unlabeled data. The proposed SAD-PL uses LOF scores obtained from both labeled and unlabeled data and then estimates the labeling probability on the unlabeled data by using the LOF scores. Because of probabilistic labeling and complementary objective function, the SAD-PL has properties of regression and classification. Through experiments, we show that the SAD-PL presents a higher performance in the average AUCs, displays tighter decision boundaries and achieves higher learning efficiency than the existing algorithms. Additionally, the SAD-PL shows an improved performance in the average AUCs as the abnormal data ratio in unlabeled data increases, whereas the existing algorithms show performance degradation. Therefore, the SAD-PL can be a good candidate for providing stable detection performance, regardless of the existence of abnormal samples in unlabeled data.

REFERENCES

[1] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Monatavon, W. Samek, M. Kloft, T. G. Dietterich, and K. R. Müller, “A unifying review of deep and shallow anomaly detection,” Proceedings of the IEEE, vol. 109, no. 5, pp. 756-795, Feb. 2021.

[2] H. Liao, C. H. R. Lin, Y. C. Lin, and K. Y. Tung, “Intrusion detection system: A comprehensive review,” Journal of Network and Computer Applications, vol. 36, pp. 16-24, Jan. 2013.

[3] M. Ahmed, A. N. Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, vol. 60, pp. 19-31, Jan. 2016.

[4] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, “Machine learning and deep learning methods for cybersecurity,” IEEE Access, vol. 6, pp. 35365-35381, May. 2018.

[5] J. Rabatel, S. Bringay, and P. Poncelet, “Anomaly detection in monitoring sensor data for preventive maintenance,” Expert Systems with Applications, vol. 38, no. 6, pp. 7003-7015, June. 2011.

[6] L. Martí, N. Sanchez-Pi, J. M. Molina, and A. C. B. García, “Anomaly detection based on sensor data in petroleum industry applications,” Sensors, vol. 15, no. 2, Jan. 2015, to be published. DOI: 10.3390/s150202774.

[7] D. Ramotsoela, A. Abu-Mahfouz, and G. Hancke, “A survey of anomaly detection in industrial wireless sensor networks with critical water system infrastructure as a case study,” Sensors, vol. 18, no. 8, Aug. 2018, to be published. DOI: 10.3390/s18082491.

[8] A. Rababou, M. Davy, S. Rossignol, and N. Ellouze, “Using one-class SVMs and wavelets for audio surveillance,” IEEE Transactions on Forensics and Security, vol. 3, no. 4, pp. 763-775, Dec. 2008.

[9] H. Lim, J. Park, K. Lee, and Y. Han, “Rare sound detection using 1D convolutional recurrent neural networks,” in proc. Workshop Detection Classification Acoustic Scenes Events, 2017, pp. 80-84.

[10] K. Lee, and C. H. Lee, “Abnormal signal detection based on parallel autoencoders,” The Journal of the Acoustical Society of Korea, vol. 40, no. 4, pp. 337-346, Jul. 2021.

[11] H. Yang, K. Lee, Y. Choo, and K. Kim, “Underwater acoustic research trends with machine learning: general background,” Journal of Ocean Engineering and Technology, vol. 34, no. 2, pp. 147-154, April. 2020.

[12] H. Yang, K. Lee, Y. Choo, K. Kim, “Underwater acoustic research trends with machine learning: passive sonar applications,” Journal of Ocean Engineering and Technology, vol. 34, no. 3, pp. 227-236, June. 2020.

[13] H. Yang, S. H. Byun, K. Lee, Y. Choo, and K. Kim, “Underwater acoustic research trends with machine learning: active sonar applications,” Journal of Ocean Engineering and Technology, vol. 34, no. 4, pp. 277-284, Aug. 2020.

[14] B. Schölkopf, R. Williamson, A. Smola, J. S. Taylor, and J. Platt, “Support vector method for novelty detection,” in Proc. NIPS 1999, Cambridge, MA, USA, 1999, pp. 582-588.

[15] D. M. Tax, and R. P. Duin, “Support vector data description,” Machine learning, vol. 59, no. 1, pp. 45-66, Jan. 2004.

[16] L. Ruff, R. A. Vandermeulen, N. Gürölt, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, “Deep one-class classification,” in Proc. ICML 2018, Stockholm, Sweden, 2018, pp. 4393-4402.

[17] S. Ishtiaq, and R. E. Yaniv, “Deep anomaly detection using geometric transformations,” in proc. NIPS 2019, Montreal, Canada, 2018, pp. 9781-9791.

[18] P. Perera, and V. M. Patel, “Learning deep features for one-class classification,” IEEE Transactions on Image Processing, vol. 28, no. 11, pp. 5450-5463, May. 2019.

[19] H. Song, Z. Jiang, A. Men, and B. Yang, “A hybrid semi-supervised anomaly detection model for high-dimensional data,” Computational Intelligence and Neuroscience, vol. 2017, Nov. 2017.

[20] S. Akvay, A. A. Abarghoushel, and T. P. Breckon, “GANomaly: semi-supervised anomaly detection via adversarial training,” in proc. ACCV 2018, Perth, Australia, 2018, pp. 622-637.

[21] L. Ruff, R. A. Vandermeulen, N. Gürölt, A. Binder, E. Müller, K. R. Müeller, and M. Kloft, “Deep semi-supervised anomaly detection,” in proc. ICLR 2020, Addis Ababa, Ethiopia, 2020.

[22] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” in proc. ICLR 2019, New Orleans, LA, USA, 2019.

[23] L. Ruff, R. A. Vandermeulen, B. J. Franks, K. R. Müller, and M. Kloft, (2021, July). Rethinking assumptions in deep anomaly detection. Presented at ICML 2021 workshop. [Online]. Available: https://sites.google.com/view/udlworkshop2021/accepeted-papers/UDL2021-paper-01.pdf

[24] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers,” in proc. ACM SIGMOD 2000, New York, NY, USA, 2000.

[25] J. T. Barron, “A general and adaptive robust loss function,” in proc. IEEE/CVF CVPR 2019, Long Beach, CA, USA, 2019.
[26] Y. Lecun, C. Cortes, and C. Burges, “MNIST handwritten digit database,” AT&T Labs, June, 28, 2010. [Online]. Available: https://yann.lecun.com/exdb/mnist/

[27] A. Krizhevsky, “Learning multiple layers of features from tiny images,” April, 8, 2009. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.9220 &rep=rep1&type=pdf

[28] N. Mu, and J. Gilmer. (2019, June). MNIST-C: robustness benchmark for computer vision. Presented at ICML 2019 workshop. [Online]. Available: http://www.gatsby.ucl.ac.uk/~balaji/udl2019/accepted-papers/UDL2019-paper-37.pdf

[29] A. F. Emmott, S. Das, T. Dietterich, A. Fern, and W. K. Wong, “Systematic construction of anomaly detection benchmarks from real data,” in proc. ACM SIGKDD workshop 2013, Chicago, IL, USA, 2013.

Kibae Lee (Student Member, IEEE) received the B.S. degree and M.S. degree in Ocean System Engineering from Jeju National University, in 2016 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Department of Ocean System Engineering, Jeju National University. His research interests include machine learning, acoustic signal processing and sonar system.

Chong Hyun Lee (Member, IEEE) received the B.S. degree in electronic engineering from Hanyang University, in 1985, the M.S. degree from Michigan Technological University, in 1987, and the Ph.D. degree in electronic engineering from Korea Institute of Science and Technology (KAIST), in 2002. He is currently working as a professor with the Department of Ocean System Engineering, Jeju National University. His research interests include machine learning, sonar signal processing and smart RF sensor design.

Jongkil Lee received the B. S. degree in mechanical engineering from Pusan National University, in 1984, the M.S. degree from the University of Utah, in 1990, and the Ph.D. degree in mechanical engineering from the University of Utah in 1993. He is currently working as a professor with the Department of Mechanical Engineering Education, Andong National University. His research interests include underwater acoustics, noise and vibration, fiber optic sensor, dynamics, energy harvesting system, etc.