Cross-Domain Fault Diagnosis of Rotating Machinery Using Discriminative Feature Attention Network

GYE-BONG JANG, (Member, IEEE), JIN-YOUNG KIM, (Member, IEEE), AND SUNG-BAE CHO, (Senior Member, IEEE)
Department of Computer Science, Yonsei University, Seoul 03722, South Korea
Corresponding author: Sung-Bae Cho (sbcho@cs.yonsei.ac.kr)
This work was supported in part by the Institute for Information and Communication Technology Promotion (IITP) Grant by the Korean Government through the Ministry of Science and ICT (MSIT) [AI Graduate School Program (Yonsei University)] under Grant 2020-0-01361 and through Doosan Infracore, Inc., Seoul, South Korea.

ABSTRACT In recent industrial applications, machine learning technology is proving useful in preventing equipment failures in advance through early failure diagnosis. In particular, we show that different domains can be linked through adversarial learning with data available in different working conditions to facilitate the training of the model, as it is impractical to acquire data for all conditions in real-world applications. Nevertheless, the initial failure is a difficult problem to diagnose because it does not show a significant difference from the normal data between different conditions. Moreover, if only the domain discriminator is judged when adapting the domain, it tends to easily cause misclassification, so the reliability of the detection result needs to be improved. In this study, we propose a new learning method that improves classification performance by sharing the classification characteristics of the classifier for each task with the target domain characteristic generator. The proposed mechanism uses spatial attention to extract the focused partial information of the feature generator and discriminator, and further enhances task-specific features using the attention mechanism between the two extracted information types. Addresses the challenge of implementing both domain adaptation and classification. Extensive experimentation demonstrates efficiency and improved classification performance on benchmark and real-world application datasets. In real machine cases, the classification accuracy is improved by almost 4%. In addition, the negative impact on false alarms was lowered by increasing the classification accuracy of minimum failures. Convincingly demonstrate model effectiveness by performing an empirical analysis of the method through ablation analysis and visualization.

INDEX TERMS Attention mechanism, domain adaptation, fault diagnosis, reliability, rotating machines, vibration measurement.

I. INTRODUCTION

With the success of many machines learning tasks, machine learning has become widely used in fault diagnosis for real applications [1], [2]. A key factor in achieving this development is the availability of vast amounts of labeled data in the workspace of interest. However, in real-world applications, collecting training instances given the enormous domain or labeling costs is challenging. Domain adaptation is performed using available labeled source data to address the lack of labeled target domain data [3]. The data-collection range for fault diagnosis becomes wider with dynamic machine operation (vehicle, heavy equipment, etc.). Even if the model is designed to share the same label space for the source and target operations, domain adaptation still suffers from data distribution changes. Moreover, maintaining the classifier performance while satisfying domain adaptation is a challenging problem [4]. Thus, the main goal of domain adaptation is to extract the domain-invariant features so that task classifiers can learn from the source data and be readily applied to the target domain.

The most recent domain adaptation methods are based on modern deep architecture. They rely on the superior model capacity of the networks to learn hierarchical features that have been empirically proven to transfer between domains better than previous methods [5], [6]. One of the transfer
learning methods is a domain adaptation method based on the maximum mean discrepancy (MMD). The MMD-based domain adaptation method adjusts the source and target domain feature space by minimizing feature discrepancies based on distance. A domain adaptation method was proposed to extract common features between bearing data collected under various working conditions by combining the MMD normalization with a sparse autoencoder network [7]. A diagnostic method was proposed to reduce the discrepancy between the two domains by applying a pseudo-label learning strategy to the unlabeled target domain data [8].

Intuitively, these methods achieved domain adaptation by reducing the mean value discrepancy between the source and target domains. However, the data distribution properties have several differences, such as the mean, median, and standard deviation. Therefore, the minimization of the MMD distance method cannot guarantee that a normalized feature set for fault diagnosis can be obtained [9].

Domain adversarial training-based models have been proposed to overcome above deficiency [10], [11], achieving state-of-the-art performance for images. Using a domain adaptation neural network (DANN) [10], the classification network can be further extended with an additional domain discriminator. The network is trained on differentiate instances of two domains by taking features as input from the classification network feature generator. With the application of the minimax game [12], adversarial training can learn domain-invariant features. These domain adversarial networks can significantly reduce domain mismatch.

However, designing the task classifier and domain discriminator separately has the following drawbacks. First, the domain discriminator distinguishes between the source and target domains without considering the task-specific decision boundaries between classes. In particular, for fault diagnosis using a vibration signal, the problem of changing the characteristics of the original signal using strong ambient noise is more pronounced. Thus, the trained feature generator produces ambiguous features near class boundaries. Moreover, because only the loss value for domain classification is backpropagated to the feature generator, the model does not know which information to focus on to determine the difference between the two domains. This situation has the disadvantage that the classification performance is not guaranteed even if domain adaptation is successful.

Second, the data affected by the mode collapse problem (i.e., the co-distribution of features and categories) are not properly aligned in the source and target domains [13], [14]. The first step toward solving the above shortcomings was co-parameterizing the task and domain classifier into an integrated classifier. However, due to the noise added to the vibration signal, it is easy to encounter extracting features by focusing only on a region that is easy to adapt to the source domain.

Thus, we propose a new learning method that can ensure the categorical prediction performance for all input instances for fault-relevant feature extraction, as presented in Fig. 1. Intuitively, training the proposed model is primarily performed by sharing the attention score between the encoding neurons (input) and real category neurons (output). Unlike the somewhat implicit work [15], the feature generator and discriminator share key information about the hidden layer of the classifier through the spatial attention mechanism.

![FIGURE 1. Overall architecture of the proposed method. The attention map from the classifier is shared to the feature generator and discriminator for the target domain dataset. The source domain data is trained through blue and red line. But target domain data is trained through only red line.](image-url)
to induce the feature focused on fault classification to be judged as the source domain data. We introduce a category-level transferable feature sharing module to learn an adaptive classifier boundary for the target domain, which can further improve the performance. The classification performance for the target domain can be guaranteed through this method.

Thus, our main contributions are summarized as follows:

1) In this study, we propose a model that can focus on the remarkable spatial information related to fault characteristics and an information-sharing mechanism that can maintain classification performance between classes. The model compensates for the relationship between categorization and domain adaptation to reduce the misclassification of target domains.

2) Misclassification due to information unrelated to classification performance (e.g., noise and peripheral disturbances) can be minimized through a combined model using feature extraction and an attention mechanism for classification models.

3) The proposed method was quantitatively verified using three datasets, the rolling bearing benchmark dataset and the gearbox of a real machine. Promising results were obtained by comparing the method with the existing method and related studies.

4) An additional experiment validates the robustness of the classifier in noisy environments. This aids in better understanding of the proposed method.

The paper is organized as follows. The related work is described in Section II. Section III combines the overall ideas and proposes a new diagnostic method employing the sharing of classification information using spatial attention as a discriminative domain adaptation. Section IV describes the performance verification results using the public dataset and actual equipment dataset and analyzes the effectiveness of the proposed model. Finally, Section V presents the conclusion.

II. RELATED WORKS

In [16] and [17], shallow invariant feature extraction method proposed to extract domain-invariant features. The input signal is preprocessed through the spectral envelope extraction method, and the time domain synchronization averaging principle using transfer component analysis (TCA) was used for cross-domain feature extraction [16]. The domain-adaptive neural network (DaNN) is a pioneer algorithm that uses shallow neural networks to calculate the distance between the source and target domains [17]. However, according to the latest research, deep learning architectures for domain adaptation are more promising and can learn more transferable features [18], [19].

A domain adaptation method using a deep learning framework was applied to extract domain-invariant features [18]. In addition, a model that generalizes and optimizes the distance between and within classes by applying a deep distance metric learning method to diagnose rolling bearing defects was proposed [19]. A deep convolutional transfer learning network with adversarial training was proposed to learn domain-invariant features across domains to diagnose disabilities [20]. A multi-sensor fusion method was proposed using a deep belief network for ball screw fault diagnosis [21]. An image-based convolutional neural network diagnostic architecture was proposed for ball screw spindle failure [22]. A method to minimize the MMD between the source and target domain features based on a sparse autoencoder structure was proposed [7].

A. DOMAIN ADVERSARIAL NETWORK

A domain adversarial network [10] is an adversarial training approach to minimizing domain discrepancy. Inspired by the generative adversarial network [23], the approach aims to find the Nash equilibrium in which the domain discriminator cannot distinguish each domain through adversarial learning between the feature generator and domain discriminator. The feature generator learns to fool the discriminators, and the discriminators struggle to avoid being deceived. The adaptive generative network consists of a feature generator $G(\cdot)$, discriminator $D(\cdot)$, and classifier $C(\cdot)$.

Formally, this method is called a minimax game. Thus, the overall objective is defined as follows:

$$C (\theta_G, \theta_C, \theta_D) = \frac{1}{n_s} \sum_{s\in \mathbb{S}} \mathcal{L}_C (C (G_S (x_s)), y_s) - \frac{\beta}{n_s + n_t} \times \sum_{s\in \mathbb{S}, t\in \mathbb{T}, d\in \mathbb{D}, y \in \mathbb{Y}} \mathcal{L}_D (D (G_S (x_s), G_T (x_t), d), d)$$

where $\theta_G$, $\theta_C$, and $\theta_D$ are the parameters of $G_S$, $G_T$, $C$, and $D$, respectively, and $d$ denotes the domain label of $x_i$. In addition, $\mathcal{L}_C (\cdot)$ is the label prediction loss, and $\mathcal{L}_D (\cdot)$ is the domain classification loss. Further, $\beta$ is a hyperparameter to trade off the two objectives $\mathcal{L}_C$ and $\mathcal{L}_D$. Through minimax adversarial optimization training, the parameters $\hat{\theta}_G$, $\hat{\theta}_C$, and $\hat{\theta}_D$ deliver a saddle point of (1):

$$\left( \hat{\theta}_G, \hat{\theta}_C \right) = \arg \min \max C(\theta_G, \theta_C, \theta_D), \quad \hat{\theta}_D = \arg \max C (\hat{\theta}_G, \hat{\theta}_C, \theta_D).$$

B. ATTENTION MECHANISM

The traditional research methods primarily start with a single label for the entire fault signal. Therefore, training in-depth models generally require a wealth of domain knowledge and a vast volume of experimental data to increase the accuracy of the fault signal classification. However, collecting these data for practical applications is difficult.

Unlike traditional domain-invariant feature extraction for the entire input signal, the attention layer focuses on the signal directly related to the discriminative failure signal, allowing the model to adapt easily even if the domain changes. Inspired by the field of image processing [24], we propose a model allowing users to focus on the fault area. The attention
mechanism was initially proposed for learning the alignment between source and target tokens to improve the performance of neural machine translation [25], [26]. The mechanism was primarily used in the language and image fields for object detection. Several studies were done on basic and modified attention mechanisms for time series data. The long short-term memory fully convolutional neural network with an attention mechanism was proposed for multivariate time series data [27]. The transition functions of the attention mechanism are described in (3)–(5), where $H$ is a matrix representing the features extracted by the predictive model. The matrix $\{h_1, h_2, \ldots, h_n\} \cdot e_n \in \mathbb{R}^n$ is the input vector, and $v_a$ denotes the embedded aspect of the attention mechanism. In addition, $\alpha$ is a vector representing the attention weight for feature $H$, and $r$ is the attentive neural network representing the weighted sum of feature $H$:

$$M = \tanh \left[ \frac{W_h H}{W_v v_a \cdot e_n} \right]$$  \hspace{1cm} (3)

$$\alpha = \text{softmax} \left( w^T M \right)$$  \hspace{1cm} (4)

$$r = H \alpha^T.$$  \hspace{1cm} (5)

III. METHODS

Subsections A, B, C, and D provide the problem formulation, overall idea, domain discriminator with attention, and spatial attention mechanism, respectively.

A. PROBLEM FORMULATION

We have a labeled dataset $\{X_s, Y_s\}$ and an unlabeled target domain data $X_T$. Fig. 2 shows the obvious differences between the four types of faults between the source domain and the target domain, regardless of the raw signal or frequency spectrum of the actual machine data. As shown in Fig. 2 (a), it can be seen that the 0.5mm crack case, which can be said to be an initial fault between domains, has a similar type and amplitude to the normal data of different domains. This acts as a factor that increases the misclassification rate with domain adaptation. Also, in Fig. 2 (b), the monolayer spectrum of 0.5 mm failure is similar to that of 1.0 mm failure. This can be viewed as a sensitive condition that is easily misclassified.

Fig. 3 illustrates the misclassified area that can occur while performing domain adaptation by focusing on the area where the existing adversarial domain adaptation model can easily adapt. In addition, the figure presents a method of minimizing the misclassified part by focusing on discriminative feature extraction through the proposed model. The goal is to maintain fault classification performance through a trained model even in the target domain without a label. In this paper, the method was verified for the datasets on three cases.

In Fig. 4, the distributions of the source and target domains of each dataset are different. The prefixes S_ and N_ indicate the source and target, respectively. For the actual machine data collected under real driving conditions, the alignment of the source and target data distribution is not uniform due to the surrounding noise and environmental factors compared
to the data collected in the experimental environment. This factor makes domain adaptation more difficult.

**B. OVERALL IDEA**

Inspired by the generative adversarial network of the image processing domain, we introduced the attention module to the domain adaptation framework to learn a discriminative feature. We propose a model that can maintain fault classification performance even in the target domain by combining the adversarial domain adaptation model and attention mechanism to share the discriminative feature region directly related to the fault of the input signal. This model focuses on the tooth signal where the actual failures occurred, focusing on the failure signal even at different rotational speeds. Therefore, the characteristics of the target domain data are extracted with emphasis on key information for failure classification. The feature generator of the target domain creates a feature combined with the fault classification feature concentrated in the classifier and learns the feature so that it cannot be distinguished from the source domain data through a domain discriminator.

Through adversarial training between the generator ($G$) and domain discriminator ($D$), the source and target domain instances are reflected in the same learned feature subspace. In addition, through a new algorithm sharing the information’s discriminatory characteristics of the source data with the feature extraction space of the target domain data, the defect diagnostic knowledge is shared, obtaining better classification performance in the target domain.

The attention layer is inserted into the corresponding layer between the feature generator and classifier, as depicted in Fig. 5, to create an attention model that learns the correlation to discriminate the fault signal. The spatial attention focuses on the informative part for classification or prediction. Spatial attention represents the attention score on the feature map or a single cross-sectional slice of the tensor. By refining the feature maps using spatial attention, we enhance the concentration to the subsequent convolutional layers, which improves the performance for fault classification of the model.

**FIGURE 5. Architecture of the feature generator for extracting the attention feature map.**

**C. ADVERSARIAL LEARNING FOR DOMAIN ADAPTATION**

The learning model consists of two steps. First, we trained the entire system using the source domain data and labels. Next, by combining the feature generator with the extracted attention map, the target domain data were learned, and the loss of the domain discriminator was maximized so that the distribution of the target domain data is matched to the source domain. Its main purpose is to focus only on the characteristics of the error signal obtained from the original signal so that time-varying regions of different distributions can be excluded as much as possible.

Overall, the proposed method consists of a feature generator such as $G\equiv \{G_{t}, G_{f}\}$ with the parameter $\theta_{G}\equiv (\theta_{g}, \theta_{r})$. The front layers of each $G$ are shared between the source and target domains. Furthermore, the method consists of the task-specific classifier $C$ with the parameter $\theta_{c}$. In addition, $C$ classifies the features into $K$ classes, outputting a $K$-dimensional vector of logits. The method also consists of a domain discriminator $D$ with $\theta_{d}$. The proposed method aims to learn a domain-invariant discriminative feature that is accurately classified in the unlabeled target domain.

Moreover, $G \sim D$ extracts feature for optimal classification from the classifier through each hidden layer of the feature generator. Further, $G \sim C$ plays the role of a classifier to classify instances. Therefore, our goal is to extract features in a form that can accurately perform classification and domain shift for the same signal. We aim to share the focused part in each hidden layer by applying a spatial attention mechanism between the hidden layers in the feature generator and classifier. This shared attention score is input back to the feature generator to extract features while minimizing the classification loss and maximizing the domain discrimination loss.

**D. SPATIAL ATTENTION WITH CLASSIFIER**

The classifier classifies the input to each class. We propose a spatial attention method to extract the fault identification areas on which the classifier concentrates for each classification in the source domain. These fault identification areas reveal the areas where the classifier concentrates to classify the input properly. If given an input feature $f$, then a map $M(f)$ of spatial attention equal to the size of the input feature $f$ is computed by the classifier.

Moreover, $M(f)$ presents the concentration of the hidden unit to classify the input accurately. The classifier’s attention map correlates with discrete, distinguishable defect areas and is shared with the domain feature generator for focusing. In the feature generator and classifier, average pooling, maximum pooling, and a $1 \times 1$ convolution are sequentially concatenated to compute the spatial map feature representation. Inspired by convolutional block attention module [28], we applied average pooling, maximum pooling, and a $1 \times 1$ convolution to the channel dimension to effectively represent the features. Then, from the convolutional layer, a spatial attention map $M(f) \in \mathbb{R}^{1\times H\times W}$, which is focused on the salient region, is extracted. The above process applies to each feature generator and classifier. Finally, each attention map is connected to the element-wise dot production in the feature generator of the target domain.
The spatial attention map is generated by applying average pooling, maximum pooling, and a $1 \times 1$ convolution, and the convolution layers are as large as $7 \times 7$. The operation for each layer is performed sequentially, and the output of each layer is expressed as $f_{\text{avg}}, f_{\text{max}}, f_{1 \times 1}, f \in \mathbb{R}^{1 \times H \times W}$, respectively. Extracted spatial attention maps of the feature generator and classifier are denoted by $M^S_C(f_s)$ and $M_G(f)$, respectively. The merged spatial attention map $M(f)$ is calculated by a nonlocal neural network [25] to apply the attention computation. The feature maps $f_s$ and $f_T$ branch out into three copies corresponding to the concepts of the key, value, and query. Additionally, $M^S_C(f_s)$ is the query, and $M_G(f)$ is the key and value. Thus, the dot-product attention is applied to output the attention feature maps:

$$M^S_C(f_s) = (f_s)_{7 \times 7} \left[ (f_{\text{avg}}(f_s), f_{\text{max}}(f_s))_{1 \times 1} \right],$$

$$M_G(f) = (f)_{7 \times 7} \left[ (f_{\text{avg}}(f), f_{\text{max}}(f))_{1 \times 1} \right],$$

$$a_{f, f_s} = \sigma(M_G(f)^T M^S_C(f_s)), \quad \text{where } \sigma \text{ is the softmax activation.}$$

The final output is as follows:

$$M(f) = a_{f, f_s} M_G(f).$$

As presented in Fig. 5, the attention maps for the target domain are a combination of spatial attentive information from the feature generator of the target domain data and classifier. The attention vector contains spatial information that applies the feature generator and low-level feature responses to the classification [29].

E. Training Strategy

The flowchart of the proposed fault diagnostic method is illustrated in Fig. 6. First, the labeled source domain data and unlabeled target domain data are collected as raw vibration signals from the working equipment. The target domain data may be exposed to additional environmental noise from the source domain and may be collected under different working conditions than the training data. Second, feature extraction and domain adaptation training are performed on the source and target domain data. The fault classifier learns only from the source domain data and shares the source domain classification feature when extracting the target domain feature. Finally, we used the trained model to perform a final defect classification on the newly collected data.

The classifier was trained using the standard supervised loss below with source domain data, $x_s, y_s$:

$$\min_{G, C} \mathbb{E}_{(x_s, y_s)} [ - \log C(G(x_s))] + \lambda_G L_D,$$

where $G(x)$ generates the input $x$, whereas $D$ performs domain discrimination and $C$ classifies.

The objective can be expressed as follows:

$$L_{\text{total}} = \lambda_C L_C + \lambda_D L_D, \quad \text{where } L_D \text{ and } L_C \text{ are losses for the discriminator network } D \text{ and the classifier network } C, \text{ respectively.}$$

Furthermore, $\lambda_C$ and $\lambda_D$ are the hyperparameters used to balance the relative importance of different terms. During training, hyperparameters are obtained experimentally. To meet the same conditions as the previous methods, we set $\lambda_C = 0.01$, where $\lambda_D = 0.01$. The calculation algorithm of the loss function is explained in Algorithm 1.

Finally, the total objective function incorporating the adversarial loss and modified classification loss with the attentive feature is expressed as follows:

$$\min_{G, C} \max_D \mathbb{E}_{(x_s, y_s)} [ - \log C(G(x_s))] + \lambda_D L_D, \quad \text{where } \lambda_C L_C + \lambda_D L_D. \quad \text{(14)}$$

IV. Experiments

The proposed model was validated using a rolling element bearing dataset (two benchmark dataset) and a dataset of heavy equipment that works dynamically with various loads. Moreover, the proposed model was compared to state-of-the-art domain-adaptive architectures, including TCA [16],
Algorithm 1 Learning Algorithm

Input: Input data $X = x^1, x^2, \cdots, x^n$
Output: Trained models $D, C, G$

1: begin
2: Initialize parameters for $D, C, G$
3: for $i = 1 \rightarrow n$ do
4: # extract latent vector and classification with domain discrimination at each timestep
5: for $t = 0 \rightarrow T$ do
6: Compute $M(f)$ according to (9)
7: $\hat{x}_i^t = [M(f); x_i^t]$,
8: $\tilde{x}_i^t = D(\hat{x}_i^t)$
9: Compute $\tilde{x}_i^t$ with $C$ and $\hat{x}_i^t$
10: end for
11: $\hat{x}_i^t \leftarrow \{\hat{x}_i^t\}_{t=1}^T$
12: $\tilde{x}_i^t \leftarrow \{\tilde{x}_i^t\}_{t=1}^T$
13: Compute $L_D$ with $\tilde{x}_i^t$ and $x_i^t$
14: Compute $L_C$ with $C$ and $\hat{x}_i^t$
15: Update $G$ with $L_G$ formulated in (10), (11)
16: Update $D$ with $L_D$ formulated in (12)
17: Update $C$ with $L_C$ formulated in (10)
18: end for
19: end

DaNN [17], DANN [10], domain adaptation fault diagnosis (DAFD) [15], and discriminative adversarial domain adaptation (DADA) [30]. In addition, a model in which the attention mechanism was removed was also added to verify the effectiveness of the proposed attention architecture. The experiment for the second case was set to the actual operating conditions, including noise and other environmental factors, to verify the feasibility and practicality of heavy equipment failure diagnosis. The DAFD architecture introduced domain adaptation for fault diagnosis for the motor application. The DADA architecture used a domain adaptation method to maintain discriminative performance in the image field. Adversarial multiple-target domain adaptation (AMDA) was proposed to transform multiple-target domain feature learning [31].

A. IMPLEMENTATION DETAILS

The comparison methods were verified using the same implementation strategy proposed in the original paper. The data for each driving condition were randomly divided into training and testing data. Training and testing data were used for model training and validation using a specific domain as the source domain. The rest of the domains were defined as target domains, and the training data were used as the target domain data while training. The testing data defined as the target domain were used for the final model verification. The detailed parameters for each experiment are listed in Table 1.

Table 2 presents the details of the network in the experiment. Multiple overlapping samples were extracted using a sliding window approach. Each sample length is 1500 data points. The network consists of convolutional, attention, normalization, and fully connected layers. A leaky rectified linear unit (LReLU) was used as an activation function. The LReLU layer uses a backpropagation learning method to make it easier to learn the weights of the shallow layer when adjusting its parameters.

Considering that the vibration signal repeatedly occurs with positive and negative values, the LReLU function was adopted to minimize the limitations of the rectified linear unit (ReLU) activation function that stops learning at negative values. The LReLU is an improved variant of the ReLU, which is also suitable for feature extraction for negative values of the input signal by providing a small positive value of $\alpha$ when the input is negative, allowing the neurons to continue working [32]:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \leq 0. \end{cases}$$  

(15)
Network training and inferencing were performed on a workstation with a single AMD 7742, V100 GPU, and an Ubuntu 18.04 operating system. The learning model was implemented in the Keras library using Python 3.7. Each sample was divided based on its length and was used as input. We adopted a cross-entropy loss function during training with a learning rate of 0.0001 and a batch size of 300 using the Adam optimization algorithm.

B. CASE STUDY 1: CWRU DATASET

We validated the model using the 12-k drive end bearing fault data in the Case Western Reserve University (CWRU) bearing dataset [33]. The four types of bearing defects are normal, ball, internal race, and external race. Each fault type has three sizes: 0.007, 0.014, and 0.021 inches. Therefore, 10 fault label types exist. Each fault label contains four load types (0, 1, 2, and 3 HP) measured from motor speeds of 1,797, 1,772, 1,750, and 1,730 rpm, respectively. Each sample was extracted at each sampling point, as depicted in Fig. 7. Datasets A, B, and C are defined with different rotational speed conditions, each containing 12,300 training samples and 6,059 testing samples from Table 3. Each measurement dataset is divided into 70% training and 30% testing sets. One of them is defined as the source data, and one of the other datasets is defined as the target domain (e.g., A: source domain and C: target domain).

![FIGURE 7. Measuring position of each vibration signal in CWRU dataset.](image)

| TABLE 3. Property of four datasets in the CWRU dataset. |
|---------------------------------|---------|---------|---------|---------|
| Dataset                        | A       | B       | C       | D       |
| Speed (rpm)                    | 1797    | 1772    | 1750    | 1730    |
| Fault Location                 | None, Inner Race, Outer Race |          |         |         |
| Load                           | 0       | 1       | 2       | 3       |
| Fault diameter (inch)          | 0.07, 0.014, 0.021 |         |         |         |
| Train                          | 12300 (Respectively) |         |         |         |
| Test                           | 5271 (Respectively) |         |         |         |

The results of comparative experiments are shown in Table 6. The comparative results prove that the proposed method is more effective than the others. In particular, C→A shows much higher results than other data distribution methods. The proposed method contributes to the improvement of disability classification performance through the sharing of classification-oriented domain information.

C. CASE STUDY 2: XJTU-SY DATASET

Unlike the dataset CWRU, which collects data with apparent fault characteristics, the XJTU-SY bearing dataset collects run-to-failure bearing information with different operation mode such as inner race, outer race, and mixed fault [35]. The testbed illustrated in Fig. 8. In particular, data set C not only contains data on obvious defect characteristics similar to data sets A and B, but also collects early-stage lifecycle data with weak defect characteristics. The vibration signal was sampled at 25.6 kHz, and condition A with the fewest samples was met to match the data length. The XJTU-SY data set details are summarized in Table 5. Each data set was divided into A, B, and C, and experiments were performed by defining each as a source domain or a target domain.

![FIGURE 8. Testbed of rolling element bearings of XJTU-SY dataset.](image)

The results of comparative experiments are shown in Table 6. The comparative results prove that the proposed method is more effective than the others. In particular, C→A shows much higher results than other data distribution methods. The proposed method contributes to the improvement of classification performance through the sharing of classification-oriented domain information.

D. CASE STUDY 3: REAL MACHINE DATASET

Vibrations occurring in the gearbox of actual heavy equipment are very complex. Heavy equipment operates dynamically at various speeds and loads. The sensor is mounted on the outside of the gearbox and collects the gearbox vibration and the vibration (noise) from other equipment components in Fig. 9. There are three classes of failures, determined according to their crack sizes: 0.5, 1.0, and 2.5 mm. Smaller crack sizes make it more difficult to classify the failure.
Unlike the CWRU dataset, the data were collected from actual heavy equipment operating under various conditions (speed, load, etc.) with exposure to different noise environments. The corresponding data did not contain accurate speed information (no tachometer), and the tests were conducted at speeds of about 100%, 75%, 50%, and 25% of max rotation speed about 9.1 rpm based on user manipulation. In addition, the data were collected only for some classes. Table 7 lists the operating conditions, number of fault classes, and count of datasets for training and testing. In this experiment, each training dataset is also used as the target domain data.

Because only Datasets A and C contain all task data, the experiment was conducted using Dataset A or C for training and other datasets for verification. The confusion matrix of the classification result of real machine, using the proposed model, is displayed in Fig. 10. The confusion matrix displays the predicted results of the samples on the columns and the actual labels of the samples on the rows. Approximately 88% to 90% accuracy was achieved for the new domain data. The performance was significantly reduced when the attention layer was not applied. Moreover, we confirmed that the recognition of minimum failure (abnormal 0.5/0.5 mm) and the rate of false alarm are significantly reduced. We compared the performance of the proposed model to that of several other of the latest deep learning methods, using the conditions in Table 8 as the accuracy and standard deviation values. In all cases, the proposed model outperformed other models.

### E. INFERENCE TIME COSTS

A key consideration in the fault diagnostic method is the cost of time required to generate the classification results. An important factor in the practical implementation of an intelligent fault diagnostic system is inference time. The reason is that model training can be done offline, whereas inferences about sequences must be performed onboard in real time. The result of comparing the inference time cost was added to each experiment result table, such as

| Task (source—target) | TCA | DaNN | DAFD | DANN | DADA | AMDA | Proposed model/wo attention | Proposed model |
|----------------------|-----|------|------|------|------|------|-------------------------------|---------------|
| A→B                  | 51.37±0.02 | 64.87±0.01 | 82.38±0.03 | 75.76±0.54 | 99.27±0.08 | 99.35±0.08 | 98.78±0.21 | 99.99±0.01 |
| A→C                  | 49.87±0.03 | 63.41±0.03 | 83.04±0.21 | 74.96±0.04 | 99.86±0.12 | 99.70±0.10 | 98.11±0.12 | 99.89±0.10 |
| A→D                  | 50.38±0.04 | 62.56±0.10 | 84.48±0.12 | 74.81±0.19 | 99.97±0.03 | 99.82±0.10 | 98.56±0.08 | 98.77±0.03 |
| B→A                  | 51.50±0.02 | 61.33±0.08 | 80.88±0.01 | 73.73±0.22 | 99.82±0.16 | 98.56±0.09 | 98.67±0.15 | 99.43±0.13 |
| B→C                  | 58.89±0.02 | 62.12±0.05 | 80.99±0.02 | 73.46±0.04 | 99.55±0.06 | 99.95±0.03 | 99.12±0.37 | 98.42±0.18 |
| B→D                  | 60.29±0.03 | 60.37±0.03 | 81.47±0.13 | 72.65±0.15 | 98.27±0.16 | 99.31±0.03 | 98.12±0.44 | 99.47±0.04 |
| C→A                  | 60.13±0.03 | 64.81±0.04 | 82.11±0.16 | 71.70±0.25 | 98.79±0.07 | 99.10±0.05 | 97.23±0.32 | 99.86±0.08 |
| C→B                  | 59.34±0.03 | 62.35±0.03 | 80.26±0.14 | 72.40±0.05 | 97.75±0.10 | 98.62±0.06 | 98.32±0.47 | 99.56±0.12 |
| C→D                  | 58.23±0.02 | 65.21±0.03 | 82.43±0.14 | 73.82±0.13 | 97.62±0.11 | 99.65±0.03 | 98.13±0.53 | 99.38±0.15 |
| D→A                  | 55.38±0.01 | 70.81±0.04 | 81.21±0.20 | 70.62±0.13 | 97.23±0.13 | 98.27±0.01 | 97.45±0.87 | 98.82±0.09 |
| D→B                  | 53.83±0.03 | 68.27±0.06 | 81.95±0.12 | 71.41±0.09 | 97.53±0.25 | 97.58±0.03 | 98.23±0.12 | 99.23±0.07 |
| D→C                  | 54.10±0.02 | 68.12±0.05 | 82.01±0.08 | 71.98±0.02 | 96.90±0.07 | 99.97±0.06 | 97.12±0.21 | 99.09±0.27 |
| Average              | 55.27   | 64.60   | 81.87   | 74.20   | 98.54   | 99.13   | 98.15   | 99.32   |
| Inference Time (ms)  | 2.10    | 2.11    | 2.89    | 3.12    | 3.28    | 5.21    | 2.33    | 3.15    |

| Dataset | A | B | C |
|---------|---|---|---|
| Speed (rpm) | 2100 | 2250 | 2400 |
| Fault location | None, Inner Race, Outer Race |
| Load | 12 | 11 | 10 |
| Train | 9426 | 10520 | 11248 |
| Test | 4040 | 4508 | 4820 |
Tables 6, 7 and 8. No increase in inference time was found due to the proposed method because no additional preprocessor or feature extraction module exists. Therefore, the proposed method improves the reliability of the discrimination, and the discrimination time is effective for use in the industrial field in a similar way.

**F. EFFECTIVENESS OF ATTENTIONAL GUIDANCE**

In this section, we investigate the robustness of the proposed model to a noisy environment. The CWRU data were collected under limited conditions in a laboratory environment; thus, the data did not reflect the noise conditions for an actual industrial site. Considering that the target equipment is used in a harsh environment, the model performance was verified under noisy conditions.

We verified the performance of the trained model under noisy conditions using varying Gaussian white noise (−3 dB or −6 dB). The SNR is defined as follows:

$$SNR_{dB} = 10 \log_{10} \frac{P_{signal}}{P_{noise}},$$

where $P_{signal}$ and $P_{noise}$ are the power of the signal and power of the noise, respectively. The comparison result is represented in Table 9.

The accuracy of the proposed model is better than the conventional model in all SNR scenarios. When SNR is −6dB, the proposed model still reaches a diagnostic accuracy of 94.89%, which is almost 8% improvement over the DADA, the best of the comparison models. This result reveals that the proposed model extracts feature focused on fault classification characteristics without additional noise removal preprocessing, which indicates that the proposed model is robust for a wide range of operating conditions in real applications.

**G. ABLATION STUDY**

The ablation study was performed to investigate the effect of classification information sharing the attention mechanism. We visualized the spatial attention score and map in each attention layer to further understand the feature learning mechanism in the generator, as illustrated in Fig. 11. All attention scores have the same length as the input signals. Colors
represent relative attention weights, and yellow represents high levels.

In general, vibrational peaks can represent defects, and variations and subtle differences in peaks can represent the failure type. By comparing before (attention layer of the generator in Fig. 10) and after (attention map in Fig. 10) sharing the classification properties, we confirmed that the attention regions are more attended and move according to the captured pattern in the classifier. Fig. 12 shows the probability density before and after domain adaptation in feature space by various domain adaptation learning methods. After domain adaptation, features were extracted with a similar distribution between the two domains. Among these, the proposed method is superior to the comparison method.

The attention mechanism focuses on specific information in the input signal, which indicates that it can learn differential features that are beneficial for diagnosing various bearing conditions. As depicted in Fig. 13, if the classifier’s attentive information is not applied using the attention mechanism, the incorrect classification data increase despite the well-adapted domain (a left figure in each (a–d)). By sharing informative classification information using attention mechanism, the network can selectively learn more important feature information (a right figure in each (a–d)). As a result, the efficiency and ability of feature learning are improved, which improves the accuracy of fault classification in the target domain and effectively focuses on the attention fault signal segment.

In the case of misclassification data, the classification accuracy was lowered by extracting features from other areas or noise focused only on domain classification rather than features indicating actual failure. This is because the focused
TABLE 8. Comparison of classification accuracy obtained with four operating conditions using real machine dataset.

| Task (source) | TCA    | DaNN   | DAFD   | DANN   | DADA   | AMDA   | Proposed model | Proposed model |
|---------------|--------|--------|--------|--------|--------|--------|----------------|----------------|
| A→B           | 48.21±0.02 | 63.14±0.08 | 85.72±1.46 | 85.92±1.11 | 84.93±1.37 | 86.32±0.57 | 86.61±0.76 | 90.60±0.75   |
| A→C           | 45.35±0.03 | 63.94±0.03 | 84.82±1.79 | 86.91±1.84 | 84.83±1.98 | 87.94±0.78 | 85.13±0.88 | 90.46±0.69   |
| A→D           | 46.34±0.01 | 62.23±0.07 | 83.09±1.93 | 85.02±1.66 | 83.84±1.94 | 84.82±0.89 | 84.70±1.21 | 91.77±0.55   |
| C→A           | 46.33±0.01 | 61.56±0.07 | 83.85±1.22 | 84.38±1.72 | 82.01±1.67 | 86.33±0.61 | 85.25±0.89 | 88.23±0.57   |
| C→B           | 42.56±0.02 | 62.11±0.05 | 84.92±1.90 | 84.81±2.83 | 82.98±1.37 | 85.27±1.13 | 85.18±0.72 | 89.92±0.92   |
| C→D           | 43.16±0.01 | 57.23±0.03 | 83.67±1.87 | 85.98±2.02 | 82.27±1.56 | 85.98±1.08 | 83.83±0.89 | 88.27±0.82   |
| AVG           | 45.32   | 61.70   | 84.34   | 85.50   | 83.47   | 86.09   | 85.11   | 89.87        |
| Inference time (ms) | 2.03 | 1.87 | 2.40 | 2.82 | 3.15 | 5.77 | 2.43 | 3.04 |

TABLE 9. Comparison of classification accuracy of noisy CWRU dataset.

| Noise (SNR) | Task (source) | TCA    | DaNN   | DAFD   | DANN   | DADA   | AMDA   | Proposed model | Proposed model |
|-------------|---------------|--------|--------|--------|--------|--------|--------|----------------|----------------|
| ~6 dB       | A→B           | 42.63±0.61 | 58.84±0.20 | 74.83±1.01 | 68.42±0.84 | 87.89±0.51 | 86.32±0.57 | 86.61±0.90 | 94.89±1.52 |
| ~3 dB       | A→B           | 48.32±0.07 | 60.26±0.07 | 76.35±0.23 | 70.08±0.65 | 90.19±0.39 | 92.10±0.77 | 91.58±0.31 | 96.35±0.36 |
| 0 dB        | A→B           | 51.37±0.02 | 64.87±0.01 | 82.38±0.03 | 75.76±0.54 | 99.27±0.08 | 99.35±0.08 | 98.78±0.21 | 99.99±0.01 |

FIGURE 14. Illustration of the misclassified feature and re-weighted transferable feature with attention region.

area for domain classification may be concentrated on an area unrelated to task classification. However, it can be confirmed that when the re-weighting is applied to the failure classification feature area that the classifier concentrates on, feature extraction for failure classification is induced in Fig. 14.

V. CONCLUSION

This paper presents a domain adaptation approach capable of maintaining performance to classify and diagnose rotating machine failures under various operating conditions. The proposed algorithm solves the difficult problem of obtaining sufficient samples for various operating conditions, one of the main challenges in domain adaptation for fault diagnosis. The feature extraction method can share spatial information between the discriminator and hidden layer of the feature generator by applying an attention mechanism, extracting the characteristics of the attentive fault signal.

In addition, through a domain discriminator, the approach learns to minimize the discrepancy in the distribution between the source and target domains. The proposed model was validated through extensive experimentation by comparing its performance with that of the latest deep learning models using three datasets such as the CWRU, XJTU-SY and real machine datasets. The proposed model maintains a classification performance with an accuracy of over 89% in the target domain. The ablation study results indicate that our model effectively diagnoses faults while maintaining the desired classification performance under the new operating conditions.

The proposed model suggests a new feature extraction method for domain adaptation for the same task. Therefore, in future work, we will further refine the model to realize diagnostic tasks in which the target testing and source domains have different label categories.

REFERENCES

[1] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, “Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review,” 2019, arXiv:1901.08247. [Online]. Available: http://arxiv.org/abs/1901.08247
[2] J.-Y. Kim and S.-B. Cho, “Deep CNN transferred from VAE and GAN for classifying irritating noise in automobile,” Neurocomputing, vol. 452, pp. 395–403, Sep. 2021.
[3] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[4] H. Tang and K. Jia, “Discriminative adversarial domain adaptation,” in Proc. 34th AAAI Conf. Artif. Intell., Feb. 2020, pp. 5940–5947.
[5] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 3320–3328.
[6] Y. Zhang, H. Tang, K. Jia, and M. Tan, “Domain-symmetric networks for adversarial domain adaptation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 5031–5040.
[7] L. Wen, L. Gao, and X. Y. Li, “A new deep transfer learning based on sparse auto-encoder for fault diagnosis,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 49, no. 1, pp. 136–144, Jan. 2019.
[8] B. Yang, Y. Lei, F. Jia, and S. Xing, “An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings,” Mech. Syst. Signal Process., vol. 122, pp. 692–706, May 2019.
[9] W. Mao, Y. Liu, L. Ding, A. Safian, and X. Liang, “A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions,” IEEE Trans. Instrum. Meas., vol. 70, 2021, Art. no. 3500013, doi: 10.1109/TIM.2020.3038596.
[10] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” J. Mach. Learn. Res., vol. 17, no. 1, pp. 2030–2096, 2016.

[11] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, “Transferable attention for domain adaptation,” in Proc. AAAI Conf. Artif. Intell., vol. 33, no. 1, 2019, pp. 5345–5352. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/4472

[12] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[13] V. K. Kurmi and V. P. Namboodiri, “Looking back at labels: A class based domain adaptation technique,” 2019, arXiv:1904.01341. [Online]. Available: https://arxiv.org/abs/1904.01341

[14] L. Tran, K. Sohn, X. Yu, X. Liu, and M. K. Chandralekha, “Gotta adapt ‘Em all: Joint pixel and feature-level domain adaptation for recognition in the wild,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 2672–2681.

[15] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, “Deep model based domain adaptation for fault diagnosis,” IEEE Trans. Ind. Electron., vol. 64, no. 3, pp. 2296–2305, Mar. 2017.

[16] J. Xie, L. Zhang, L. Duan, and J. Wang, “On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis,” in Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM), Jun. 2016, pp. 1–6.

[17] M. Ghifary, W. Kleijn, and M. Zhang, “Domain adaptive neural networks for object recognition,” in Proc. Pacific Rim Int. Conf. Artif. Intell. Cham, Switzerland: Springer, 2014, pp. 899–904.

[18] Z. Cao, M. Long, J. Wang, and M. I. Jordan, “Partial transfer learning with selective adversarial networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 2724–2732.

[19] X. Li, W. Zhang, and Q. Ding, “A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning,” Neurocomputing, vol. 310, pp. 77–95, Oct. 2018.

[20] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, “Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data,” IEEE Trans. Ind. Electron., vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[21] L. Zhang, H. Gao, J. Wen, S. Li, and Q. Liu, “A deep learning-based multi-sensor data fusion method for degradation monitoring of ball screws,” Microelectron. Reliab., vol. 75, pp. 215–222, Oct. 2017.

[22] T. Schlagenhauf, C. Feuring, J. Hillenbrand, and J. Fleischer, “Camera based ball screw spindle defect classification system,” in Proc. 9th Congr. German Academic Assoc. Prod. Technol. (WGP), Hamburg Germany; Springer Nature, 2019, pp. 503–524.

[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural. Inf. Process. Syst., 2014, pp. 2672–2680.

[24] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, “A theory of learning from different domains,” Mach. Learn., vol. 79, nos. 1–2, pp. 151–175, May 2010.

[25] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” presented at the 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, 2015. [Online]. Available: https://nycscholars.nyu.edu/en/publications/natural-machine-translation-by-jointly-learning-to-align-and-translate-2

[26] T. Luong, H. Pham, and C. D. Manning, “Effective approaches to attention-based neural machine translation,” in Proc. Conf. Empirical Natural Lang. Process., 2015, pp. 1412–1421.

[27] F. Karim, S. Majumdar, H. Darabi, and S. Chen, “LSTM fully convolutional networks for time series classification,” IEEE Access, vol. 6, pp. 1662–1669, 2017, doi: 10.1109/ACCESS.2017.2779939.

[28] S. Woo, J. Park, Y. J. Lee, and S. I. Kweon, “CBAM: Convolutional block attention module,” in Proc. Eur. Conf. Comput. Vis. Munich, Germany, Sep. 2018, pp. 3–19.

[29] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7794–7803.

[30] H. Tang and K. Jia, “Discriminative adversarial domain adaptation,” 2019, arXiv:1911.12036. [Online]. Available: http://arxiv.org/abs/1911.12036

[31] M. Ragab, Z. Chen, M. Wu, H. Li, C.-K. Kwoh, R. Yan, and X. Li, “Adversarial multiple-target domain adaptation for fault classification,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, 2021.

[32] A. Maas, A. Hannun, and A. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in Proc. 30th Int. Conf. Mach. Learn. (ICML), Atlanta, GA, USA, 2013, pp. 1–6.

[33] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study,” Mech. Syst. Signal Process., vol. 64–65, pp. 100–131, Dec. 2015, doi: 10.1016/j.ymssp.2015.04.021.

[34] A. Klausen and K. G. Robbersmyr, “Cross-correlation of whitened vibration signals for low-speed bearing diagnostics,” Mech. Syst. Signal Process., vol. 118, pp. 226–244, Mar. 2019.

[35] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” IEEE Trans. Rel., vol. 69, no. 1, pp. 401–412, Mar. 2020.

GYE-BONG JANG (Member, IEEE) received the M.S. degree in signal processing from Korea University, Seoul, South Korea. He is currently pursuing the Ph.D. degree in computer science with Yonsei University, Seoul. His research interests include prognostic health management and deep learning for autonomous equipment and smart factories. In particular, he is interested in transfer learning and cyber physical systems.

JIN-YOUNG KIM (Member, IEEE) received the B.S. degree in mathematics from Yonsei University, Seoul, South Korea, in 2018, where he is currently pursuing the M.S. degree with the Department of Computer Science. His current research interests include neural networks, artificial intelligence, and controllable representation learning.

SUNG-BAE CHOI (Senior Member, IEEE) received the B.S. degree in computer science from Yonsei University, Seoul, South Korea, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He was an Invited Researcher with the Human Information Processing Research Laboratories, Advanced Telecommunications Research Institute Research Institute, Kyoto, Japan, from 1993 to 1995, and a Visiting Scholar with the University of New South Wales, Canberra, Australia, in 1998. He was also a Visiting Professor with The University of British Columbia, Vancouver, Canada, from 2005 to 2006. Since 1995, he has been a Professor with the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, and evolutionary computation. He is a member of the Korea Information Science Society, INNS, the IEEE Computational Intelligence Society, and the IEEE Systems, Man, and Cybernetics Society. He was awarded Outstanding Paper Prizes from the IEEE Korea Section, in 1989 and 1992 and an Outstanding Paper Prize from the Korea Information Science Society, in 1990. He was a recipient of the Richard E. Merwin prize from the IEEE Computer Society, 1989 and 1992 and an Outstanding Paper Prize from the International Association for Pattern Recognition, in 1994 and received the Best Paper Awards at the International Conference on Soft Computing, in 1996 and 1998. Moreover, he received the Best Paper Award at the World Automation Congress, in 1998 and was listed in Marquis Who’s Who in Science and Engineering, in 2000 and in Marquis Who’s Who in the World, in 2001.