# Ordinal Aligned Domain Generalization for Sensor-based Time Series Regression

Yunchuan Shi, Wei Li (✉), and Albert Y. Zomaya

The University of Sydney, Sydney, Australia
{yshi7084, weiwilson.li, albert.zomaya}@sydney.edu.au

**Abstract.** Time series data powers sensor systems in health, cities, and beyond, demanding robust analysis for real-world impact. While deep learning models excel in this field, their performance degrades in new environments due to data distribution shifts. Domain generalization (DG) aims to enhance model performance in new environments, but current methods primarily focus on discrete data, assuming a discrete, fixed label space, and addressing distribution shifts by extracting common features from inputs across all source domains. However, sensor-based tasks involve real-valued data with diverse input and label spaces. Existing approaches overlook the continuity between data and labels, mapping input data with similar labels to scattered feature spaces, making models susceptible to distribution shifts. Additionally, variations in the label space cause predictive features to change across domains, complicating the identification of stable, generalizable features. This work introduces a new DG framework tailored for sensor-based tasks, operating without access to target domain data or post-deployment adjustments. Our approach learns Ordinal-Aligned Task-Specific (OATS) features that capture stable relationships between continuous labels and input features while maintaining domain independency under input and label space shift. This enables the model to make accurate predictions across unseen domains and continuous label spaces. Experiments on multiple real-world time series regression datasets show that our method outperforms 14 baselines, reducing prediction error by 13% on average.

**Keywords:** Domain Generalization, · Time Series Regression, · Ordinal Alignment, · Label Space Shift.

## 1 Introduction

Time series analysis is crucial for understanding and predicting sequential data in sensor-based systems, with applications in smart grids [3], environmental monitoring [28], and healthcare [7]. Deep neural networks excel in capturing complex patterns in such data [27][15]. However, building these models requires large, labeled datasets [14], which are costly and impractical to collect in all deployment environments. Traditional model-building process assumes that the training and testing data share the same statistical distributions, called independent and identically distributed (i.i.d.) data [3]. Under this assumption, the trained models can generalize to new, unseen data in new environments. In practice, sensor data vary across geographic locations and environments due to factors like sensor configurations and environmental conditions [13], forming distinct *domains*. For example, seismic sensor data in one area forms a domain, while data collected in a different area forms another domain. These variations lead to shifts in both input and the label distributions, known as *domain shifts* [33]. As a result, models trained on training (source) domains may perform poorly on new, unseen (target) domains, leading to a significant drop in predictive accuracy and limiting the model's reliability in real-world applications. To mitigate distribution shifts, transfer learning [24] and domain adaptation [8] have been proposed, but they require auxiliary data from target domains for model adjustment before or after deployment, which is impractical in time-sensitive and resource-limited sensor settings such as earthquake monitoring and emergency health monitoring [15].

Domain generalization (DG) [39] offers a solution that enables models to learn knowledge from multiple source domains that generalize well to unseen domains without requiring auxiliary data or post-deployment adjustments [2][33]. While DG has shown promise in sensor-based classification tasks like activity recognition and fault diagnosis, its application to regression remains underexplored. Classification tasks typically have a fixed label space where all domains share the same discrete labels. Many DG methods primarily address distribution shifts in the input space, learning domain-invariant features by identifying stable predictive patterns across source domains. However, many sensor-based applications, such as environmental monitoring, energy management, and health assessment, have real-valued data with diverse continuous input and label spaces across domains, leading to *label space shifts*. This variability affects model generalization in two ways. First, it complicates the identification of stable predictive features, as label space shifts alter the distribution of predictive patterns across domains. Crucial label-associated features may only appear in certain domains. For example, seismic activity in low-intensity regions lacks high-intensity earthquake signatures. This challenges DG methods that assume a shared predictive structure across all domains. In extreme cases, label space shifts create domain-exclusive label ranges, preventing label overlap and hindering the extraction of stable, transferable patterns. As a result, learning informative features while eliminating domain dependencies becomes difficult. Second, the target domain may contain unseen labels with no corresponding features in the source domain, making accurate predictions in the new label space difficult. Existing DG methods are often ineffective under these conditions, leading to inaccurate or suboptimal features for generalization. Furthermore, they overlook the continuous nature of label spaces and the subtle relationships between labels and input patterns. In regression tasks, small changes in labels correspond to gradual, continuous changes in input patterns. Methods designed for discrete labels fail to capture these relationships, mapping continuous patterns into discrete features [34]. This limitation becomes particularly problematic under domain shifts, where small variations in input features can cause large prediction deviations, making models that learn fragmented features highly sensitive to domain shifts.

To address these challenges, we propose a new DG framework for time series regression tasks in sensor-based applications. Our approach learns **O**rdinal **A**ligned **T**ask **S**pecific (OATS) to enhance generalization across unseen domains by learning stable features aligned with labels while remaining independent of domain variations. *Ordinal aligned* ensures that features capture continuous label relationships, reflecting smooth, gradual changes in regression tasks. *Task-specific* removes domain-dependent information, enabling models to generalize effectively across different environments. Together, these properties improve robustness to distribution shifts and allow extrapolation to unseen domains. Our framework leverages contrastive learning [37], which helps the model distinguish relevant features without additional annotation by contrasting similar and dissimilar examples. We extract two key feature types using regression labels as a reference: (1) Ordinal aligned features, which consistently correspond with labels across domains and capture continuous variations as labels change. We encourage an ordered and consistent alignment by enforcing signal pairs with closer regression labels to have higher similarity across domains. This structured alignment improves robustness to distribution shifts and helps the model discern which features are truly associated with label variations and how changes in feature space correspond to label variations, and vice versa. Using this relationship, the model can recognize patterns in unseen label ranges that lie near or between familiar patterns from training by extrapolating from learned continuous structures. (2) Domain-dependent features, which encode domain variations unrelated to the regression task by contrasting pairs of signal from same domain with different labels. These features help eliminate domain dependencies from the ordinal aligned features, ensuring task specificity through a loss function designed to explicitly minimize the mutual information between ordinal aligned features and domain-dependent features. This approach addresses the challenge of removing domain-dependent information from predictive representations when label space shifts occur. The contributions of this paper are summarized as follows:

1. We propose a new DG framework for sensor-based time series regression that aligns features with the ordinal nature of continuous labels, moving beyond the distributional assumptions of existing methods.
2. Our framework learns OATS features that capture essential predictive information while stable with respect to labels and independent of domain variations, enhancing generalization across domains.
3. We conducted extensive experiments on four real-world sensor applications, comparing our approach with 14 DG methods. The results demonstrate that our framework outperforms SOTA approaches, achieving improved performance and reliability under domain shifts.

## 2 Preliminaries and Problem Setting

### 2.1 Domain Generalization

DG is a subfield of transfer learning that aims to develop models capable of performing well on unseen domains with distinct data distributions from the training domain [33]. Unlike domain adaptation [3][35], DG does not rely on target domain information during training, making it particularly challenging. Initially prominent in computer vision, DG now gains traction in time-series and signal processing, such as human activity recognition [25], industrial automation [22], and healthcare [7]. Existing DG methods fall into three categories [33]. 1) Data augmentation approaches, such as Mixup and GAN-based techniques [13][11] generate synthetic data to improve generalization. 2) Learning strategies include meta-learning [18][22], which simulates domain shifts by splitting source tasks to enhance adaptability, invariant risk minimization (IRM) [10], which encourages consistent predictive rules across varying domains , and distributionally robust optimization (DRO) [9], which optimizes model performance in worst-case scenarios. 3) Representation learning seeks domain-invariant features through distribution alignment methods by reducing maximum mean discrepancy(MMD) [13] or correlation [1], domain-adversarial learning(DANN) [15] [16], and feature disentanglement [2][4][25] to separate domain-specific factors from robust, transferable representations. Despite these advances, most DG methods assume discrete, consistent label spaces and focus primarily on classification. This limits their effectiveness in time-series regression, where label space shifts introduce challenges that disrupt stable predictive patterns across domains. As a result, existing approaches often degrade when label distributions vary between source and target domains. Some recent studies have explored regression, [16] tackles time-series forecasting with temporal shifts where distributions evolve within a single domain but do not address domain shifts across environments. [8][14][30][35] tackle regression task under label drift, they rely on extra target-domain information, such as unlabeled samples, do not strictly conform to the DG setting where the target domain is entirely unknown. As a result, developing robust regression models that handle continuously shifting label distributions in unseen domains remains an open and pressing challenge in DG research. In Appendix A, we further discuss the limitations of DG methods in regression with continuous label space shifts. Propositions 1 and 2 show that domain-invariant feature extraction can lose critical predictive information and fail to remove domain dependency.

### 2.2 Problem Setting of DG Regression

Let $\mathcal{X} \subseteq \mathbb{R}^{M \times T}$ be the input space and $\mathcal{Y} \subseteq \mathbb{R}$ be the label space of real numbers. where $T$ represents the length of the time series, and $M$ is the number of dimensions in each input sample. A domain is defined over the product space $\mathcal{X} \times \mathcal{Y}$ and is represented by a joint probability distribution $\mathcal{P}(X, Y)$. For each domain $d$, we have a set of data samples $D_d = \{(x_i^d, y_i^d)\}_{i=1}^{N_d}$, where each sample

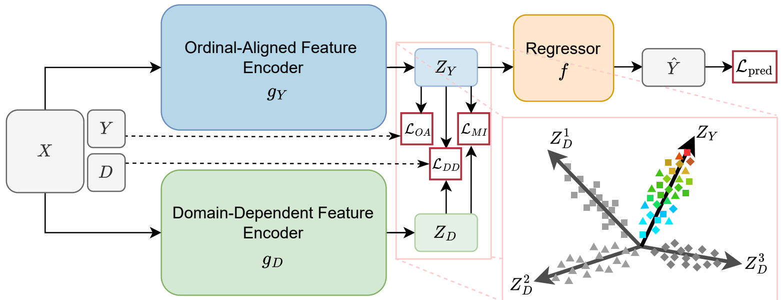



Fig. 1: The overview of our proposed framework

$(x_i^d, y_i^d)$ is drawn from the distribution $\mathcal{P}_d(X, Y)$. Here, $N_d$ is the number of samples in domain $d$. We use $X_d$ and $Y_d$ to represent the input set and label set, respectively, in domain $d$. To represent multiple source domains, we define a set $\mathcal{S} = \{1, \ldots, S\}$, where each element corresponds to a source domain with its joint distribution $\mathcal{P}_s(X, Y)$. Let $D_S$ represent the combined data from all source domains, with $X_S$ and $Y_S$ denoting the full set of inputs and labels across these domains. The unseen target domain, where we will evaluate the model's generalization ability, is represented by a separate distribution, $\mathcal{P}_t(X, Y)$. Label space shift may exist between domains, which is defined as occurring between domains $j$ and $k$ when the possible values or ranges of their label distributions, denoted as $\mathcal{P}(Y_j)$ and $\mathcal{P}(Y_k)$, differ. This shift is characterized by the inequality of their support sets, $\mathcal{S}_{Y_j} \neq \mathcal{S}_{Y_k}$. To measure the generalization capability of a model $f : \mathcal{X} \to \mathcal{Y}$, we calculate its empirical risk on a given domain $d$, defined as:

$$R(f, \mathcal{P}(X, Y)) = \mathbb{E}_{x,y \sim \mathcal{P}(X,Y)} \|y - f(x)\|_2^2 \tag{1}$$

This risk measure calculates the average Mean Squared Error (MSE) loss between the model's predictions $f(x)$ and the actual labels $y$ for samples from domain $d$. The goal of DG is to train a model $f$ using data from the source domains that minimizes the risk in the unseen target domain, denoted $R(f, \mathcal{P}_t(X, Y))$.

## 3 Our Approach

We propose a framework for time series regression that learns Ordinal Aligned, Task-Specific (OATS) features by extracting both ordinal-aligned features and domain-dependent features, and minimizing the mutual information between them to enforce task specificity. Figure 1 provides an overview of our framework, which includes three main components: an ordinal-aligned feature encoder $g_Y$, a domain-dependent feature encoder $g_D$, and a regressor $f$. Both encoders, $g_Y$ and $g_D$, share the same network architecture but are trained to extract different feature representations from the input data $x$. Specifically, $g_Y$ extracts ordinal-aligned features $z_Y = g_Y(x)$ that align with the labels in an ordinal way, capturing information directly related to the regression task. In contrast, the domain-dependent encoder $g_D$ extracts features $z_D = g_D(x)$ that capture unique characteristics of each domain and assist $g_Y$ in filtering out domain-dependent influences. We use $Z_D$ to denote the set of domain-dependent features, and $Z_Y$ to denote the set of domain-dependent features. $Z_D^i$ and $Z_Y^i$ represent the subsets of domain-dependent and ordinal-aligned features for the domain $i$, respectively. The ordinal-aligned features $Z_Y$ are then inputted into the regressor $f$, which outputs the final prediction $\hat{y} = f(Z_Y)$. See Appendix B for the pseudocode of training and inference. Below, we formally define ordinal aligned features and domain dependent features:

**Definition 1.** ***Ordinal Aligned Features****: Let $x_i, x_j, x_k$ be three data samples from different source domains, with labels $y_i, y_j, j_k$ satisfying $|y_i - y_j| < |y_i - y_k|$, indicating that $y_i$ is closer to $y_j$ than to $y_k$. Ordinal aligned features are $z_i, z_j, z_k \in Z_Y$ consistent across domain given label and satisfying $sim(z_i, z_j) < sim(z_i, z_k)$, where $sim(\cdot, \cdot)$ represents a similarity measure between two features (i.e. euclidean distance or cosine similarity).*

These features encode the smooth, gradual relationships inherent in regression tasks. For instance, in seismic analysis, waveform characteristics change gradually from a magnitude 3 to a magnitude 6 earthquake. A waveform for magnitude 4 falls between those for magnitudes 3 and 6 but more closely resembles magnitude 3. Similarly, in clinical pain assessments, physiological signals respond gradually as pain levels increase. Moving from "no pain" to "severe pain" involves intermediate stages like "mild pain" and "moderate pain," with each stage marked by subtle changes in physiological signals. Such gradual shifts in labels produce a smooth, continuous effect on the associated features. This ordinal continuity implies that inputs contain features reflecting a progression of label values, where adjacent labels are associated with more similar features. By identifying these features, the model captures representations that encode the stable relationship between labels and feature variations across domains, effectively modeling this relationship based on label distances. This enables the model to approximate and align features smoothly across neighboring labels, improving generalization to unseen label ranges by leveraging the learned similarity structure among features. To remove domain-dependency from ordinal aligned features while preserving task specificity, we identify these features from each domain that remain stable even as labels change, formally defined as follows:

**Definition 2.** ***Domain Dependent Features****: Let $x_i, x_j$ be two data samples from the same domain $X_d$ with different labels $y_i, y_j$, and let $x_k$ be a sample from a different domain $X_{d'}$. The Domain Dependent Features $z_i, z_j, z_k \in Z_D$ capture domain-specific variations while remaining independent of the labels by maximizing the similarity between $z_i$ and $z_j$, while minimizing their similarity to $z_k$ from other domains and the ordinal aligned features $z_y \in Z_Y$.*

To extract these features effectively, our framework uses contrastive learning [4][37] to identifies shared patterns in data. We design contrastive objectives for both encoders, $g_Y$ and $g_D$, helping them specialize in their respective roles. As shown in Figure 1 (bottom right), our framework organizes the feature space by separating ordinal-aligned features $Z_Y$ from domain-dependent features $Z_D$ and aligning $Z_Y$ features according to label values to form an ordinal structure.

### 3.1 Ordinal Aligned Feature for Regression

To achieve ordinal alignment of features, as defined in Definition 1, we aim to train the ordinal-aligned encoder $g_Y$ to prioritize similarity between feature pairs based on the closeness of their labels. Specifically, for any two features $g_Y(x_i)$ and $g_Y(x_j)$ with labels $y_i$ and $y_j$, we aim for their similarity to be greater than that of any pair $g_Y(x_i)$ and $g_Y(x_k)$, where $|y_i - y_j| < |y_i - y_k|$. This approach adapts recent advances in learning features through contrastive learning [37] to address the unique challenges of multi-domain generalization with shifting continuous label spaces. To guide $g_Y$ in learning these ordinal aligned features, we introduce a loss function $\mathcal{L}_{OA}$, which increases the similarity between features with closer labels. The loss function is defined as:

$$\mathcal{L}_{OA} = -\frac{1}{N} \sum_{x_i, x_j \in X_S, i \neq j} \log \frac{e^{\mathrm{sim}(g_Y(x_i), g_Y(x_j))/\tau_{OA}}}{\sum_{x_k \in \Phi(i,j)} e^{\mathrm{sim}(g_Y(x_i), g_Y(x_k))/\tau_{OA}}} \tag{2}$$

In this setup, we evaluate all pairs of input features in the source domain. For each pair $x_i$ and $x_j$, the numerator measures the similarity between their encoded features, while the denominator sums similarities between $g_Y(x_i)$ and all features $x_k$ in $\Phi(i,j)$. Here, $\Phi(i,j) = \{x_k \in X_S \mid |y_i - y_j| < |y_i - y_k|\}$ includes features with label distances to $y_i$ greater than the distance between $y_i$ and $y_j$, ensuring closer labels are prioritized. The temperature parameter $\tau_{oa}$ adjusts the sensitivity to similarity: lower values sharpen the focus on more similar pairs, while higher values smooth the distribution. The exponential scaling further emphasizes pairs with higher similarity scores, amplifying their influence on the loss function. By minimizing $\mathcal{L}_{OA}$, the encoder is encouraged to produce feature pairs with smaller label distances that are more similar than pairs with larger label distances. This optimization enforces ordinal alignment across the feature space, helping the encoder capture the continuous nature of the labels. Additionally, this objective aligns features with similar labels from different domains, allowing the model to learn consistent patterns in diverse environments. By focusing on stable, label-consistent features, our approach enhances generalization across domains by capturing elements that remain invariant despite domain variations.

### 3.2 Minimizing Domain Dependency

To address the presence of non-overlapping label ranges that may exist exclusively in a single source domain, and to ensure that the ordinal-aligned encoder $g_Y$ learns features free from domain-dependent information, we introduce a domain-dependent encoder, $g_D$, along with two contrastive loss functions. The domain-dependent encoder captures label unrelated characteristics unique to each domain, allowing us to remove these features from the ordinal-aligned ones, enforcing task-specificity. The encoder $g_D$ is designed to ensure that features with different labels from the same domain are similar, maximizing their intra-domain similarity. Specifically, for two data points $x_i^s$ and $x_j^s$ from the same domain $s$, we aim to maximize their similarity in the feature space. Conversely, for data points $x^s$ and $x^t$ from different domains $s$ and $t$, we minimize their similarity. To encourage the removal of domain-dependent features from ordinal-aligned feature, we also minimize the similarity between the features generated by ordinal-aligned encoder $g_Y$ and domain-dependent encoder $g_D$. These optimization objectives are collectively achieved through a contrastive loss function, $\mathcal{L}_{DD}$. In this setting, $Z_Y$, the features extracted by $g_Y$, represent one category, while $Z_D^s$, the domain-dependent features extracted by $g_D$ for each domain $s$, form separate categories. The aim is to ensure high similarity within each category and low similarity across categories. The contrastive loss function is defined as:

$$\mathcal{L}_{\mathrm{DD}} = -\sum_{z_i \in Z_Y \cup Z_D} \frac{1}{|\mathring{P}(z_i)|} \sum_{z_j \in \mathring{P}(z_i)} \log \frac{e^{\mathrm{sim}(z_i, z_j)/\tau_{\mathrm{DD}}}}{\sum_{z_k \in \mathring{N}(z_i)} e^{\mathrm{sim}(z_i, z_k)/\tau_{\mathrm{DD}}}} \tag{3}$$

$\mathring{P}(z_i) = \{z_j \in Z_Y \cup Z_D \mid \mathrm{category}(z_j) = \mathrm{category}(z_i),\ z_j \neq z_i$ denotes features from the same category as $z_i$, and $\mathring{N}(z_i) = \{z_k \in Z_Y \cup Z_D \mid \mathrm{category}(z_k) \neq \mathrm{category}(z_i)\}$ denotes features from different category. The temperature parameter $\tau_{\mathrm{DD}}$ controls the sensitivity to similarity differences. Minimizing $\mathcal{L}_{\mathrm{DD}}$ promotes high similarity among features within the same category (numerator) while reducing similarity among features from different categories (denominator). This encourages $g_D$ to identify domain-dependent features within each domain. However, there remains a possibility that $Z_Y$ and $Z_D$ may still be correlated. We aim to minimize their mutual information to remove their dependency: $I(Z_Y, Z_D) = D_{KL}(\mathcal{P}(Z_Y, Z_D) \parallel \mathcal{P}(Z_Y)P(Z_D))$. Minimizing mutual information is equivalent to minimizing the Kullback-Leibler (KL) divergence between the joint probability distribution $\mathcal{P}(Z_Y, Z_D)$ and the product of the marginal distributions $\mathcal{P}(Z_Y)\mathcal{P}(Z_D)$ by definition. However, directly computing the KL divergence between high-dimensional distributions is challenging due to its intractability. To address this challenge, we propose an contrastive objective and a discriminating head $h$ to approximate the KL divergence with feature similarity. We construct a set of feature pairs $\{(Z_Y, Z_D)\}$ by concatenating ordinal-aligned and domain-dependent feature along the feature dimension to represent joint distribution $\mathcal{P}(Z_Y, Z_D)$. Additionally, we create another set of feature pairs $\{(Z_Y, Z_D')\}$, where $Z_D'$ is obtained by shuffling the indices of $Z_D$, ensuring independence between concatenated ordinal-aligned and domain-dependent features to simulate samples from the product of the marginals $\mathcal{P}(Z_Y)\mathcal{P}(Z_D)$. The discriminating head $h$ is a network designed to project $\{(z_Y, z_D)\}$ and $\{(z_Y, z_D')\}$

into distinct feature spaces. The proposed objective function is defined as:

$$\begin{aligned}\mathcal{L}_{\mathrm{MI}} = &-\frac{1}{N}\sum_{z_i,z_j\in Z_J, i\neq j}\log\frac{e^{\mathrm{sim}(h(\mathrm{GRL}(z_i)),h(z_j))/\tau_{\mathrm{MI}}}}{\sum_{z'_k\in Z_M} e^{\mathrm{sim}(h(z_i),h(z'_k))/\tau_{\mathrm{MI}}}}\\ &-\frac{1}{N}\sum_{z'_i,z'_j\in Z_M, i\neq j}\log\frac{e^{\mathrm{sim}(h(z'_i),h(z'_j))/\tau_{\mathrm{MI}}}}{\sum_{z_k\in Z_J} e^{\mathrm{sim}(h(z'_i),h(z_k))/\tau_{\mathrm{MI}}}}\end{aligned} \tag{4}$$

Here, $Z_J$ represent the set of features drawn from the joint distribution $\mathcal{P}(Z_Y, Z_D)$, and $Z_M$ represent the set of features drawn from the marginal distribution $\mathcal{P}(Z_Y)\mathcal{P}(Z_D)$. Minimizing this objective function encourages the discriminative head $h$ to map feature pairs from the same distribution to be closer together while pushing apart those from different distributions, enable $h$ effectively distinguish between concatenated feature pairs drawn from the joint versus marginal distributions, thus assessing the similarity between the two distributions and approximating the KL divergence. To minimize the KL divergence and reduce the mutual information, we employ adversarial learning by applying a gradient reversal layer (GRL) to $Z_Y$ before it is passed to $h$. This approach reverse the gradient sign been prepackaged, guiding the encoder to learn patterns that make $h$ unable to differentiate the features belonging to the joint distribution or the marginal distributions, thereby minimizing the KL divergence and reducing their mutual information. By integrating these strategies, we ensure that the ordinal-aligned encoder $g_Y$ focuses on learning features pertinent to the task while being robust to domain-dependent variations, thus enhancing generalization across domains.

### 3.3 Selection of Similarity Measurements

The choice of similarity measurements is critical to optimizing our framework. Specifically, $\mathcal{L}_{OA}$ aims for ordinal-aligned features to be more similar when conditioned on their corresponding labels, while $\mathcal{L}_{DD}$ seeks to minimize the similarity between ordinal-aligned dependent and domain-dependent features from each domain. This setup introduces a potential conflict in contrasting ordinal-aligned features, as they are simultaneously pushed to align closely with each other and to diverge from domain-dependent features. To resolve this potential conflict, we employ distinct similarity metrics for each objective: L2 distance for $\mathcal{L}_{OA}$ and cosine similarity for $\mathcal{L}_{DD}$. This strategic choice positions the ordinal-aligned features in different orientations within a high-dimensional space relative to domain-dependent features, while aligning the ordinal-aligned features along a similar axis but at varying positions (see Figure 1 - bottom right).

### 3.4 Overall Loss Function

The overall objective function for training the model is:

$$\mathcal{L}_{\mathrm{ALL}} = \mathbb{E}_{x,y\sim\mathcal{P}_s(X,Y)}\|y - f(g_Y(x))\|_2^2 + \lambda_1\mathcal{L}_{OA} + \lambda_2\mathcal{L}_{DD} + \lambda_3\mathcal{L}_{MI} \tag{5}$$

The first term represents the mean squared error between the predicted outputs and the actual labels. The remaining terms are weighted components of the loss function that manage domain-dependent and ordinal label aligned representation constraints. The parameters $\lambda_1$, $\lambda_2$, $\lambda_3$ are regularization weights that balance the influence of constraints relative to the primary regression task. By jointly optimizing these losses, the model learns to extract OATS features that align with label order while simultaneously removing domain-dependent characteristics. This enables the model to generalize effectively across different domains and make accurate predictions for the regression task. Our generalization method operates solely during training, using $f$, $g_Y$ and $g_D$ with their respective loss functions to guide learning. During inference, only the ordinal-aligned encoder $g_Y$ and regressor $f$ are used to make predictions on unseen data. A few more discussions on our design can be found in Appendix C.

## 4 Experiments

### 4.1 Datasets and Setup

We tested our method on real-world time series data from biomedical, seismology, and energy systems with distribution shifts. The Pain Assessment dataset (BioVid) [32] predicts pain intensity levels using physiological sensor data collected from 87 subjects in response to heat-induced pain. Pain intensity includes levels ranging from 0 (no pain) to 4 (severe pain). Each subject was treated as a distinct domain due to individual biological and psychological variations that affect data distribution, following established practices [36]. This setup allowed us to evaluate the ability of the model to generalize to new patients. To assess the impact of label space shifts, we modified the original dataset, which features uniformly distributed labels, by randomly removal some range of labels from each subject's label distribution. These conditions tested the robustness of our method and other DG approaches under varying degrees of label space shift. The Air Quality Prediction dataset (PRSA) [5], consists of hourly air pollution index measurements and meteorological data collected from multiple air monitoring stations in Beijing. The objective is to predict future air pollution index variations based on meteorological data. The Earthquake Detection dataset (LEN-DB) [21], contains three-component seismic data captured along vertical, north-south, and east-west axes. Each recording is labeled with earthquake magnitude. Data from ten countries were treated as separate domains. The magnitude distributions vary across stations due to geological and geographical factors. This dataset includes 22,207 sequences, providing a diverse evaluation setting for domain generalization under label space shifts. The Energy Disaggregation dataset (REFIT) [23] records household energy use at 8-second intervals over two years. The data includes whole-house aggregate consumption and individual appliance loads. Our task involves short-term Non-Intrusive Load Monitoring (NILM), where the goal is to predict the power consumption of a target appliance at the midpoint of a sequence based on the overall household energy usage. We selected eight houses and targeted four appliances: washing machines, microwaves, fridges, and dishwashers. Each appliance is treated as a distinct regression task, and data from each house represents a unique domain. Each appliance-specific dataset includes 64,000 sequences. For a comprehensive description of the used datasets, refer to Appendix D. In our experiments, we employed a Leave-One-Domain-Out (LODO) strategy to evaluate our method. Under this approach, one domain was held out as the target domain while the remaining domains were used for training. Each domain was sequentially excluded, ensuring that all domains were evaluated as target domains. We conducted 15 trials per experiment, averaging the results to ensure a robust evaluation. Appendix E gives details on implementation and hyperparameter settings. The code is available at https://github.com/yshi22/OATSDG.

### 4.2 Baselines

We evaluated our model against 14 recent and widely used DG approaches for sensor-based data to assess its effectiveness comprehensively. Empirical Risk Minimization (ERM)[31] serves as a baseline, representing conventional training by minimizing loss on source domains without additional generalization strategies. We compared our model with widely adopted DG methods, including Coral [29], MMD[26], DANN[6], MLDG[12], DRO[9], and Mixup[38], which have been applied to many time series tasks [1][13][15][18]. Additionally, we included recent methods that perform well in image-based applications but remain underexplored for time series data. VREx [10] applies variance regularization to stabilize performance across diverse environments, improving domain resilience. mDSDI[2] combines meta-learning with feature disentanglement, and CDDG[4] incorporates contrastive learning to learn invariant features. We also evaluated recent approaches that address time series data. Diversify[16] leverages domain adversarial learning to extract domain-invariant features from time sequences. GILE[25] employs feature disentanglement to identify stable components in sensor signals. Fixed[17] enhances generalization by augmenting stable features, with evaluations conducted explicitly on time series data. MAMR[19] applies weighted meta-learning tailored for regression tasks.

### 4.3 Results

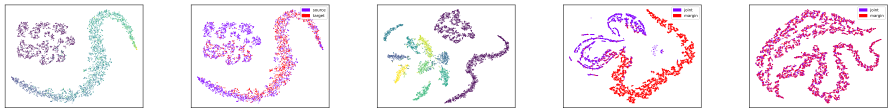
The experiments evaluated the robustness and effectiveness of our DG framework across the datasets, with a particular focus on the challenges posed by domain shifts in regression tasks. As detailed in Table 1, our framework outperforms others, achieving the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) in most target domains. For Biovid tests, existing domain-invariant representation learning methods, MMD, Coral, DANN and Diversify outperformed ERM when label spaces are consistent across domains (reduced generalization errors). However, their performance degraded significantly under label space shifts, consistent with our Proposition 1. In some cases, their performance was even worse than ERM. This is likely because ERM, while

| | Biovid | | | | REFIT | | | |
|---|---|---|---|---|---|---|---|---|
| | No LSS | | Under LSS | | Microwave | | Washing Machine | |
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| ERM[31] | 1.132±0.004 | 1.569±0.004 | 1.136±0.010 | 1.585±0.010 | 34.56±01.50 | 158.26±03.25 | 41.16±02.03 | 184.03±05.19 |
| MMD[26] | 1.127±0.014 | 1.540±0.012 | 1.143±0.001 | 1.558±0.002 | 44.65±00.89 | 158.32±01.52 | 52.64±01.33 | 188.15±02.98 |
| Coral[29] | 1.131±0.016 | 1.541±0.017 | 1.139±0.008 | 1.560±0.006 | 32.53±00.31 | 159.73±00.94 | 42.65±00.73 | 191.19±02.45 |
| DANN[6] | 1.147±0.004 | 1.560±0.003 | 1.178±0.010 | 1.598±0.013 | 30.74±02.91 | **143.31±05.84** | 58.52±04.18 | 193.51±09.66 |
| Mixup[38] | 1.116±0.009 | 1.533±0.013 | 1.135±0.013 | 1.557±0.012 | 33.15±01.50 | 156.26±03.69 | 40.00±02.31 | 181.10±07.02 |
| MLDG[12] | 1.152±0.010 | 1.577±0.029 | 1.167±0.010 | 1.600±0.012 | 49.12±13.34 | 150.66±13.94 | 48.48±18.40 | 176.69±23.35 |
| DRO[9] | 1.120±0.005 | 1.515±0.004 | 1.167±0.018 | 1.799±0.020 | 41.18±02.21 | 171.46±05.09 | 44.60±01.85 | 192.65±06.02 |
| VREx[10] | 1.130±0.011 | 1.544±0.012 | 1.136±0.012 | 1.558±0.013 | 35.97±01.64 | 160.82±04.03 | 41.71±01.63 | 187.17±04.98 |
| mDSDI[2] | 1.116±0.006 | 1.527±0.007 | 1.126±0.006 | 1.543±0.009 | 32.45±01.79 | 146.85±04.99 | 42.36±02.81 | 179.92±06.49 |
| GILE[25] | 1.228±0.008 | 1.606±0.009 | 1.251±0.013 | 1.637±0.016 | 39.64±09.19 | 157.06±10.42 | 47.68±11.18 | 185.75±25.20 |
| Fixed[17] | 1.172±0.007 | 1.588±0.005 | 1.193±0.013 | 1.618±0.010 | 31.00±01.85 | 143.63±01.55 | 55.43±01.92 | 190.61±05.49 |
| MAMR[19] | 1.183±0.012 | 1.618±0.019 | 1.182±0.013 | 1.612±0.012 | 43.50±09.14 | 150.18±12.04 | 53.70±14.54 | 179.62±16.91 |
| Diversify[16] | 1.124±0.004 | 1.532±0.005 | 1.138±0.011 | 1.550±0.014 | 31.27±02.58 | 152.69±03.74 | 43.53±03.78 | 176.06±03.55 |
| CDDG[4] | 1.129±0.005 | 1.547±0.007 | 1.158±0.003 | 1.575±0.005 | 38.07±01.97 | 158.88±04.33 | 41.08±01.58 | 180.61±05.48 |
| Ours | **1.107±0.008** | **1.504±0.009** | **1.123±0.002** | **1.540±0.004** | **29.26±02.90** | 146.71±05.16 | **35.78±02.91** | **174.43±07.13** |

| | LENDB | | PRSA | | REFIT | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Fridge | | Dishwasher | |
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| ERM[31] | 0.422±0.035 | 0.707±0.046 | 30.79±4.05 | 45.97±5.21 | 37.97±00.84 | 50.66±00.86 | 36.99±01.40 | 202.27±04.16 |
| MMD[26] | 0.449±0.037 | 0.698±0.047 | 30.40±1.15 | 45.57±1.19 | 39.67±00.20 | 50.32±00.25 | 47.06±00.49 | 194.53±01.41 |
| Coral[29] | 0.489±0.008 | 0.744±0.009 | 31.48±0.09 | 45.77±0.20 | **36.94±00.26** | 58.38±00.18 | 51.31±00.33 | 197.34±00.77 |
| DANN[6] | 0.455±0.024 | 0.706±0.032 | 33.85±1.62 | 48.92±2.54 | 41.90±00.64 | 53.72±07.80 | 37.06±02.18 | 193.21±05.85 |
| Mixup[38] | 0.464±0.028 | 0.710±0.036 | 38.47±1.39 | 56.19±1.96 | 37.84±00.72 | 50.40±00.72 | 36.67±01.04 | 201.47±02.23 |
| MLDG[12] | 0.506±0.109 | 0.794±0.120 | 31.45±1.59 | 46.85±2.41 | 37.56±02.14 | 56.46±02.32 | 51.58±13.85 | 196.57±11.88 |
| DRO[9] | 0.480±0.029 | 0.730±0.034 | 31.11±2.37 | 45.53±2.97 | 41.93±00.98 | 57.74±01.40 | 39.88±01.39 | 207.55±03.41 |
| VREx[10] | 0.474±0.035 | 0.720±0.043 | 31.14±2.67 | 45.84±2.88 | 38.20±00.88 | 51.03±00.94 | 38.30±01.67 | 204.39±04.03 |
| mDSDI[2] | 0.415±0.025 | 0.697±0.032 | 31.04±4.49 | 46.02±5.41 | 39.90±01.37 | 49.98±01.17 | 39.73±02.20 | 201.73±04.78 |
| GILE[25] | 0.440±0.029 | 0.695±0.036 | 30.55±1.32 | 44.35±1.99 | 41.52±00.81 | 49.89±00.85 | 39.11±05.18 | 199.68±09.19 |
| Fixed[17] | 0.492±0.047 | 0.747±0.056 | 38.81±0.80 | 56.70±1.40 | 41.67±00.30 | 51.03±03.36 | 37.68±01.93 | 193.36±05.44 |
| MAMR[19] | 0.542±0.102 | 0.824±0.116 | 38.47±4.15 | 55.19±5.96 | 42.44±03.77 | 53.34±04.36 | 51.46±17.26 | **193.05±18.63** |
| Diversify[16] | 0.457±0.050 | 0.698±0.054 | 30.81±0.85 | 46.55±2.03 | 39.69±01.17 | 52.46±05.15 | 36.59±03.36 | 205.30±01.32 |
| CDDG[4] | 0.423±0.031 | 0.707±0.041 | 33.34±7.90 | 49.10±9.73 | 38.03±01.01 | 50.97±01.11 | 39.42±01.38 | 204.59±03.93 |
| Ours | **0.386±0.025** | **0.666±0.032** | **28.98±0.72** | **43.72±1.37** | 38.35±01.48 | **49.84±01.12** | **35.83±00.61** | 202.86±01.34 |

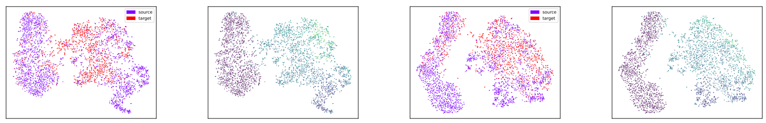
Table 1: The performance comparison of various DG methods on BioVid, LEN-DB, PRSA and REFIT datasets using LODO testing. The table shows average MAE and RMSE across all target domains for each method (mean ± standard deviation). Results are in bold if our model has the best performance.

retaining some domain-dependent noise, also preserves essential predictive features of different label spaces. Conversely, mDSDI, which attempts to adapt domain-specific information after learning domain-invariant features, preserve more label-related information, making it less susceptible to label shifts compared to other baselines. However, it struggles to align features from different domains with labels and fails to capture continuous relationships between features and labels, resulting in lower performance than our approach. Our framework shows robust, stable performance under label shifts, with consistently low standard deviations. It also outperforms across most metrics on the LEN-DB, PRSA and REFIT datasets. MLDG, MAMR, DRO, and VREx, while different in their treatment of domain shifts, struggle to meet the diverse predictive challenges across different labels within domains. Their performance is limited under variable domain conditions, impacted by the distinct prediction difficulties and variances among labels. Data augmentation techniques also fail to bridge the domain gap effectively, largely due to the inherent noise and discrepancies present in time-series data across domains. These results demonstrate the effectiveness of our framework in handling the complexities of domain generalization in re-



(a) $Z_Y$ w.r.t. label (b) $Z_Y$ w.r.t. $D_s D_T$ (c) $Z_Y$ and $Z_D$ (d) Before $L_{MI}$ (e) After $L_{MI}$

Fig. 2: t-SNE visualization of features extracted by our approach.



(a) w.r.t. $D_S/D_T$ (b) w.r.t. label (c) w.r.t. $D_S/D_T$ (d) w.r.t. label

Fig. 3: t-SNE visualization of features extracted by ERM (a,b) and Domain-Invariant Feature Learning (c,d)

gression tasks, particularly its ability to adapt to diverse real-world scenarios. See Appendix F for the significance analysis and detailed results of each method across the leave-out domains for each dataset.

### 4.4 Visualization of Extracted Features

We used t-SNE[20] to visualize the extracted features and validate the effectiveness of ordinal alignment. Figures 2a and 2b showcase the alignment of ordinal-aligned features with label values (blue for lower, yellow for higher) and between source and target domains (purple and red, respectively). The visualizations highlight our ability to maintain a correct ordinal relationship with respect to label values (monotonous blue-to-yellow gradient), and the overlapping colors in Figure 2b indicate successful domain generalization. Figure 2c demonstrates effective learning of domain-dependent features, with each domain represented by distinct colors and clear separation, without overlap. Figure 2d and 2e validates the effectiveness of minimizing domain-dependency in Section 3.2, shows the mutual information between ordinal-aligned and domain dependent features via distributional divergence between $\mathcal{P}(Z_Y, Z_D)$(purple points) and $\mathcal{P}(Z_Y)\mathcal{P}(Z_D)$(red points) before and after applying $\mathcal{L}_{MI}$. Before applying $\mathcal{L}_{MI}$ (Figure 2d), a large distance between samples of $\mathcal{P}(Z_Y, Z_D)$ and $\mathcal{P}(Z_Y)\mathcal{P}(Z_D)$ indicates high KL divergence, suggesting considerable mutual information between ordinal-aligned and domain-dependent features. After applying $\mathcal{L}_{MI}$ (Figure 2e), the distributions are aligned, indicating a lower KL divergence and reduced mutual information between the ordinal-aligned and domain-dependent features. We also provide visualization results of features learned through ERM and domain-invariant feature learning methods. Figure 3a and 3b shows the alignment results for ERM, where target domain features (red points) do not align well with source domain features (purple points), as indicated by a significant number of red points (target) located outside the purple points (source). This indicates that ERM fails

| | REFIT | | Biovid | | LEN-DB | | PRSA | |
|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| ERM | 37.67±01.44 | 148.77±03.37 | 1.136±0.010 | 1.585±0.010 | 0.422±0.035 | 0.707±0.046 | 30.79±4.05 | 45.97±5.21 |
| $\mathcal{L}_{\text{OA}}$ | 35.34±01.29 | 145.79±03.67 | 1.133±0.008 | 1.549±0.007 | 0.391±0.034 | 0.670±0.018 | 30.04±1.21 | 44.98±1.78 |
| $\mathcal{L}_{\text{ALL}}$ | **34.71±01.91** | **143.33±03.78** | **1.126±0.005** | **1.536±0.006** | **0.386±0.025** | **0.666±0.032** | **28.98±0.72** | **43.72±1.37** |
| Cosine | 35.71±02.17 | 144.42±04.48 | 1.129±0.005 | 1.545±0.007 | 0.392±0.027 | 0.676±0.032 | 29.69±1.41 | 44.56±1.83 |
| L2 | 35.90±02.18 | 145.22±04.15 | 1.130±0.006 | 1.549±0.006 | 0.389±0.024 | 0.674±0.027 | 29.06±0.92 | 43.86±1.65 |

Table 2: Ablation study showing the effect of different loss components and distance metrics across four datasets.

to learn features that generalize across domains. Additionally, the learned features do not capture the ordinal relationships between labels, as the distribution of label values around the features does not show a gradient transition (label values are indicated by color). Figure 3c and 3d shows the alignment results using DANN [6]. The target and source domain distributions align, with red points mostly covered by purple areas. However, the features still fail to capture the ordinal relationships between labels, and label relationships are incorrectly learned. This issue is observed in areas with smaller label values (darker points), where unexpectedly high label values (brighter points) appear. These t-SNE visualizations support the effectiveness and robust generalization capabilities of our approach. More visualization results and comparisons with different DG approaches are given in Appendix G.

### 4.5 Ablation Study

Our ablation study examines the contributions of different components in our framework, as shown in Table 2. The baseline ERM model (without DG), shows high MAE and RMSE, indicating its poor generalization. Each additional component in our framework progressively improves the model's ability to generalize. The inclusion of ordinal alignment $\mathcal{L}_{OA}$ shows a definite improvement in performance, and combining $\mathcal{L}_{DD}$ and $\mathcal{L}_{MI}$ to minimize domain-dependency further enhances performance ($\mathcal{L}_{all}$ in the table). These results demonstrate the effectiveness of our method. We also analyzed the effect of different distance metrics, Cosine and L2, used within the contrastive objectives $\mathcal{L}_{DD}$ and $\mathcal{L}_{OA}$. We compared separate uses of each metric, as well as a hybrid approach: Cosine similarity for $\mathcal{L}_{DD}$ and L2 distance for $\mathcal{L}_{OA}$. The hybrid approach ($\mathcal{L}_{all}$ in the table) achieved the best performance and stability. It effectively resolves potential conflicts between metrics and positions ordinal-aligned features optimally within the feature space. This ensures that while these features are aligned, their magnitudes can vary, ultimately leading to improved model generalization.

## 5 Discussion

Our framework is broadly applicable to various sensor-based applications, including environmental monitoring, medical diagnostics, and industrial process control. Importantly, it introduces no additional computational cost during inference. The latency, memory footprint, and overall complexity of the deployed model remain identical to a model with the same architecture trained using conventional methods. The only added overhead arises during training, where the $g_D$ requires one additional forward pass, and the contrastive losses require pairwise feature similarity computations, leading to quadratic complexity with respect to batch size. In practice, this burden is mitigated by performing training on dedicated high-performance machines, while deployment is carried out on resource-constrained edge or embedded devices. A key assumption of the framework is the existence of stable label-aligned features across domains. In cases where label-feature relationships vary significantly, such as when similar labels correspond to distinct sensor reading patterns due to concept drift, or when non-linear or non-monotonic dependencies violate the assumption of gradual label transitions, the effectiveness of ordinal alignment may be reduced.

## 6 Conclusion and Future Work

This paper presented a new framework for improving domain generalization in time-series regression by learning ordinal-aligned, task-specific features. Our method aligns representations with label order while explicitly disentangling domain-dependent variations. Extensive experiments across diverse real-world sensor datasets demonstrate that the framework reliably models subtle variations in sensor readings, preserves ordinal continuity across domains, and consistently outperforms existing approaches. For future work, we plan to conduct pilot studies in real-world deployment settings to further evaluate and refine the framework's performance. A key objective is to extend the method to accommodate complex, non-linear, and non-monotonic relationships between features and labels, which may violate the assumption of gradual label–feature alignment. We will also optimize the model architecture and training pipeline for improved efficiency, introduce a more compact loss formulation to simplify hyperparameter tuning, and adapt the framework to multi-dimensional regression targets. These developments will expand its applicability across a broader range of sensor reading characteristics and label distributions, advancing robust domain generalization for complex regression tasks in dynamic, real-world environments.

## Acknowledgment

The authors acknowledge support from Australia-China Centre for Energy Informatics and Demand Response Technologies through Department of Industry, Innovation and Science, Australia (ACRIII000004). Dr. Wei Li acknowledges the support of the Australian Research Council (ARC) through the Discovery Early Career Researcher Award (DE210100263). Professor Zomaya and Dr. Wei Li acknowledge the support of an ARC Discovery Project (DP200103494).

## References

1. Avendano, D.N., Deschrijver, D., Van Hoecke, S.: Unsupervised transfer learning across different data modalities for bearing's speed identification. International Journal of Acoustics & Vibration **29**(2) (2024)
2. Bui, M.H., Tran, T., Tran, A., Phung, D.: Exploiting domain-specific features to enhance domain generalization. Advances in Neural Information Processing Systems **34**, 21189–21201 (2021)
3. Chang, X., Li, W., Shi, Y., Zomaya, A.Y.: Taming the domain shift in multi-source learning for energy disaggregation. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 3805–3816 (2023)
4. Chen, H., Zhang, Q., Huang, Z., Wang, H., Zhao, J.: Towards domain-specific features disentanglement for domain generalization. arXiv preprint arXiv:2310.03007 (2023)
5. Du, Y., Wang, J., Feng, W., Pan, S., Qin, T., Xu, R., Wang, C.: Adarnn: Adaptive learning and forecasting of time series. In: Proceedings of the 30th ACM international conference on information & knowledge management. pp. 402–411 (2021)
6. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
7. Guan, Z., Yu, J., Shi, Z., Liu, X., Yu, R., Lai, T., Yang, C., Dong, H., Chen, R., Wei, L.: Dynamic graph transformer network via dual-view connectivity for autism spectrum disorder identification. Computers in Biology and Medicine **174**, 108415 (2024)
8. He, H., Queen, O., Koker, T., Cuevas, C., Tsiligkaridis, T., Zitnik, M.: Domain adaptation for time series under feature and label shifts. In: International Conference on Machine Learning. pp. 12746–12774. PMLR (2023)
9. Huang, Z., Zhu, M., Xia, X., Shen, L., Yu, J., Gong, C., Han, B., Du, B., Liu, T.: Robust generalization against photon-limited corruptions via worst-case sharpness minimization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16175–16185 (2023)
10. Krueger, D., Caballero, E., Jacobsen, J.H., Zhang, A., Binas, J., Zhang, D., Le Priol, R., Courville, A.: Out-of-distribution generalization via risk extrapolation. In: International Conference on Machine Learning. pp. 5815–5826. PMLR (2021)
11. Lee, B.T., Kwon, J.m., Jo, Y.Y.: Tada: Temporal adversarial data augmentation for time series data. arXiv preprint arXiv:2407.15174 (2024)
12. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
13. Li, D., Liu, L., Qi, Y., Liu, S., Luo, Z.: A cross-domain short-term power prediction method based on multiple environmental sensors in photovoltaic systems and domain generalisation theory. IEEE Sensors Journal (2024)
14. Li, J., Yang, Y., Chen, Y., Zhang, J., Lai, Z., Pan, L.: Dwlr: domain adaptation under label shift for wearable sensor. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. pp. 4425–4433 (2024)
15. Liu, Z., Yang, X.: A generalization model for arrhythmia classification based on spectral feature extraction and domain generalization. In: the 5th International Symposium on Artificial Intelligence for Medicine Science. pp. 378–384 (2024)

16. Lu, W., Wang, J., Sun, X., Chen, Y., Ji, X., Yang, Q., Xie, X.: Diversify: A general framework for time series out-of-distribution detection and generalization. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
17. Lu, W., Wang, J., Yu, H., Huang, L., Zhang, X., Chen, Y., Xie, X.: Fixed: Frustratingly easy domain generalization with mixup. In: Conference on Parsimony and Learning. pp. 159–178. PMLR (2024)
18. Luo, C., Xie, Z., Huang, Y., Chen, G., Yao, H., Zhang, J., Cheng, L., Xu, W., Li, J.: Laserkey: Eavesdropping keyboard typing leveraging vibrational emanations via laser sensing. IEEE Transactions on Mobile Computing (2025)
19. Ma, N., Liu, F., Wang, H., Zhou, S., Bu, J., Han, B.: Domain generalization in regression (2024)
20. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
21. Magrini, F., Jozinović, D., Cammarano, F., Michelini, A., Boschi, L.: Len-db - local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale. Zenodo (2020), `https://zenodo.org/records/3648232`, data set
22. Meng, Y., Dong, Z., Lu, K.C., Li, S., Shao, C.: Meta-learning-based domain generalization for cost-effective tool condition monitoring in ultrasonic metal welding. IEEE Transactions on Industrial Informatics (2024)
23. Murray, D., Stankovic, L., Stankovic, V.: An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. Scientific data **4**(1), 1–12 (2017)
24. Otović, E., Njirjak, M., Jozinović, D., Mauša, G., Michelini, A., Stajduhar, I.: Intra-domain and cross-domain transfer learning for time series data—how transferable are the features? Knowledge-Based Systems **239**, 107976 (2022)
25. Qian, H., Pan, S.J., Miao, C.: Latent independent excitation for generalizable sensor-based cross-person activity recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 11921–11929 (2021)
26. Qin, T., Wang, S., Li, H.: Evolving domain generalization via latent structure-aware sequential autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(12), 14514–14527 (2023)
27. Shi, Y., Li, W., Chang, X., Yang, T., Sun, Y., Zomaya, A.Y.: On enabling collaborative non-intrusive load monitoring for sustainable smart cities. Scientific Reports **13**(1), 6569 (2023)
28. Shi, Y., Li, W., Zomaya, A.Y.: Domain generalization for time-series forecasting via extended domain-invariant representations. In: 2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT). pp. 110–116. IEEE (2024)
29. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: Computer Vision–ECCV Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. pp. 443–450. Springer (2016)
30. Sun, Q., Murphy, K.P., Ebrahimi, S., D'Amour, A.: Beyond invariance: test-time label-shift adaptation for addressing" spurious" correlations. Advances in Neural Information Processing Systems **36**, 23789–23812 (2023)
31. Vapnik, V.N., Vapnik, V., et al.: Statistical learning theory (1998)
32. Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H.C., Werner, P., Al-Hamadi, A., Crawcour, S., Andrade, A.O., da Silva, G.M.: The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In: IEEE international conference on cybernetics. pp. 128–131. IEEE (2013)

33. Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., Yu, P.: Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering (2022)
34. Wang, Y., Liu, F., Chen, Z., Wu, Y.C., Hao, J., Chen, G., Heng, P.A.: Contrastive-ace: Domain generalization through alignment of causal mechanisms. IEEE Transactions on Image Processing **32**, 235–250 (2022)
35. Yang, H.R., Ren, C.X., Luo, Y.W.: Cod: Learning conditional invariant representation for domain adaptation regression. arXiv preprint arXiv:2408.06638 (2024)
36. Zamzmi, G., Paul, R., Salekin, M.S., Goldgof, D., Kasturi, R., Ho, T., Sun, Y.: Convolutional neural networks for neonatal pain assessment. IEEE Transactions on Biometrics, Behavior, and Identity Science **1**(3), 192–200 (2019)
37. Zha, K., Cao, P., Son, J., Yang, Y., Katabi, D.: Rank-n-contrast: Learning continuous representations for regression. Advances in Neural Information Processing Systems **36** (2024)
38. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
39. Zhang, W., Deng, L., Zhang, L., Wu, D.: A survey on negative transfer. IEEE/CAA Journal of Automatica Sinica **10**(2), 305–329 (2022)