NeuCoReClass AD: Redefining Self-Supervised Time Series Anomaly Detection

Aitor Sánchez-Ferrera, Usue Mori, Borja Calvo, Jose A. Lozano

Abstract—**Time series anomaly detection plays a critical role in a wide range of real-world applications. Among unsupervised approaches, self-supervised learning has gained traction for modeling normal behavior without the need of labeled data. However, many existing methods rely on a single proxy task, limiting their ability to capture meaningful patterns in normal data. Moreover, they often depend on handcrafted transformations tailored specific domains, hindering their generalization accross diverse problems. To address these limitations, we introduce Neu-CoReClass AD, a self-supervised multi-task time series anomaly detection framework that combines contrastive, reconstruction, and classification proxy tasks. Our method employs neural transformation learning to generate augmented views that are informative, diverse, and coherent, without requiring domain-specific knowledge. We evaluate NeuCoReClass AD across a wide range of benchmarks, demonstrating that it consistently outperforms both classical baselines and most deep-learning alternatives. Furthermore, it enables the characterization of distinct anomaly profiles in a fully unsupervised manner. Our code is publicly available at https://github.com/Aitorzan3/NeuCoReClass-AD.**

Index Terms—**Time series analysis, anomaly detection, self-supervised learning, contrastive learning, neural networks.**

I. INTRODUCTION

Time series refer to collections of measurements indexed in chronological order, describing the behavior of a system or entity over time [1]. Among the various tasks in time series data mining, such as classification, forecasting, and clustering, anomaly detection has gained increasing attention in recent years. Time series anomalies—also referred to as outliers, novelties, or out-of-distribution samples—are abnormal events where the behavior of the system described by the time series deviates from an expected normality [2]. Time series anomaly detection (TSAD) aims to identify these events and has broad applications in intrusion detection [3], financial fraud detection [4], and healthcare [5], among other fields.

In the literature, two main types of TSAD problems are distinguished: (a) identifying abnormal points and subsequences within an individual time series, and (b) detecting anomalous time series in a database comprising multiple complete time series [6]. In this work we address the latter.

Due to the difficulties of employing supervised learning in anomaly detection contexts (i.e. cost of labeling, class unbalance), unsupervised techniques have gained attention in the past few years [7]. In this setting, machine learning models are trained on a dataset that is assumed to contain only normal samples or, at worst, a negligible number of anomalous samples. Once trained, they make use of the learnt

representations about normal data to compute an abnormality degree—commonly referred to as the anomaly score—for new samples. A high anomaly score indicates a high likelihood that a sample is anomalous, according to the model [8].

Within the unsupervised proposals, recent approaches seek to improve representation learning to enhance anomaly detection, with self-supervised learning becoming increasingly prominent [9]. In self-supervised anomaly detection, models are trained on one or more pretext tasks designed to extract relevant characteristics of the normal data. These proxy tasks involve predicting specific data parts, attributes, or relationships that are present on normal data [10]. By training on these pretext tasks, models are expected to develop a better understanding of the structure and properties of the normal training data. Once models are trained, the learnt data representations are leveraged to perform the downstream task, which represents the final objective—in this case, anomaly detection.

The main difference among the self-supervised works in the literature lies in the choice of the pretext tasks for training the models [11]. Typically, normal training samples are transformed to generate diverse versions of them, commonly referred to in the literature as augmented views. Proxy tasks then focus on predicting specific attributes of these augmented views in relation to the original samples.

One of the most widely used proxy tasks in the literature is *self-supervised reconstruction* via autoencoders [12]. Autoencoders map normal samples into a more compact latent space, and then reconstruct them by minimizing the reconstruction error between the model's outputs and the original input samples. During inference, the reconstruction error of the trained autoencoder serves as the anomaly score for new samples.

A major limitation of standard autoencoders is their tendency to overfit to the normal training data, learning to reconstruct seen samples with high accuracy but failing to generalize to unseen variations of normality [6]. To mitigate this issue, some studies apply a transformation to each original input sample to generate an augmented view that introduces variations with respect to the original sample. The autoencoder is then trained to reconstruct the original sample using its augmented view as input. The key criterion for selecting an appropriate transformation for self-supervised reconstruction is ensuring that it generates augmented samples that reflect diverse variations of the original data, encouraging the model to focus on more robust and generalizable features [6]. Common transformations in this approach include masking (removing parts of the input) [13], [14] and injecting Gaussian noise [15], [16].

In addition to reconstructing normal samples from their transformed versions, other works in TSAD build upon a type of proxy task termed as *self-supervised classification* [11]. In this approach, a fixed set of transformations is applied to each normal sample, producing one augmented view per transformation, and a classifier is trained to predict which transformation was applied to the original sample to generate each of its augmented views [17]. After training, the same set of transformations is applied to each new input sample, and the anomaly score is computed as the model's classification error across its associated augmented views.

When using this type of proxy task, the transformations must disrupt the normality of the data by converting normal samples into abnormal ones, ideally ensuring that the distributions of the augmented views from different transformations do not overlap. For example, the authors in [8] propose generating augmented views of the original sequences by upscaling them with different magnitudes for water leak detection. In this context, upscaling increases the water flow represented in the sequences, which reflects the type of anomaly targeted in this scenario. Similarly, [18] introduces transformations that alter both the amplitude and frequency of sequences to facilitate the detection of seizures in EEG signals.

Although the methods aligned with self-supervised reconstruction and classification achieve strong results, they commonly rely on manually defined transformations. Their main drawback is that designing hand-crafted transformations that are suitable for these proxy tasks requires specialized knowledge about each specific problem. This restricts the general applicability of the models, thus limiting their capacity to be effectively transferred across different anomaly detection tasks [6].

To alleviate this, recent studies have explored the use of learnable neural transformations in a methodology known as *neural transformation learning* [19]. Learnable neural transformations are neural networks that replace manually defined transformations to generate augmented views for the pretext tasks. These transformations are automatically learnt from normal training data through contrastive learning ensuring that they satisfy specific criteria in the model's latent space. In contrastive learning, the input data is mapped into a structured latent space via an encoder network. The objective is to produce representations where similar data samples or parts (positive pairs) are positioned close together in the latent space, while dissimilar ones (negative pairs) are pushed apart [20]. In the context of neural transformation learning for anomaly detection, the positive and negative pairs are defined in such a way that the neural transformations generate augmented views that meet a predefined set of conditions.

One of the most prominent methods in neural transformation learning is NeuTraL AD [19]. NeuTraL AD employs Convolutional Neural Networks (CNNs) as neural transformations, alongside a convolutional encoder, both learnt through contrastive learning to enforce specific properties of the augmented views in the latent space. The authors argue that the transformations and the encoder should generate diverse views of each sample without redundant information. In addition, they argue that the augmented views should retain relevant semantic information from the original sample. To achieve this, NeuTraL AD introduces the *Deterministic Contrastive Loss* (DCL), which defines the original sample and its associated augmented views as the positive pairs, while considering the augmented views as negative with respect to each other. The DCL has been further extended to other studies in the field of neural transformation learning for TSAD, where it is combined with additional contrastive learning-based proxy tasks [21], [22].

NeuTraL AD is a highly relevant study, as it is the first to introduce neural transformation learning for anomaly detection across different types of data and addresses the limitations of manually defined transformations. However, the DCL does not fully ensure that the augmented views fulfill the criteria proposed in NeuTraL AD. One major limitation stems from the inherent trade-off between diversity and information preservation in the latent space. Merely pulling augmented views closer to the original instance does not necessarily ensure the retention of relevant information of the original sample. Additionally, enforcing similarity between the original sample and all its augmented views causes them to converge toward the same point in the latent space. This effect reduces transformation diversity, limiting their separation and restricting the model's ability to learn distinct and informative variations of the original data. As a consequence, the learnt latent space may become overly compact, making the model less effective in detecting diverse types of anomalies.

These issues highlight the need for more flexible learning strategies that balance diversity and information preservation without collapsing the latent space. Additionally, most TSAD methods that identify anomalous sequences focus on producing highly discriminative anomaly scores to flag them, but they do not enable the characterization of distinct anomalous profiles among the detected anomalies. As a result, characterizing different types of anomalies in an unsupervised way remains a largely unexplored problem. This task requires identifying and separating heterogeneous anomaly patterns without any prior knowledge about them.

In this paper, we propose a novel self-supervised TSAD method that addresses the limitations of existing approaches. Our method builds on prior research suggesting that combining different proxy tasks enhances representation learning [23], thereby improving the characterization of normal data patterns and the subsequent anomaly detection. Specifically, we integrate neural transformation learning via contrastive learning, self-supervised reconstruction, and self-supervised classification into a multi-task framework to leverage their complementary strengths, enabling better identification of different types of anomalies. Based on the proxy tasks that we combine in our method, we call it **NeuCoReClass AD**.

The contributions of this work are the followings:

- We present a self-supervised TSAD method that combines multiple proxy tasks to improve representation learning, leading to better characterization of normal data patterns and enhanced anomaly detection performance.
- We introduce a flexible loss function and architecture that ensure that the generated augmented views better satisfy

the conditions required by the proxy tasks in our method, improving their effectiveness in anomaly detection.

- We propose a method that not only detects anomalies with high performance, but also enables the characterization of different anomaly types in an unsupervised manner.
- We release our implementation as open-source code to facilitate reproducibility and future research.

The rest of the paper is structured as follows. Section II introduces our method, NeuCoReClass AD. Section III describes the experimental setup for evaluating our approach against state-of-the-art baselines. Section IV presents the experimental results and additional analyses of NeuCoReClass AD's properties. Finally, Section V summarizes our conclusions and outlines directions for future research.

II. PROPOSED METHOD: NEUCORECLASS AD

In this paper, we propose NeuCoReClass AD. Our method considers neural transformation learning, reconstruction, and classification to learn transformations that generate informative and diverse augmented views of the data, suitable for the different proxy tasks considered to perform TSAD. The pipeline and learning objectives of NeuCoReClass AD are depicted in Figure 1.

In Section II-A, we describe the architecture of NeuCoRe-Class AD. Following, Section II-B defines the desired properties that the generated augmented views should satisfy to be effective across the proxy tasks considered in our approach. Section II-C details the training process of NeuCoReClass AD. Finally, II-D presents how the representations of the augmented views of new samples can be used to compute their anomaly scores.

A. Architecture of NeuCoReClass AD

Our model's pipeline consists of four key components: (i) a neural transformation module, (ii) a feature extraction module, (iii) a reconstruction module, and (iv) a linear classification module. These modules enable the model to learn to perform the the three proxy tasks.

1) Neural Transformation Module: The neural transformation module comprises a set of K learnable neural transformations $\mathcal{T} = \{T_1, ..., T_K\}$. These are neural networks representing parameterized functions whose parameters $\theta_{\mathcal{T}} = \{\theta_{T_1}, ..., \theta_{T_K}\}$ are learnt via gradient-based optimization. When applied to an input sample x_i, these transformations generate a set of augmented views associated to the original sample $\mathcal{T}(x_i) = \{T_1(x_i), ..., T_K(x_i)\} = \{x_i^1, ..., x_i^K\}$, where x_i^k denotes the augmented view that is generated by applying the k-th transformation $T_k(\cdot)$ to that sample. Following the existing literature on self-supervised classification for anomaly detection [6], the first transformation $T_1(\cdot)$ in the neural transformation module is the identity function, i.e., $T_1(x_i) = x_i$.

2) Feature Extraction Module: The feature extraction module consists of an encoder $\phi(\cdot)$ that maps the augmented views, generated by the neural transformation module, into a more compact and abstract latent space. We denote the parameters of the encoder $\phi(\cdot)$ as θ_{ϕ}. After generating the augmented views of a sample x_i, their corresponding latent representations are computed using the encoder $\phi(\cdot)$, resulting in $\phi(\mathcal{T}(x_i)) = \{z_i^1, ..., z_i^K\}$.

3) Reconstruction Module: The reconstruction module is based on a decoder $\psi(\cdot)$, which is added atop the encoder $\phi(\cdot)$. The parameters of the decoder are denoted as θ_{ψ}. This reconstructs the original sample x_i from each latent view z_i^k, such that ideally $\psi(z_i^k) = x_i$, for $1 \leq k \leq K$.

4) Linear Classification Module: Finally, the linear classification module employs a linear classifier $f(\cdot)$, whose parameters are denoted as θ_f. The classifier takes the latent representations of the augmented views associated with a sample $\{z_i^1, ..., z_i^K\}$ as inputs and performs self-supervised classification. In other words, it predicts which transformation was applied to the original input sample for generating each of its augmented views, i.e., ideally $f(z_i^k) = k$ for $1 \leq k \leq K$.

B. Revisiting Neural Transformation Learning for TSAD

The conditions that the augmented views of the samples must satisfy in the latent space are explicitly derived from those established in previous works for each of the respective proxy tasks [6].

We assume that each transformation should be coherent, meaning that it generates a latent subspace where the representations of its associated augmented views are positioned closely together. In addition, we require the latent representations of the augmented views to fulfill the following rules:

1) **Disruption**. The augmented views must represent 'corrupted' versions of their corresponding original samples, thereby disrupting the normal structure of the data.
2) **Diversity**. They must also reflect diverse variations of their associated original samples, ideally avoiding redundancy in the information they retain.
3) **Preservation**. The augmented views must preserve relevant semantic information from their corresponding original samples.

The disruption rule is derived from prior work in the field of self-supervised classification [24]. The diversity rule is a shared requirement across all three proxy tasks considered in our method. Finally, the preservation rule is inherited from the conditions established for neural transformation learning in NeuTraL AD [19]. Figure 1b shows an example latent space structure that meets the criteria below.

C. Training NeuCoReClass AD

We consider three learning objectives that enable NeuCoReClass AD to jointly optimize the contrastive, reconstruction, and classification proxy tasks. These objectives are designed to enforce the rules introduced in Section II-B: the contrastive and classification losses promote the disruption and diversity rules, while the reconstruction loss supports the preservation rule.

During the forward pass, a minibatch of B normal samples is randomly sampled from the training set. For each sample x_i, the neural transformation module and the encoder are used to generate latent representations of its K augmented views, denoted as $\{z_i^1, ..., z_i^K\}$, for $1 \leq i \leq B$ (see Figure 1a).
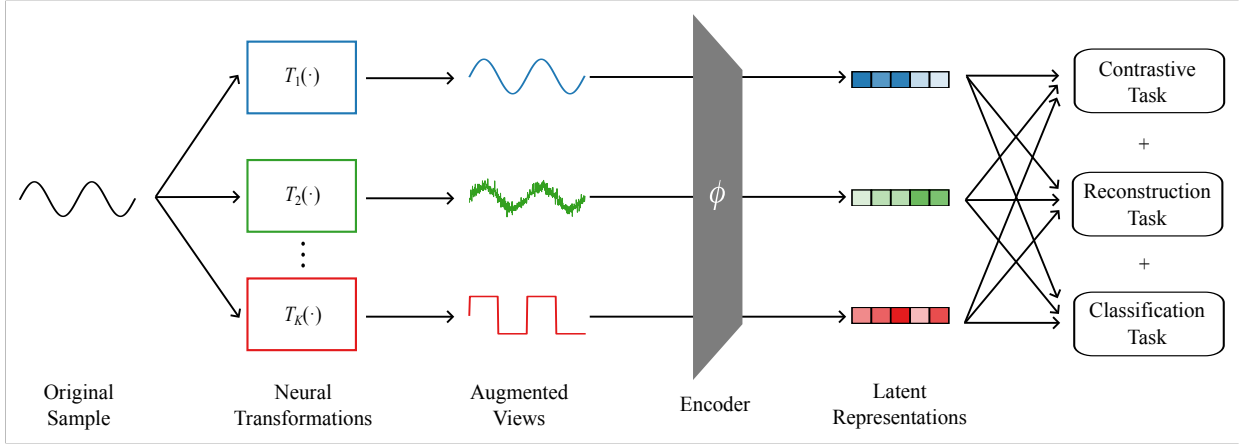
(a) Pipeline of NeuCoReClass AD. Each sequence is processed by a set of K neural transformations to generate diverse augmented views, with the first one representing the identity. These views are then encoded into latent representations, which serve as the input for three self-supervised proxy tasks: contrastive learning, reconstruction, and classification.
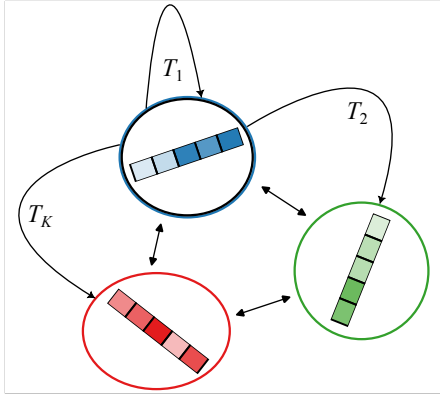
(b) Contrastive task. The representations of the augmented views generated by the same transformation are pulled together, while pushing apart the rest.
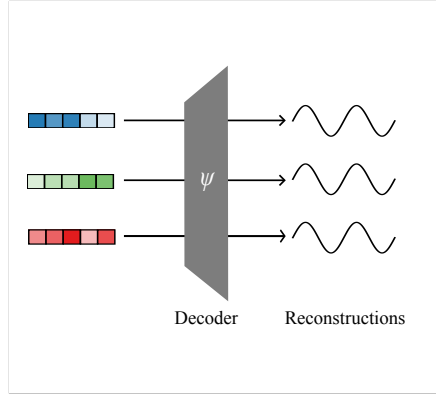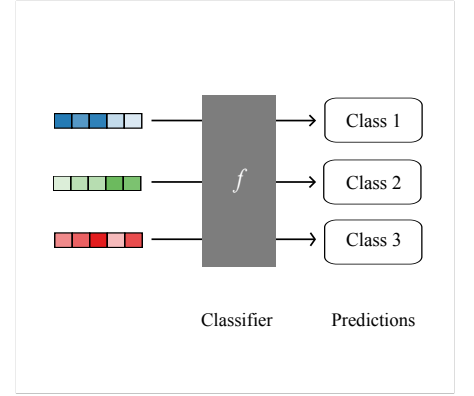
(c) Reconstruction task. The latent representations of each augmented view are used to reconstruct the original input sequence.

(d) Classification task. The latent representations of the augmented views are used to predict the transformation that generated them.

Fig. 1: Overview of NeuCoReClass AD. The top panel shows the generation process of the augmented views and their codification into the latent space. The bottom row illustrates the three self-supervised learning tasks used to guide the training of the method. Different colors denote different transformations.

These latent representations serve as the basis for learning the three proxy tasks. The losses corresponding to the three objectives are then computed and combined in a multi-task learning fashion. Finally, during the backward pass, gradients are propagated and model parameters are updated using the mean of the loss for all the minibatch.

1) Contrastive Loss: To encourage coherent representations of augmented views, the embedding of an augmented view z_i^k should be close to the embeddings of other samples' augmented views z_j^k that are generated by the same transformation $T_k(\cdot)$, with $j \neq i$ and $1 \leq j \leq B$. Simultaneously, to promote diversity and disrupt normality, each representation z_i^k is pushed away from the embeddings of augmented views from all the samples in the minibatch that are generated using different transformations z_j^q, with $q \neq k$ (see Figure 1b).

Following prior work on contrastive learning [25], we define the similarity score between two augmented views in the latent

space as follows:

$$h(x_i^k, x_j^k) = \exp\left(\text{sim}(z_i^k, z_j^k)/\tau\right), \quad (1)$$

where $\text{sim}(\cdot, \cdot)$ denotes the cosine similarity measure, and τ is a temperature hyperparameter that controls the sharpness of the contrastive loss. Lower values of τ amplify the distinction between positive and negative samples, while higher values produce a smoother separation.

To enforce that the latent space of our model is structured according to both the disruption and diversity criteria, we define the following contrastive loss for each sample x_i in the minibatch:

$$\ell_{\text{con}}(x_i) = -\frac{1}{K}\sum_{k=1}^{K}\log\left(\frac{\text{Pos}_k}{\text{Pos}_k + \text{Neg}_k}\right) \quad (2)$$

where

$$\text{Pos}_k = \sum_{j=1}^{B} h(x_i^k, x_j^k), \quad \text{Neg}_k = \sum_{q \neq k}^{K} \sum_{j=1}^{B} h(x_i^k, x_j^q). \quad (3)$$

Here, Pos_k aggregates the similarities between x_i^k and the augmented views of all other samples in the minibatch associated with the k-th transformation (the positive pairs). Conversely, Neg_k aggregates the similarities of x_i^k and the augmented views of the samples in the minibatch generated by transformations $q \neq k$ (negative pairs). During the backward pass, this loss is used to update the parameters of the neural transformation module and the feature extraction module $\theta_{\mathcal{T}}, \theta_\phi$.

2) Reconstruction loss: The proposed contrastive loss encourages that the latent representations of the augmented views meet the rules of disruption and diversity in the latent space generated by the encoder $\phi(\cdot)$. However, these representations may converge to a trivial solution as all augmented views associated with the same transformation may be mapped to a constant embedding, a phenomenon commonly referred to in the literature as 'hypersphere collapse' [26]. In this situation, the resulting augmented views lose all the information they obtained from the original sample, which violates the third rule about preservation proposed in Section II-B.

To avoid this, our decoder reconstructs the original sample x_i from each latent view z_i^k, such that ideally $\psi(z_i^k) = x_i$, for $1 \leq k \leq K$ (see Figure 1c). This is enfoced by the reconstruction loss:

$$\ell_{\text{rec}}(x_i) = \frac{1}{K} \sum_{k=1}^{K} (x_i - \psi(z_i^k))^2. \quad (4)$$

In the backward pass, This loss is used to update the parameters of the neural transformation module, the feature extraction module, and the reconstruction module, $T_{\theta_k}, \theta_\phi, \theta_\psi$, for $1 \leq k \leq K$.

3) Classification loss: Finally, the linear classification module $f(\cdot)$ is trained to perform the self-supervised classification (see Figure 1d). This is accomplished by minimizing the following cross-entropy loss for every sample x_i in the minibatch:

$$\ell_{\text{class}}(x_i) = -\frac{1}{K} \sum_{k=1}^{K} y_i^k \log\big(f(z_i^k)\big), \quad (5)$$

where y_i^k is a one-hot vector indicating the true class of z_i^k, and $f(z_i^k)$ outputs its associated predicted probability distribution over the K classes. In the backward pass, this objective is used to update the parameters of the neural transformation module, the feature extractor and the classification module $T_{\theta_k}, \theta_\phi, \theta_f$, for $1 \leq k \leq K$.

4) Multi-task learning in NeuCoReClass AD: We train our model concurrently on the three proxy tasks described above, combining their loss functions in a multi-task learning setting. To prevent the optimization process from being dominated by losses with larger magnitudes, potentially leading to the neglect of others, we adopt the uncertainty-based weighting scheme proposed in [27]. This method automatically adjusts the contribution of each task based on its predictive uncertainty, promoting a more balanced and adaptive optimization across tasks.

To implement this scheme, we introduce a learnable uncertainty parameter for each loss: σ_{con} for the contrastive loss, σ_{rec} for reconstruction, and σ_{class} for classification. These parameters are learnt together with the model's other parameters and determine the contribution of each loss to the overall learning objective. Accordingly, we optimize the following general loss for every sample x_i in a minibatch:

$$\ell(x_i) = \frac{1}{2\sigma_{\text{con}}^2} \ell_{\text{con}}(x_i) + \frac{1}{2\sigma_{\text{rec}}^2} \ell_{\text{rec}}(x_i) + \frac{1}{2\sigma_{\text{class}}^2} \ell_{\text{class}}(x_i)$$
$$+ \ln(1 + \sigma_{\text{con}}) + \ln(1 + \sigma_{\text{rec}}) + \ln(1 + \sigma_{\text{class}}). \quad (6)$$

D. Anomaly Detection in NeuCoReClass AD

Once trained, NeuCoReClass AD reproduces the same three proxy tasks with new samples and uses a variation of the training loss to compute an anomaly score. The idea is that the model performs worse on anomalous samples, since it was trained only with normal data.

Each proxy task gives an anomaly score. For a new sample x_{new}, the scores from the reconstruction $\text{AS}_{\text{rec}}(x_{\text{new}})$ and classification $\text{AS}_{\text{class}}(x_{\text{new}})$ tasks are the same as their respective losses (see Equations 4 and 5). However, the original contrastive loss needs a minibatch of samples to create positive and negative pairs, which is not suitable for scoring one sample at a time. Therefore, we define a modified contrastive anomaly score $\text{AS}_{\text{con}}(x_{\text{new}})$ based on Equation 2, but without relying on minibatches:

$$\text{AS}_{\text{con}}(x_{\text{new}}) = -\frac{1}{K} \sum_{k=1}^{K} \log \frac{1}{1 + \sum_{q \neq k}^{K} h(x_i^k, x_j^q)}. \quad (7)$$

This score considers only the augmented views of x_{new}, measuring their separation in latent space.

The overall anomaly score for a new sample, $\text{AS}(x_{\text{new}})$, combines the three task-specific anomaly scores. We weight each score by the uncertainty parameters learnt during training. Thus, the overall anomaly score for a new sample is computed as:

$$\begin{aligned} \text{AS}(x_{\text{new}}) = {} & \frac{1}{2\sigma_{\text{con}}^2} \text{AS}_{\text{con}}(x_{\text{new}}) \\ & + \frac{1}{2\sigma_{\text{rec}}^2} \text{AS}_{\text{rec}}(x_{\text{new}}) \\ & + \frac{1}{2\sigma_{\text{class}}^2} \text{AS}_{\text{class}}(x_{\text{new}}) \end{aligned} \quad (8)$$

III. EXPERIMENTAL SETUP

In this section, we describe the experimental setup used to evaluate the performance of NeuCoReClass AD. We assess the method across a diverse set of benchmark datasets and compare it against both shallow and deep anomaly detection approaches under diverse evaluation settings.

A. Datasets

For the experimentation, we consider datasets from the UCR Time Series Classification Archive [28], encompassing both univariate and multivariate time series from various domains. We adopt a one-class classification framework: for each dataset with N classes, n classes are designated as normal, and the remaining $N - n$ are treated as anomalies. Following prior work [19], we evaluate performance under two settings: *one-vs-rest*, where one class is considered normal and the remaining $(N - 1)$ classes are treated as anomalies; and $(N-1)$-*vs-rest*, where $(N - 1)$ classes are considered normal and the remaining one is treated as anomalous. This latter scenario is more challenging, as it requires models to handle greater intra-class variability within the normal category.

Considering these two settings, we derive a total of $2N$ anomaly detection problems from each dataset with N classes. Since the UCR archive contains a large number of benchmarks, we select the first 10 univariate and 10 multivariate datasets that satisfy the following criteria:

- The dataset contains at most 10 classes.
- Each class includes at least 30 sequences.
- All sequences in the dataset are of equal length.
- No sequence contains missing values.

On this basis, we conduct experiments on a total of 144 TSAD problems. The characteristics of the datasets considered in our experimentation are presented in Appendix A.

B. Baselines for Time Series Anomaly Detection

We evaluate NeuCoReClass AD by comparing it with standard shallow and deep unsupervised methods.

1) Shallow Baselines: We select three popular shallow anomaly detection baselines widely used in the literature: One-Class SVM (OCSVM) [29], which learns a compact decision boundary around the normal data; Isolation Forest (IF) [30], which detects anomalies by recursively partitioning the feature space and isolating outliers; and Local Outlier Factor (LOF), which flags samples with substantially lower local density than their neighbors [31].

2) Deep Baselines: Among common deep baselines, we include an autoencoder [32], which computes anomaly scores based on reconstruction errors; Deep SVDD, a deep variant of One-Class SVM that operates in the latent space of an autoencoder [33]; and DAGMM, which combines a deep autoencoder with a Gaussian mixture model to estimate density in the latent space of the model [34]. We also consider deep self-supervised baselines, including NeuTraL AD [19], the first neural transformation learning approach for anomaly detection, and FixedTS, a self-supervised classification scheme for TSAD based on fixed transformations, inspired by [35].

C. Implementation and Experimental Details

Here we describe the implementation choices and training protocols used for NeuCoReClass AD and all baseline methods.

1) Implementation of Methods: Shallow baselines are implemented using the scikit-learn library [36], adopting their predefined default hyperparameters without further tuning.

For NeuCoReClass AD and all deep learning baselines, we adopt the encoder architecture proposed in TS2Vec [37]: a linear projection layer that maps the input channels to hidden channels of size 64, a dilated CNN module containing 10 hidden blocks of "GELU $\rightarrow$ DilatedConv $\rightarrow$ GELU $\rightarrow$ DilatedConv" with skip connections between adjacent blocks. For the i-th block, the dilation of the convolution is set to 2^i. The kernel size is set to 3. Each hidden dilated convolution has a channel size of 64. Finally, an output residual block maps the hidden channels to the output channels, resulting in a latent space with 320 dimensions [37].

For NeuCoReClass AD and NeuTraL AD, learnable transformations are modeled following the architecture of NeuTraL AD [19]. Each transformation consists of three residual blocks of 1D convolutional layers with instance normalization and ReLU activations, followed by a final 1D convolutional layer. The resulting augmented views preserve the dimensionality of the original data. Finally, all transformations are implemented as feed-forward networks, without recurrence, for simplicity and efficiency.

2) Training and Validation Protocol: For shallow models, the entire set of normal samples in the training set is used for model fitting. In the case of deep methods, we split the normal samples into training (90%) and validation (10%) subsets. Models are trained on the training set for up to 10,000 epochs, and the validation set is used to monitor overfitting and adjust the learning rate.

Specifically, we employ a `ReduceLROnPlateau` scheduler that decreases the learning rate when the validation anomaly score plateaus (patience of 10 epochs), and apply early stopping when the learning rate drops below 10^{-6}.

We use a batch size of 32 for all datasets, except for *Motor Imagery* and *Ethanol Concentration*, where we reduce it to 16 to accommodate the longer sequences and higher dimensionality of these datasets, which increase memory requirements.

3) Additional Hyperparameters: Certain hyperparameters specific to the considered deep methods must be fixed for all experiments. Below, we describe these hyperparameters along with the values we have chosen for them.

The key hyperparameter of DAGMM is the number of Gaussian components in its mixture-density network—that is, the number of GMM mixtures modeling the latent-space energy distribution. Based on [34], we fix this to 3 components.

Following the methodology employed in the experimentation of NeuTraL AD [19], we set the fixed transformations in FixedTS to be all the possible combinations of flipping along the time axis (true/false), flipping along the channel axis (true/false), and shifting along the time axis by 0.25 of its time length (forward/backward/none). This leads to a total of 12 transformations. For consistency, we also use 12 transformations in NeuTraL AD and NeuCoReClass AD.

Finally, the temperature parameter for the contrastive loss in NeuTraL AD and NeuCoReClass AD is fixed at 0.1, a standard choice in prior work in contrastive learning [25].

D. Evaluation Protocol

For each problem considered, each method is trained exclusively on the normal samples from its corresponding training set and then used to compute anomaly scores on the unseen instances of the evaluation set.

Following prior work, model performance is assessed using two widely adopted metrics in anomaly detection: the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPR). [38]. The AUROC evaluates the model's ability to distinguish between normal and anomalous samples across varying thresholds, while the AUPR is particularly suited to imbalanced scenarios, summarizing the trade-off between precision and recall. High scores in both metrics indicate that the model effectively prioritizes anomalous instances over normal ones, while maintaining precision in its detections.

IV. EXPERIMENTAL RESULTS

In this section, we present the experimental evaluation of NeuCoReClass AD. Section IV-A reports the quantitative performance of all methods across all the anomaly detection problems considered under both *one-vs-rest* and *(N-1)-vs-rest* protocols. Section IV-B illustrates how the generated augmented views satisfy the disruption, diversity, and preservation criteria. Finally, Section IV-C shows how per-transformation contributions enable unsupervised characterization of distinct anomaly types.

A. Results and Discussion

We evaluate each method in each problem over five runs with different random seeds to mitigate random variability. Tables I and II summarize the results under the one-vs-rest and (N-1)-vs-rest settings, respectively.

Each table summarizes the mean AUROC and AUPR across all datasets for each method; the signed difference (Δ) between each method's mean and NeuCoReClass AD's mean for each metric; and its average rank (lower is better). The values are computed by averaging first over the random seeds, and then across the problems. In the Δ columns, a negative value indicates the method performs worse than NeuCoReClass AD, while a positive value indicates it performs better. The best results in each column are underlined and bolded, and NeuCoReClass AD's entries are shaded in dark gray whenever it is not the top-performing method. The complete results of each method in each problem considered for the experimentation are presented in Appendix B.

Under the *one-vs-rest* setting (see Table I), NeuCoReClass AD achieves a mean AUROC of 73.22% and a mean AUPR of 81.42%. It outperforms all self-supervised and traditional baselines (IF, LOF, OC-SVM, AutoAD, FixedTS) by 1.9–19.3 points in AUROC and 0.8–4.1 points in AUPR. Against NeuTraL AD—the neural transformation learning competitor—it wins by 3.02% in AUROC and 1.38% in AUPR. The method with the closest performance is DeepSVDD, which edges it by a mere +0.40% in AUROC and +0.23% in AUPR, an essentially negligible margin.

Now we focus on the *(N-1)-vs-rest setting*, which is more challenging because it must handle greater intra-class variance among the normal class (see Table II). NeuCoReClass AD achieves a mean AUROC of 66.63% and a mean AUPR of 46.77%, outperforming every other method. It beats NeuTraL AD by 5.55 points in AUROC and 8.81 points in AUPR, and even surpasses DeepSVDD by 6.22 points in AUROC and 8.56 points in AUPR under this tougher protocol. Among the baselines, LOF is the strongest alternative, with a mean AUROC of 62.62% (-4.01%) and AUPR of 45.26% (-1.52%). Moreover, NeuCoReClass AD attains the best average ranks (3.65 in AUROC vs. 4.97 for DeepSVDD and 4.64 for LOF; 4.06 in AUPR vs. 5.06 and 4.92, respectively).

Appendices C and D report the results of the ablation and sensitivity studies, respectively. These analyze the impact of the training objectives and the effect of the key hyperparameters on the performance of NeuCoReClass AD.

B. Illustration of the Properties of the Augmented Views

We analyze how jointly training on the proposed learning objectives encourages the augmented views to satisfy the principles of disruption, diversity, and preservation, as defined in Section II-B.

This analysis is conducted using the *Epilepsy* dataset, which includes four human action recognition activities: *sawing*, *walking*, *running* and *epilepsy*. We train NeuCoReClass AD to characterize the *sawing* class as the normal class, while the remaining three classes are treated as anomalous. The posterior analysis is conducted using all the samples in the evaluation set.

1) Evidence of Disruption and Diversity: The contrastive and classification objectives are designed to guide the model to organize the latent space into separated subspaces, each associated with a specific transformation. To assess whether this structure emerges in practice, we apply t-SNE [39] to visualize the latent representations of the augmented views generated from both normal and anomalous samples in the evaluation set.

As shown in Figure 2, the proxy tasks guide the encoder to produce a structured latent space in which the augmented views of normal samples form well-separated clusters, each corresponding to a different transformation. This behavior reflects the fulfillment of the corruption and diversity principles. In contrast, the augmented anomalous samples result in latent representations that are more scattered and entangled, suggesting a failure to align with the learnt transformation structure and thereby exposing their atypical nature.

2) Evidence of Semantic Preservation: To assess whether the augmented views retain meaningful information from their original time series, we analyze the model's ability to reconstruct the original samples from their augmented representations. Specifically, we evaluate how well the decoder reconstructs unseen normal and anomalous samples by computing the reconstruction error for each input dimension. These errors are then visualized using box plots, allowing us to observe whether normal samples exhibit systematically lower reconstruction errors—a key indicator that semantic information is preserved through the transformations.

TABLE I: Average AUROC and AUPR scores, Δ vs NeuCoReClass AD, and rankings over all problems under the *one-vs-rest* setting.

Método	Mean AUROC (%)	Δ AUROC (%)	Rank	Mean AUPR (%)	Δ AUPR (%)	Rank
IF	66.32	−6.89	6.07	78.53	−2.89	5.92
LOF	70.19	−3.02	4.69	80.34	−1.07	4.81
OC-SVM	70.63	−2.59	4.43	80.32	−1.10	4.47
AutoAD	71.28	−1.93	4.56	80.58	−0.84	4.62
DAGMM	53.94	−19.27	7.28	77.56	−3.86	6.50
DeepSVDD	**73.62**	**+0.40**	**3.33**	**81.65**	**+0.23**	**3.44**
FixedTS	66.59	−6.63	5.88	78.38	−3.03	6.10
NeuTraL AD	70.20	−3.01	5.17	80.04	−1.38	5.36
NeuCoReClass AD	73.22	—	3.60	81.42	—	3.78

TABLE II: Average AUROC and AUPR scores, Δ vs NeuCoReClass AD, and rankings over all problems under the *(N-1)-vs-rest* setting.

Método	Mean AUROC (%)	Δ AUROC (%)	Rank	Mean AUPR (%)	Δ AUPR (%)	Rank
IF	57.55	−9.08	5.72	36.60	−10.18	6.00
LOF	62.62	−4.01	4.64	45.26	−1.52	4.92
OC-SVM	58.97	−7.66	5.33	38.34	−8.43	5.42
AutoAD	59.87	−6.76	5.14	37.91	−8.86	5.22
DAGMM	55.29	−11.34	6.08	39.90	−6.88	4.51
DeepSVDD	60.41	−6.22	4.97	38.22	−8.56	5.06
FixedTS	63.33	−3.30	4.53	44.87	−1.90	4.82
NeuTraL AD	61.08	−5.55	4.93	37.97	−8.81	4.99
NeuCoReClass AD	**66.63**	—	**3.65**	**46.77**	—	**4.06**

As shown in Figure 3, the reconstruction error for normal samples is consistently low across all input dimensions, suggesting that the semantic content of the original time series is effectively preserved through the augmented views. In contrast, anomalous samples yield substantially higher reconstruction errors, reflecting their misalignment with the semantic structure learnt from normal data.

C. Characterization of Different Anomaly Types

While the previous analysis focused on how NeuCoReClass-AD distinguishes between normal and anomalous samples, a complementary question emerges: can the model also capture different types of anomalies based on the contribution of each transformation to the anomaly score? To explore this, we analyze the individual contribution of each transformation to the anomaly score for every sample in the evaluation set of the *Epilepsy* dataset.

We compare the structure of the samples in the evaluation set in two different spaces: the original input space and the transformation-contribution space, where each sample is represented by a vector of K elements, and each element in that vector denotes the anomaly score of the sample derived from each transformation. In both cases, we use t-SNE to project the high-dimensional data into two dimensions for visualization.

As shown in Figure 4, the projection based on the original features (Figure 4a) exhibits no clear structure or clustering that corresponds to the true anomaly classes. In contrast, the

projection based on the transformation-wise anomaly contributions (Figure 4b) reveals well-separated clusters that align with the different types of anomalies. This highlights the capability of NeuCoReClass-AD not only to detect anomalous behavior, but also to uncover structure within the anomaly space. Therefore, the contribution patterns of the transformations act as interpretable descriptors, enabling the model to differentiate between types of anomalies in a completely unsupervised manner.

V. CONCLUSIONS AND FUTURE WORK

This paper presents NeuCoReClass AD, a self-supervised time series anomaly detection method that integrates contrastive, reconstructive, and classification proxy tasks within a unified multi-task framework. This yields augmented views that are informative, diverse, and semantically coherent. NeuCoReClass AD outperforms traditional baselines and most deep-learning methods under both *one-vs-rest* and *(N-1)-vs-rest* settings. While DeepSVDD slightly outperforms it in the *one-vs-rest* setting (+0.40 % AUROC, +0.23 % AUPR), NeuCoReClass AD achieves significantly better results under the more challenging *(N-1)-vs-rest* scenario (+6.22 % AUROC, +8.56 % AUPR).

Latent-space visualizations and reconstruction-error analyses confirm that the generated augmented views fulfill the disruption, diversity, and preservation criteria, enabling both clear cluster separation and semantic retention. Additionally, analyzing transformation-specific anomaly scores reveals Neu-

(a) Latent representations of the augmented views of normal samples.

(b) Latent representations of the augmented views of anomalous samples.
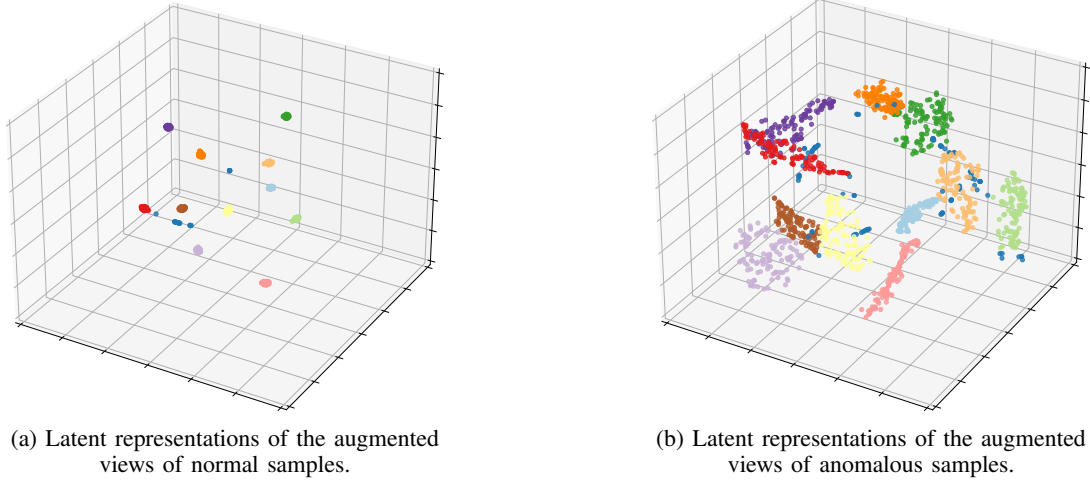
Fig. 2: Visualization of the latent space induced by the joint training with reconstruction, contrastive, and classification objectives. Different colors represent different classes.

(a) Reconstruction error per dimension for normal samples.

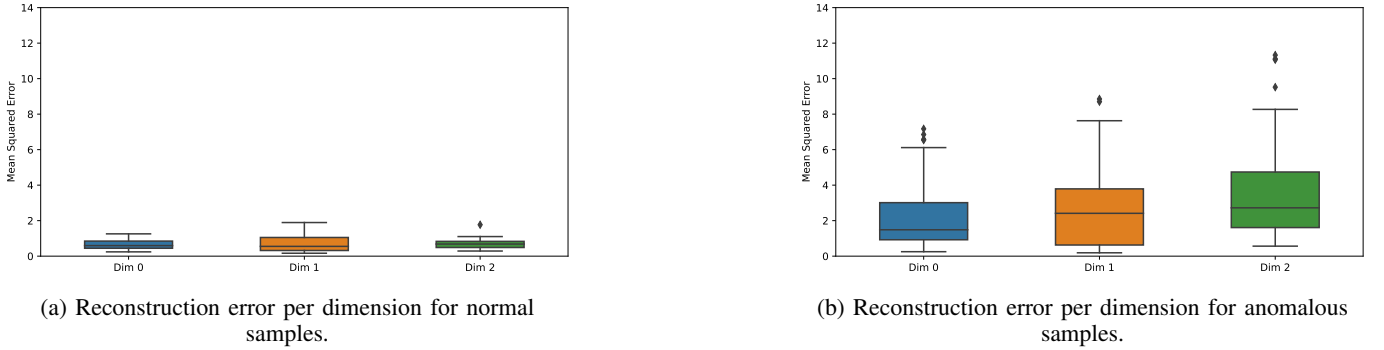(b) Reconstruction error per dimension for anomalous samples.

Fig. 3: Comparison of the reconstruction errors per input dimension between normal and anomalous test samples. Different colors represent different dimensions.

CoReClass AD's ability to identify and differentiate distinct types of anomalies in a fully unsupervised manner.

For future work, we plan to explore hybrid transformation modules (e.g., attention-based, or frequency-based) to better capture different anomaly profiles. In addition, we propose to combine this with pruning strategies to retain only the most informative transformations, improving performance without added cost. Lastly, learning anomaly-subspace prototypes and mapping them to the input space may enhance explainability, offering interpretable representations of each anomaly type directly in the original data domain.

REFERENCES

[1] J. D. Hamilton, *Time series analysis*. Princeton university press, 1994.

[2] A. Carreño, I. Inza, and J. A. Lozano, "Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework," *Artificial Intelligence Review*, vol. 53, pp. 3575–3594, 2020.

[3] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for iot time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2019.

[4] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial fraud: a review of anomaly detection techniques and recent advances," *Expert systems With applications*, vol. 193, p. 116429, 2022.

[5] J. Pereira and M. Silveira, "Learning representations from healthcare time series data for unsupervised anomaly detection," in *2019 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 2019, pp. 1–7.

[6] A. Sánchez-Ferrera, B. Calvo, and J. A. Lozano, "A review on self-supervised learning for time series anomaly detection: Recent advances and open challenges," *arXiv preprint arXiv:2501.15196*, 2025.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[8] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "Water leak detection using self-supervised time series classification," *Information Sciences*, vol. 574, pp. 528–541, 2021.

[9] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," *ACM Computing Surveys*, vol. 57, no. 1, pp. 1–42, 2024.
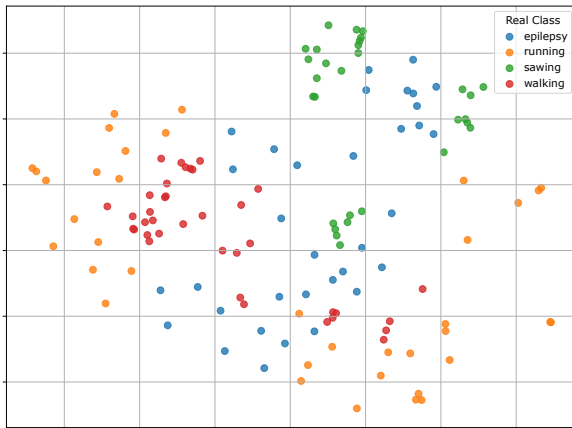
(a) Dimensionality reduction over the original samples in the evaluation set.
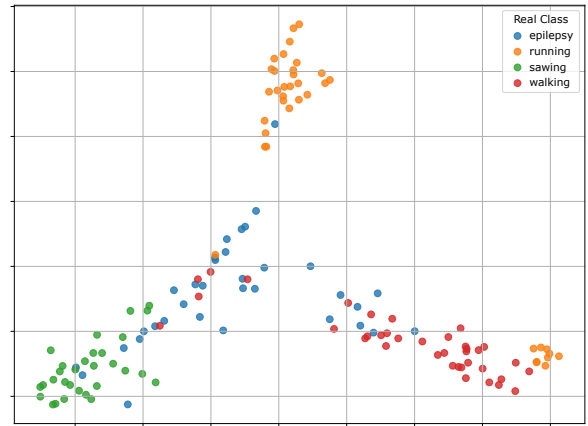
(b) Dimensionality reduction over the transformation-wise anomaly contributions.

Fig. 4: Class separability in the *Epilepsy* dataset across different feature spaces. Each point corresponds to a sample from the evaluation set, and each color indicates one of the ground-truth classes.

[10] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 1, pp. 857–876, 2021.

[11] H. Hojjati, T. K. K. Ho, and N. Armanfard, "Self-supervised anomaly detection: A survey and outlook," *arXiv preprint arXiv:2205.05173*, 2022.

[12] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *arXiv preprint arXiv:2003.05991*, 2020.

[13] Y. Fu and F. Xue, "Mad: Self-supervised masked anomaly detection task for multivariate time series," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–8.

[14] M. Liu, Z. Xu, and Q. Xu, "Deepfib: Self-imputation for time series anomaly detection," *arXiv preprint arXiv:2112.06247*, 2021.

[15] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, 2014, pp. 4–11.

[16] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2015, pp. 1996–2000.

[17] J. Zhang, K. Saleeby, T. Feldhausen, S. Bi, A. Plotkowski, and D. Womble, "Self-supervised anomaly detection via neural autoregressive flows with active learning," Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), Tech. Rep., 2021.

[18] Y. Zheng, Z. Liu, R. Mo, Z. Chen, W.-s. Zheng, and R. Wang, "Task-oriented self-supervised learning for anomaly detection in electroencephalography," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 193–203.

[19] C. Qiu, T. Pfrommer, M. Kloft, S. Mandt, and M. Rudolph, "Neural transformation learning for deep anomaly detection beyond images," in *International conference on machine learning*. PMLR, 2021, pp. 8703–8714.

[20] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 1, p. 2, 2020.

[21] K. Chen, M. Feng, and T. S. Wirjanto, "Time-series anomaly detection via contextual discriminative contrastive learning," *arXiv preprint arXiv:2304.07898*, 2023.

[22] T. Schneider, C. Qiu, M. Kloft, D. A. Latif, S. Staab, S. Mandt, and M. Rudolph, "Detecting anomalies within time series using local neural transformations," *arXiv preprint arXiv:2202.03944*, 2022.

[23] C. Doersch and A. Zisserman, "Multi-task self-supervised visual learning," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2051–2060.

[24] J. Yoo, T. Zhao, and L. Akoglu, "Data augmentation is a hyperparameter: Cherry-picked self-supervision for unsupervised anomaly detection is creating the illusion of success," *arXiv preprint arXiv:2208.07734*, 2022.

[25] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[26] R. Wang, C. Liu, X. Mou, K. Gao, X. Guo, P. Liu, T. Wo, and X. Liu, "Deep contrastive one-class time series anomaly detection," in *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, 2023, pp. 694–702.

[27] L. Liebel and M. Körner, "Auxiliary tasks in multi-task learning," *arXiv preprint arXiv:1805.06334*, 2018.

[28] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, "The ucr time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293–1305, 2019.

[29] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, 1999.

[30] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.

[31] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[32] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, "Autoencoder-based network anomaly detection," in *2018 Wireless telecommunications symposium (WTS)*. IEEE, 2018, pp. 1–5.

[33] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *International conference on machine learning*. PMLR, 2018, pp. 4393–4402.

[34] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[35] S. Wang, Y. Zeng, X. Liu, E. Zhu, J. Yin, C. Xu, and M. Kloft, "Effective end-to-end unsupervised outlier detection via inlier priority of discriminative network," *Advances in neural information processing systems*, vol. 32, 2019.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[37] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "Ts2vec: Towards universal representation of time series," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 8, 2022, pp. 8980–8987.

[38] D. Huang, L. Shen, Z. Yu, Z. Zheng, M. Huang, and Q. Ma, "Effi-

cient time series anomaly detection by multiresolution self-supervised discriminative network," *Neurocomputing*, vol. 491, pp. 261–272, 2022.

[39] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[40] X. Fan, Y. Yue, P. Sarkar, and Y. R. Wang, "On hyperparameter tuning in general clustering problemsm," in *International conference on machine learning*.   PMLR, 2020, pp. 2996–3007.

APPENDIX A
DATASETS

We summarize the key characteristics of each dataset used in our experiments. Table III provides an overview of each dataset, including the number of classes, train/test sample counts, fixed sequence length, dimensionality, and data source. The source column denotes the origin or nature of the data. Note that the train size refers to the number of samples in the original training set of the time series classification problem. In each experiment, we only consider samples from the class(es) designated as normal for models' training.

APPENDIX B
COMPLETE RESULTS

This appendix presents the complete results of the experiments conducted for each dataset and method. We report the mean AUROC and AUPR performance scores for all problems considered under both evaluation settings.

Specifically, we include four tables: (i) Table IV shows the mean AUROC scores under the *one-vs-rest* setting; (ii) Table V presents the corresponding AUPR results for the same setting; (iii) Table VI displays the AUROC scores under the *(N-1)-vs-rest* setting; and (iv) Table VII contains the corresponding AUPR results. Each value represents the average over five independent runs with different seeds. These detailed results complement the summary statistics reported in the main paper and provide a comprehensive view of the performance of each method across all problems.

APPENDIX C
ABLATION STUDIES

To assess the contribution of each self-supervised objective in NeuCoReClass AD, we conduct an ablation study by comparing the full model against six reduced variants. These variants selectively consider one or two of the three training objectives: self-supervised reconstruction, transformation classification, and contrastive learning. The evaluated combinations include: (1) Classification only, (2) Contrastive only, (3) Reconstruction only, (4) Contrastive + Classification, (5) Reconstruction + Classification, and (6) Reconstruction + Contrastive. The results are reported in Table VIII and Table IX, which correspond to the two evaluation settings used in our experiments: *one-vs-rest* and *(N-1)-vs-rest*, respectively.

Each table shows the average AUROC and AUPR scores across all problems for each ablation setting, together with the performance of the full NeuCoReClass AD model. We also include the relative performance gap (in parentheses) between each variant and the full model.

Interestingly, we observe that the configuration combining only reconstruction and classification achieves slightly better scores than the full model. This can be attributed to the fact that both contrastive and classification losses encourage similar latent space structure—fulfilling the disruption and diversity properties described in Section II-B. However, despite this marginal improvement, we argue that the contrastive loss plays a key role in enabling semantic characterization of the anomalies.

To support this claim, we repeat the characterization experiment introduced in Section IV-C using the ablated model that combines only the reconstruction and classification objectives. We conduct this analysis using the *Epilepsy* dataset with the class *sawing* defined as normal and the rest as anomalies. We compare two feature spaces: (i) the original input space, and (ii) the transformation-contribution space, where each sample is represented by the anomaly scores assigned to each transformation. We use t-SNE to project both spaces into two dimensions for visualization, as shown in Figure 5.

In the original experiment conducted considering the three learning objectives of NeuCoReClass AD, the projection based on the original features reveals no apparent class structure, while the projection based on transformation contributions uncovers clear and well-separated clusters that correspond to distinct anomaly types. In contrast, when repeating this experiment with the ablated model that excludes the contrastive loss, such structure is no clearly observed. This indicates that the contrastive loss is essential for enriching the characterization of different anomaly types in an unsupervised way.

APPENDIX D
SENSITIVITY TO HYPERPARAMETERS

Fixing hyperparameter values in unsupervised scenarios can be challenging, as ground truth annotations are lacking for validation [40]. Therefore, we fix the key hyperparameters of NeuCoReClass AD to popular values from the literature—the number of neural transformations K and the temperature τ in the contrastive loss—and evaluate its robustness with respect to them. This analysis shows that our method maintains stable performance across different values, thereby demonstrating its robustness and its practical applicability.

A. Sensitivity to Number of Transformations

We vary the number of learnable transformations K with $K \in \{3, 4, 7, 11, 15\}$ on the *Electric Devices* (univariate) and *Epilepsy* (multivariate) datasets under both the *one-vs-rest* and the $(N-1)$-vs-rest settings. For each value of K, we train NeuCoReClass AD with five random seeds for each problem derived from each dataset and report the mean AUROC and AUPR in percentage terms. Figures 6 and 7 show that both metrics improve slightly as K increases, yet remain largely stable, indicating robustness to the exact choice of transformations. This trend holds across both the one-vs-rest and $(N-1)$-vs-rest settings.

B. Sensitivity to Temperature

To verify robustness with respect to the contrastive loss temperature τ, we try different popular values for τ in the literature, $\tau \in \{0.05, 0.07, 0.1, 0.2, 0.5\}$, under both evaluation settings on the Electric Devices (univariate) and Epilepsy (multivariate) datasets. For each τ, NeuCoReClass AD is trained with five seeds in each anomaly detection problem derived from the dataset, and we record mean AUROC and AUPR. Figures 8 and 9 plot both metrics versus τ (log scale), showing that the method performs consistently for all the values considered.

TABLE III: Overview of the 20 UCR time series datasets used in the experiments.

Dataset	Classes	Train	Test	Length	Dim	Source
ChlorineConcentration	3	467	3840	166	1	Simulated
Computers	2	250	250	720	1	Device
DistalPhalanxOutlineCorrect	2	600	276	80	1	Image
DistalPhalanxOutlineAgeGroup	3	400	139	80	1	Image
Earthquakes	2	322	139	512	1	Sensor
ECG200	2	100	100	96	1	ECG
ElectricDevices	7	8926	7711	96	1	Device
EthanolLevel	4	504	500	1751	1	Spectro
FordA	2	3601	1320	500	1	Sensor
FordB	2	3636	810	500	1	Sensor
Epilepsy	4	137	138	206	3	HAR
EthanolConcentration	4	261	263	1751	3	Spectro
FaceDetection	2	5890	3524	62	144	EEG
FingerMovements	2	316	100	50	28	EEG
HandMovementDirection	4	160	74	400	10	EEG
Heartbeat	2	204	205	405	61	Audio
MotorImagery	2	278	100	3000	64	EEG
NATOPS	6	180	180	51	24	HAR
PEMS-SF	7	267	173	144	963	Other
PenDigits	10	7494	3498	8	2	Motion
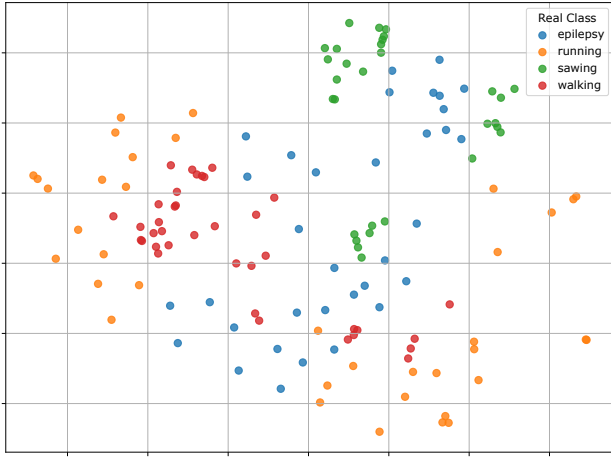
(a) Dimensionality reduction over the original samples in the evaluation set.
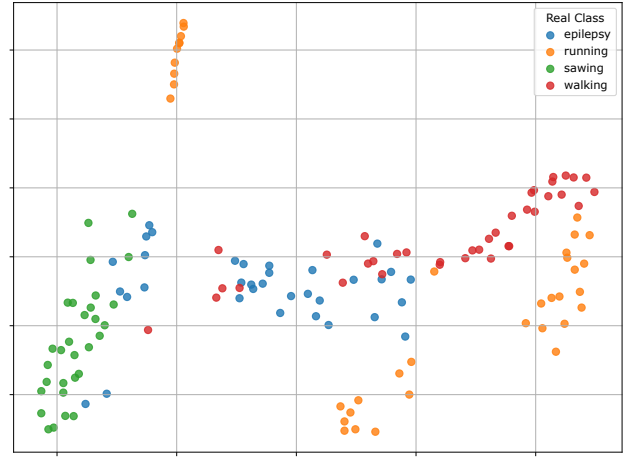
(b) Dimensionality reduction over the transformation-wise anomaly contributions of the ablated model.

Fig. 5: Class separability in the *Epilepsy* dataset across different feature spaces for the ablated model. Each point corresponds to a sample from the evaluation set, and each color indicates one of the ground-truth classes.
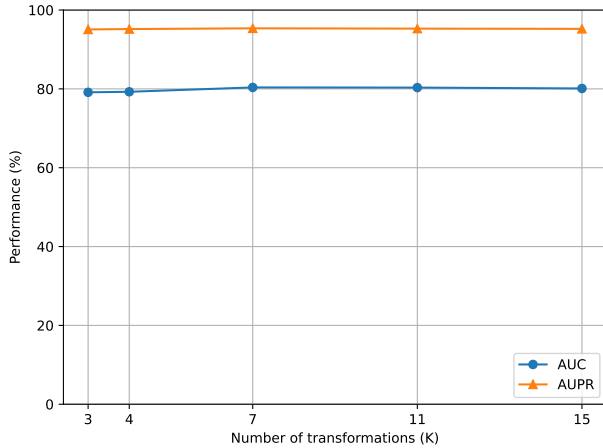
(a) one-vs-rest setting.

(b) $(N-1)$-vs-rest setting.

Fig. 6: Sensitivity of NeuCoReClass AD to the number of transformations K on the *Electric Devices* dataset.

TABLE IV: Mean AUROC scores and rankings for all problems under the *one-vs-rest* setting

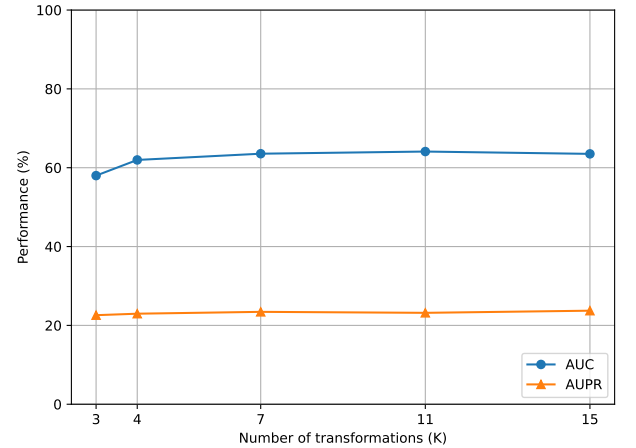| Dataset_Normality	IF	LOF	OCSVM	AutoAD	DAGMM	DeepSVDD	FixedTS	NeuTraL AD	NeuCoReClass AD
ChlorineConcentration_0	41.10 (7.00)	38.47 (8.00)	36.76 (9.00)	41.75 (5.00)	50.83 (3.00)	41.22 (6.00)	46.07 (4.00)	52.38 (2.00)	**58.35 (1.00)**
ChlorineConcentration_1	53.26 (6.00)	52.17 (8.00)	56.91 (3.00)	54.88 (4.00)	53.74 (5.00)	52.33 (7.00)	51.64 (9.00)	59.74 (2.00)	**66.92 (1.00)**
ChlorineConcentration_2	60.04 (5.00)	56.34 (9.00)	66.39 (2.00)	58.98 (6.00)	56.78 (8.00)	61.78 (4.00)	57.07 (7.00)	61.87 (3.00)	**68.54 (1.00)**
Computers_0	30.28 (9.00)	62.00 (2.00)	54.21 (5.00)	**63.26 (1.00)**	45.52 (7.00)	50.42 (6.00)	35.40 (8.00)	57.24 (4.00)	61.91 (3.00)
Computers_1	**69.17 (1.00)**	49.56 (8.00)	54.59 (5.00)	52.17 (7.00)	48.79 (9.00)	55.90 (4.00)	61.43 (2.00)	59.27 (3.00)	54.41 (6.00)
DistalPhalanxOutlineAgeGroup_0	68.32 (8.00)	**79.09 (1.00)**	73.39 (6.00)	76.44 (3.00)	75.97 (4.00)	78.86 (2.00)	49.24 (9.00)	71.29 (7.00)	75.02 (5.00)
DistalPhalanxOutlineAgeGroup_1	**75.57 (1.00)**	74.92 (2.00)	73.79 (3.00)	73.68 (4.00)	31.10 (8.00)	61.01 (5.00)	26.44 (9.00)	34.34 (7.00)	41.12 (6.00)
DistalPhalanxOutlineAgeGroup_2	81.45 (3.00)	68.48 (8.00)	79.63 (5.00)	78.20 (6.00)	53.59 (9.00)	82.17 (2.00)	80.13 (4.00)	73.95 (7.00)	**83.56 (1.00)**
DistalPhalanxOutlineCorrect_0	50.29 (2.00)	40.74 (8.00)	46.97 (4.00)	**52.48 (1.00)**	46.25 (6.00)	47.87 (3.00)	32.81 (9.00)	46.83 (5.00)	43.29 (7.00)
DistalPhalanxOutlineCorrect_1	71.85 (6.00)	**83.08 (1.00)**	71.96 (5.00)	67.27 (8.00)	48.97 (9.00)	70.80 (7.00)	72.27 (3.00)	73.73 (2.00)	72.11 (4.00)
ECG200_0	52.16 (6.00)	22.68 (9.00)	56.51 (2.00)	51.39 (7.00)	42.77 (8.00)	54.02 (4.00)	**59.67 (1.00)**	52.40 (5.00)	56.48 (3.00)
ECG200_1	86.80 (6.00)	89.11 (3.00)	87.41 (5.00)	85.11 (7.00)	89.68 (2.00)	**89.91 (1.00)**	68.63 (9.00)	73.63 (8.00)	87.53 (4.00)
Earthquakes_0	65.55 (4.00)	**66.18 (1.00)**	64.59 (5.00)	41.52 (9.00)	50.39 (6.00)	47.71 (8.00)	65.96 (2.00)	48.27 (7.00)	65.55 (3.00)
Earthquakes_1	35.95 (9.00)	58.26 (6.00)	53.93 (7.00)	67.78 (3.00)	61.86 (4.00)	**70.71 (1.00)**	48.33 (8.00)	68.17 (2.00)	61.31 (5.00)
ElectricDevices_0	**75.95 (1.00)**	67.06 (7.00)	72.03 (3.00)	70.61 (4.00)	41.90 (9.00)	69.55 (5.00)	73.52 (2.00)	66.72 (8.00)	67.51 (6.00)
ElectricDevices_1	13.36 (9.00)	83.67 (5.00)	37.90 (8.00)	84.49 (4.00)	62.37 (7.00)	**96.34 (1.00)**	76.71 (6.00)	93.61 (2.00)	93.51 (3.00)
ElectricDevices_2	75.05 (7.00)	74.24 (8.00)	78.88 (6.00)	81.39 (5.00)	52.21 (9.00)	88.63 (3.00)	81.73 (4.00)	**93.49 (1.00)**	90.20 (2.00)
ElectricDevices_3	**89.05 (1.00)**	85.08 (5.00)	85.31 (4.00)	70.86 (8.00)	47.19 (9.00)	89.04 (2.00)	82.66 (7.00)	83.62 (6.00)	89.04 (3.00)
ElectricDevices_4	48.16 (9.00)	67.99 (4.00)	65.77 (5.00)	**80.69 (1.00)**	48.43 (8.00)	71.18 (3.00)	55.23 (7.00)	61.91 (6.00)	73.13 (2.00)
ElectricDevices_5	83.61 (4.00)	78.53 (8.00)	81.29 (6.00)	80.58 (7.00)	64.41 (9.00)	**86.65 (1.00)**	84.38 (3.00)	85.76 (2.00)	83.57 (5.00)
ElectricDevices_6	44.86 (9.00)	69.78 (2.00)	62.65 (4.00)	**77.30 (1.00)**	55.61 (7.00)	60.64 (6.00)	48.62 (8.00)	66.84 (3.00)	61.45 (5.00)
Epilepsy_0	75.53 (9.00)	90.38 (5.00)	87.19 (6.00)	93.48 (3.00)	91.83 (4.00)	**97.43 (1.00)**	61.29 (9.00)	81.62 (7.00)	93.84 (2.00)
Epilepsy_1	21.70 (8.00)	25.21 (6.00)	20.63 (9.00)	23.69 (7.00)	88.63 (3.00)	**92.59 (1.00)**	80.98 (4.00)	89.99 (2.00)	74.71 (5.00)
Epilepsy_2	75.96 (8.00)	84.89 (7.00)	92.35 (6.00)	96.14 (4.00)	94.79 (5.00)	**99.22 (1.00)**	69.68 (9.00)	99.02 (3.00)	99.09 (2.00)
Epilepsy_3	98.07 (6.00)	98.58 (4.00)	98.77 (3.00)	99.03 (2.00)	50.00 (9.00)	**99.56 (1.00)**	77.80 (8.00)	85.70 (7.00)	98.58 (5.00)
EthanolConcentration_0	61.98 (4.00)	62.66 (3.00)	**65.31 (1.00)**	63.85 (2.00)	54.54 (7.00)	60.02 (5.00)	48.64 (8.00)	45.00 (9.00)	58.60 (6.00)
EthanolConcentration_1	61.26 (2.00)	**62.04 (1.00)**	58.54 (3.00)	52.01 (6.00)	49.95 (8.00)	53.05 (5.00)	50.09 (7.00)	45.82 (9.00)	55.32 (4.00)
EthanolConcentration_2	52.10 (4.00)	53.50 (2.00)	52.27 (3.00)	**53.63 (1.00)**	48.49 (8.00)	51.13 (5.00)	48.47 (9.00)	50.23 (6.00)	49.94 (7.00)
EthanolConcentration_3	42.10 (8.00)	48.35 (6.00)	41.20 (9.00)	43.11 (7.00)	49.67 (4.00)	48.49 (5.00)	51.96 (3.00)	**56.71 (1.00)**	52.02 (2.00)
EthanolLevel_0	48.47 (8.00)	49.04 (7.00)	51.98 (5.00)	54.13 (4.00)	49.34 (6.00)	45.40 (9.00)	**66.90 (1.00)**	57.94 (3.00)	59.77 (2.00)
EthanolLevel_1	56.11 (5.00)	53.39 (7.00)	56.22 (4.00)	52.58 (8.00)	51.28 (9.00)	56.37 (3.00)	**60.82 (1.00)**	53.53 (6.00)	59.66 (2.00)
EthanolLevel_2	56.13 (2.00)	54.94 (5.00)	**56.25 (1.00)**	56.00 (3.00)	50.05 (8.00)	51.06 (6.00)	50.26 (7.00)	49.01 (9.00)	55.91 (4.00)
EthanolLevel_3	52.91 (7.00)	52.80 (8.00)	55.39 (2.00)	54.14 (3.00)	48.06 (9.00)	52.98 (6.00)	**56.77 (1.00)**	53.69 (4.00)	53.62 (5.00)
FaceDetection_0	51.64 (4.00)	**52.25 (1.00)**	52.08 (2.00)	50.90 (6.00)	49.44 (9.00)	51.30 (5.00)	50.31 (8.00)	50.64 (7.00)	52.04 (3.00)
FaceDetection_1	48.91 (5.00)	48.38 (8.00)	48.67 (9.00)	49.33 (4.00)	49.82 (2.00)	48.05 (9.00)	49.43 (3.00)	**49.84 (1.00)**	48.70 (6.00)
FingerMovements_0	62.67 (2.00)	**63.63 (1.00)**	61.78 (3.00)	61.37 (4.00)	54.79 (7.00)	57.72 (5.00)	45.30 (9.00)	46.06 (8.00)	57.04 (6.00)
FingerMovements_1	38.70 (7.00)	39.98 (6.00)	37.49 (9.00)	41.55 (5.00)	**50.50 (1.00)**	37.71 (8.00)	49.62 (2.00)	48.27 (3.00)	41.89 (4.00)
FordA_0	52.11 (5.00)	56.96 (3.00)	56.99 (2.00)	38.31 (9.00)	51.34 (6.00)	**62.05 (1.00)**	46.31 (7.00)	52.71 (4.00)	40.37 (8.00)
FordA_1	52.80 (8.00)	59.80 (6.00)	58.02 (7.00)	69.71 (5.00)	49.18 (9.00)	**74.64 (1.00)**	73.51 (2.00)	71.52 (3.00)	70.85 (4.00)
FordB_0	50.31 (4.00)	**58.66 (1.00)**	57.57 (2.00)	44.78 (9.00)	49.43 (5.00)	50.83 (3.00)	47.74 (6.00)	47.27 (7.00)	47.00 (8.00)
FordB_1	49.65 (8.00)	52.41 (7.00)	54.38 (6.00)	62.27 (5.00)	48.87 (9.00)	64.09 (3.00)	64.41 (2.00)	**66.97 (1.00)**	63.31 (4.00)
HandMovementDirection_0	53.76 (5.00)	49.04 (7.00)	53.67 (6.00)	**60.75 (1.00)**	48.97 (8.00)	57.60 (2.00)	46.42 (9.00)	54.12 (4.00)	55.75 (3.00)
HandMovementDirection_1	57.56 (6.00)	63.86 (3.00)	62.12 (4.00)	**68.76 (1.00)**	50.21 (9.00)	61.58 (5.00)	53.95 (7.00)	53.83 (8.00)	66.70 (2.00)
HandMovementDirection_2	46.13 (9.00)	50.17 (3.00)	49.60 (4.00)	46.98 (7.00)	48.14 (6.00)	**56.20 (1.00)**	53.22 (2.00)	49.20 (5.00)	46.73 (8.00)
HandMovementDirection_3	48.12 (7.00)	48.57 (5.00)	45.71 (9.00)	50.00 (2.00)	48.38 (6.00)	**58.36 (1.00)**	45.98 (8.00)	48.98 (4.00)	49.93 (3.00)
Heartbeat_0	63.25 (3.00)	64.91 (2.00)	**65.55 (1.00)**	60.84 (4.00)	44.44 (9.00)	57.49 (6.00)	51.47 (8.00)	53.99 (7.00)	60.36 (5.00)
Heartbeat_1	39.67 (7.00)	37.34 (8.00)	36.96 (9.00)	40.34 (6.00)	52.60 (2.00)	52.15 (3.00)	**57.97 (1.00)**	47.99 (4.00)	41.66 (5.00)
MotorImagery_0	40.65 (7.00)	41.76 (6.00)	40.04 (8.00)	43.50 (5.00)	50.00 (3.00)	43.50 (4.00)	54.77 (2.00)	**55.83 (1.00)**	39.18 (9.00)
MotorImagery_1	58.16 (4.00)	58.16 (5.00)	60.70 (2.00)	58.40 (3.00)	52.10 (7.00)	57.07 (6.00)	44.18 (8.00)	44.02 (9.00)	**60.90 (1.00)**
NATOPS_0	76.77 (7.00)	77.29 (6.00)	79.93 (5.00)	82.29 (2.00)	64.44 (9.00)	**82.44 (1.00)**	74.27 (8.00)	81.09 (4.00)	81.46 (3.00)
NATOPS_1	80.24 (7.00)	81.83 (6.00)	84.00 (4.00)	83.35 (5.00)	67.83 (8.00)	**85.46 (1.00)**	60.38 (9.00)	84.42 (3.00)	84.77 (2.00)
NATOPS_2	84.15 (7.00)	88.57 (4.00)	88.36 (5.00)	**89.80 (1.00)**	68.14 (8.00)	89.31 (3.00)	49.92 (9.00)	85.37 (6.00)	89.66 (2.00)
NATOPS_3	85.24 (7.00)	87.69 (6.00)	92.31 (4.00)	95.02 (3.00)	74.74 (9.00)	96.67 (2.00)	80.75 (8.00)	88.12 (5.00)	**97.52 (1.00)**
NATOPS_4	85.07 (9.00)	91.76 (6.00)	98.24 (2.00)	97.97 (4.00)	86.40 (8.00)	98.18 (3.00)	89.54 (7.00)	91.89 (5.00)	**98.91 (1.00)**
NATOPS_5	89.49 (7.00)	96.27 (6.00)	**98.78 (1.00)**	96.17 (5.00)	79.95 (8.00)	96.54 (3.00)	54.90 (9.00)	89.91 (6.00)	98.39 (2.00)
PEMS-SF_0	92.80 (7.00)	97.86 (3.00)	**98.09 (1.00)**	97.01 (5.00)	64.67 (9.00)	97.28 (4.00)	94.26 (6.00)	89.54 (8.00)	97.95 (2.00)
PEMS-SF_1	68.73 (9.00)	87.32 (6.00)	90.95 (2.00)	89.57 (5.00)	73.60 (8.00)	**92.24 (1.00)**	81.38 (7.00)	90.36 (4.00)	90.80 (3.00)
PEMS-SF_2	67.17 (8.00)	79.57 (5.00)	82.63 (2.00)	77.92 (6.00)	43.78 (9.00)	**86.67 (1.00)**	76.44 (7.00)	79.67 (4.00)	82.58 (3.00)
PEMS-SF_3	62.28 (8.00)	80.06 (6.00)	**86.61 (1.00)**	83.70 (4.00)	37.33 (9.00)	86.32 (3.00)	79.58 (7.00)	81.33 (5.00)	86.47 (2.00)
PEMS-SF_4	63.55 (8.00)	73.83 (7.00)	79.56 (5.00)	75.27 (6.00)	49.37 (9.00)	**85.47 (1.00)**	82.07 (3.00)	82.29 (2.00)	80.43 (4.00)
PEMS-SF_5	62.20 (9.00)	75.14 (8.00)	84.32 (4.00)	81.18 (6.00)	75.82 (7.00)	**91.59 (1.00)**	86.91 (3.00)	87.47 (2.00)	83.53 (5.00)
PEMS-SF_6	93.18 (8.00)	98.20 (4.00)	**99.35 (1.00)**	97.84 (6.00)	18.77 (9.00)	98.90 (2.00)	96.73 (7.00)	98.08 (5.00)	98.78 (3.00)
PenDigits_0	94.83 (8.00)	98.19 (3.00)	97.72 (5.00)	98.17 (4.00)	39.20 (9.00)	**98.72 (1.00)**	97.01 (6.00)	95.61 (7.00)	98.69 (2.00)
PenDigits_1	95.37 (6.00)	97.31 (5.00)	95.21 (7.00)	97.60 (4.00)	39.77 (9.00)	98.56 (2.00)	97.76 (3.00)	92.93 (8.00)	**99.29 (1.00)**
PenDigits_2	98.36 (7.00)	**99.40 (1.00)**	98.79 (6.00)	99.24 (4.00)	62.40 (9.00)	99.37 (2.00)	99.02 (5.00)	92.52 (8.00)	99.37 (3.00)
PenDigits_3	99.17 (6.00)	**99.57 (1.00)**	99.25 (5.00)	99.38 (3.00)	26.02 (9.00)	99.43 (2.00)	95.23 (7.00)	88.45 (8.00)	99.31 (4.00)
PenDigits_4	98.64 (6.00)	99.88 (4.00)	99.95 (3.00)	99.97 (2.00)	44.19 (9.00)	**99.99 (1.00)**	97.96 (7.00)	95.41 (8.00)	99.84 (5.00)
PenDigits_5	99.14 (5.00)	**99.86 (1.00)**	99.32 (4.00)	98.71 (6.00)	68.44 (9.00)	99.33 (3.00)	98.62 (7.00)	97.98 (8.00)	99.80 (2.00)
PenDigits_6	99.72 (6.00)	**100.00 (1.00)**	99.91 (4.00)	99.92 (4.00)	48.85 (9.00)	99.97 (2.00)	98.38 (7.00)	92.85 (8.00)	99.97 (3.00)
PenDigits_7	94.39 (6.00)	96.03 (3.00)	95.70 (5.00)	95.71 (4.00)	41.97 (9.00)	96.08 (2.00)	90.42 (7.00)	88.46 (8.00)	**97.07 (1.00)**
PenDigits_8	99.09 (6.00)	**99.95 (1.00)**	99.53 (4.00)	98.19 (7.00)	28.78 (9.00)	99.65 (3.00)	99.43 (5.00)	94.81 (8.00)	99.81 (2.00)
PenDigits_9	97.45 (6.00)	**99.25 (1.00)**	98.52 (3.00)	97.89 (5.00)	26.10 (9.00)	97.91 (4.00)	97.01 (7.00)	91.52 (8.00)	99.14 (2.00)

TABLE V: Mean AUPR scores and rankings for all problems under the *one-vs-rest* setting

Dataset_Normality	IF	LOF	OCSVM	AutoAD	DAGMM	DeepSVDD	FixedTS	NeuTraL AD	NeuCoReClass AD
ChlorineConcentration_0	69.38 (7.00)	68.20 (9.00)	68.42 (8.00)	70.64 (5.00)	**77.92 (1.00)**	69.77 (6.00)	73.44 (4.00)	76.35 (3.00)	77.28 (2.00)
ChlorineConcentration_1	79.85 (7.00)	79.36 (8.00)	81.85 (3.00)	80.77 (5.00)	80.80 (4.00)	80.07 (6.00)	78.62 (9.00)	83.21 (2.00)	**85.41 (1.00)**
ChlorineConcentration_2	60.93 (6.00)	58.49 (8.00)	67.77 (2.00)	61.28 (5.00)	58.58 (8.00)	63.42 (4.00)	59.40 (7.00)	63.88 (3.00)	69.26 (1.00)
Computers_0	37.67 (9.00)	57.70 (4.00)	54.30 (5.00)	**64.09 (1.00)**	51.65 (7.00)	52.43 (6.00)	40.45 (8.00)	58.01 (3.00)	59.60 (2.00)
Computers_1	**74.13 (1.00)**	49.09 (9.00)	51.46 (8.00)	51.57 (7.00)	56.87 (5.00)	56.88 (4.00)	61.37 (2.00)	58.81 (3.00)	53.36 (6.00)
DistalPhalanxOutlineAgeGroup_0	92.35 (8.00)	95.22 (3.00)	94.29 (7.00)	94.98 (5.00)	**96.44 (1.00)**	95.75 (2.00)	89.15 (9.00)	95.21 (4.00)	94.66 (6.00)
DistalPhalanxOutlineAgeGroup_1	**80.42 (1.00)**	78.81 (2.00)	74.57 (4.00)	75.77 (3.00)	48.05 (8.00)	64.96 (5.00)	44.84 (9.00)	48.63 (7.00)	55.52 (6.00)
DistalPhalanxOutlineAgeGroup_2	76.85 (4.00)	72.48 (8.00)	75.71 (5.00)	73.93 (6.00)	71.11 (9.00)	78.76 (3.00)	**82.28 (1.00)**	73.75 (7.00)	80.70 (2.00)
DistalPhalanxOutlineCorrect_0	55.74 (2.00)	49.14 (8.00)	52.83 (6.00)	**57.81 (1.00)**	55.70 (3.00)	53.58 (5.00)	46.23 (9.00)	53.60 (4.00)	50.95 (7.00)
DistalPhalanxOutlineCorrect_1	64.17 (6.00)	**78.23 (1.00)**	67.39 (3.00)	63.16 (7.00)	46.82 (9.00)	66.69 (4.00)	61.13 (8.00)	65.52 (5.00)	68.30 (2.00)
ECG200_0	60.18 (6.00)	47.97 (9.00)	61.89 (3.00)	58.79 (7.00)	57.17 (8.00)	60.76 (5.00)	**67.69 (1.00)**	62.56 (2.00)	61.68 (4.00)
ECG200_1	79.70 (6.00)	82.24 (3.00)	81.23 (4.00)	77.34 (7.00)	83.43 (2.00)	**84.10 (1.00)**	62.48 (9.00)	63.25 (8.00)	80.32 (5.00)
Earthquakes_0	33.07 (4.00)	34.70 (3.00)	34.49 (3.00)	20.86 (9.00)	28.91 (6.00)	23.81 (7.00)	**37.36 (1.00)**	22.60 (8.00)	33.00 (5.00)
Earthquakes_1	70.40 (9.00)	81.50 (6.00)	78.18 (7.00)	86.29 (3.00)	83.40 (5.00)	**87.66 (1.00)**	76.97 (8.00)	87.42 (2.00)	84.69 (4.00)
ElectricDevices_0	**96.53 (1.00)**	94.76 (5.00)	95.99 (2.00)	94.91 (4.00)	90.61 (9.00)	94.38 (6.00)	95.69 (3.00)	93.55 (8.00)	93.86 (7.00)
ElectricDevices_1	57.12 (9.00)	93.70 (5.00)	73.78 (8.00)	94.89 (4.00)	85.42 (6.00)	**98.68 (1.00)**	81.73 (7.00)	97.31 (2.00)	97.07 (3.00)
ElectricDevices_2	97.00 (7.00)	96.48 (8.00)	97.29 (6.00)	97.65 (5.00)	91.33 (9.00)	98.71 (3.00)	97.79 (4.00)	**99.27 (1.00)**	98.82 (2.00)
ElectricDevices_3	**97.89 (1.00)**	96.35 (4.00)	94.60 (7.00)	89.77 (8.00)	87.27 (9.00)	97.43 (2.00)	96.14 (5.00)	95.37 (6.00)	96.82 (3.00)
ElectricDevices_4	77.21 (8.00)	86.35 (4.00)	82.96 (5.00)	**91.48 (1.00)**	75.36 (9.00)	86.91 (2.00)	79.22 (7.00)	82.70 (6.00)	86.62 (3.00)
ElectricDevices_5	98.02 (4.00)	94.64 (9.00)	96.89 (5.00)	96.68 (6.00)	94.81 (8.00)	**98.22 (1.00)**	98.04 (3.00)	98.12 (2.00)	95.65 (7.00)
ElectricDevices_6	94.12 (5.00)	96.93 (2.00)	95.83 (6.00)	**97.46 (1.00)**	95.13 (7.00)	96.12 (4.00)	93.69 (9.00)	96.78 (3.00)	96.04 (5.00)
Epilepsy_0	90.44 (8.00)	96.94 (5.00)	95.75 (6.00)	98.02 (3.00)	96.97 (4.00)	**99.29 (1.00)**	83.77 (9.00)	91.88 (7.00)	98.15 (2.00)
Epilepsy_1	59.71 (7.00)	59.98 (6.00)	58.87 (8.00)	58.24 (9.00)	95.71 (3.00)	**96.72 (1.00)**	91.78 (4.00)	96.04 (2.00)	90.76 (5.00)
Epilepsy_2	93.01 (8.00)	95.97 (7.00)	98.11 (6.00)	98.97 (4.00)	98.56 (5.00)	**99.80 (1.00)**	89.63 (9.00)	99.75 (3.00)	99.76 (2.00)
Epilepsy_3	99.26 (6.00)	99.56 (4.00)	99.57 (3.00)	99.67 (2.00)	86.59 (9.00)	**99.86 (1.00)**	89.93 (8.00)	94.01 (7.00)	99.54 (5.00)
EthanolConcentration_0	84.62 (3.00)	**86.09 (1.00)**	84.92 (2.00)	83.58 (4.00)	80.24 (7.00)	80.40 (6.00)	73.45 (8.00)	72.90 (9.00)	80.89 (5.00)
EthanolConcentration_1	80.49 (2.00)	**81.69 (1.00)**	80.11 (3.00)	78.63 (5.00)	77.70 (8.00)	78.83 (4.00)	77.73 (7.00)	74.01 (9.00)	77.98 (6.00)
EthanolConcentration_2	77.27 (4.00)	**78.92 (1.00)**	77.46 (3.00)	76.65 (5.00)	77.51 (2.00)	76.52 (6.00)	72.10 (9.00)	74.73 (8.00)	76.43 (7.00)
EthanolConcentration_3	72.84 (7.00)	74.14 (5.50)	72.04 (9.00)	72.49 (8.00)	78.20 (2.00)	74.14 (5.50)	77.33 (3.00)	**80.34 (1.00)**	76.91 (4.00)
EthanolLevel_0	72.83 (8.00)	74.02 (7.00)	75.06 (6.00)	76.62 (4.00)	75.21 (5.00)	72.76 (9.00)	**85.50 (1.00)**	79.59 (2.00)	78.77 (3.00)
EthanolLevel_1	79.20 (3.00)	76.73 (7.00)	78.45 (4.00)	77.96 (6.00)	76.65 (8.00)	78.38 (5.00)	**81.18 (1.00)**	76.27 (9.00)	79.62 (2.00)
EthanolLevel_2	**78.32 (1.00)**	76.78 (5.00)	77.74 (3.00)	77.46 (4.00)	75.57 (7.00)	75.68 (6.00)	73.80 (8.00)	73.57 (9.00)	77.83 (2.00)
EthanolLevel_3	77.25 (5.00)	**78.04 (1.00)**	77.61 (3.00)	77.64 (2.00)	75.12 (9.00)	76.45 (8.00)	76.76 (7.00)	77.26 (4.00)	77.02 (6.00)
FaceDetection_0	51.64 (3.00)	**51.86 (1.00)**	51.67 (2.00)	50.89 (6.00)	49.88 (9.00)	51.07 (5.00)	50.24 (8.00)	50.43 (7.00)	51.50 (4.00)
FaceDetection_1	49.02 (5.00)	48.74 (8.00)	48.92 (7.00)	49.43 (4.00)	**50.34 (1.00)**	48.60 (9.00)	49.25 (4.00)	49.96 (2.00)	48.98 (6.00)
FingerMovements_0	**64.68 (1.00)**	61.86 (5.00)	63.03 (2.00)	62.83 (3.00)	62.43 (4.00)	59.83 (6.00)	50.17 (8.00)	48.86 (9.00)	58.07 (7.00)
FingerMovements_1	40.72 (7.00)	41.64 (5.00)	40.39 (9.00)	41.44 (6.00)	**55.67 (1.00)**	40.53 (8.00)	48.41 (3.00)	48.57 (2.00)	42.76 (4.00)
FordA_0	49.28 (3.00)	48.01 (6.00)	48.28 (5.00)	38.80 (9.00)	48.66 (4.00)	**57.84 (1.00)**	43.66 (7.00)	49.34 (2.00)	39.53 (8.00)
FordA_1	53.62 (9.00)	70.87 (5.00)	68.95 (6.00)	78.76 (3.00)	55.19 (8.00)	**81.76 (1.00)**	68.69 (7.00)	76.15 (4.00)	81.14 (2.00)
FordB_0	51.37 (4.00)	53.80 (2.00)	**54.87 (1.00)**	44.33 (9.00)	52.35 (3.00)	49.61 (5.00)	48.07 (6.00)	46.56 (7.00)	45.15 (8.00)
FordB_1	49.93 (9.00)	58.69 (6.00)	59.90 (5.00)	66.79 (4.00)	52.74 (8.00)	68.67 (3.00)	56.74 (7.00)	69.30 (2.00)	**69.63 (1.00)**
HandMovementDirection_0	83.13 (6.00)	83.77 (5.00)	84.15 (3.00)	**86.18 (1.00)**	80.46 (9.00)	84.08 (4.00)	78.31 (9.00)	80.83 (7.00)	84.26 (2.00)
HandMovementDirection_1	66.22 (6.00)	72.93 (3.00)	72.27 (4.00)	**76.83 (1.00)**	67.88 (5.00)	64.67 (7.00)	62.17 (9.00)	64.45 (8.00)	75.28 (2.00)
HandMovementDirection_2	76.35 (9.00)	76.61 (8.00)	76.68 (7.00)	79.82 (4.00)	80.15 (3.00)	**85.08 (1.00)**	82.78 (2.00)	78.81 (5.00)	77.32 (6.00)
HandMovementDirection_3	81.52 (8.00)	83.19 (3.00)	82.25 (6.00)	84.00 (2.00)	82.28 (5.00)	**86.12 (1.00)**	79.49 (9.00)	82.86 (4.00)	81.53 (7.00)
Heartbeat_0	39.30 (2.00)	38.94 (3.00)	**39.91 (1.00)**	35.84 (5.00)	28.60 (9.00)	32.78 (6.00)	29.19 (7.00)	28.89 (8.00)	36.25 (4.00)
Heartbeat_1	66.64 (7.00)	66.17 (9.00)	66.20 (8.00)	67.29 (6.00)	**81.36 (1.00)**	74.87 (3.00)	77.49 (2.00)	71.32 (4.00)	67.60 (5.00)
MotorImagery_0	48.21 (4.00)	47.16 (6.00)	46.45 (8.00)	46.89 (7.00)	**75.00 (1.00)**	47.66 (5.00)	52.63 (3.00)	56.32 (2.00)	46.27 (9.00)
MotorImagery_1	52.51 (7.00)	53.82 (6.00)	56.46 (2.00)	55.66 (3.00)	**67.99 (1.00)**	54.45 (5.00)	46.22 (9.00)	47.56 (8.00)	54.60 (4.00)
NATOPS_0	94.62 (7.00)	95.31 (5.00)	95.79 (4.00)	96.28 (2.00)	93.19 (8.00)	**96.32 (1.00)**	92.61 (9.00)	95.04 (6.00)	96.10 (3.00)
NATOPS_1	95.21 (7.00)	96.36 (5.00)	96.69 (3.00)	96.51 (4.00)	93.67 (8.00)	**97.05 (1.00)**	88.99 (9.00)	96.04 (6.00)	96.87 (2.00)
NATOPS_2	96.66 (7.00)	97.69 (5.00)	97.70 (4.00)	**97.96 (1.00)**	93.80 (8.00)	97.82 (3.00)	86.44 (9.00)	96.97 (6.00)	97.91 (2.00)
NATOPS_3	94.49 (9.00)	95.67 (6.00)	97.93 (4.00)	98.95 (3.00)	94.80 (8.00)	99.37 (2.00)	95.50 (7.00)	97.75 (5.00)	**99.50 (1.00)**
NATOPS_4	95.71 (9.00)	98.21 (6.00)	99.64 (3.00)	99.58 (4.00)	96.91 (8.00)	99.64 (2.00)	97.29 (7.00)	98.39 (5.00)	**99.77 (1.00)**
NATOPS_5	97.38 (7.00)	99.21 (3.00)	**99.75 (1.00)**	99.02 (5.00)	96.00 (8.00)	99.17 (4.00)	86.30 (9.00)	97.58 (6.00)	99.64 (2.00)
PEMS-SF_0	98.34 (6.00)	99.55 (3.00)	**99.59 (1.00)**	99.37 (5.00)	93.55 (9.00)	99.42 (4.00)	98.10 (7.00)	94.37 (8.00)	99.56 (2.00)
PEMS-SF_1	89.49 (9.00)	96.58 (6.00)	97.26 (3.00)	96.61 (5.00)	95.91 (8.00)	**97.61 (1.00)**	96.32 (7.00)	97.35 (2.00)	97.24 (4.00)
PEMS-SF_2	89.10 (8.00)	94.92 (6.00)	95.55 (3.00)	93.61 (7.00)	83.15 (9.00)	**97.18 (1.00)**	95.45 (4.00)	95.14 (5.00)	95.63 (2.00)
PEMS-SF_3	90.27 (8.00)	95.84 (6.00)	**96.60 (1.00)**	95.95 (5.00)	82.67 (9.00)	96.41 (4.00)	96.44 (3.00)	95.52 (7.00)	96.59 (2.00)
PEMS-SF_4	91.68 (8.00)	94.89 (6.00)	95.72 (5.00)	94.19 (7.00)	90.13 (9.00)	**97.04 (1.00)**	96.88 (2.00)	96.46 (3.00)	95.96 (4.00)
PEMS-SF_5	85.85 (9.00)	93.23 (8.00)	95.71 (5.00)	94.20 (7.00)	95.88 (4.00)	**97.78 (1.00)**	97.22 (2.00)	96.86 (3.00)	95.57 (6.00)
PEMS-SF_6	97.17 (8.00)	99.77 (4.00)	**99.91 (1.00)**	99.72 (6.00)	80.26 (9.00)	99.85 (2.00)	99.57 (7.00)	99.72 (5.00)	99.84 (3.00)
PenDigits_0	98.28 (8.00)	99.69 (3.00)	99.38 (6.00)	99.66 (4.00)	88.34 (9.00)	**99.79 (1.00)**	99.57 (5.00)	99.25 (7.00)	99.77 (2.00)
PenDigits_1	99.43 (6.00)	99.66 (6.00)	99.41 (7.00)	99.71 (3.00)	87.97 (9.00)	99.83 (2.00)	99.71 (4.00)	99.06 (8.00)	**99.91 (1.00)**
PenDigits_2	99.81 (7.00)	**99.93 (1.00)**	99.86 (6.00)	99.91 (3.50)	94.10 (9.00)	99.93 (2.00)	99.88 (5.00)	98.85 (8.00)	99.91 (3.50)
PenDigits_3	99.89 (3.00)	99.87 (4.50)	99.87 (4.50)	99.89 (2.00)	85.99 (9.00)	**99.91 (1.00)**	99.48 (7.00)	97.87 (8.00)	99.80 (6.00)
PenDigits_4	99.83 (6.00)	99.99 (3.50)	99.99 (3.50)	100.00 (2.00)	88.30 (9.00)	**100.00 (1.00)**	99.75 (7.00)	99.34 (8.00)	99.98 (5.00)
PenDigits_5	99.91 (6.00)	**99.99 (1.00)**	99.92 (4.00)	99.86 (6.00)	96.76 (9.00)	99.93 (3.00)	99.85 (7.00)	99.71 (6.00)	99.96 (2.00)
PenDigits_6	99.97 (6.00)	**100.00 (1.50)**	99.99 (4.00)	99.99 (5.00)	91.14 (9.00)	**100.00 (1.50)**	99.82 (7.00)	98.92 (8.00)	100.00 (3.00)
PenDigits_7	99.10 (6.00)	99.25 (5.00)	99.34 (3.00)	99.30 (4.00)	90.20 (9.00)	99.37 (2.00)	98.74 (7.00)	97.92 (8.00)	**99.54 (1.00)**
PenDigits_8	99.88 (6.00)	**99.99 (1.00)**	99.94 (4.00)	99.80 (7.00)	85.30 (9.00)	99.96 (2.00)	99.92 (5.00)	99.04 (8.00)	99.95 (3.00)
PenDigits_9	99.67 (5.00)	99.83 (2.00)	99.77 (3.00)	99.64 (7.00)	84.56 (9.00)	99.70 (4.00)	99.64 (6.00)	98.92 (8.00)	**99.88 (1.00)**

TABLE VI: Mean AUROC scores and rankings for all problems under the *(N-1)-vs-rest* setting

Dataset_Normality	IF	LOF	OCSVM	AutoAD	DAGMM	DeepSVDD	FixedTS	NeuTraL AD	NeuCoReClass AD
ChlorineConcentration_0	64.54 (6.00)	63.35 (7.00)	**73.80 (1.00)**	64.59 (5.00)	55.81 (9.00)	65.15 (4.00)	60.74 (8.00)	66.97 (3.00)	71.67 (2.00)
ChlorineConcentration_1	47.88 (6.00)	45.99 (9.00)	46.50 (8.00)	46.54 (7.00)	49.81 (5.00)	50.18 (3.00)	49.86 (4.00)	52.16 (2.00)	**57.24 (1.00)**
ChlorineConcentration_2	45.50 (6.00)	45.32 (7.00)	42.26 (9.00)	46.41 (5.00)	47.78 (4.00)	43.03 (8.00)	47.86 (3.00)	51.02 (2.00)	**60.82 (1.00)**
Computers_0	**69.17 (1.00)**	49.56 (8.00)	54.59 (6.00)	52.46 (7.00)	44.28 (9.00)	55.90 (4.00)	60.97 (2.00)	59.27 (3.00)	54.81 (5.00)
Computers_1	30.28 (9.00)	62.00 (2.00)	54.21 (5.00)	**63.51 (1.00)**	45.68 (7.00)	50.42 (6.00)	35.72 (8.00)	57.24 (4.00)	60.49 (3.00)
DistalPhalanxOutlineAgeGroup_0	83.76 (5.00)	94.25 (2.00)	79.73 (7.00)	71.87 (8.00)	59.89 (9.00)	83.13 (6.00)	**94.40 (1.00)**	92.45 (4.00)	93.46 (3.00)
DistalPhalanxOutlineAgeGroup_1	69.87 (4.00)	47.25 (9.00)	70.68 (3.00)	**71.51 (1.00)**	55.11 (7.00)	71.25 (2.00)	55.09 (8.00)	61.12 (6.00)	69.25 (5.00)
DistalPhalanxOutlineAgeGroup_2	**72.58 (1.00)**	64.91 (4.00)	67.57 (3.00)	70.67 (2.00)	47.33 (6.00)	52.29 (5.00)	20.01 (9.00)	31.39 (7.00)	25.39 (8.00)
DistalPhalanxOutlineCorrect_0	71.85 (6.00)	**83.08 (1.00)**	71.96 (5.00)	67.09 (8.00)	58.36 (9.00)	70.80 (7.00)	72.79 (3.00)	73.44 (2.00)	72.02 (4.00)
DistalPhalanxOutlineCorrect_1	50.29 (2.00)	40.74 (8.00)	46.97 (5.00)	**52.58 (1.00)**	49.62 (3.00)	47.43 (4.00)	33.16 (9.00)	46.83 (6.00)	43.40 (7.00)
ECG200_0	86.80 (6.00)	89.11 (3.00)	87.41 (4.00)	85.16 (7.00)	89.72 (2.00)	**89.91 (1.00)**	67.79 (9.00)	73.63 (8.00)	87.22 (5.00)
ECG200_1	52.16 (6.00)	22.68 (9.00)	56.51 (2.00)	51.97 (7.00)	42.77 (8.00)	54.02 (4.00)	**59.63 (1.00)**	52.40 (5.00)	56.45 (3.00)
Earthquakes_0	35.95 (9.00)	58.26 (6.00)	53.93 (7.00)	67.51 (3.00)	62.19 (4.00)	**70.71 (1.00)**	48.29 (8.00)	68.17 (2.00)	58.48 (5.00)
Earthquakes_1	65.55 (4.00)	66.18 (2.00)	64.59 (5.00)	41.26 (9.00)	51.11 (6.00)	47.71 (8.00)	65.88 (3.00)	47.99 (7.00)	**66.70 (1.00)**
ElectricDevices_0	35.11 (9.00)	56.19 (3.00)	50.06 (7.00)	63.75 (2.00)	52.57 (6.00)	54.60 (5.00)	37.72 (8.00)	55.89 (4.00)	**64.79 (1.00)**
ElectricDevices_1	91.64 (2.00)	91.32 (3.00)	86.15 (5.00)	45.25 (8.00)	40.32 (9.00)	90.58 (4.00)	80.03 (7.00)	81.24 (6.00)	**91.78 (1.00)**
ElectricDevices_2	28.98 (9.00)	67.35 (2.00)	44.76 (5.00)	**79.92 (1.00)**	53.76 (4.00)	43.36 (6.00)	38.63 (7.00)	36.55 (8.00)	56.24 (3.00)
ElectricDevices_3	17.77 (9.00)	29.51 (7.00)	29.95 (6.00)	**79.64 (1.00)**	49.03 (2.00)	43.79 (3.00)	40.97 (5.00)	36.11 (4.00)	34.22 (5.00)
ElectricDevices_4	63.35 (5.00)	60.22 (6.00)	53.73 (8.00)	27.60 (9.00)	58.96 (7.00)	70.45 (4.00)	**79.17 (1.00)**	72.43 (3.00)	74.21 (2.00)
ElectricDevices_5	17.88 (9.00)	31.71 (8.00)	43.09 (6.00)	**83.25 (1.00)**	54.59 (3.00)	46.12 (5.00)	33.89 (7.00)	48.90 (4.00)	61.96 (2.00)
ElectricDevices_6	60.54 (5.00)	62.76 (3.00)	46.61 (7.00)	33.68 (9.00)	50.93 (6.00)	45.59 (8.00)	**84.90 (1.00)**	62.54 (4.00)	65.48 (2.00)
Epilepsy_0	44.22 (9.00)	45.56 (8.00)	46.01 (7.00)	61.20 (4.00)	66.81 (4.00)	**77.01 (1.00)**	76.25 (2.00)	59.15 (6.00)	70.50 (3.00)
Epilepsy_1	96.68 (6.00)	99.25 (5.00)	99.57 (4.00)	99.87 (2.00)	95.24 (7.00)	**99.90 (1.00)**	57.61 (9.00)	79.66 (8.00)	99.65 (3.00)
Epilepsy_2	60.49 (7.00)	49.32 (9.00)	57.07 (8.00)	62.96 (6.00)	66.01 (5.00)	80.34 (3.00)	**83.95 (1.00)**	78.93 (4.00)	82.88 (2.00)
Epilepsy_3	22.50 (7.00)	35.32 (6.00)	11.75 (9.00)	22.29 (8.00)	80.47 (3.00)	69.16 (5.00)	**92.76 (1.00)**	83.62 (2.00)	79.36 (4.00)
EthanolConcentration_0	45.95 (8.00)	50.53 (3.00)	46.71 (7.00)	44.62 (9.00)	47.34 (5.00)	50.31 (4.00)	46.74 (6.00)	**56.15 (1.00)**	50.72 (2.00)
EthanolConcentration_1	42.08 (9.00)	50.78 (3.00)	46.19 (6.00)	45.97 (7.00)	47.43 (5.00)	47.88 (4.00)	52.54 (2.00)	**52.85 (1.00)**	44.15 (8.00)
EthanolConcentration_2	**55.00 (1.00)**	48.98 (7.00)	52.88 (2.00)	50.33 (3.00)	49.82 (5.00)	49.41 (6.00)	50.06 (4.00)	48.63 (8.00)	48.24 (9.00)
EthanolConcentration_3	**65.65 (1.00)**	45.10 (8.00)	60.88 (3.00)	63.65 (2.00)	49.69 (7.00)	56.27 (5.00)	53.86 (6.00)	41.57 (9.00)	57.64 (4.00)
EthanolLevel_0	57.29 (3.00)	56.80 (4.00)	57.47 (2.00)	52.79 (6.00)	50.27 (8.00)	**59.10 (1.00)**	49.40 (9.00)	52.08 (7.00)	56.22 (5.00)
EthanolLevel_1	47.40 (5.00)	47.22 (6.00)	47.63 (3.00)	**50.51 (1.00)**	48.66 (2.00)	45.71 (9.00)	45.71 (8.00)	47.41 (4.00)	46.87 (7.00)
EthanolLevel_2	43.89 (8.00)	44.10 (7.00)	43.85 (9.00)	45.74 (6.00)	**53.39 (1.00)**	48.75 (4.00)	53.00 (2.00)	51.13 (3.00)	45.88 (5.00)
EthanolLevel_3	57.85 (3.00)	57.26 (6.00)	57.66 (5.00)	57.66 (4.00)	48.10 (9.00)	48.78 (8.00)	63.28 (2.00)	56.40 (7.00)	**66.17 (1.00)**
FaceDetection_0	48.91 (5.00)	48.38 (8.00)	48.67 (7.00)	49.36 (3.00)	**51.02 (1.00)**	48.05 (9.00)	49.28 (4.00)	49.92 (2.00)	48.90 (6.00)
FaceDetection_1	51.64 (4.00)	**52.25 (1.00)**	52.08 (2.00)	50.97 (6.00)	49.10 (9.00)	51.30 (5.00)	50.15 (8.00)	50.45 (7.00)	51.92 (3.00)
FingerMovements_0	38.70 (7.00)	39.98 (6.00)	37.49 (9.00)	41.52 (5.00)	**51.15 (1.00)**	37.57 (8.00)	49.51 (3.00)	49.85 (2.00)	42.47 (4.00)
FingerMovements_1	62.67 (2.00)	**63.63 (1.00)**	61.78 (3.00)	61.38 (4.00)	54.72 (7.00)	58.02 (5.00)	46.97 (8.00)	45.96 (9.00)	57.37 (6.00)
FordA_0	52.80 (9.00)	59.80 (6.00)	58.02 (7.00)	69.74 (5.00)	57.30 (8.00)	74.74 (2.00)	73.96 (3.00)	**77.21 (1.00)**	70.55 (4.00)
FordA_1	52.11 (5.00)	56.96 (3.00)	56.99 (2.00)	38.14 (9.00)	49.91 (6.00)	**62.05 (1.00)**	47.83 (7.00)	52.93 (4.00)	42.62 (8.00)
FordB_0	49.65 (8.00)	52.41 (7.00)	54.38 (6.00)	62.27 (5.00)	45.78 (9.00)	64.09 (3.00)	64.29 (2.00)	**72.39 (1.00)**	63.35 (4.00)
FordB_1	50.31 (4.00)	**58.66 (1.00)**	57.57 (2.00)	44.69 (9.00)	48.90 (7.00)	51.28 (3.00)	50.27 (5.00)	47.15 (8.00)	49.26 (6.00)
HandMovementDirection_0	50.67 (7.00)	**55.37 (1.00)**	53.11 (4.00)	52.61 (5.00)	45.67 (9.00)	52.05 (6.00)	54.64 (2.00)	49.65 (8.00)	53.29 (3.00)
HandMovementDirection_1	43.26 (6.00)	45.00 (5.00)	43.03 (7.00)	39.49 (9.00)	49.67 (3.00)	52.15 (2.00)	46.24 (4.00)	**56.95 (1.00)**	41.65 (8.00)
HandMovementDirection_2	56.23 (5.00)	56.38 (4.00)	57.85 (2.00)	**66.08 (1.00)**	53.97 (6.00)	52.54 (7.00)	35.95 (9.00)	40.63 (8.00)	56.90 (3.00)
HandMovementDirection_3	52.98 (6.00)	51.55 (7.00)	54.40 (4.00)	**59.93 (1.00)**	53.17 (5.00)	54.67 (3.00)	51.52 (8.00)	51.17 (9.00)	59.17 (2.00)
Heartbeat_0	39.67 (7.00)	37.34 (8.00)	36.96 (9.00)	40.33 (6.00)	52.36 (2.00)	52.15 (3.00)	**57.38 (1.00)**	47.80 (4.00)	41.55 (5.00)
Heartbeat_1	63.25 (3.00)	64.91 (2.00)	**65.55 (1.00)**	60.90 (4.00)	44.83 (9.00)	57.49 (6.00)	51.59 (8.00)	53.78 (7.00)	60.82 (5.00)
MotorImagery_0	58.16 (4.00)	58.16 (5.00)	60.70 (2.00)	58.54 (3.00)	50.00 (7.00)	57.18 (6.00)	43.96 (8.00)	43.78 (9.00)	**60.89 (1.00)**
MotorImagery_1	40.65 (7.00)	41.76 (6.00)	40.04 (8.00)	43.10 (5.00)	49.24 (3.00)	44.19 (4.00)	52.23 (2.00)	**56.02 (1.00)**	39.17 (9.00)
NATOPS_0	42.50 (8.00)	52.60 (5.00)	44.13 (7.00)	52.39 (6.00)	41.27 (9.00)	71.49 (3.00)	75.86 (2.00)	**77.10 (1.00)**	59.84 (4.00)
NATOPS_1	24.59 (9.00)	27.58 (8.00)	30.49 (7.00)	33.35 (6.00)	48.36 (4.00)	51.95 (3.00)	**53.74 (1.00)**	53.63 (2.00)	37.71 (5.00)
NATOPS_2	19.33 (9.00)	19.87 (8.00)	26.04 (7.00)	26.43 (6.00)	50.85 (3.00)	45.85 (4.00)	**54.49 (1.00)**	53.46 (2.00)	32.70 (5.00)
NATOPS_3	94.82 (6.00)	98.93 (2.00)	97.69 (3.00)	95.06 (4.00)	44.08 (9.00)	92.12 (7.00)	94.84 (5.00)	64.17 (8.00)	**99.21 (1.00)**
NATOPS_4	90.53 (5.00)	**95.67 (1.00)**	93.20 (4.00)	87.89 (6.00)	40.32 (9.00)	83.74 (7.00)	95.48 (2.00)	63.58 (8.00)	94.16 (3.00)
NATOPS_5	60.71 (7.00)	80.29 (4.00)	54.62 (8.00)	73.75 (5.00)	46.69 (9.00)	73.23 (6.00)	**98.20 (1.00)**	87.69 (3.00)	89.38 (2.00)
PEMS-SF_0	**92.32 (1.00)**	60.28 (8.00)	82.24 (5.00)	55.04 (9.00)	75.58 (6.00)	82.70 (4.00)	90.65 (2.00)	83.51 (3.00)	64.87 (7.00)
PEMS-SF_1	43.20 (9.00)	51.57 (8.00)	53.95 (7.00)	61.30 (6.00)	67.54 (4.00)	72.43 (3.00)	**82.99 (1.00)**	77.61 (2.00)	63.04 (5.00)
PEMS-SF_2	40.05 (9.00)	63.32 (7.00)	68.45 (6.00)	74.68 (2.00)	69.25 (4.00)	**77.76 (1.00)**	60.61 (8.00)	68.88 (5.00)	74.34 (3.00)
PEMS-SF_3	39.91 (9.00)	61.45 (5.00)	64.00 (4.00)	**74.45 (1.00)**	57.84 (7.00)	72.54 (3.00)	46.43 (8.00)	60.43 (6.00)	73.34 (2.00)
PEMS-SF_4	51.74 (7.00)	61.80 (5.00)	68.90 (4.00)	74.12 (2.00)	50.78 (9.00)	73.44 (3.00)	51.36 (8.00)	56.71 (6.00)	**74.15 (1.00)**
PEMS-SF_5	74.60 (7.00)	89.12 (2.00)	88.76 (3.00)	88.13 (4.00)	59.39 (9.00)	84.81 (6.00)	62.89 (8.00)	86.26 (5.00)	**89.61 (1.00)**
PEMS-SF_6	73.16 (2.00)	71.31 (4.00)	67.25 (6.00)	60.56 (8.00)	60.99 (7.00)	73.12 (3.00)	**76.11 (1.00)**	60.03 (9.00)	67.54 (5.00)
PenDigits_0	97.16 (4.00)	99.71 (2.00)	97.74 (3.00)	87.27 (6.00)	62.81 (9.00)	70.01 (7.00)	96.91 (5.00)	67.96 (8.00)	**99.83 (1.00)**
PenDigits_1	71.73 (6.00)	**95.07 (1.00)**	80.37 (4.00)	80.77 (3.00)	47.31 (9.00)	63.23 (8.00)	74.69 (5.00)	70.78 (7.00)	91.91 (2.00)
PenDigits_2	71.59 (4.00)	87.95 (2.00)	60.44 (6.00)	50.84 (8.00)	54.47 (7.00)	43.35 (9.00)	72.63 (3.00)	61.54 (5.00)	**90.01 (1.00)**
PenDigits_3	48.74 (6.00)	95.41 (2.00)	23.49 (9.00)	38.95 (8.00)	64.79 (4.00)	44.48 (7.00)	**96.14 (1.00)**	62.33 (5.00)	94.75 (3.00)
PenDigits_4	79.27 (5.00)	97.30 (3.00)	58.15 (8.00)	54.54 (9.00)	66.42 (6.00)	65.91 (7.00)	**98.74 (1.00)**	87.11 (4.00)	98.58 (2.00)
PenDigits_5	77.37 (5.00)	**97.92 (1.00)**	74.94 (7.00)	77.73 (4.00)	70.05 (8.00)	54.54 (9.00)	91.59 (3.00)	75.57 (6.00)	96.20 (2.00)
PenDigits_6	91.41 (4.00)	**99.42 (1.00)**	86.48 (6.00)	82.11 (7.00)	78.10 (8.00)	58.77 (9.00)	97.73 (3.00)	88.32 (5.00)	98.58 (2.00)
PenDigits_7	71.41 (4.00)	**98.92 (1.00)**	50.68 (7.00)	49.05 (8.00)	71.00 (5.00)	41.28 (9.00)	92.74 (3.00)	70.32 (6.00)	96.30 (2.00)
PenDigits_8	95.71 (4.00)	**99.75 (1.00)**	94.06 (5.00)	84.25 (6.00)	63.55 (7.00)	52.16 (9.00)	99.15 (2.00)	58.43 (8.00)	99.09 (3.00)
PenDigits_9	79.66 (7.00)	**99.07 (1.00)**	88.15 (4.00)	85.02 (5.00)	59.94 (8.00)	52.09 (9.00)	95.73 (3.00)	81.19 (6.00)	98.08 (2.00)

TABLE VII: Mean AUPR scores and rankings for all problems under the *(N-1)-vs-rest* setting

Dataset_Normality	IF	LOF	OCSVM	AutoAD	DAGMM	DeepSVDD	FixedTS	NeuTraL AD	NeuCoReClass AD
ChlorineConcentration_0	48.39 (7.00)	48.57 (6.00)	**57.80 (1.00)**	49.26 (5.00)	34.27 (9.00)	51.66 (3.00)	41.70 (8.00)	50.45 (4.00)	57.32 (2.00)
ChlorineConcentration_1	21.37 (6.00)	20.74 (9.00)	20.87 (8.00)	20.99 (7.00)	**28.39 (1.00)**	22.16 (5.00)	24.48 (3.00)	24.38 (4.00)	26.29 (2.00)
ChlorineConcentration_2	47.75 (6.00)	47.67 (7.00)	46.89 (8.00)	49.01 (5.00)	51.44 (3.00)	46.83 (9.00)	50.11 (4.00)	52.04 (2.00)	**58.15 (1.00)**
Computers_0	**74.13 (1.00)**	49.09 (9.00)	51.46 (8.00)	51.75 (7.00)	52.38 (6.00)	56.88 (4.00)	61.09 (2.00)	58.81 (3.00)	54.05 (5.00)
Computers_1	37.67 (9.00)	57.70 (4.00)	54.30 (5.00)	**64.18 (1.00)**	51.94 (7.00)	52.43 (6.00)	40.34 (8.00)	58.01 (3.00)	59.46 (2.00)
DistalPhalanxOutlineAgeGroup_0	29.85 (8.00)	65.13 (2.00)	32.58 (6.00)	26.92 (9.00)	29.92 (7.00)	33.98 (5.00)	58.25 (3.00)	**65.33 (1.00)**	53.82 (4.00)
DistalPhalanxOutlineAgeGroup_1	53.82 (4.00)	38.36 (9.00)	55.24 (3.00)	56.87 (2.00)	52.02 (6.00)	**57.49 (1.00)**	44.31 (8.00)	49.90 (7.00)	53.74 (5.00)
DistalPhalanxOutlineAgeGroup_2	**69.07 (1.00)**	55.75 (4.00)	57.11 (3.00)	64.08 (2.00)	46.97 (5.00)	45.21 (6.00)	31.97 (9.00)	35.56 (7.00)	34.53 (8.00)
DistalPhalanxOutlineCorrect_0	64.17 (6.00)	**78.23 (1.00)**	64.53 (5.00)	63.03 (7.00)	53.69 (9.00)	66.69 (3.00)	61.38 (8.00)	65.13 (4.00)	68.00 (2.00)
DistalPhalanxOutlineCorrect_1	55.74 (3.00)	49.14 (8.00)	52.83 (6.00)	**57.72 (1.00)**	56.82 (2.00)	53.26 (5.00)	46.34 (9.00)	53.60 (4.00)	50.93 (7.00)
ECG200_0	79.70 (6.00)	82.24 (3.00)	81.23 (4.00)	77.48 (7.00)	83.52 (2.00)	**84.10 (1.00)**	61.50 (9.00)	63.25 (8.00)	80.28 (5.00)
ECG200_1	60.18 (6.00)	47.97 (9.00)	61.89 (3.00)	59.10 (7.00)	57.16 (8.00)	60.76 (5.00)	**67.61 (1.00)**	62.56 (2.00)	61.67 (4.00)
Earthquakes_0	70.40 (9.00)	81.50 (6.00)	78.18 (7.00)	86.24 (3.00)	83.82 (4.00)	**87.66 (1.00)**	76.94 (8.00)	87.42 (2.00)	83.19 (5.00)
Earthquakes_1	33.07 (5.00)	34.70 (2.00)	34.49 (3.00)	20.72 (9.00)	29.72 (6.00)	23.81 (7.00)	**37.11 (1.00)**	22.48 (8.00)	33.89 (4.00)
ElectricDevices_0	8.56 (8.00)	10.09 (6.00)	9.99 (7.00)	**14.38 (1.00)**	12.04 (5.00)	12.26 (4.00)	6.37 (9.00)	14.07 (2.00)	13.36 (3.00)
ElectricDevices_1	64.51 (4.00)	**73.27 (1.00)**	51.08 (7.00)	21.18 (9.00)	36.90 (8.00)	51.65 (6.00)	67.31 (2.00)	53.80 (5.00)	65.47 (3.00)
ElectricDevices_2	6.54 (9.00)	13.53 (3.00)	8.76 (5.00)	21.94 (2.00)	**23.71 (1.00)**	7.92 (6.00)	7.25 (7.00)	7.18 (8.00)	10.37 (4.00)
ElectricDevices_3	8.95 (9.00)	10.03 (7.00)	10.43 (6.00)	**30.64 (1.00)**	16.68 (2.00)	13.36 (3.00)	9.26 (8.00)	11.27 (4.00)	10.81 (5.00)
ElectricDevices_4	28.85 (6.00)	27.18 (7.00)	25.57 (8.00)	16.32 (9.00)	38.32 (4.00)	38.14 (5.00)	**49.60 (1.00)**	39.57 (2.00)	39.24 (3.00)
ElectricDevices_5	5.64 (9.00)	6.45 (8.00)	14.14 (2.00)	**33.01 (1.00)**	10.44 (4.00)	8.82 (6.00)	6.67 (7.00)	9.72 (5.00)	13.40 (3.00)
ElectricDevices_6	8.12 (6.00)	8.82 (5.00)	6.16 (8.00)	4.94 (9.00)	8.99 (3.00)	6.47 (7.00)	**23.66 (1.00)**	8.84 (4.00)	9.80 (2.00)
Epilepsy_0	20.37 (9.00)	21.33 (7.00)	21.29 (8.00)	27.28 (6.00)	38.02 (4.00)	**63.88 (1.00)**	49.08 (2.00)	37.78 (5.00)	43.52 (3.00)
Epilepsy_1	92.05 (6.00)	98.27 (5.00)	99.02 (4.00)	99.68 (2.00)	91.79 (7.00)	**99.74 (1.00)**	47.95 (9.00)	59.59 (8.00)	99.23 (3.00)
Epilepsy_2	23.54 (7.00)	19.37 (9.00)	22.31 (8.00)	24.53 (6.00)	35.26 (5.00)	45.47 (3.00)	**53.69 (1.00)**	44.74 (4.00)	53.28 (2.00)
Epilepsy_3	17.18 (8.00)	19.92 (6.00)	15.73 (9.00)	17.19 (7.00)	65.13 (3.00)	40.78 (5.00)	**88.40 (1.00)**	72.53 (2.00)	58.62 (4.00)
EthanolConcentration_0	22.70 (9.00)	25.65 (4.00)	23.83 (8.00)	23.88 (7.00)	25.08 (5.00)	26.48 (2.00)	24.66 (6.00)	**28.75 (1.00)**	26.14 (3.00)
EthanolConcentration_1	23.99 (4.00)	**28.51 (1.00)**	23.83 (5.00)	22.25 (8.00)	23.38 (6.00)	23.04 (7.00)	25.18 (3.00)	27.63 (2.00)	21.71 (9.00)
EthanolConcentration_2	**31.06 (1.00)**	26.73 (6.00)	27.92 (2.00)	27.57 (4.00)	26.86 (5.00)	26.41 (7.00)	27.62 (3.00)	25.84 (9.00)	26.12 (8.00)
EthanolConcentration_3	**33.70 (1.00)**	24.29 (8.00)	32.03 (3.00)	33.62 (2.00)	28.73 (7.00)	29.28 (6.00)	29.73 (5.00)	20.62 (9.00)	31.67 (4.00)
EthanolLevel_0	**30.94 (1.00)**	28.98 (5.00)	29.76 (3.00)	28.30 (6.00)	27.78 (7.00)	30.87 (2.00)	24.24 (9.00)	26.97 (8.00)	28.98 (4.00)
EthanolLevel_1	23.41 (5.00)	**25.07 (1.00)**	23.82 (4.00)	24.01 (3.00)	24.40 (2.00)	23.29 (6.00)	22.84 (9.00)	23.05 (7.00)	23.03 (8.00)
EthanolLevel_2	22.28 (9.00)	23.15 (7.00)	22.45 (8.00)	24.28 (6.00)	27.18 (2.00)	24.38 (4.00)	**28.47 (1.00)**	26.50 (3.00)	24.31 (5.00)
EthanolLevel_3	29.21 (3.00)	27.70 (7.00)	28.72 (4.00)	28.53 (5.00)	24.18 (8.00)	23.97 (9.00)	**34.27 (1.00)**	27.81 (6.00)	33.95 (2.00)
FaceDetection_0	49.02 (6.00)	48.74 (8.00)	48.92 (7.00)	49.40 (3.00)	**51.65 (1.00)**	48.60 (9.00)	49.16 (4.00)	50.15 (2.00)	49.15 (5.00)
FaceDetection_1	51.64 (3.00)	**51.86 (1.00)**	51.67 (2.00)	50.94 (6.00)	49.95 (9.00)	51.07 (5.00)	50.01 (8.00)	50.37 (7.00)	51.42 (4.00)
FingerMovements_0	40.72 (7.00)	41.64 (5.00)	40.39 (8.00)	41.43 (6.00)	**56.70 (1.00)**	40.37 (9.00)	48.71 (3.00)	50.21 (2.00)	42.84 (4.00)
FingerMovements_1	**64.68 (1.00)**	61.86 (5.00)	63.03 (2.00)	62.92 (3.00)	62.37 (4.00)	59.90 (6.00)	50.73 (8.00)	48.72 (9.00)	58.09 (7.00)
FordA_0	53.62 (9.00)	70.87 (5.00)	68.95 (7.00)	78.78 (4.00)	59.98 (8.00)	**81.86 (1.00)**	69.13 (6.00)	80.79 (3.00)	80.90 (2.00)
FordA_1	49.28 (3.00)	48.01 (5.00)	48.28 (4.00)	38.68 (9.00)	47.11 (6.00)	**57.84 (1.00)**	44.84 (7.00)	49.85 (2.00)	40.64 (8.00)
FordB_0	49.93 (8.00)	58.69 (6.00)	59.90 (5.00)	66.78 (4.00)	47.54 (9.00)	68.67 (3.00)	56.27 (7.00)	**73.88 (1.00)**	69.83 (2.00)
FordB_1	51.37 (4.00)	53.80 (3.00)	**54.87 (1.00)**	44.34 (9.00)	53.97 (2.00)	49.87 (6.00)	49.98 (5.00)	47.12 (7.00)	45.90 (8.00)
HandMovementDirection_0	20.33 (9.00)	21.32 (7.00)	20.43 (8.00)	22.09 (5.00)	**36.90 (1.00)**	23.40 (4.00)	24.82 (3.00)	24.83 (2.00)	21.66 (6.00)
HandMovementDirection_1	35.85 (9.00)	38.32 (5.00)	38.33 (4.00)	37.14 (7.00)	**57.37 (1.00)**	45.89 (3.00)	37.88 (6.00)	47.82 (2.00)	36.79 (8.00)
HandMovementDirection_2	31.37 (5.00)	32.28 (4.00)	32.83 (3.00)	37.83 (2.00)	**41.98 (1.00)**	21.13 (7.00)	17.41 (8.00)	16.62 (9.00)	23.16 (6.00)
HandMovementDirection_3	20.70 (4.00)	18.81 (9.00)	19.42 (7.00)	22.05 (3.00)	**42.51 (1.00)**	20.66 (5.00)	20.21 (6.00)	19.34 (8.00)	24.68 (2.00)
Heartbeat_0	66.64 (7.00)	66.17 (9.00)	66.20 (8.00)	67.29 (6.00)	**81.50 (1.00)**	74.87 (3.00)	77.30 (2.00)	71.18 (4.00)	67.54 (5.00)
Heartbeat_1	39.30 (2.00)	38.94 (3.00)	**39.91 (1.00)**	35.93 (5.00)	34.43 (6.00)	32.78 (7.00)	29.57 (8.00)	28.80 (9.00)	36.69 (4.00)
MotorImagery_0	52.51 (7.00)	53.82 (6.00)	56.46 (2.00)	55.90 (3.00)	**75.00 (1.00)**	54.56 (5.00)	45.41 (9.00)	47.28 (8.00)	54.62 (4.00)
MotorImagery_1	48.21 (4.00)	47.16 (6.00)	46.45 (8.00)	46.65 (7.00)	**69.29 (1.00)**	47.84 (5.00)	52.02 (3.00)	56.41 (2.00)	46.29 (9.00)
NATOPS_0	13.31 (9.00)	16.79 (6.00)	13.68 (8.00)	16.09 (7.00)	22.72 (4.00)	32.02 (3.00)	42.06 (2.00)	**45.85 (1.00)**	19.28 (5.00)
NATOPS_1	10.48 (9.00)	11.11 (8.00)	11.21 (7.00)	15.69 (5.00)	18.10 (3.00)	**20.34 (1.00)**	17.79 (4.00)	20.19 (2.00)	14.94 (6.00)
NATOPS_2	10.00 (9.00)	10.01 (8.00)	10.70 (7.00)	10.70 (6.00)	17.42 (2.00)	14.41 (4.00)	17.36 (3.00)	**17.55 (1.00)**	11.69 (5.00)
NATOPS_3	68.30 (5.00)	94.31 (2.00)	87.42 (3.00)	74.50 (4.00)	34.09 (8.00)	67.25 (6.00)	66.48 (7.00)	25.73 (9.00)	**96.18 (1.00)**
NATOPS_4	54.26 (5.00)	80.84 (2.00)	65.76 (4.00)	52.86 (6.00)	22.79 (8.00)	46.38 (7.00)	72.70 (3.00)	21.16 (9.00)	**81.59 (1.00)**
NATOPS_5	18.27 (8.00)	30.70 (4.00)	16.23 (9.00)	25.64 (6.00)	18.80 (7.00)	26.56 (5.00)	**92.25 (1.00)**	49.81 (2.00)	49.33 (3.00)
PEMS-SF_0	**61.52 (1.00)**	23.60 (9.00)	43.31 (5.00)	25.13 (8.00)	59.52 (3.00)	42.86 (6.00)	61.24 (2.00)	44.07 (4.00)	29.54 (7.00)
PEMS-SF_1	12.36 (9.00)	19.50 (7.00)	18.82 (8.00)	24.44 (5.00)	51.51 (2.00)	30.92 (4.00)	**51.91 (1.00)**	35.22 (3.00)	24.11 (6.00)
PEMS-SF_2	12.46 (9.00)	33.68 (7.00)	41.48 (5.00)	45.28 (2.00)	**55.99 (1.00)**	42.59 (4.00)	31.16 (8.00)	34.60 (6.00)	44.49 (3.00)
PEMS-SF_3	10.55 (9.00)	23.06 (7.00)	24.95 (5.00)	33.26 (2.00)	**51.85 (1.00)**	28.23 (4.00)	13.32 (8.00)	23.21 (6.00)	32.38 (3.00)
PEMS-SF_4	12.45 (9.00)	26.29 (6.00)	30.52 (4.00)	33.36 (2.00)	**38.22 (1.00)**	29.69 (5.00)	18.34 (8.00)	22.79 (7.00)	32.90 (3.00)
PEMS-SF_5	37.18 (8.00)	57.37 (2.00)	**57.74 (1.00)**	53.99 (4.00)	49.90 (5.00)	43.97 (6.00)	31.45 (9.00)	42.40 (7.00)	56.10 (3.00)
PEMS-SF_6	19.31 (4.00)	19.62 (3.00)	16.66 (7.00)	14.38 (8.00)	**35.73 (1.00)**	18.23 (5.00)	26.53 (2.00)	14.32 (9.00)	16.72 (6.00)
PenDigits_0	75.40 (5.00)	95.44 (2.00)	80.26 (3.00)	36.94 (6.00)	20.37 (8.00)	29.09 (7.00)	76.85 (4.00)	20.24 (9.00)	**97.40 (1.00)**
PenDigits_1	20.35 (8.00)	**68.78 (1.00)**	37.04 (5.00)	40.80 (3.00)	12.70 (9.00)	23.59 (7.00)	38.02 (4.00)	24.75 (6.00)	60.15 (2.00)
PenDigits_2	15.65 (4.00)	**40.56 (1.00)**	12.08 (7.00)	11.20 (8.00)	13.72 (5.00)	9.57 (9.00)	23.58 (3.00)	13.28 (6.00)	39.83 (2.00)
PenDigits_3	8.52 (7.00)	53.83 (2.00)	5.85 (9.00)	7.34 (8.00)	19.75 (4.00)	13.14 (6.00)	**66.17 (1.00)**	15.15 (5.00)	49.30 (3.00)
PenDigits_4	21.15 (7.00)	79.66 (3.00)	12.57 (8.00)	10.88 (9.00)	25.09 (5.00)	22.85 (6.00)	**89.47 (1.00)**	45.64 (4.00)	80.85 (2.00)
PenDigits_5	39.80 (5.00)	**78.32 (1.00)**	39.00 (6.00)	48.60 (4.00)	27.11 (8.00)	11.27 (9.00)	55.91 (3.00)	27.62 (7.00)	72.65 (2.00)
PenDigits_6	39.03 (5.00)	**92.89 (1.00)**	28.13 (7.00)	24.82 (8.00)	35.49 (6.00)	13.70 (9.00)	80.41 (2.00)	48.79 (4.00)	78.09 (3.00)
PenDigits_7	15.45 (6.00)	**89.18 (1.00)**	10.35 (8.50)	10.94 (7.00)	26.27 (4.00)	10.35 (8.50)	61.48 (3.00)	17.52 (5.00)	63.50 (2.00)
PenDigits_8	65.74 (4.00)	**96.98 (1.00)**	56.51 (5.00)	39.52 (6.00)	19.61 (8.00)	20.22 (7.00)	85.70 (2.00)	16.58 (9.00)	83.77 (3.00)
PenDigits_9	21.40 (8.00)	**88.53 (1.00)**	35.09 (5.00)	44.43 (4.00)	30.78 (6.00)	15.26 (9.00)	73.26 (3.00)	28.53 (7.00)	74.44 (2.00)

TABLE VIII: Ablation results under the *one-vs-rest* setting.

Task Combination	AUROC (%)	AUPR (%)
Classification	61.19 (-12.03)	77.12 (-4.30)
Contrastive	67.71 (-5.51)	79.23 (-2.19)
Reconstruction	72.82 (-0.40)	81.33 (-0.09)
Contrastive + Classification	65.38 (-7.84)	78.49 (-2.93)
Reconstruction + Classification	73.18 (-0.04)	**81.50 (+0.08)**
Reconstruction + Contrastive	73.06 (-0.16)	81.38 (-0.04)
NeuCoReClass AD	**73.22**	81.42

TABLE IX: Ablation results under the *(N-1)-vs-rest* setting.

| Task Combination	AUROC (%)	AUPR (%)
Classification	54.68 (-11.95)	34.22 (-12.55)
Contrastive	56.74 (-9.89)	34.85 (-11.92)
Reconstruction	66.76 (+0.13)	46.70 (-0.07)
Contrastive + Classification	55.32 (-11.31)	34.17 (-12.60)
Reconstruction + Classification	**66.85 (+0.22)**	**47.10 (+0.33)**
Reconstruction + Contrastive	66.68 (+0.05)	46.67 (-0.10)
NeuCoReClass AD	66.63	46.77
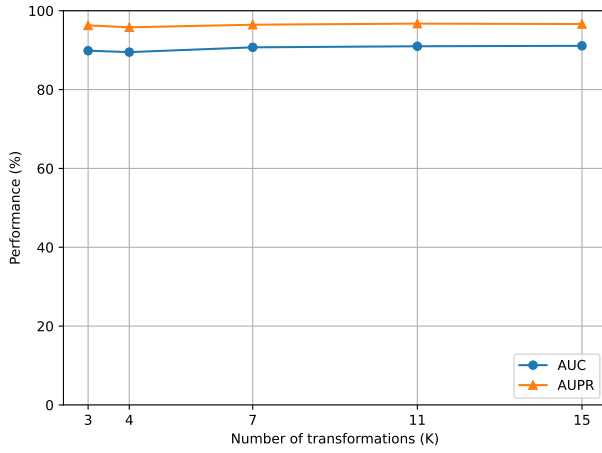
(a) one-vs-rest setting.
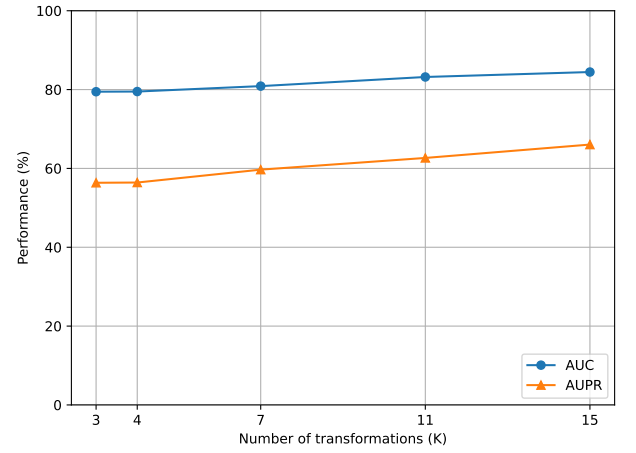
(b) $(N-1)$-vs-rest setting.

Fig. 7: Sensitivity of NeuCoReClass AD to the number of transformations K on the *Epilepsy* dataset.
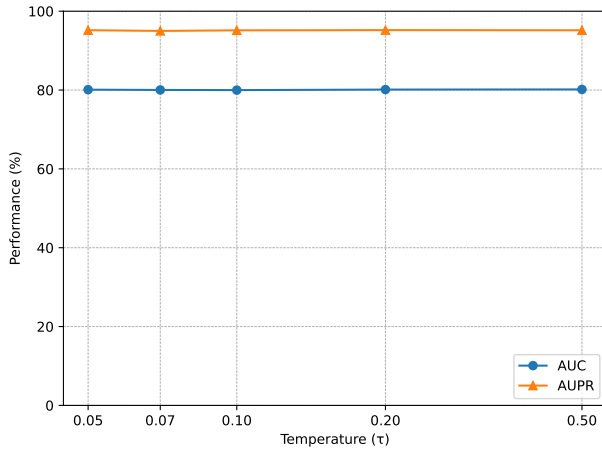
(a) one-vs-rest setting.
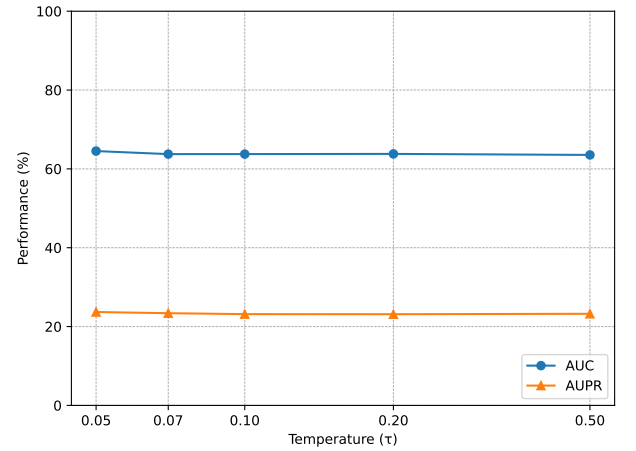
(b) $(N-1)$-vs-rest setting.

Fig. 8: Sensitivity of NeuCoReClass AD to the temperature parameter τ on the *Electric Devices* dataset.
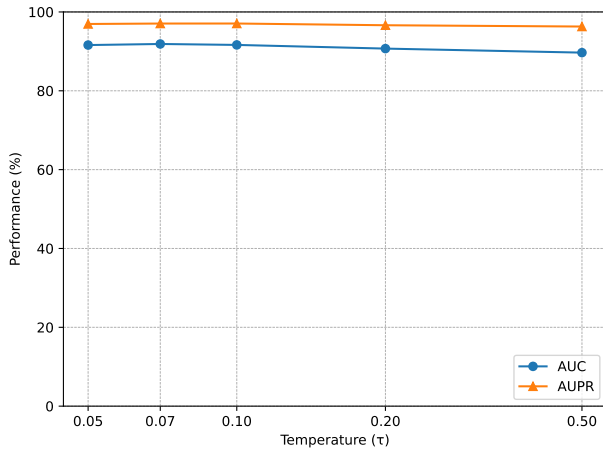
(a) one-vs-rest setting.
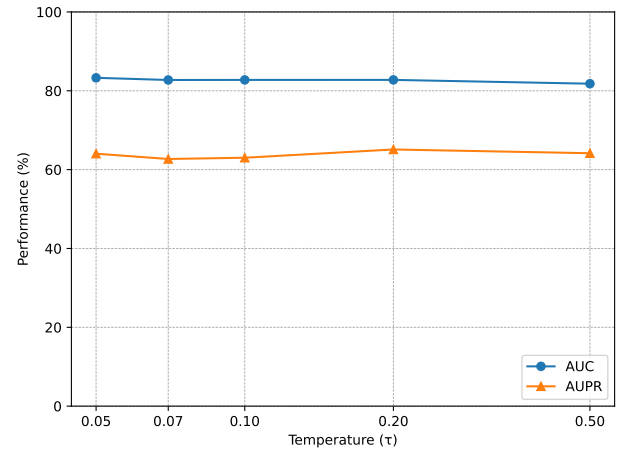
(b) $(N-1)$-vs-rest setting.

Fig. 9: Sensitivity of NeuCoReClass AD to the temperature parameter τ on the *Epilepsy* dataset.