This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

Fuzzy Domain Adaptation via Variational Inference for Evolving Concept Drift

Chuang Wang, Zidong Wang, *Fellow*, *IEEE*, Weiguo Sheng, *Member*, *IEEE*, Qingqiang Liu, and Hongli Dong, *Senior Member*, *IEEE*

***Abstract*—The concept of fuzzy domain adaptation (FDA) is focused on transferring a model trained in a source domain to a target domain, where intrinsic distribution discrepancies exist in non-stationary and non-deterministic environments. In this paper, a novel drift decoupling-based variational adaptation network (DD-VAN) is proposed for FDA, allowing for the learning of intra-domain evolutionary patterns and inter-domain uncertainties. The DD-VAN algorithm is implemented in three main steps: (1) an intra-domain evolutionary trend modeling module is first employed to capture unknown temporal variations through an autoencoder architecture with variational inference; (2) a prototype-assisted fuzzy clustering module is used to estimate the membership degree of the target data, characterizing the inherent uncertainty and imprecision present in real-world distributions; and (3) a membership-aware domain fuzzy matching module is utilized to learn the gradual transitions between category-related data pairs in the source and target domains by introducing uncertainties. Furthermore, it is theoretically demonstrated that the inferred posterior distributions of latent codes can be optimized to align with the corresponding prior distributions by minimizing the Kullback-Leibler divergence. Extensive experiments are conducted on cross-domain tasks involving both synthetic and real-world datasets, and the experimental results suggest that the DD-VAN algorithm outperforms existing state-of-the-art methods.**

***Index Terms*—Fuzzy domain adaptation, concept drift, fuzzy pseudo-label estimation, variational inference, fault diagnosis.**

I. Introduction

Over the past decades, impressive achievements have been made by deep neural networks (DNNs) in various application areas including biology [19], [20], industry [28], [32], [45], and finance [35]. A fundamental assumption underlying the effectiveness of these traditional DNNs is that both training and test data must obey the independent and identically distributed (I.I.D.) principle, which is difficult to satisfy due to the uncertainties and variations in real-world environments [41]. As a result, when trained on non-IID data, DNNs inevitably encounter the distribution discrepancy issue, leading to a decrease in inference or prediction performance [6]. To address this issue, unsupervised domain adaptation (UDA) has been introduced, as described in [36], which focuses on learning a robust model from the provided training data (i.e., the source domain) that can generalize well to the unlabeled test data (i.e., the target domain) [39], [44], [49].

Despite the significant progress achieved by UDA, two pivotal challenges have not been fully addressed in previous works. The first challenge can be summarized as the issue of *non-deterministic covariate shift*. Specifically, to enhance transfer performance, numerous well-regarded UDA algorithms have adopted a conditional alignment strategy to tackle the covariate shift issue. Unfortunately, these UDA algorithms are invariably constrained to a discrete environment setting, where samples from each category are assumed to possess distinctive features. In real-world scenarios, however, domains often encompass intrinsic uncertainties and imprecisions, which can ultimately result in suboptimal performance when addressing covariate shift [2]. For example, pipeline failure is a gradual process in which varying degrees of leakage occur over time, leading to unclear decision boundaries between each conditional distribution. Under such circumstances, traditional pseudo-label algorithms for conditional alignment in UDA inevitably lead to performance degradation. On one hand, previous UDA works have rarely considered the inherent noise embedded in pseudo labels due to the covariate shift, which may compromise subsequent learning. On the other hand, pseudo labels with crisp values are incapable of describing the gradual transition process between different categories, which may result in negative transfer.

The second challenge in UDA research is associated with the phenomenon of data distribution changing consecutively over time, referred to as the *evolving concept drift* issue. Most existing UDA algorithms operate under the assumption that knowledge is transferred within a stationary environment, where both training and test data are arbitrarily sampled. This assumption can prove problematic in practical applications, as working conditions and service environments may continuously evolve in unforeseen ways over time, leading to the issue of evolving concept drift [51], [56]. When evolving concept drift occurs, the static domain knowledge induced from past data may no longer be relevant to new data, resulting in

This work was supported in part by the National Natural Science Foundation of China under Grants 62403119 and U21A2019, the Hainan Province Science and Technology Special Fund of China under Grant ZDYF2022SHFZ105, the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation under Grant GZB20240136, the China Postdoctoral Foundation under Grant 2024MD753911, the Heilongjiang Provincial Postdoctoral Science Foundation of China under Grant LBH-TZ2405, the Engineering and Physical Sciences Research Council (EPSRC) of the UK, the Royal Society of the UK, and the Alexander von Humboldt Foundation of Germany. *(Corresponding author: Hongli Dong.)*

Chuang Wang and Hongli Dong are with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing 163318, China, also with the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Northeast Petroleum University, Daqing 163318, China, and also with Sanya Offshore Oil & Gas Research Institute, Northeast Petroleum University, Sanya 572025, China. (Emails: wangchuang64@126.com, shiningdhl@vip.126.com)

Zidong Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (Email: Zidong.Wang@brunel.ac.uk)

Weiguo Sheng is with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China. (Email: weiguouk@hotmail.com)

Qingqiang Liu is with the School of Electrical & Information Engineering, Northeast Petroleum University, Daqing 163318, China. (Email: petroboy@163.com)

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

poor decision-making outcomes. In the ever-changing big data landscape, evolving concept drift has been identified as a root cause of the degraded effectiveness observed in many data-driven information systems, such as data-driven early warning systems and fault diagnosis systems [5], [12], [55]. The recent developments in concept drift have focused on (1) gradual self-training; (2) domain manifold; (3) incremental learning; and (4) meta-learning. However, most of these methods are borrowed from static or stationary environments, leading to some limitations when they are applied to changing environments, such as computational instability, catastrophic forgetting, and so on [9], [14], [22], [42], [50].

Motivated by the aforementioned analysis of the non-deterministic covariate shift issue and the evolving concept drift issue, a novel fuzzy domain adaptation (FDA) model, termed drift decoupling-based variational adaptation network (DD-VAN), is developed to address these two challenges. The proposed DD-VAN consists of the following three modules:

1) The first is an intra-domain evolutionary trend modeling module, which employs a variational inference-based intra-domain evolution pattern modeling (VIIE) mechanism to infer the changing dynamics of temporal features in the latent space. Technically, VIIE models the prior distribution of transformation properties in source samples through a trainable posterior distribution, with the objective of maximizing the evidence lower bound (ELBO). In comparison to continuous learning and online adaptation techniques, the VIIE mechanism offers a more principled approach to capturing temporal dependencies between samples and adapting to new patterns without the need for frequent retraining.
2) The second is a prototype-assisted fuzzy clustering module, which estimates pseudo labels of target data using fuzzy logic, a mathematical technique designed to handle inherent uncertainty and imprecision [3], [17]. Most existing fuzzy domain adaptation (FDA) algorithms are based on Takagi-Sugeno fuzzy rules designed for deterministic and stationary distributions, which implies that the concept drift issue remains unsolved. Furthermore, fuzzy techniques, such as the fuzzy C-Means algorithm [4], are highly flexible in modeling complex, nonlinear, and intricate relationships in data. However, existing fuzzy clustering approaches for UDA tasks typically operate independently of distribution discrepancy optimization, which can lead to error accumulation. To address this, a prototype-assisted fuzzy clustering (PFC) method is proposed within this module, aiming to correct pseudo labels through a persistent prediction-correction process.
3) The third is a membership-aware domain fuzzy matching module, addresses the non-deterministic covariate shift issue by introducing a novel metric, Fuzzy Maximum Mean Discrepancy (FMMD). FMMD aids in capturing domain-invariant features from imprecise data by learning fuzzy decision boundaries between different conditional distributions. Unlike existing variants of MMD, FMMD acknowledges the inherent uncertainty and multi-label nature of target data, thus enhancing generalization performance in real-world environments.

The core contributions of this paper are highlighted as follows.

1) A more realistic and challenging setting, termed FDA, is addressed in this paper, which extends beyond traditional UDA. In many real-world applications, the inherent uncertainty and imprecision in data make solving the deterministic UDA optimization problem suboptimal or even infeasible. Therefore, it is imperative to design an effective method to analyze and adapt to these uncertainties. This paper aims to enhance the classification performance of fuzzy-valued target data by leveraging knowledge from source data with evolutionary properties. In this paper, we propose a novel FDA technique, namely drift decoupling-based variational adaptation network (DD-VAN), which aims to improve the classification evolution performance on target data with fuzzy categories by utilizing the changing source data knowledge over time.
2) To address the concept drift problem, a variational inference-based intra-domain evolution pattern modeling (VIIE) mechanism is first designed to train a time-adaptive model by capturing unknown features from the underlying data distribution over time. The key insight of this module lies in explicitly modeling the process of the target variable changing over time in the training stage to enable the model to dynamically adjust and adapt to the evolutionary trend of labels in the testing stage.
3) To solve the issue of non-deterministic covariate shift, prototype-assisted fuzzy clustering method and membership-aware domain fuzzy matching method are successively adopted, which have the advantage of describing and learning the multilevel categorical relations between the source and target domains during the dynamic adaptation process, thus avoiding the issue of negative transfer.
4) Extensive experiments conducted on eight synthetic datasets and three real-world datasets confirm that the proposed DD-VAN model outperforms state-of-the-art algorithms in its category. Additionally, findings from the ablation study and stability test provide empirical evidence supporting the underlying principles of variational inference and fuzzy logic utilized in the DD-VAN model.

The remaining sections of this paper are organized as follows: In Section II, related works regarding UDA, FDA, and Evolving DA (EDA) are discussed. Section III presents a challenge analysis and problem definition in relation to the DD-VAN. In Section IV, the implementation procedure of each module within the DD-VAN is introduced in detail. Section V presents the experimental results and relevant analysis. Conclusions are drawn in Section VI. Finally, an appendix displaying additional experimental results is provided.

II. Related Work

A. Unsupervised Domain Adaptation

The task setting of UDA assumes that the labeled source domain and the unlabeled target domain exhibit different feature distributions, while their learning objectives remain the same. To enhance transfer performance, existing UDA algorithms have addressed distribution differences in various ways,

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

broadly categorized into: (1) metric learning, (2) adversarial training, and (3) zero-shot generation. Metric learning aims to mitigate domain shift by adjusting statistical moments, such as MMD and its variants [36], Kullback-Leibler divergence [15], and entropy optimization [30]. For instance, Sun et al. [38] introduced deep CORAL, which aligns the second-order statistics (correlation) of source and target distributions to minimize domain shift.

Adversarial training, on the other hand, employs a domain discriminator and a gradient inversion layer to achieve UDA without relying on an explicit discrepancy metric [8]. Additionally, zero-shot domain adaptation (ZSDA) represents a special case of UDA, where the target-domain data for the task of interest are unavailable. A classic work in ZSDA is [23], which proposed a novel coupled generative adversarial network. The CoGAN model learns the joint distribution of multi-domain images by enforcing a simple weight-sharing constraint, allowing it to generate pairs of images in different domains that share the same high-level features.

Traditional UDA algorithms primarily focus on adapting marginal distributions, often neglecting category information. More recent UDA algorithms, however, have introduced fine-grained discriminative information into optimization objectives to avoid negative transfer. This is typically achieved through subdomain alignment, also known as conditional distribution matching [25], [54]. Nevertheless, these algorithms often overlook the challenges posed by imprecise and non-stationary data.

B. Fuzzy Domain Adaptation

FDA has emerged as a well-advanced framework in recent years, integrating fuzzy rules or fuzzy relations into classical DA models to address uncertainty in the knowledge transfer process. Preliminary attempts have been made by introducing Takagi-Sugeno fuzzy rules [53], [57] to explore transfer uncertainty caused by unlabeled data in target domains [18], [27], [31], [37], [48]. For instance, Xu et al. [48] presented a transfer representation learning approach using the Takagi-Sugeno-Kang fuzzy system, which facilitates distribution discrepancy minimization and feature transformation through fuzzy mapping. Li et al. [18] proposed a source-free multi-domain adaptation method based on fuzzy rule-based deep neural networks. Additionally, Ma et al. [31] introduced a combination of Takagi-Sugeno fuzzy rules with a self-supervised pseudo-labeling strategy to address uncertainties in both source and target domains.

Most existing FDA algorithms have been designed for deterministic and stationary distributions, meaning that the concept drift issue remains unsolved. In this paper, a novel DD-VAN algorithm is proposed to simultaneously tackle non-deterministic covariate shift and evolving concept drift under fuzzy classification decision boundaries. First, an intra-domain evolutionary trend modeling module is designed to model the changing dynamics of temporal features in the latent space using variational inference. Second, a prototype-assisted fuzzy clustering module is used to estimate pseudo-labels for target conditional distributions using fuzzy logic to describe the degree of multi-level information belonging to multiple categories. Finally, a membership-aware domain fuzzy matching module is proposed to learn fuzzy decision boundaries between different conditional distributions via FMMD to capture domain invariant features from non-deterministic and non-stationary data streams.

C. Evolving Domain Adaptation

Previous UDA methods generally assume that training and test data are derived from static domains with different distributions but the same semantics, which often does not hold in real-world applications where environments and operation conditions evolve unpredictably. The above phenomenon refers to concept drift, which can be formalized as:

$$\exists X : P_t\left(X, y\right) \neq P_{t+1}\left(X, y\right) \tag{1}$$

where t represents time, X is feature space, y is label, $P_t\left(X, y\right)$ and $P_{t+1}\left(X, y\right)$ denote the data distributions at time t and $t+1$, respectively. By definition, the key to solving the concept drift issue is to enable the model to continuously adapt to varying data distributions in the time dimension. Evolving domain adaptation (EDA), also known as continuous domain adaptation, aims to address knowledge transfer for continuously varying domains over time. Most existing EDA approaches can be broadly categorized into four main groups.

a. *Gradual self-training* resorts to an extra sequence of continuous unlabeled samples as intermediate domains to adapt the source classifier to the target domain by self-training instead of feature alignment. For example, Kumar et al. [14] established the first non-vacuous upper bound on self-training error under gradual shifts and demonstrated that regularization and label sharpening are crucial even with infinite data. However, gradual self-training suffers from two limitations. Firstly, determining the appropriate number of intermediate domains remains an open question. Secondly, self-training using multiple pseudo-labeling iterations may lead to unstable training.
b. *Domain manifold* treats each target sample as potentially coming from a different subspace on the domain manifold and aims to learn continuous manifolds of the evolving target domain. For example, Hoffman et al. [9] introduced a continuous manifold adaptation algorithm to tackle the problem where test data not only differ from training data but also evolve continuously. Nevertheless, the issue of fuzzy class distribution decision boundaries, which is the concern of this paper, may not be applicable to manifold learning, as it assumes that the data are distributed on a low-dimensional manifold that is characterized by relatively clear and consistent boundaries of the data.
c. *Incremental learning* simplifies concept drift by dividing the domain alignment process into smaller incremental transfer processes. For instance, Korycki et al. [13] proposed a centroid-driven experience replay method with a reactive subspace buffer that preserves relevant information and adapts to changing category distributions. Incremental learning requires models to be able to quickly adapt to new data while maintaining memory and adaptability to old data in an ever-changing domain,

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

which can be negatively affected by the issue of catastrophic forgetting.

d. *Meta-learning* aims to learn adaptable representations during the training phase, thus new target data that is continuously changing can be adapted by adapters during the meta-testing phase without forgetting the previous target. For example, Wu et al. [47] proposed an adaptive compositional continuous meta-learning algorithm that utilizes a compositional framework to share meta-knowledge across heterogeneous tasks for continuous adaptation. Meta-learning methods typically assume that training and test tasks share a prior distribution, however, if concept drift exceeds this assumption, the model may fail to generalize.

In contrast to the approaches mentioned above, our research focuses on continuous adaptation to real-time data streams, which presents a more challenging yet realistic task for real-world applications. This task requires the model to be dynamically updated in response to the constant influx of new data, necessitating the development of a methodology that can immediately incorporate valid knowledge of new behaviors without the need for extensive retraining.

III. Challenge Analysis and Problem Definition

A. Challenge Analysis

In this section, we will analyze the research approaches to address the aforementioned challenges.

For the first challenge, data uncertainty typically arises due to overlapping categories or the presence of noise in the data. This type of uncertainty is inherent in the data distribution itself and is not a property of the model, making it irreducible. Therefore, rather than attempting to eliminate these uncertainties, it is essential for the proposed method to adapt to them. A potential solution involves incorporating uncertainty estimation into the pseudo-label generation process, and then using these pseudo labels to provide fine-grained discriminant information for subdomain matching. Based on this analysis, we intuitively employ fuzzy logic to generate pseudo labels, which are subsequently used to represent overlapping decision boundaries.

For the second challenge, addressing the non-stationary problem in UDA requires continuous learning and adaptation to new data streams while retaining valuable information from the initial training phase. Incremental learning is a common technique used in existing UDA methods to address the concept drift problem. However, the drawbacks of incremental learning (such as catastrophic forgetting and high computational resource consumption) significantly hinder model performance. To overcome these limitations, we propose decoupling the evolving factors from the domain and modeling their prior distribution during the training phase. In the testing phase, time-dependent variables can then be extracted from this distribution and combined with time-constant features (learned by a shared feature extractor) to obtain the final outcomes.

B. Problem Definition

In this section, the problem of the FDA is first formulated. Consider a set of related domains $\{\mathcal{S}, \mathcal{T}\}$, where $\mathcal{S}$ and $\mathcal{T}$ represent source domain and target domain, respectively. In the source domain $\mathcal{S}$, each data-label pair in the sample space follows the joint distribution $P(X^s, Y^s)$, where X^s and Y^s denote the source feature and source label, respectively. In the target domain $\mathcal{T}$, each data point in the sample space follows the marginal distribution $P(X^t)$, where X^t represents the target feature.

Suppose that we are given T sequentially arriving source samples $\mathbf{x}^s_{1:T} = \{\mathrm{x}^s_1, \mathrm{x}^s_2, \ldots, \mathrm{x}^s_T | \mathrm{x}^s_t \in \mathbb{R}^{M\times N}\}$, where M represents the number of series, N is the length of each series, and $t \in \{1, 2, \ldots, T\}$. Conventional DA methods only address the issue of covariate shift across domains (i.e., when $P(X)$ varies), which may result in suboptimal performance when dealing with evolving covariate shift and concept drift (i.e., when both $P(X)$ and $P(Y|X)$ vary over time). As a result, the goal of FDA is to train a classification model on the source domain $\mathcal{S}$ that can generalize to L sequentially arriving target samples $\mathbf{x}^t_{1:L} = \{\mathrm{x}^t_1, \mathrm{x}^t_2, \ldots, \mathrm{x}^t_L | \mathrm{x}^t_\delta \in \mathbb{R}^{M\times N}\}$ by exploring the evolving patterns of covariate shift across domains and concept shift across samples, where $\delta \in \{1, 2, \ldots, L\}$.

To ensure the solvability of the above problem setting, we characterize the evolving nature of two sequential source samples by defining the distance between them as $0 \leq \mathrm{Dis}(\mathrm{x}^s_t, \mathrm{x}^s_{t+1}) \leq \epsilon$ under some distribution distance function Dis (e.g., Kullback-Leibler divergence, Jensen-Shannon divergence, MMD), where ϵ is a constant with a finite value.

IV. A Drift Decoupling-Based Variational Adaptation Network

In this section, a novel DD-VAN algorithm for FDA is introduced, which aims to learn evolving, transferable, and discriminative features through three customized modules. The first module, called *the intra-domain evolutionary trend modeling module*, is designed to train a time-adaptive model by capturing unknown variations over time in underlying data distributions using variational inference. The second module, referred to as *the prototype-assisted fuzzy clustering module*, is employed to derive membership degree information, allowing the model to learn the inherent uncertainty and imprecision present in real-world distributions. The third module, termed *the membership-aware domain fuzzy matching module*, focuses on learning the gradual transitions between category-related data pairs in the source and target domains by introducing fuzzy weight vectors, enabling a more nuanced adaptation process.

A. Intra-Domain Evolutionary Trend Modeling Module

Variational inference can disentangle time-invariant and time-variant variables under more relaxed constraints, enabling the exploration of underlying factors associated with concept drift. Therefore, the module described in this subsection is designed to achieve robust adaptation to evolving data distributions through variational inference. The architecture of the intra-domain evolutionary trend modeling module is illustrated in Fig. 1.

To model the evolving nature of source samples at different time stamps, we consider two independent factors, C and V, which represent the static behavior in the sample space and the dynamic behavior in the category space (i.e., concept drift),

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

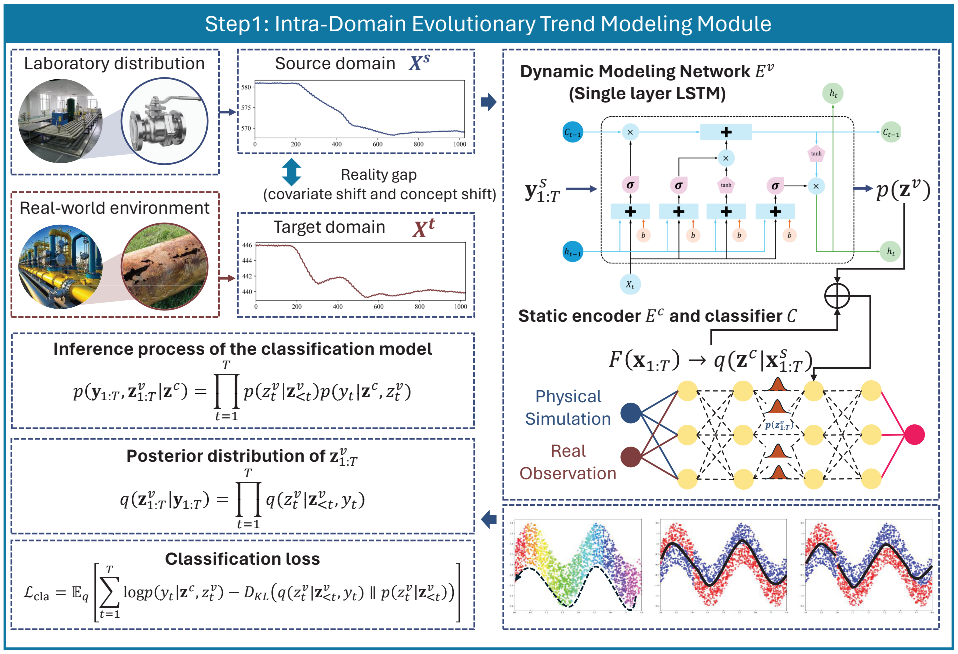


Fig. 1: Architecture of intra-domain evolutionary trend modeling module. The dynamic modeling network takes $\mathbf{y}_{1:T}$ as input and outputs $q(\mathbf{z}^v_{1:T}|\mathbf{y}_{1:T})$ to determine $p(\mathbf{z}^v)$. In addition, the classifier outputs the predicted label $\hat{\mathbf{y}}_{1:T}$ based on the latent variables $\mathbf{z}^v$ and $\mathbf{z}^c$. Notably, only source-domain data are available during the modeling stage, with the aim of generalizing the model well to unseen target domains.

respectively. Specifically, for the data-label pair $\{\mathrm{x}^s_t, \mathrm{y}^s_t\}$ that arrives at time stamp t, it can be decoupled into time-constant latent variables z^c_t for C and time-variant latent variables z^v_t for V. For simplicity, the notation s is omitted when modeling concept drift.

The inference process of the classification model can be formulated using a Markov chain model as follows:

$$p(\mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}|\mathbf{z}^c) = \prod_{t=1}^{T} p(\mathrm{z}^v_t|\mathbf{z}^v_{<t})p(\mathrm{y}_t|\mathbf{z}^c, \mathrm{z}^v_t), \tag{2}$$

where $p(\mathrm{z}^v_t|\mathbf{z}^v_{<t}) = \mathrm{Cat}\left(\pi\left(\mathbf{z}^v_{<t}\right)\right)$ is a learnable categorical distribution indicating that $\mathbf{z}^v_{<t}$ is sampled from a finite set of categories with probability determined by $\pi(\cdot)$. The term $p(\mathrm{y}_t|\mathbf{z}^c, z^v_t)$ is the classification model, which addresses covariate shift through $\mathbf{z}^c$ and concept drift through z^v_t for FDA.

The variables $\mathbf{z}^c$ and $\mathbf{z}^v$ are inferred from the observable data as $p(\mathbf{z}^c|\mathbf{x})$ and $p(\mathbf{z}^v|\mathbf{y})$, respectively. Consequently, the joint distribution of $\mathbf{z}^c$ and $\mathbf{z}^v$ can be described as

$$p(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T}) = p(\mathbf{z}^c|\mathbf{x}_{1:T})p(\mathbf{z}^v_{1:T}|\mathbf{y}_{1:T}). \tag{3}$$

To approximate the prior distribution $p(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})$, a learnable posterior distribution $q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})$ over the latent variables of provided source-domain data is introduced using variational inference, which allows for the calculation of the Kullback-Leibler Divergence described as follows:

$$\begin{aligned} D_{\mathrm{KL}}(q\|p) =& \mathbb{E}_q\left[\log\frac{q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}{p(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}\right] \\ =& \mathbb{E}_q\left[\log\frac{q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}{p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}, \mathbf{z}^c)}p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T})\right] \\ =& \mathbb{E}_q\left[\log\frac{q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}{p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}, \mathbf{z}^c)}\right] \\ &+ \log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}). \end{aligned} \tag{4}$$

Based on (4), we obtain

$$\begin{aligned} \log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}) =& D_{\mathrm{KL}}(q\|p) \\ &+ \mathbb{E}_q\left[\log\frac{p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}, \mathbf{z}^c_{1:T})}{q(\mathbf{z}^c_{1:T}, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}\right]. \end{aligned} \tag{5}$$

As $D_{\mathrm{KL}}(q\|p) \geq 0$, the variational lower bound for $\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T})$ is

$$\mathcal{L} = \mathbb{E}_q\left[\log\frac{p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}, \mathbf{z}^c)}{q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}\right], \tag{6}$$

and the above formulation can be reorganized as

$$\begin{aligned} \mathcal{L} =& \mathbb{E}_q\left[\log\frac{p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}^v_{1:T}, \mathbf{z}^c)}{q(\mathbf{z}^c, \mathbf{z}^v_{1:T}|\mathbf{x}_{1:T}, \mathbf{y}_{1:T})}\right] \\ =& \mathbb{E}_q\left[\log\frac{\prod_{t=1}^{T} p\left(\mathrm{x}_t, \mathrm{y}_t|\mathbf{z}^c, \mathrm{z}^v_t\right) p\left(\mathbf{z}^c\right) p\left(\mathrm{z}^v_t|\mathbf{z}^v_{<t}\right)}{\prod_{t=1}^{T} q\left(\mathbf{z}^c|\mathrm{x}_t\right) q\left(\mathrm{z}^v_t|\mathbf{z}^v_{<t}, \mathrm{y}_t\right)}\right] \\ =& \mathbb{E}_q\left[-\sum_{t=1}^{T}\log\frac{q\left(\mathbf{z}^c|\mathrm{x}_t\right)}{p\left(\mathbf{z}^c\right)} - \sum_{t=1}^{T}\log\frac{q\left(\mathrm{z}^v_t|\mathbf{z}^v_{<t}, \mathrm{y}_t\right)}{p\left(\mathrm{z}^v_t|\mathbf{z}^v_{<t}\right)}\right. \\ &\left.+\sum_{t=1}^{T}\log\left(p\left(\mathrm{x}_t|\mathbf{z}^c\right) p\left(\mathrm{y}_t|\mathbf{z}^c, \mathrm{z}^v_t\right)\right)\right]. \end{aligned} \tag{7}$$

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

Since $D_{\mathrm{KL}}(q\|p) = \sum_x q(x)\log\frac{q(x)}{p(x)}$, (7) can be further reformulated by using Jensen's inequality as follows:

$$\begin{aligned}\mathcal{L} \geq & \mathbb{E}_q\Bigg[\sum_{t=1}^{T}\log\left(p\left(\mathrm{x}_t|\mathbf{z}^c\right)p\left(\mathrm{y}_t|\mathbf{z}^c,\mathrm{z}_t^v\right)\right) \\ & - D_{\mathrm{KL}}\left(q\left(\mathbf{z}^c|\mathrm{x}_t\right)\|p\left(\mathbf{z}^c\right)\right) \\ & - D_{\mathrm{KL}}\left(q\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v,\mathrm{y}_t\right)\|p\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v\right)\right)\Bigg]. \end{aligned} \tag{8}$$

Based on the above derivation, we are ready to give the Evidence Lower Bound (ELBO) as follows:

$$\begin{aligned}\mathcal{L} = & \mathbb{E}_q\Bigg[\sum_{t=1}^{T}\log\left(p\left(\mathrm{x}_t|\mathbf{z}^c\right)p\left(\mathrm{y}_t|\mathbf{z}^c,\mathrm{z}_t^v\right)\right) \\ & - D_{\mathrm{KL}}\left(q\left(\mathbf{z}^c|\mathrm{x}_t\right)\|p\left(\mathbf{z}^c\right)\right) \\ & - D_{\mathrm{KL}}\left(q\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v,\mathrm{y}_t\right)\|p\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v\right)\right)\Bigg], \end{aligned} \tag{9}$$

where $\mathcal{L}$ consists of the classification loss $\mathcal{L}_{\mathrm{cla}}$ and the reconstruction loss $\mathcal{L}_{\mathrm{res}}$. Here, $\mathcal{L}_{\mathrm{cla}}$ is formulated as

$$\begin{aligned}\mathcal{L}_{\mathrm{cla}} = & \mathbb{E}_q\Bigg[\sum_{t=1}^{T}\log p\left(\mathrm{y}_t|\mathbf{z}^c,\mathrm{z}_t^v\right) \\ & - D_{\mathrm{KL}}\left(q\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v,\mathrm{y}_t\right)\|p\left(\mathrm{z}_t^v|\mathbf{z}_{<t}^v\right)\right)\Bigg], \end{aligned} \tag{10}$$

and $\mathcal{L}_{\mathrm{res}}$ is described as

$$\mathcal{L}_{\mathrm{res}} = \mathbb{E}_q\left[\sum_{t=1}^{T}\log p\left(\mathrm{x}_t|\mathbf{z}^c\right) - D_{\mathrm{KL}}\left(q\left(\mathbf{z}^c|\mathrm{x}_t\right)\|p\left(\mathbf{z}^c\right)\right)\right]. \tag{11}$$

Remark 1: It is important to note that the intra-domain evolutionary trend modeling module learns the evolutionary patterns between samples exclusively on the source domain, without involving the target domain. This approach considers two potential scenarios: (1) the target domain arrives sequentially after the source domain, and (2) the target domain exists independently of the source domain in terms of a sample-time relationship. Concept drift in either scenario can be effectively addressed by the generalizable feature extractor and classifier, which are trained to adapt to evolving patterns across domains.

B. Prototype-Assisted Fuzzy Clustering Module

Previous works on UDA have utilized pseudo labels obtained through methods such as self-training [21], k-nearest neighbor search, and clustering algorithms [43] for training. These approaches typically employ an alternative optimization strategy, involving two stages: a label generation stage that assigns pseudo labels, and a training stage that uses these pseudo labels to train a generalizable model. However, inherent noise and error accumulation during pseudo label generation can significantly degrade the performance of these unsupervised methods. To address this issue, the PFC method is proposed, which leverages fuzzy set theory to assign each data point to multiple categories with varying degrees of membership, thereby enhancing robustness against random noise and label errors. In addition, PFC provides detailed membership information, which is crucial for domain matching. The architecture of the prototype-assisted fuzzy clustering module is depicted in Fig. 2.

In this work, the clustering prototypes $\left\{v_{k,0}^t\right\}_{k=1}^K$ of the target domain are initialized as the mean values of the samples from the same grouping category in the source domain, which can be expressed as:

$$v_{k,0}^t = \frac{\sum_{i=1}^{n_k}\mathbb{I}_{y_i^s\in Y_k}q(\mathbf{z}^c|x_i^s)}{\sum_{i=1}^{n_k}\mathbb{I}_{y_i^s\in Y_k}}, \tag{12}$$

where K is the category number; $x_i^s \in \mathbb{R}^{T\times M}$ represents each source sample in the training set without considering time t; $v_{k,0}^t$ denotes initialized clustering prototype for target domain in first epoch; and n_k denotes the number of source samples in Y_k.

The membership degree of target samples $x_i^t \in \mathbb{R}^{L\times M}$ is defined as

$$u_{i,k}^t = \frac{1}{\sum_{j=1}^{K}\left(\frac{\|q(\mathbf{z}^c|\mathbf{x}_{1:L}^t)-v_k^t\|}{\|q(\mathbf{z}^c|\mathbf{x}_{1:L}^t)-v_j^t\|}\right)^{\frac{2}{m-1}}}, \tag{13}$$

where m is the fuzzy factor greater than 1; $q\left(\mathbf{z}^c|\mathbf{x}_{1:L}^t\right)$ is updated on the basis of the feature extractor. After the first epoch, the clustering prototypes $\{v_k^t\}_{k=1}^K$ are updated as the training process continues using the following equation:

$$v_k^t = \frac{\sum_{i=1}^{n_t}\left(u_{i,k}^t\right)^m q\left(\mathbf{z}^c|x_i^t\right)}{\sum_{i=1}^{n_t}\left(u_{i,k}^t\right)^m} \tag{14}$$

where $n_t = L\times M$ denotes the number of total target samples.

Different from the conventional fuzzy C-means algorithm, which updates clustering centers based on a distance function, the PFC method is designed to capture continuous and gradual transitions between different categories of distributions in the shared feature space by utilizing the global distribution information contained in adaptive prototypes. First, the categorical distribution $P^{s\to t}$ is obtained using source prototypes as follows:

$$P_{i,k}^{s\to t} = \frac{\exp\left(\left(q\left(\mathbf{z}^c|x_i^t\right)\cdot v_k^s\right)/\tau\right)}{\sum_{j=1}^{K}\exp\left(\left(q\left(\mathbf{z}^c|x_i^t\right)\cdot v_j^s\right)/\tau\right)}, \tag{15}$$

where $q\left(\mathbf{z}^c|x_i^t\right)$ can be parameterized by neural networks; τ represents the temperature coefficient (defaulting to 1) and "$\cdot$" denotes the inner product.

The term v_k^s represents the k-th category prototype derived from the common feature space and is defined as follows:

$$v_k^s = \frac{\sum_{i=1}^{n_k}\mathbb{I}_{y_i^s\in Y_k}q(\mathbf{z}^c|x_i^s)}{\sum_{i=1}^{n_k}\mathbb{I}_{y_i^s\in Y_k}}. \tag{16}$$

Furthermore, the fuzzy categorical distribution P^t is obtained using clustering prototypes as follows:

$$P_{i,k}^t = \frac{\exp\left(\left(q\left(\mathbf{z}^c|x_i^t\right)\cdot v_k^t\right)/\tau\right)}{\sum_{j=1}^{K}\exp\left(\left(q\left(\mathbf{z}^c|x_i^t\right)\cdot v_j^t\right)/\tau\right)}. \tag{17}$$

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

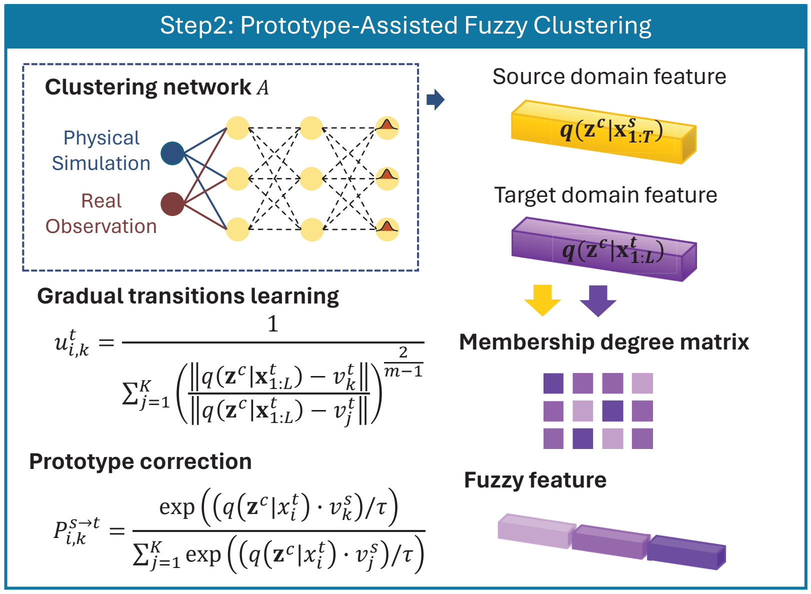


Fig. 2: Architecture of prototype-assisted fuzzy clustering module.

As the PFC method is less sensitive to initial source annotation errors, the optimal solution provided by the source prototype becomes comparable to a fully supervised scenario in eliminating cross-domain covariate shift. To further stabilize the learning process of the fuzzy clustering module, a prototype-driven objective function is proposed as follows:

$$\mathcal{L}_{\text{pfc}} = -\sum_{i=1}^{n_t}\sum_{k=1}^{K} u^t_{i,k}\zeta\left(P^{s\to t}_{i,k}\right)\log\left(P^t_{i,k}\right), \tag{18}$$

where $\zeta\left(P^{s\to t}_{i,k}\right)$ represents the conversion from predicted probabilities to soft targets, which is formulated as follows

$$\zeta\left(P^{s\to t}_{i,k}\right) = \begin{cases} 1 & \text{if } k = \arg\max_{k'} P^{s\to t}_{i,k} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Remark 2: The primary role of the PFC method is to generate a set of rough solutions that drive the model forward through multiple epochs. The DD-VAN model then leverages these rough solutions to obtain a more precise set of solutions by training the feature extractor and classifier. This group of refined solutions serves two purposes: it not only facilitates further training of the overall model at the end of the i-th epoch but also updates the clustering prototypes of the PFC for the subsequent epoch. We refer to this iterative approach as the "prediction-correction process." In this process, the rough solutions generated by the PFC in the early stages are progressively refined through repeated predictions and corrections, leading to more accurate domain adaptation and model performance over time.

C. Membership-Aware Domain Fuzzy Matching Module

Most existing UDA algorithms focus on minimizing conditional distribution discrepancies across domains to address covariate shift, operating under the assumption of deterministic category relationships. However, in real-world distributions, complex relationships, nonlinearities, and intricate patterns often lead to imprecise features and category uncertainties due to gradual transitions caused by external environmental factors or internal changes. Under such conditions, the model's transferability and discriminative ability can stagnate or even degrade. To overcome this limitation, the *membership-aware domain fuzzy matching* method is introduced, which mitigates the impact of covariate shift by ensuring the model updates the most confident fuzzy cluster embeddings. By incorporating fuzzy matching, the model can account for the uncertainties inherent in real-world data, enhancing its adaptability and robustness in non-deterministic environments.

The architecture of membership-aware domain fuzzy matching module is shown in Fig. 3. This module helps maintain model performance by refining the domain matching process through the use of fuzzy logic, allowing for smoother transitions between different categories and more accurate cross-domain adaptation.

Membership-aware domain fuzzy matching is achieved through FMMD, which compensates the information of fuzzy category boundaries for the distribution alignment via the fuzzy weight vectors generated by the PFC. The formulation of FMMD is expressed as

$$\begin{aligned}
&\mathrm{FMMD}^2\left(\mathcal{S},\mathcal{T};\mathbf{k}\right) \\
=&\frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{n_s(n_s-1)}\sum_{i=1}^{n_s}\sum_{j\neq i}^{n_s}\mathbf{k}\left(x^s_{i,k},x^s_j\right)\right. \\
&\qquad+\frac{1}{n_t(n_t-1)}\sum_{i=1}^{n_t}\sum_{j\neq i}^{n_t}u^t_{i,k}u^t_{j,k}\mathbf{k}\left(x^t_{i,k},x^t_j\right) \\
&\qquad\left.-\frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j\neq i}^{n_t}u^t_{j,k}\mathbf{k}\left(x^s_{i,k},x^t_j\right)\right],
\end{aligned} \tag{20}$$

where $\mathbf{k}$ is a kernel function. Note that (13) helps to match similar conditional distributions across domains when consid-

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

ering membership uncertainty in the target domain.

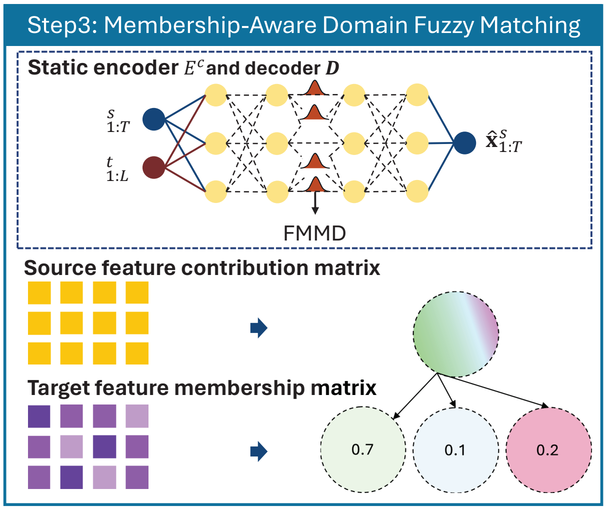


Fig. 3: The architecture of membership-aware domain fuzzy matching module.

D. Optimization

The network framework of the DD-VAN consists of two stages: (1) the training stage composed of the evolutionary trend modeling module and the membership-aware domain fuzzy matching module; and (2) the label generation stage composed of the prototype-assisted fuzzy clustering module. This two-stage approach enables the DD-VAN to learn both evolving trends and handle category uncertainties effectively, resulting in improved domain adaptation and classification accuracy.

The training stage consists of a static variational encoder E^c for $q\left(\mathbf{z}^c|\mathbf{x}^s_{1:T}\right)$, a decoder D responsible for $p\left(\mathrm{x}^s_t|\mathbf{z}^c\right)$, a dynamic modeling network E^v which takes the one-hot label as input and outputs the categorical distribution $q(\mathrm{z}^v_t|\mathbf{z}^v_{<t}, \mathrm{y}_t)$, a classifier C that takes $\mathbf{z}^c$ and $\mathbf{z}^v$ as input to output label for $p(\mathrm{y}_t|\mathbf{z}^c, z^v_t)$, a dynamic prior network F^v working for $p(\mathrm{z}^v_t|\mathbf{z}^v_{<t})$. Specifically, the encoder E^c, based on one-dimensional convolutional neural networks (1DCNN), is used as a feature extractor to capture domain-invariant features during the training and testing stages. The decoder D mirrors the architecture of the encoder. E^v is implemented using a single-layer LSTM network and several linear layers, while the classifier C is a single linear layer. The dynamic prior network F^v is also an LSTM network, which outputs categorical distribution $\mathrm{Cat}\left(\pi\left(\mathbf{z}^v_{<t}\right)\right)$.

Similar to VAE, the reparameterization trick is used to update E^c, E^v, and F^v. Based on (5), (6), and (20), the optimization objective for the training module is defined as

$$\mathcal{L}_{\mathrm{train}} = \mathcal{L}_{\mathrm{res}} + \mathcal{L}_{\mathrm{cla}} + \lambda_1 \mathrm{FMMD}^2\left(\mathcal{S}, \mathcal{T}; \mathbf{k}\right), \tag{21}$$

where λ_1 is a trade-off parameter that balances the contributions of the various loss terms.

The prototype-assisted fuzzy clustering module is implemented by clustering network A, which is constructed based on E^c with frozen network parameters and multiple affine transformation layers.

Algorithm 1 Optimization Procedure of the DD-VAN Model

Input: Sequential source samples $\mathbf{x}^s_{1:T}$ and target samples $\mathbf{x}^t_{1:L}$; source labels $\mathbf{y}^s_{1:T}$; static feature extractor E^c; dynamic modeling network E^v and its corresponding dynamic prior network F^v; classifier C, decoder D, and clustering network A.
Output: Static feature extractor E^c, dynamic prior network F^v, and classifier C.

Initialize E^c, E^v, F^v, C, D, A
Initialize clustering prototypes $v^t_{k,0}$ for initial epoch by (12)
Assign $\mathbf{u} \leftarrow 1$
for $(i_0 \leftarrow 1; i_0 \leq M_0; i_0 \leftarrow i_0 + 1)$ **do**
  Generate prior distribution $p(\mathrm{z}^v_t|\mathbf{z}^v_{<t})$ through F^v
  for $(i_1 \leftarrow 1; i_1 \leq M_1; i_1 \leftarrow i_1 + 1)$ **do**
    Train E^c, D, F^v, C, E^v with mini-batches from $\{\mathcal{S}, \mathcal{T}\}$ to minimize $\mathcal{L}_{\mathrm{train}}$.
  end for
  for $(i_2 \leftarrow 1; i_2 \leq M_2; i_2 \leftarrow i_2 + 1)$ **do**
    Calculate membership degree by (13)
    Calculate clustering prototypes by (14)
    Update clustering network by minimizing $\mathcal{L}_{\mathrm{pfc}}$
  end for
end for

V. Experiments

In this section, the performance of the proposed DD-VAN algorithm is evaluated on both synthetic and real datasets and compared with state-of-the-art algorithms.

A. Implementation Details

1) **Model Architecture and Hyperparameters**. The neural network architectures for the synthetic and real datasets are given in Tabs. I and II. Both classifiers for synthetic and real datasets are implemented by fully connected linear layers followed by a softmax function. All models in this paper are optimized by the Adam optimizer. For synthetic datasets, the hyperparameters used in this paper include: learning rate for E^c, C, D, and A is $3e-5$, learning rate for E^v and F^v is $2e-6$, batch size is 64, and λ_1 is 0.2. For real datasets, the hyperparameters used in this paper include: learning rate for E^c, C, D, and A is $1e-4$, learning rate for E^v and F^v is $1e-5$, batch size is 64, and λ_1 is 0.2.
2) **Comparison Algorithms**. The advanced comparison algorithms used in this paper include: (1) ERM; (2) DTL [46]; (3) Deep Coral [38]; (4) DASAN [24]; (5) DRMEA [29]; (6) IAST [33]; and (7) DIVA [16]. All experiments are implemented using PyTorch on NVIDIA GEFORCE RTX 3090 GPU, Intel(R) Core(TM) i9-10900k, 3.70-GHz CPU. For a fair comparison, all the same parts of the baseline network architecture are identical for different benchmarks.

B. Experimental Setting

To gain insight into the performance improvement of the DD-VAN algorithm, we first conduct experiments on two

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

TABLE II: Implementation of ConvNet and ConvTranNet.

#	Encoder Layer	Decoder Layer
1	Conv1d(in=1, out=32)	Linear(in=128, output=256)
2	ReLU	ReLU
3	MaxPool1d(kernel_size=2, stride=2)	Linear(in=256, output=128*128)
4	Conv1d(in=32, out=64)	Reshape(1, 128, 128)
5	ReLU	ConvTranspose1d(in=128, output=64, kernel=3)
6	MaxPool1d(kernel_size=2, stride=2)	ReLU
7	Conv1d(in=64, out=128)	BatchNorm
8	ReLU	ConvTranspose1d(in=64, output=32, kernel=3)
9	MaxPool1d(kernel_size=2, stride=2)	ReLU
10	Linear(in=128 * 128, output=256)	BatchNorm
11	ReLU	ConvTranspose1d(in=32, output=1, kernel=3)
12	Dropout(0.5)	BatchNorm
13	Linear(in=256, output=128)	—

TABLE I: Implementation of non-linear encoder and decoder.

#	Encoder Layer	Decoder Layer
1	Linear(in=d, output=512)	Linear(in=512, output=128)
2	ReLU	BatchNorm
3	Linear(in=512, output=512)	LeakyReLU(0.2)
4	ReLU	Linear(in=128, output=64)
5	Linear(in=512, output=512)	BatchNorm
6	ReLU	LeakyReLU(0.2)
7	Linear(in=512, output=512)	Linear(in=64, output=d)

synthetic datasets Circle/-C and Sine/-C equipped with manual concept drifts.

(1) Circle/-C [34]. This dataset consists of evolving 30 domains where the instance is distributed in [0,1]. For Circle-C, concept drift is synthesized by changing the center (x_c, y_c) and radius r_c of the decision boundary in a gradual manner over time. The instances inside the circle are categorized as positives, otherwise negatives. The domains in Circle/-C are represented by different colors in the first row of Fig. 4, and the categorical distributions in the second row are represented in red and blue, respectively. During the experiments, the 30 domains are divided into 9, 6, 5, and 10 domains in order of source, target, intermediate, and unknown domains for model training, selection, and testing respectively.

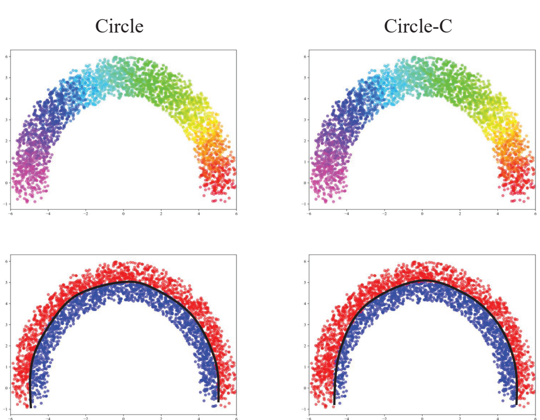


Fig. 4: The visualization of domains and decision boundaries in the Circle/-C.

(2) Sine/-C [26]. This dataset consists of evolving 24 domains where the instance is distributed in [0,1] with classification function $\sin(x)$. Before the first drift, instances under the curve are classified as positive and others as negative. When the drift point is reached, the categorization is reversed. The domains in Sine/-C are represented by different colors in the first row of Fig. 5, and the categorical distributions in the second row are represented in red and blue, respectively. During the experiments, the 24 domains are divided into 6, 6, 4, and 8 domains in order of source, target, intermediate, and unknown domains for model training, selection, and testing respectively.

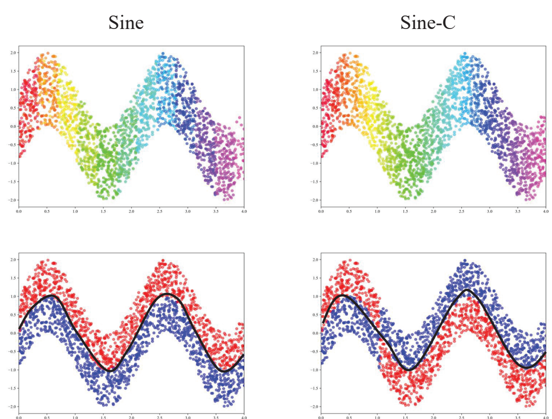


Fig. 5: The visualization of domains and decision boundaries in the Sine/-C.

(a) (b)

Fig. 6: (a) Console. (b) Pipeline platform.

(3) NPWD. To validate the practicability of the algorithm, we deploy natural gas pipeline fault diagnosis as the test platform. Pipeline fault diagnosis plays a crucial role in ensuring high reliability and long-term stability of long-distance oil and gas transportation systems. For intelligent

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

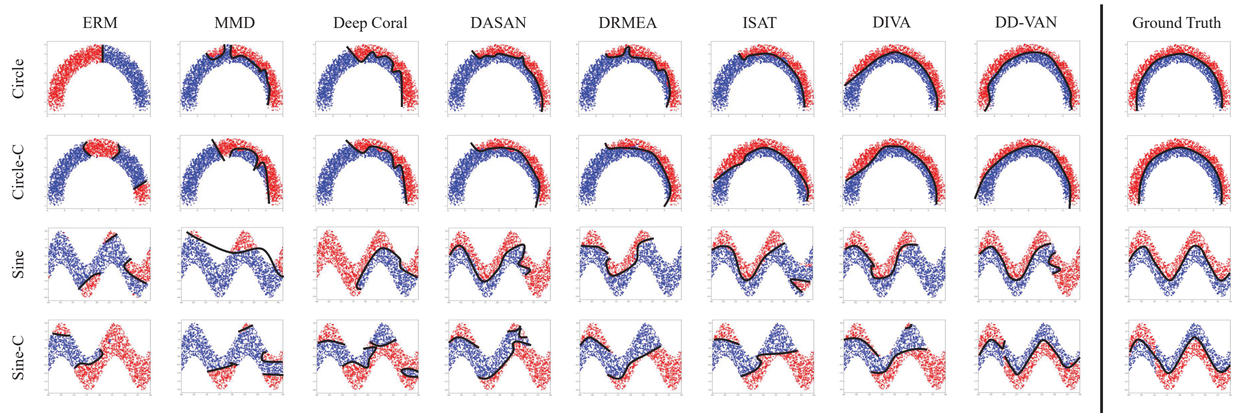


Fig. 7: Decision boundary visualization for the Circle/-C and Since/-C datasets. The last column shows the positive labels in each dataset with red dots, the negative labels with blue dots, and the black lines representing the decision boundaries. The other columns show the results of the baselines and the proposed DD-VAN.

TABLE III: Accuracy results (%) of DD-VAN and other UDA baselines on four synthetic datasets.

Algorithm	**Without Concept Drift**			**With Concept Drift**		
	Circle	**Sine**	**Avg.**	**Circle-C**	**Sine-C**	**Avg.**
ERM	56.25	51.26	53.76	56.16	54.42	55.29
MMD	70.31	68.75	69.53	69.70	64.31	69.21
Deep Coral	71.88	68.49	70.18	70.83	71.53	67.57
DASAN	73.44	69.87	71.66	74.23	71.88	73.05
DRMEA	73.25	72.01	72.43	76.54	72.30	74.42
IAST	75.00	73.61	74.31	75.31	73.52	74.42
DIVA	76.56	75.99	76.28	76.74	75.28	76.01
DD-VAN(ours)	**83.16**	**82.39**	**82.78**	**83.68**	**80.61**	**82.15**

fault diagnosis tasks, obtaining large-scale, well-characterized datasets is not even practical in real industrial scenarios, and most existing diagnostic algorithms tend to suffer from poor accuracy and limited generalization capability [7], [40]. The pipeline datasets used in this paper are obtained from the ZJ-CSGD-type simulation platform, as shown in Fig. 6. The length of the pipeline is $180.2\mathrm{m}$, the flow rate is $10\mathrm{m}^3/\mathrm{h}$, and the sampling frequency is $1024\mathrm{Hz}$. The pipeline dataset contains instances of three working conditions, which can be categorized as high pressure (HP), medium pressure (MP), and low pressure (LP). Instances of each domain can be classified into four categories, including large leakage, medium leakage, small leakage, and normal state. Each condition contains 24 domains, which are divided into 6, 9, and 9 domains in the order of source domain, intermediate domain, and unknown domain for model training, selection, and testing, respectively. To implement the FDA, 6 domains are selected as target domains from another working condition.

C. Performance Analysis on Synthetic Data

The evaluation results of the proposed DD-VAN algorithm and baselines on synthetic data are presented in Tab. III. Considering that the comparison algorithms are not specialized in dealing with concept drift, this paper presents the experimental results in two parts according to concept drift and without concept drift for fairness. As can be seen from the table, the conventional DA algorithms or UDA algorithms with pseudo-labeling strategies cannot perform satisfactorily when dealing with the concept drift issue. Compared to the baselines, our DD-VAN algorithm exhibits stronger advantages in capturing time-dependent long-distance evolving patterns.

To better explore the merits of the proposed DD-VAN algorithm, we visualize the decision boundaries of the baseline and the DD-VAN algorithm on two synthetic datasets. The visualization results are shown in Fig. 7. As can be observed in Fig. 7, the predicted labels depicted in the first column are equivalent to random discrimination, which means that the ERM algorithm without any adaptation operation cannot fit the target domain well. Furthermore, the conventional DA algorithms depicted in the second through fifth columns of Fig. 7 are incapable of dealing with the long-range concept drift issue of regarding labels. However, as for Circle/-C and Sine/-C, our DD-VAN algorithm shows a superior capability in capturing the underlying evolving patterns across sequential time stages.

D. Performance Analysis on Real Data

The diagnosis results of all algorithms are listed in Tab. IV. In summary, the DD-VAN algorithm consistently outperforms other comparison algorithms in a variety of FDA settings, achieving 89.42% average accuracy. Specifically, while ERM adapts well to the source domain, its generalizability to the target domain is compromised due to leaving covariate shift out of consideration. Additionally, the conventional DA algorithms focus solely on tackling the covariate shift issue when concept drift negatively impacts the model's performance, although the ISAT algorithm enhances the transfer performance to a certain extent due to adopting a pseudo-label generation strategy to

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

TABLE IV: Accuracy results (%) of DD-VAN and other UDA baselines on six real-world pipeline evolving datasets.

Algorithm	**Pipeline datasets**						
	HM	**HL**	**MH**	**ML**	**LH**	**LM**	**Avg.**
ERM	68.75	62.50	60.93	62.50	65.63	59.38	63.28
MMD	71.33	73.85	73.18	72.98	72.05	70.63	72.33
Deep Coral	74.15	75.28	71.80	73.05	70.98	75.88	73.52
DASAN	72.00	74.43	73.63	71.90	73.03	75.85	73.47
DRMEA	72.65	74.50	72.35	75.43	75.90	73.93	74.13
IAST	73.60	73.65	76.85	74.75	77.63	75.18	75.28
DIVA	76.56	74.88	78.40	73.53	73.60	75.00	75.33
DD-VAN(ours)	**88.47**	**89.47**	**88.34**	**89.78**	**92.86**	**87.57**	**89.42**

generate reliable pseudo labels, its capability is not sufficient to deal with the uncertainty and imprecision in the real world. Moreover, we remold the DIVA algorithm from a multi-source domain DG task to a UDA task, since it can handle the concept drift problem by learning three independent latent subspaces corresponding to domain-invariant feature, categorical information, and other variation information, respectively. Nevertheless, since the DD-VAN algorithm can explore, adapt, and exploit evolving drift patterns to achieve FDA, it is reasonable to achieve significantly better results than existing UDA methods in pipeline fault diagnosis applications.

E. Parameter Sensitivity Analysis

The loss function for DD-VAN described in (21) deploys a hyperparameter λ_1, which is used to control the contribution of the FMMD. To verify the impact of the hyperparameter on the performance of the proposed algorithm, a sensitivity analysis is implemented by varying the λ_1 value according to $\{0.1, 0.2, 0.3, 0.5, 0.7, 0.9\}$. The experimental results are depicted in Fig. 8, where the dashed line represents the best mean value among baseline methods and the solid line represents the results of our DD-VAN algorithm.

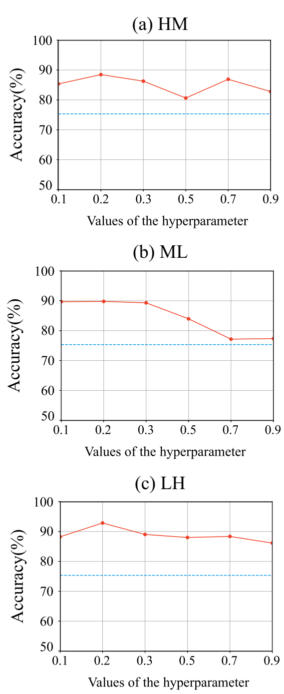


Fig. 8: The hyper-parameter sensitivity analysis. (1) HM; (2) ML; and (3) LH.

It can be noticed that the accuracy trend of DD-VAN fluctuates slightly under different values of λ_1, but it always achieves better or at least competitive results than the optimal baseline. Furthermore, the best performance is achieved when $\lambda_1 = 0.2$.

F. Ablation Study

To investigate the effectiveness of the proposed algorithm in handling FDA tasks, we conduct an ablation study on Sine-C and MH by removing the PFC and FMMD, respectively. The results are shown in Fig. 9. For both MH and Sine-C, removing z^v would decrease the accuracy, which means that z^v can be used to model and learn the evolutionary trend of concept drift. Furthermore, removing PFC and FMMD significantly degrades the performance, which implies that ignoring the uncertainty of real-world data leads to overconfident training.

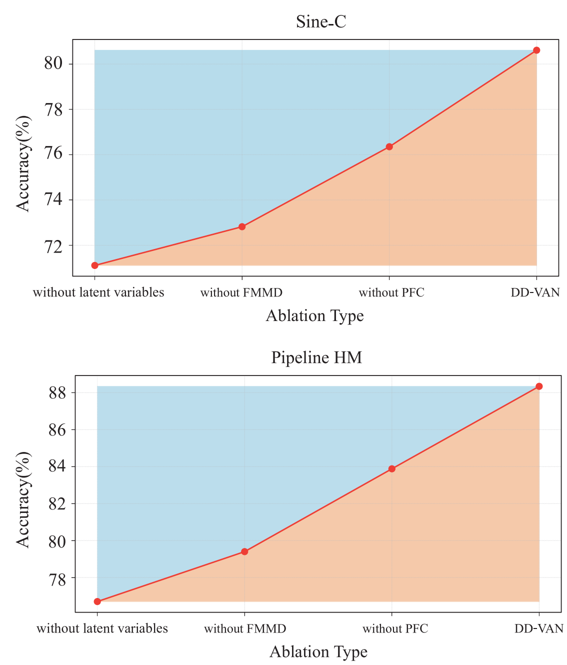


Fig. 9: The ablation study for the z^v, PFC, and FMMD.

VI. Conclusion

In this paper, a realistic and challenging DA setting, FDA, has been investigated to address the issues of covariate shift and concept drift varying over time. To tackle these challenges, the DD-VAN algorithm has been proposed, which learns

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

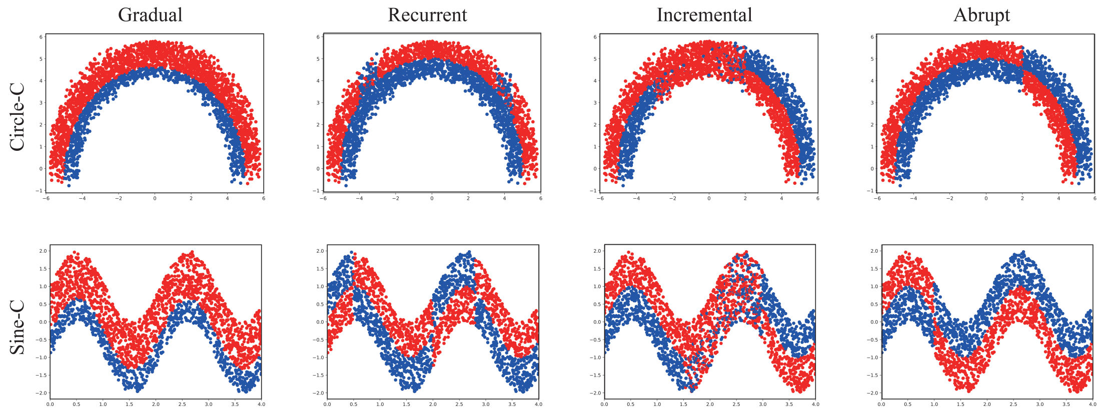


Fig. 10: Visualization of three concept drift types in Circle and Sine synthetic datasets.

time-varying, transferable, and distinguishable domain-sharing features by leveraging variational inference and fuzzy logic. Extensive experiments on both synthetic datasets and real-world pipeline datasets have been conducted, and the results have demonstrated that the proposed DD-VAN algorithm effectively captures evolving patterns and generalizes well to unseen domains compared to state-of-the-art algorithms. In future work, we will focus on introducing advanced control strategies to handle evolving covariate shift and concept drift in more challenging industrial scenarios, such as domain generalization or multi-source domain generalization.

APPENDIX

A. SIMULATION EXPERIMENTS

The implementation details and experimental results of the simulation experiments are specified as follows.

- **Model and method setups:** To reduce computational consumption while maintaining performance, we try to use $4\times$ fully-connected layers in the static variational encoder E^c, decoder D, and clustering network A, single-layer LSTM network and $3\times$ fully-connected layers for dynamic modeling network E^v, $1\times$ fully-connected layers for classifier C, and $5\times$ layers LSTM network for dynamic prior network F^v. The runtime complexity of the DD-VAN used for the synthetic datasets is $\mathcal{O}\left(T \cdot d^2\right)$.
- **Data usage:** We use synthetic datasets Circle and Sine equipped with four types of concept drifts, including gradual drift, recurrent drift, incremental drift, and abrupt drift [52]. Gradual drift implies that both sampling sources are active at a given moment. Over time, the probability of sampling from one source decreases while the probability of sampling from another source increases. Recurrent drift is characterized by a data distribution that returns to its original distribution over time after drifting. Incremental drift represents a change in the data distribution from one to the other over a period of time. Abrupt drift indicates a variation in the data distribution at a precise point in time. The synthetic datasets are shown in Fig. 10.
- **Comparison algorithms:** We use comparison algorithms that include the second-best algorithm DIVA and the third-best algorithm IAST in Tab. III, with the addition of the state-of-the-art MCMO algorithm [11] and CADM+ algorithm [10].
- **Evaluation**: All simulation results are derived by averaging the set of 5 random seeds.
- textbfNumerical results and relevant analysis of simulation experiments: We report the simulated classification results of the DD-VAN and other comparative algorithms under five random seeds in Tab. V. As shown in the Table, DD-VAN always achieves the best performance to different concept drifts, which proves its effectiveness in dealing with changing target variables. Specifically, the DD-VAN exhibits superior performance for gradual and recurrent drifts where evolving patterns significantly exist. In addition, the DD-VAN also achieves sharp performance improvement compared to state-of-the-art algorithms in dealing with incremental and abrupt drifts on Circle and Sine, demonstrating its potential for rapid adaptation to new variations.

B. REAL-WORLD EXPERIMENTS

In addition to the same configuration as the simulation experiments, we summarize the implementation details of the real-world experiments as follows.

- **Model and method setups:** Different from the simulation experiments, we adopt 1D-CNN for the encoder, decoder, and clustering network. Therefore, the runtime complexity is increased to $\mathcal{O}\left(k \cdot T \cdot d^2\right)$, but the balance between computational efforts and performance can also be maintained.
- **Data usage:** We extensively evaluate model performance using the following public datasets to ensure reproducibility.
  a. Jiaolong_DSMS_V2 [10] is collected and provided by the National Deep Sea Center located in Qingdao, Shandong, China. The initial data is collected during the Jiaolong deep-sea manned submersible exploration mission on March 19, 2017. The number of data is 30,000 in the form of multivariate time series with about 24 features. The dataset has three safety levels, including (1) Level I, which indicates no safety risks from the external environment and the system is in a healthy state; (2) Level II, which

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

TABLE V: Classification accuracy (%) and relevant analysis of Circle-C and Sine-C datasets. The best performance is highlighted in **bold**.

Dataset	Shift type	DIVA	IAST	MCMO	CADM+	DD-VAN
Circle-C	Gradual	76.92$\pm$1.27	74.67$\pm$1.42	79.01$\pm$0.31	81.26$\pm$0.59	**83.54**$\pm$0.63
	Recurrent	73.62$\pm$1.64	73.43$\pm$2.33	77.52$\pm$1.69	77.84$\pm$1.40	**81.32**$\pm$0.47
	Incremental	73.07$\pm$1.04	70.51$\pm$2.59	75.18$\pm$0.96	76.24$\pm$0.57	**79.69**$\pm$1.22
	Abrupt	69.53$\pm$2.39	65.68$\pm$2.25	71.56$\pm$1.23	75.18$\pm$0.82	**78.83**$\pm$0.98
Sine-C	Gradual	79.06$\pm$0.73	76.61$\pm$2.21	82.55$\pm$0.56	86.40$\pm$1.53	**90.54**$\pm$1.30
	Recurrent	77.86$\pm$1.56	73.63$\pm$1.55	81.32$\pm$1.32	84.10$\pm$1.36	**85.99**$\pm$0.71
	Incremental	71.41$\pm$2.60	69.68$\pm$3.27	75.15$\pm$0.90	76.59$\pm$0.99	**80.03**$\pm$0.97
	Abrupt	75.07$\pm$1.19	73.31$\pm$1.15	77.68$\pm$0.73	76.14$\pm$0.44	**80.38**$\pm$0.70

indicates that the current operation is mildly unsafe state and there may be security risks in the external environment or controllable anomalies within the system; and (3) Level III, which indicates that the current operation is in an unsafe state and there are certain safety risks in the external environment or dangerous abnormalities inside the system. Concept drift exists in this dataset due to variations in the evaluation criteria used to determine the safety of the current state at different depths.

b. Electricity market dataset (ELEC) [1] is a widely used dataset described by M. Harries and analyzed by Gama. These data are collected from the Australian New South Wales Electricity Market. In this market, prices are not fixed but fluctuate based on supply and demand dynamics. The price is determined every five minutes. The ELEC dataset comprises 45,312 instances. The class labels indicate the variation in price relative to the moving average over the last 24 hours.

- **Numerical results and relevant analysis of real-world experiments:** As shown in Tab. VI, DD-VAN achieves state-of-the-art performance on two real-world datasets compared to four competitors. Specifically, variational inference and fuzzy technique collectively improve the DD-VAN and achieve 7.62% (75.42% $\rightarrow$ 83.04%) and 4.06% (84.04% $\rightarrow$ 79.98%) average absolute improvements over the second-best algorithms, respectively. The experimental results demonstrate the effectiveness of the DD-VAN in dealing with unknown concept drift patterns.

TABLE VI: Classification accuracy (%) and relevant analysis of real-world datasets. The best performance is highlighted in **bold**.

Methods	Jiaolong	ELEC
DIVA	71.40$\pm$1.62	63.24$\pm$1.40
IAST	66.28$\pm$3.30	68.00$\pm$1.18
MCMO	72.05$\pm$1.34	79.98$\pm$2.55
CADM+	75.42$\pm$0.58	75.54$\pm$1.70
DD-VAN	**83.04**$\pm$0.54	**84.04**$\pm$0.87

REFERENCES

[1] M. Baena-García, J. Campo-Ávila, R. Fidalgo, A. Bifet, R. Gavaldà, and R. Morales-Bueno, Early drift detection method, In: *Fourth international workshop on knowledge discovery from data streams*, 2006, vol. 6, pp. 77–86.

[2] Y. Cao, M. Jia, X. Zhao, X. Yan, and K. Feng, Complex augmented representation network for transferable health prognosis of rolling bearing considering dynamic covariate shift, *Reliability Engineering & System Safety*, vol. 241, art. no. 109692, 2024.

[3] H. Chen, R. Wu, C. Tao, W. Xu, H. Liu, C. Xu, and M. Jian, Multi-scale class attention network for diabetes retinopathy grading, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100012, 2024.

[4] J. Fang, Z. Wang, W. Liu, S. Lauria, N. Zeng, C, Prieto, et al., A new particle swarm optimization algorithm for outlier detection: Industrial data clustering in wire arc additive manufacturing, *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 2, pp. 1244–1257, 2024.

[5] K. Feng, J. C. Ji, Q. Ni, and M. Beer, A review of vibration-based gear wear monitoring and prediction techniques, *Mechanical Systems and Signal Processing*, vol. 182, art. no. 109605, 2023.

[6] K. Feng, J. C. Ji, Y. Zhang, Q. Ni, Z. Liu, and M. Beer, Digital twin-driven intelligent assessment of gear surface degradation, *Mechanical Systems and Signal Processing*, vol. 186, art. no. 109896, 2023.

[7] C. Guan, R. Shang, F. Yang, A. Shao, and S. Zhang, A priori information-guided generative adversarial network for data augmentation: application to pipeline fault diagnosis, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2343311, 2024.

[8] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, and V. Lempitsky, Domain-adversarial training of neural networks, *Journal of Machine Learning Research*, vol. 17, pp. 1–35, 2016.

[9] J. Hoffman, T. Darrell, and K. Saenko, Continuous manifold based adaptation for evolving visual domains, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 867–874.

[10] S. Hu, Z. Liu, M. Li, and X. He, CADM+: Confusion-based learning framework with drift detection and adaptation for real-time safety assessment, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 3, pp. 5126–5139, 2024.

[11] B. Jiao, Y. Guo, S. Yang, J. Pu, and D. Gong, Reduced-space multistream classification based on multiobjective evolutionary optimization, *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 4, pp. 764–777, 2023.

[12] H. Jin, Z. Zhang, S. X. Ding, Z. Gao, Y. Wang, and Z. Zuo, Fault diagnosis for parallel lithium-ion battery packs with main current sensor

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

fault and internal resistance fault, *IEEE Transactions on Instrumentation and Measurement*, vol. 73, art. no. 3521210, 2024.

[13] L. Korycki and B. Krawczyk, Class-incremental experience replay for continual learning under concept drift, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021, pp. 3649–3658.

[14] A. Kumar, T. Ma, and P. Liang, Understanding self-training for gradual domain adaptation, In: *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, 2020, pp. 5468–5479.

[15] N. Kumar, A, Narang, and B. Lall, Kullback-leibler divergence based regularized normalization for low resource tasks, *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 6, pp. 2638–2650, 2024.

[16] M. Llse, J. M. Tomczak, C. Louizos, and M. Welling, Diva: Domain invariant variational autoencoders, In: *Medical Imaging with Deep Learning*, 2020, pp. 322–348.

[17] F. Li, H. Lv, and Q. Shen, Light but fruitful: Enhanced fuzzy inference via weight-guided selection of rules with attribute weights, *International Journal of Systems Science*, vol. 55, no. 15, pp. 3101–3113, 2024.

[18] K. Li, J. Lu, H. Zuo, and G. Zhang, Source-free multi-domain adaptation with fuzzy rule-based deep neural networks, *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 12, pp. 4180–4194, 2023.

[19] L.-F. Li, Y. Hua, Y.-H. Liu, and F.-H. Huang, Study on fast fractal image compression algorithm based on centroid radius, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2269183, 2024.

[20] W. Ehab, L. Huang, and Y. Li, UNet and variants for medical image segmentation, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100009, 2024.

[21] M. Litrico, A. Del Bue, and P. Morerio, Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7640–7650.

[22] H. Liu, M. Long, J. Wang, and Y. Wang, Learning to adapt to evolving domains, In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020, pp. 22338–22348.

[23] M. Y. Liu and O. Tuzel, Coupled generative adversarial networks, In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 469–477.

[24] Y. Liu, Y. Wang, T. W. S. Chow, and B. Li, Deep adversarial subdomain adaptation network for intelligent fault diagnosis, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6038–6046, 2022.

[25] M. Long, Z. Cao, J. Wang, and M. I. Jordan, Conditional adversarial domain adaptation, In: *Advances in Neural Information Processing Systems 31 (NeurIPS)*, 2018, pp. 1647–1657.

[26] L. Lobo and Jesús, Synthetic datasets for concept drift detection purposes, Harv. Dataverse, 2020, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5OWRGB.

[27] J. Lu, G. Ma, and G. Zhang, Fuzzy machine learning: a comprehensive framework and systematic review, *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 7, pp. 3861–3878, 2024.

[28] S. Lu, Z. Gao, P. Zhang, Q. Xu, T. Xie, and A. Zhang, Event-triggered federated learning for fault diagnosis of offshore wind turbines with decentralized data, *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 2, pp. 1271–1283, 2024.

[29] Y. W. Luo, C. X. Ren, P. Ge, K. K. Huang, and Y. F. Yu, Unsupervised domain adaptation via discriminative manifold embedding and alignment, In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34. no. 4, 2020, pp. 5029–5036.

[30] A. Ma, J. Li, K. Lu, L. Zhu, and H. T. Shen, Adversarial entropy optimization for unsupervised domain adaptation, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6263–6274, 2022.

[31] G. Ma, J. Lu, F. Liu, Z. Fang, and G. Zhang, Domain adaptation with interval-valued observations: Theory and algorithms, *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 5, pp. 3107–3120, 2024.

[32] G. Ma, S. Xu, T. Yang, Z. Du, L. Zhu, H. Ding, and Y. Yuan, A transfer learning-based method for personalized state of health estimation of lithium-ion batteries, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 759–769, 2024.

[33] K. Mei, C. Zhu, J. Zou, and S. Zhang, Instance adaptive self-training for unsupervised domain adaptation, In: *Computer Vision-ECCV 2020: 16th European Conference*, 2020, pp. 415–430.

[34] K. Nishida and K. Yamauchi, Detecting concept drift using statistical testing, *International Conference on Discovery Science*, 2007, pp. 264–269.

[35] N. Nazareth and Y. V. R. Reddy, Financial applications of machine learning: A literature review, *Expert Systems with Applications*, vol. 219, art. no. 119640, 2023.

[36] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, Domain adaptation via transfer component analysis, *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[37] K. Shi, J. Lu, Z. Fang, and G. Zhang, Unsupervised domain adaptation enhanced by fuzzy prompt learning, *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 7, pp. 4038–4048, 2024.

[38] B. Sun and K. Saenko, Deep coral: Correlation alignment for deep domain adaptation, In: *Proceedings of European Conference on Computer Vision*, Berlin, Germany: Springer, 2016, pp. 443–450.

[39] C. Wang, Z. Wang, and H. Dong, A novel prototype-assisted contrastive adversarial network for weak-shot learning with applications: Handling weakly labeled data, *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 1, pp. 533–543, 2024.

[40] C. Wang, Z. Wang, H. Liu, H. Dong, and G. Lu, An optimal unsupervised domain adaptation approach with applications to pipeline fault diagnosis: Balancing invariance and variance, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10019–10030, 2024.

[41] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, A review on transfer learning in EEG signal analysis, *Neurocomputing*, vol. 421, no. 15 pp. 1-14, 2021.

[42] H. Wang, H. He, and D. Katabi, Continuously indexed domain adaptation, In: *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, 2020, pp. 9898–9907.

[43] P. Wang, C. Ding, W. Tan, M. Gong, K. Jia, and D. Tao, Uncertainty-aware clustering for unsupervised domain adaptive object re-identification, *IEEE Transactions on Multimedia*, vol. 25, pp. 2624–2635, 2022.

[44] W. Wang, L. Ma, Q. Rui, and C. Gao, A survey on privacy-preserving control and filtering of networked control systems, *International Journal of Systems Science*, vol. 55, no. 11, pp. 2269–2288, 2024.

[45] Y. Wang, C. Wen, and X. Wu, Fault detection and isolation of floating wind turbine pitch system based on Kalman filter and multi-attention 1DCNN, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2362169, 2024.

[46] L. Wen, L. Gao, and X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.

[47] B. Wu, J. Fang, X. Zeng, S. Liang, and Q. Zhang, Adaptive compositional continual meta-learning, In: *Proceedings of the 40th International Conference on Machine Learning*, 2023, pp. 37358–37378.

[48] P. Xu, Z. Deng, J. Wang, Q. Zhang, K. S. Choi, and S. Wang, Transfer representation learning with TSK fuzzy system, *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 3, pp. 649–663, 2021.

[49] B. Yang, Y. Lei, X. Li, N. Li, and A. K. Nandi, Label recovery and trajectory designable network for transfer fault diagnosis of machines with incorrect annotation, *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 4, pp. 932–945, 2024.

[50] Q. Yang, J. Zhou, and Z. Wei, Time perspective-enhanced suicidal ideation detection using multi-task learning, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100011, 2024.

[51] E. Yu, J. Lu, B. Zhang, and G. Zhang, Online boosting adaptive learning under concept drift for multistream classification, In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 15, 2024, pp. 16522–16530.

[52] L. Yuan, H. Li, B. Xia, C. Gao, M. Liu, W. Yuan, and X. You, Recent advances in concept drift adaptation methods for deep learning, *The 34th International Joint Conference on Artificial Intelligence*, 2022, pp. 5654–5661.

[53] H. Zhang, D. Yue, C. Dou, X. Xie, K. Li, and G. P. Hancke, Resilient optimal defensive strategy of TSK fuzzy-model-based Microgrids' system via a novel reinforcement learning approach, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1921–1931, 2023.

[54] J. Zhang, X. Li, J. Tian, Y. Jiang, H. Luo, and S. Yin, A variational local weighted deep sub-domain adaptation network for remaining useful life prediction facing cross-domain condition, *Reliability Engineering & System Safety*, vol. 231, art. no. 108986, 2023.

[55] J. Zhang, K. Zhang, Y. An, H. Luo, and S. Yin, An integrated multitasking intelligent bearing fault diagnosis scheme based on representation learning under imbalanced sample condition, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6231–6242, 2024.

[56] M. Zhou, J, Lu, P. Lu, and G. Zhang, Dynamic graph regularization for multi-stream concept drift self-adaptation, *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 6016–6028, 2024.

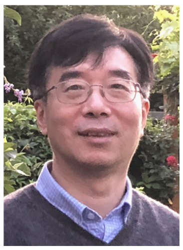
[57] X. Gao, F. Deng, W. Shang, X. Zhao, and S. Li, Attack-resilient asynchronous state estimation of interval type-2 fuzzy systems under stochastic protocols, *International Journal of Systems Science*, vol. 55, no. 13, pp. 2688–2700, 2024.

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TFUZZ.2025.3567089, IEEE Transactions on Fuzzy Systems

Chuang Wang received the B.Sc. degree in automation and the Ph.D. degree in petroleum and natural gas engineering from the Northeast Petroleum University, Daqing, China, in 2017 and 2023, respectively.

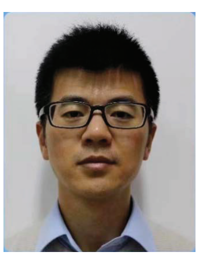
From August 2022 to August 2023, he was a Visiting Ph.D. Student with the Department of Computer Science, Brunel University London, Uxbridge, U.K. He is currently an Associate Professor with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, China. His research interests include evolutionary computation and transfer learning techniques.



Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China.

He is currently Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the U.K. Prof. Wang's research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published more than 700 papers in international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.



Weiguo Sheng (Member, IEEE) received the M.Sc. degree in information technology from the University of Nottingham, U.K., in 2002 and the Ph.D. degree in computer science from Brunel University, U.K., in 2005. Then, he worked as a Researcher at the University of Kent, U.K. and Royal Holloway, University of London, U.K. He is currently a Professor at Hangzhou Normal University. His research interests include evolutionary computation, data mining/clustering, pattern recognition and machine learning.

Qingqiang Liu received the B.Sc. degree in computer software in 2001, and the M.Sc. degree in computer application technology in 2004 and the Ph.D. degree in petroleum and natural gas engineering in 2024, all from Northeast Petroleum University, Daqing, China.

He is currently an Associate Professor with the School of Electrical & Information Engineering, Northeast Petroleum University, Daqing, China. His research interests include industrial intelligence optimization and model fusion.

Hongli Dong (Senior Member, IEEE) received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2012.

From 2009 to 2010, she was a Research Assistant with the Department of Applied Mathematics, City University of Hong Kong, Hong Kong. From 2010 to 2011, she was a Research Assistant with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. From 2011 to 2012, she was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, London, U.K. From 2012 to 2014, she was an Alexander von Humboldt Research Fellow with the University of Duisburg-Essen, Duisburg, Germany. She is currently a Professor with the Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, China. She is also the Director of the Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Daqing. Her current research interests include robust control and networked control systems.

Dr. Dong is a very active reviewer for many international journals.

Copyright © 2025 Institute of Electrical and Electronics Engineers (IEEE). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. See: https://journals.ieeeauthorcenter.ieee.org/becomean-ieee-journal-author/publishing-ethics/guidelines-and-policies/post-publication-policies/