This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

Learning with Noisy Labels for Industrial Time Series Outlier Detection: A Transformer-Embedded Contrastive Learning Framework

Jingzhong Fang, Zidong Wang, Weibo Liu, Nianyin Zeng, Yimeng He, Yu Cao, Linwei Chen, and Xiaohui Liu

***Abstract*—In many real-world industrial scenarios, acquiring accurately labeled data is often challenging due to limited resources or unexpected errors. Learning with noisy labels (LNL) has emerged as a significant research topic, aiming to develop reliable deep learning models using noisy-labeled training data. In this paper, a novel Transformer-embedded LNL framework with fuzzy-clustering-assisted contrastive learning is developed for industrial time series outlier detection under noisy labels. Specifically, a fuzzy-clustering-assisted contrastive learning strategy is proposed to enhance the robustness of the Transformer encoder against noisy labels by leveraging the intrinsic characteristics of raw data. Furthermore, a dynamic two-stage training scheme is introduced to train the outlier detector. In the first training stage, the Transformer encoder is pre-trained through data reconstruction to improve feature extraction capabilities for industrial time series. In the second stage, the outlier detector is jointly trained with the Transformer encoder, incorporating a joint learning strategy. Furthermore, a label-consistency regularization term is designed to enhance the robustness of the outlier detector against noisy labels by minimizing the discrepancy between the outputs of the outlier detector and the clustering algorithm. The proposed framework is applied to industrial time series data collected from a real-world wire arc additive manufacturing (WAAM) process. Experimental results demonstrate that the developed framework outperforms selected representative LNL approaches in WAAM outlier detection under both low and high noise ratios.**

***Index Terms*—Learning with noisy labels, outlier detection, industrial time series analysis, weakly supervised learning.**

I. Introduction

Deep learning (DL) has been extensively studied in recent years due to its remarkable ability to process large-scale data and its superior performance in feature extraction [17]. In DL, the quality of training data is crucial in determining the overall effectiveness of DL models. The accuracy of labels within the training data set is of paramount importance, as mislabeled data (commonly referred to as noisy labels) may introduce misleading information during training, thereby impairing the model's generalization ability and degrading overall performance in real-world applications.

To mitigate the adverse effects of noisy labels on DL models, learning with noisy labels (LNL) has recently been investigated extensively, leading to the proposal of numerous approaches [9], [11], [15], [23], [34], [37]. Among existing LNL methods, sample selection, label correction, and robust training have been recognized as three prominent categories due to their satisfactory performance and ease of implementation. Specifically, sample selection and label correction approaches aim to enhance the quality of training labels by identifying correctly labeled samples and rectifying noisy labels, respectively. Meanwhile, robust training approaches focus on developing noise-robust training strategies or model architectures to mitigate the influence of noisy labels during model training.

It should be noted that sample selection approaches may lead to the exclusion of potentially valuable data, thereby reducing the diversity and completeness of the training set. Meanwhile, label correction approaches can introduce further errors if the corrected labels are inaccurate, ultimately misleading the training process and potentially reinforcing incorrect patterns in the model. Furthermore, each data point in an industrial time series contains essential intrinsic and structural information, which is critical for accurately identifying outliers, preserving temporal dependencies, and maintaining feature integrity in the presence of noisy labels [9].

To establish a reliable outlier detection model based on robust training approaches, the extraction of inherent patterns and features from industrial data is of critical importance. As a powerful family of feature extraction techniques, encoder-based methods have been widely employed in robust training-based LNL approaches due to their capability of capturing complex data representations [2], [25]. Among encoder-based methods, the Transformer has been particularly effective in handling time series data, primarily due to its self-attention mechanism [27]. The Transformer encoder allows multivariate time series to be processed simultaneously while capturing dependencies across different dimensions, thereby eliminating the need for data pre-processing [30].

It is worth mentioning that the standard training scheme of the Transformer encoder typically follows a supervised approach, which is susceptible to the adverse effects of noisy labels. In contrast, self-supervised learning (SSL) has emerged

This work was supported in part by the Independent Innovation Foundation of AECC under Grant ZZCX-2023-005, the Natural Science Foundation for Distinguished Young Scholars of the Fujian Province of China under Grant 2023J06010, the National Key Research and Development Program of China under Grant 2024YFC3407000, the Royal Society of the UK, and the Alexander von Humboldt Foundation of Germany.

Jingzhong Fang, Zidong Wang, Weibo Liu, Yu Cao and Xiaohui Liu are with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (Email: `Zidong.Wang@brunel.ac.uk`)

Nianyin Zeng is with the Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen 361102, China. (Email: `zny@xmu.edu.cn`)

Yimeng He is with the State Key Laboratory of Industrial Control Technology, the College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China. (Email: `yimenghe@zju.edu.cn`)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

as a viable alternative, as it does not rely on label information during training. Among SSL techniques, contrastive learning has been particularly effective, as it aims to bring similar (i.e., positive) samples closer together while pushing dissimilar (i.e., negative) samples further apart. To date, contrastive learning has been widely adopted in robust training-based LNL approaches for constructing reliable encoders [14], [25].

Existing LNL approaches based on contrastive learning identify positive and negative sample pairs through data augmentation [13], [20]. Nevertheless, conventional data augmentation methods often fail to capture the intricate characteristics of time series data and tend to increase computational costs, particularly when handling large-scale data sets. To overcome these limitations, clustering algorithms, which group similar data points into distinct clusters based on inherent features, have been successfully employed for selecting positive and negative sample pairs in contrastive learning [4]. Recognized as a well-established clustering algorithm, the fuzzy C-means (FCM) algorithm enables data points to belong to multiple clusters with varying degrees of membership, leveraging fuzzy logic principles. Compared with hard clustering algorithms, FCM is not only easier to implement but also offers greater flexibility in handling complex data distributions. However, its performance may degrade when applied to high-dimensional data, primarily due to the presence of sparse and redundant features extracted by the Transformer encoder [8].

A seemingly natural solution for clustering high-dimensional data is to adopt the uniform manifold approximation and projection (UMAP) method, which serves as a competitive manifold-learning-based dimensionality reduction technique, in order to obtain a low-dimensional embedding from the high-dimensional input [1], [31]. The extracted low-dimensional representation can then serve as the input for the FCM algorithm, improving clustering effectiveness while mitigating the impact of irrelevant features. It is noticeable that clustering algorithms have been widely utilized in outlier detection as they are capable of uncovering the intrinsic structures and relationships within data, thereby providing valuable insights. As such, beyond determining positive and negative sample pairs, the partitions generated by the FCM algorithm can also be leveraged for model regularization, further enhancing the model's generalization ability and overall performance.

Motivated by the above discussions, this paper proposes a novel Transformer-embedded LNL framework with fuzzy-clustering-assisted contrastive learning (TFCCL) for outlier detection in industrial time series data with noisy labels. Specifically, a fuzzy-clustering-assisted contrastive learning (FCCL) approach is developed to train the Transformer encoder, where the UMAP-based FCM (UFCM) algorithm is introduced to determine positive and negative sample pairs for contrastive learning based on clustering results. A dynamic two-stage scheme is formulated for training the outlier detector. In the first training stage, the Transformer encoder is pre-trained to enhance its feature extraction capability on industrial time series data through data reconstruction. In the second stage, the outlier detector is trained jointly with the Transformer encoder, while the UFCM algorithm is dynamically updated throughout the training process. Furthermore, a joint learning strategy is proposed for the second training stage, incorporating a label-consistency regularization term to minimize the discrepancy between outputs from the outlier detector and the UFCM algorithm, thereby improving the model's robustness against noisy labels.

The main contributions of this paper are summarized as follows.

1) A novel FCCL strategy is proposed for training the Transformer encoder, where the UFCM algorithm is designed to process high-dimensional features and identify positive and negative sample pairs for contrastive learning.
2) A dynamic two-stage training scheme is introduced for the outlier detector, where the Transformer encoder is pre-trained in the first stage via data reconstruction, and in the second stage, the outlier detector is jointly trained with the Transformer encoder.
3) A joint learning strategy is developed, incorporating a label-consistency regularization term to improve the model's generalization ability and robustness against noisy labels by minimizing the discrepancy between outputs from the outlier detector and the UFCM algorithm.
4) The proposed TFCCL framework is applied to outlier detection in real-world industrial time series data with noisy labels, specifically collected from a wire arc additive manufacturing (WAAM) pilot line. Experimental results validate the effectiveness of the proposed TFCCL framework.

The remainder of this paper is organized as follows. Section II discusses the background of LNL and WAAM while presenting the relevant preliminaries. In Section III, the proposed TFCCL framework and the training scheme are introduced. Experimental results for WAAM outlier detection under noisy labels are presented in Section IV. Finally, Section V concludes the paper and provides discussions on potential future research directions.

II. Backgrounds

A. Related Work

1) Learning with Noisy Labels: Existing LNL approaches can be broadly classified into two main categories: improving the quality of training data and mitigating the impact of noisy labels during the training process.

The first category of LNL approaches focuses on selecting training samples that are likely to have correct labels (i.e., clean samples) or on correcting noisy labels within the data set. For instance, in [11], a classic approach known as Co-teaching has been proposed, where a Siamese network structure is employed, and each network selects low-loss samples to update its peer network, thereby improving training robustness. In [15], a sample selection approach called DivideMix has been developed, incorporating an improved semi-supervised training strategy that enables the model to learn from both selected clean samples and unselected samples. Furthermore, a joint training method with co-regularization (JoCoR) has been introduced in [33], where a joint loss function with a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

co-regularization term is designed to facilitate the selection of clean samples for training, thereby enhancing model stability in the presence of noisy labels.

The second category of LNL approaches focuses on mitigating the influence of noisy labels during the training process by designing robust DL models through specialized architectures or noise-resistant training strategies. These methods aim to enhance model robustness without directly modifying the training labels. For instance, in [25], a training framework has been proposed in which an encoder, combined with an SSL approach, is employed to improve the model's resistance to label noise. Similarly, in [26], a self-supervised adversarial noisy masking framework has been introduced, where adversarial noisy masking is applied to prevent the model from overfitting to noisy labels, while self-supervised reconstruction provides additional noise-free supervision, thereby improving model generalization in the presence of label noise.

2) Outlier Detection in Wire Arc Additive Manufacturing: The WAAM is a type of additive manufacturing (AM) technology that utilizes a continuous metal wire feedstock and an electric arc as the heat source to melt the wire and deposit material layer by layer, thereby constructing physical objects. Compared with other AM technologies, WAAM is particularly suitable for manufacturing large-scale components and offers relatively higher deposition rates [6], [10]. Over the past few years, WAAM has been widely adopted across various industrial sectors, including aerospace, automotive, and power systems [10].

In WAAM, the electric arc serves as the primary heat source, and its operation is governed by the total current and arc voltage, which directly influence the melting process, deposition rate, and overall manufacturing quality. It should be noted that in real-world applications, sudden fluctuations in current and voltage may occasionally occur, potentially leading to defects such as incomplete fusion or uneven material deposition [12]. Over the past few years, outlier detection has attracted increasing attention in the research community due to its critical role in various real-world applications [18], [27], [32], [39]. As a result, in the field of WAAM, numerous approaches have been developed to detect outliers, aiming to improve the WAAM process and ensure high-quality production [5], [12], [19]. For instance, in [5], a transfer learning approach has been introduced for outlier detection in WAAM data under different operational states. Furthermore, in [19], a semi-supervised method has been proposed to enable real-time anomaly detection in WAAM using current and voltage data, further enhancing the reliability and stability of the manufacturing process.

B. Preliminaries

In an M-class classification task involving noisy labels, consider a noisy data set denoted as $\tilde{\mathcal{D}} = \{(x_i, \tilde{y}_i)\}_{i=1}^{N}$, where N denotes the number of samples, x_i is the ith sample, and $\tilde{y}_i$ is the corresponding observed (potentially noisy) label. Each sample in $\tilde{\mathcal{D}}$ is assumed to be independently and identically drawn from a joint distribution $P(X, \tilde{Y})$, which is defined over the space $\mathcal{X} \times \tilde{\mathcal{Y}}$. Here, $\mathcal{X} \subset \mathbb{R}^d$ represents the input feature space, while the label space is given by $\tilde{\mathcal{Y}} = \{\tilde{y} \in \{0,1\}^M \mid \|\tilde{y}\|_1 = 1\}$, where each label $\tilde{y}$ is represented as a one-hot encoded vector, ensuring that only a single class label is assigned per sample.

III. Methodology

A. Motivation

In the presence of noisy labels, the supervised training process of the Transformer encoder may be adversely affected, leading to degraded feature extraction performance. To mitigate this issue, an effective approach is to reduce the model's reliance on ground-truth labels. As an SSL method, contrastive learning leverages the intrinsic relationships among samples to learn meaningful representations without requiring label supervision, making it a valuable technique for improving model robustness. In recent years, contrastive learning has been widely employed to assist model training, particularly in LNL. For instance, in [25], a self-supervised contrastive learning approach has been introduced for LNL to demonstrate its effectiveness in training both the encoder and classifier despite label noise.

It is worth noting that, in conventional contrastive learning, positive and negative sample pairs for each data point are typically identified using data augmentation methods. However, this approach is not well suited for large-scale time series data sets, as it may fail to capture complex temporal dependencies and could significantly increase computational costs. To overcome this limitation, a clustering-based approach for selecting positive and negative sample pairs presents a more efficient alternative. By leveraging the underlying structural information within the data, clustering enhances the correlation and diversity of positive and negative sample pairs while eliminating the need for additional label information, thereby improving the reliability of contrastive learning in noisy-label scenarios.

The FCM algorithm is a promising clustering technique for handling noisy and large-scale data, as it employs fuzzy logic to capture soft boundaries between clusters and assigns membership degrees to each data point across multiple clusters. Furthermore, the membership degrees generated by FCM can be used to regularize the classifier's output, thereby reducing the impact of label noise. To fully utilize both the inherent features and structural information in time series data for clustering, the Transformer encoder has recently been widely employed to extract effective feature representations [4]. However, directly using the high-dimensional features extracted by the Transformer encoder from large-scale time series data sets may degrade clustering performance. To overcome this challenge, applying dimensionality reduction techniques presents a natural and effective solution, as they can refine the extracted features by preserving essential structures while reducing computational complexity.

Motivated by the above discussions, in this paper, a novel TFCCL framework is developed for outlier detection on time series data with noisy labels. Specifically, an FCCL strategy is proposed to aid in training the Transformer encoder, where a UFCM algorithm is designed to process high-dimensional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

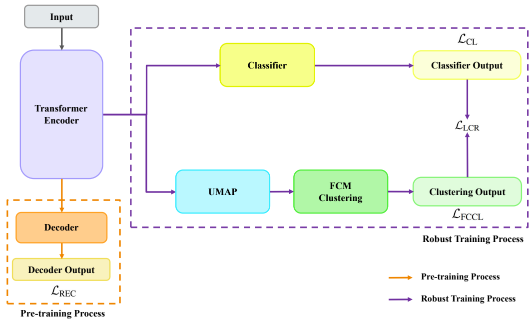


Fig. 1. The developed TFCCL framework, which consists of three key components: the Transformer encoder, the FCM clustering module, and the NN-based classifier.

features and identify positive and negative sample pairs for contrastive learning. A dynamic two-stage scheme is introduced for training the outlier detector. In the first stage, the Transformer encoder is pre-trained to enhance feature extraction capability on industrial time series data through data reconstruction. In the second stage, the outlier detector is jointly trained with the Transformer encoder, while the UFCM algorithm is dynamically updated throughout the training process. Furthermore, a joint learning strategy is proposed to improve the classifier's robustness against noisy labels, incorporating a label-consistency regularization term to minimize the discrepancy between the outputs of the outlier detector and the UFCM algorithm.

B. Overview of the TFCCL Framework

The developed TFCCL framework is illustrated in Fig. 1, which is designed to train a neural network-based (NN-based) classifier for time series outlier detection in the presence of noisy labels. The TFCCL framework consists of three key components: the Transformer encoder, the FCM clustering module, and the NN-based classifier. As shown in Fig. 2, the Transformer encoder architecture comprises multiple identical encoder layers, each consisting of a multi-head attention mechanism and a convolutional module. Residual connections and layer normalization are applied to both the attention layer and the convolutional module to mitigate gradient vanishing. Furthermore, the NN-based classifier and the decoder are implemented as multi-layer perceptrons. To enhance clustering performance, UMAP is employed for dimensionality reduction.

C. Loss Function

1) Classification Loss: The proposed TFCCL framework is designed to train a classifier for time series classification tasks. Although noisy labels may negatively impact supervised learning performance, incorporating the classification loss during training remains essential. In this paper, cross-entropy is used to compute the classification loss. The classification loss $\mathcal{L}_{\text{CL}}$ is defined as:

$$\mathcal{L}_{\text{CL}} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{j=1}^{M}\tilde{y}_{ij}\log(\hat{y}_{ij}), \tag{1}$$

where B denotes the number of samples in the current mini-batch; M is the number of classes; $\tilde{y}_{ij}$ represents the jth component of the label for the ith sample; and $\hat{y}_{ij}$ is the jth component of the the predicted output for the ith sample.

2) Reconstruction Loss: The reconstruction loss is employed for pre-training the Transformer encoder. Specifically, the mean squared error (MSE) is used to calculate the reconstruction loss $\mathcal{L}_{\text{REC}}$, which is given by:

$$\mathcal{L}_{\text{REC}} = \frac{1}{B}\sum_{i=1}^{B}\|x_i - \hat{x}_i\|^2, \tag{2}$$

where B represents the number of samples in the current mini-batch; and x_i and $\hat{x}_i$ are the true value and the predicted value of the ith sample, respectively.

3) Contrastive Loss: The contrastive loss is utilized to train the Transformer encoder. In this paper, a novel fuzzy-clustering-assisted contrastive loss $\mathcal{L}_{\text{FCCL}}$ is designed, which is shown as follows:

$$\mathcal{L}_{\text{FCCL}} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\sum_{z_j\in\mathcal{P}(z_i)}\exp(\text{sim}(z_i,z_j)/\tau)}{\sum_{z_k\in\mathcal{B}(z_i)}\exp(\text{sim}(z_i,z_k)/\tau)}, \tag{3}$$

where B is the number of samples in the current mini-batch; $z_i \in \mathbb{R}^d$ denotes the feature representation of the ith sample; $\mathcal{P}(z_i)$ is the set of positive feature representations corresponding to the ith sample; $\mathcal{B}(z_i)$ is the set of all feature representations in the current mini-batch except the feature representation of the ith sample; $\text{sim}(\cdot,\cdot)$ is the cosine similarity function between two samples; and τ represents the temperature parameter, which is a positive constant and controls the sharpness of the similarity scores.

The contrastive loss in this paper is formulated as a probability ratio that quantifies the likelihood of the model selecting the correct positive pair over all other sample pairs within the batch. The structure of the numerator and denominator reflects a probabilistic approach to pairwise discrimination, which emphasizes the relative similarity between positive and negative pairs. The exponential and logarithmic functions in (3) are combined in the contrastive loss to formulate a normalized, probabilistic objective based on the softmax and cross-entropy functions.

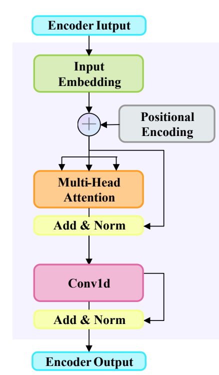


Fig. 2. The architecture of the Transformer encoder, which comprises multiple identical encoder layers. Each layer consists of a multi-head attention mechanism and a convolutional module.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

Remark 1: Traditional contrastive learning methods typically identify positive and negative sample pairs for the ith sample using augmented data or ground-truth labels [20]. However, the large volume of data and the presence of noisy labels can interfere with the sample pairing process. To reduce computational costs and mitigate the effects of label noise, this paper introduces the FCCL strategy, where positive and negative sample pairs for the ith sample are determined based on clustering results. Specifically, samples belonging to the same cluster as the ith sample are considered positive, while those in different clusters are treated as negative. The FCCL strategy enables the Transformer encoder to be trained in an unsupervised manner, thereby avoiding the adverse impact of noisy labels and improving its ability to extract meaningful feature representations.

4) Joint Learning Strategy: In this paper, a label-consistency regularization term is introduced to quantify the discrepancy between the output membership of the UFCM algorithm and the output probability of the classifier. Specifically, the Kullback-Leibler (KL) divergence is employed to compute the label-consistency regularization term $\mathcal{L}_{\text{LCR}}$, which is defined as:

$$\mathcal{L}_{\text{LCR}} = D_{\text{KL}}(P \parallel U) = \sum_{i=1}^{M} P(i) \log\left(\frac{P(i)}{U(i)}\right), \tag{4}$$

where M is the number of classes; P and U represent the distribution of the classifier's output probability and the distribution of the UFCM algorithm's output membership, respectively; and $P(i)$ and $U(i)$ represent the classifier's output probability for the ith class and the UFCM algorithm's output membership for the ith class, respectively.

Remark 2: The output membership of the UFCM algorithm is independent of the noisy labels in the training data, making it a valuable reference for mitigating the impact of label noise on the classifier. It is important to note that KL divergence is asymmetric, meaning that it quantifies the information loss incurred when one probability distribution is used to approximate another, with the divergence value varying based on the direction of comparison. Therefore, in this paper, the KL divergence $D_{\text{KL}}(P \parallel U)$ is computed instead of $D_{\text{KL}}(U \parallel P)$ as the label-consistency regularization term, aiming to better align the classifier's predictions with the UFCM algorithm's output membership.

5) Overall Loss Function: The overall loss function of the developed model consists of two parts: pre-training loss $\mathcal{L}_{\text{PRE}}$ and joint training loss $\mathcal{L}$. The pre-training loss $\mathcal{L}_{\text{PRE}}$ is given as follows:

$$\mathcal{L}_{\text{PRE}} = \mathcal{L}_{\text{REC}}. \tag{5}$$

The joint training loss $\mathcal{L}$ is given by:

$$\mathcal{L} = \mu_1 \mathcal{L}_{\text{CL}} + \mu_2 \mathcal{L}_{\text{FCCL}} + \mu_3 \mathcal{L}_{\text{LCR}}, \tag{6}$$

where μ_1, μ_2 and μ_3 are three hyper-parameters to balance the weights among the classification loss, contrastive loss and label-consistency regularization term.

D. Dynamic Two-stage Training Scheme

1) Pre-training of the Transformer Encoder in the First Stage: The feature representations extracted from the training data by the Transformer encoder play a crucial role in the training process, as they are subsequently utilized by the classifier for classification and by the UFCM algorithm for contrastive learning and regularization. Although the FCCL strategy is employed to train the Transformer encoder, its performance is highly dependent on the quality of initialization. To ensure the effectiveness of FCCL and further enhance the Transformer encoder's feature extraction capabilities, this paper employs a decoder to pre-train the encoder through input reconstruction. As illustrated in Fig. 1, during pre-training, only the encoder and decoder parameters are trainable, while all other parameters remain fixed. The reconstruction loss $\mathcal{L}_{\text{PRE}}$ is used as the loss function to guide the encoder and decoder in learning stable and robust feature representations. Pre-training allows the encoder to develop a deeper understanding of the underlying data structure, thereby improving training quality and overall model performance.

2) Dynamic Updating of the UFCM Algorithm in the Second Stage: In this paper, UMAP is employed before applying the FCM algorithm to reduce the dimensionality of the Transformer encoder output. Furthermore, cosine similarity is used to measure the similarity between data points and cluster centroids, facilitating effective clustering of feature representations.

It should be noted that the updating frequency of the UFCM algorithm during training is a critical hyperparameter. A low updating frequency may cause the algorithm to become trapped in local optima, while a high updating frequency can lead to increased computational costs and unstable clustering performance. To address this issue, a dynamic updating strategy is introduced to adjust the UFCM algorithm's update frequency throughout the training process. Specifically, the update frequency is relatively high in the early training stages and gradually decreases as training progresses. The exact epoch for performing the $(k+1)$th UFCM update, denoted by e_{UFCM}^{k+1}, is calculated by:

$$e_{\text{UFCM}}^{k+1} = \frac{k*(k+1)}{2} * e_{\text{UFCM}}^{k}. \tag{7}$$

3) Training Procedure: The training procedure of the developed TFCCL framework is presented in Algorithm 1.

IV. Outlier Detection on WAAM Data Sets with Noisy Labels

A. Experimental Setup

1) Data Description: The data used in this paper is collected from a pilot line for WAAM deployed in Sweden, which aims at developing an end-to-end digital solution by integrating automation methodologies for metal component manufacturing. Specifically, five data sets are employed in the experiments, each representing a time series corresponding to an individual manufacturing process. Each time series consists of 98,000 instances, where each instance contains measurements of welding current and voltage recorded at a

Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

Algorithm 1: Training Procedure of the Developed TFCCL Framework

Input: Noisy training data set $\tilde{\mathcal{D}} = \{(x_i, \tilde{y}_i)\}_{i=1}^{N}$, number of pre-training epochs E_P, number of training epochs E, the epoch number of the first UFCM update e^1_{UFCM}, batch size B, hyper-parameters μ_1, μ_2 and μ_3.
Output: Model parameters θ.

Randomly initialize θ.
Freeze θ except the encoder and decoder.
for $e \leftarrow 0$ to E_P **do**
  Update θ of the encoder and decoder.
end
Unfreeze θ.
Initialize UFCM using the encoder output.
for $e \leftarrow 0$ to E **do**
  for $n \leftarrow 0, 1, \cdots, \frac{|\tilde{D}|}{B}$ **do**
    Randomly draw a mini-batch $\{(x_i, \tilde{y}_i)\}_{i=1}^{B}$.
    Calculate joint loss $\mathcal{L}$ based on (6).
    Update θ.
  end
  if $e = e^k_{UFCM}$ **then**
    Update parameters of the UFCM algorithm.
    Update e^k_{UFCM} according to (7).
  end
end

sampling interval of 0.0002 seconds during the manufacturing process.

2) Data Pre-processing: In the experiments, each data set is segmented using a sliding window approach, where the window length is set to 10 and the stride to 5. Each segment is assigned a label (“Normal” or “Outlier”), determined based on the labels of individual instances within the segment. Specifically, a segment is labeled as “Normal” if all instances within it are labeled as “Normal”; otherwise, it is classified as an “Outlier”. All labels are converted into a one-hot format for ease of classification. Each segmented data set is then split into training and testing sets in a 7 : 3 ratio. Min-max normalization is subsequently applied to both sets to scale the data to a uniform range, further improving the performance of outlier detection.

3) Evaluation Metrics: To comprehensively assess the performance of the developed TFCCL framework in WAAM outlier detection, essential evaluation metrics such as accuracy, precision, recall, and F1-score are utilized. Each experiment is repeated five times, and the average values of these metrics are reported to ensure the consistency and reliability of the results.

4) Baselines: In this paper, several novel and representative approaches are selected for comparison. In addition, a vanilla outlier detection method is employed as a standard baseline for comparison. The details of each selected approach are listed as follows:

- **Co-teaching** [11], where two deep neural networks (DNNs) are trained simultaneously and select low-loss samples to update each other.
- **Co-teaching+** [38], where the disagreement-based updating strategy is combined with the original Co-teaching framework.
- **JoCoR** [33], where a co-regularization term is employed to help select clean samples for model training.
- **Co-learning** [25], where supervised learning and self-supervised learning are combined to handle noisy labels.
- **Standard Baseline**, where a classifier with the Transformer backbone is trained directly on the training data with the presence of noisy labels.

5) Implementation Details: For fair comparisons, all experiments are carried out on a server with NVIDIA RTX A6000 GPU with 48 GB memory and Intel(R) Xeon(R) Silver 4214R CPU. All approaches are implemented using Ubuntu 20.04.6, Pytorch 2.5.1, Python 3.9.21 and CUDA 12.3.

It is worth mentioning that convolutional neural networks (CNNs) are employed as the original backbone networks in selected LNL approaches. As such, in the experiments, both the CNN and the Transformer architectures are employed as the backbone networks for selected approaches for comparisons to fully investigate the competitive performance of the TFCCL framework.

To verify the effectiveness of the FCM algorithm in the TFCCL framework, the comparison experiment is conducted in this paper. Specifically, the K-means algorithm and the hierarchical clustering algorithm are chosen as baseline clustering algorithms to compare against the FCM used in the TFCCL framework. Besides, the dynamic updating strategy of the proposed UFCM algorithm is evaluated through comparative experiments. To be specific, linear and logarithmic strategies are selected as baselines for comparison.

The Transformer encoder consists of 3 encoder layers. The CNN backbone consists of 5 convolutional layers and a fully connected layer, where each convolutional layer is followed by batch normalization. The decoder and the classifier are a 4-layer MLP and a 3-layer MLP, respectively. In this paper, the hyper-parameters are selected via grid search on a held-out validation set. Each hyper-parameter is explored from a range of values and the combination that achieves the best validation performance is selected. To be specific, the temperature parameter τ in contrastive loss $\mathcal{L}_{\text{FCCL}}$ is set to 0.08. The hyper-parameters μ_1, μ_2 and μ_3 in the training loss $\mathcal{L}$ are set as 0.3, 0.8 and 0.8, respectively. In the experiments, the Adam optimizer is utilized for model training, where the learning rate is 0.0001. The pre-training and training epochs are 15 and 50, respectively. During the training process, the epoch number for the first UFCM update e^1_{UFCM} is set to 3.

The noise ratio λ is an important parameter in LNL, which indicates the ratio of the noisy labels in a data set. It should be noted that the outlier detection task in this paper is formulated as a binary classification problem with only two classes. As such, symmetrical label noise is employed in the experiment, where the label of each data point is corrupted with an equal probability defined by λ. For a comprehensive evaluation, multiple experiments are conducted on each data set with various values of λ (e.g., 30%, 40%, 50% and 60%). It should be noted that in real-world industrial scenarios, the noise ratio

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

TABLE I
OUTLIER DETECTION PERFORMANCE OF SELECTED APPROACHES ON WAAM DATA SET 1 WITH DIFFERENT NOISE RATIOS

Approaches	Accuracy				Precision				Recall				F1 Score			
	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
Co-teaching (CNN)	99.03	96.21	75.35	71.76	98.54	93.35	49.45	49.24	95.95	93.74	77.76	70.21	97.22	91.94	56.75	53.05
Co-teaching+ (CNN)	99.00	96.54	78.79	75.78	98.28	93.79	52.85	53.58	95.53	93.09	80.50	75.06	97.30	93.17	60.49	57.60
JoCoR (CNN)	98.82	98.54	83.70	14.44	99.12	98.01	53.40	14.59	94.11	93.66	70.32	81.19	96.53	95.73	60.35	24.73
Co-learning (CNN)	98.36	96.78	80.0	78.29	98.3	94.08	54.24	56.47	95.42	93.34	82.37	77.20	97.48	94.02	67.16	63.48
Co-teaching (Transformer)	99.05	97.58	77.46	76.28	98.87	94.31	46.47	63.39	95.68	92.54	81.13	72.74	97.25	93.05	57.53	55.92
Co-teaching+ (Transformer)	99.43	97.8	83.06	82.55	99.05	94.55	53.16	60.42	95.57	92.97	85.27	80.50	97.33	93.35	66.82	64.63
JoCoR (Transformer)	97.96	97.67	80.54	62.76	99.37	99.49	72.04	82.99	95.03	94.11	87.17	70.82	97.15	96.73	76.49	74.64
Co-learning (Transformer)	99.33	97.95	86.24	83.14	98.87	94.45	54.78	60.51	95.75	93.08	86.83	82.67	97.39	93.60	68.33	65.96
Standard (Transformer)	90.22	66.73	44.68	10.37	86.64	67.98	60.48	18.72	91.02	89.89	36.82	12.36	92.84	77.41	45.77	14.89
TFCCL	99.10	98.95	93.88	93.45	99.30	99.49	90.98	89.93	99.28	98.85	92.82	92.99	99.29	99.17	91.67	91.16

TABLE II
OUTLIER DETECTION PERFORMANCE OF SELECTED APPROACHES ON WAAM DATA SET 2 WITH DIFFERENT NOISE RATIOS

Approaches	Accuracy				Precision				Recall				F1 Score			
	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
Co-teaching (CNN)	99.01	96.11	73.51	71.68	98.26	91.00	54.19	53.50	97.18	90.70	79.80	78.92	97.71	90.38	60.20	59.27
Co-teaching+ (CNN)	99.16	96.33	76.73	72.08	98.39	90.93	58.64	55.34	97.13	90.27	83.72	86.87	97.68	90.6	68.21	66.28
JoCoR (CNN)	98.95	98.94	80.21	15.89	97.56	97.42	50.43	16.18	97.17	97.27	94.04	79.69	97.36	97.34	65.41	26.88
Co-learning (CNN)	98.94	96.12	79.88	74.51	98.52	91.10	61.34	57.93	97.05	90.09	85.47	88.43	97.79	90.84	70.12	68.14
Co-teaching (Transformer)	99.17	96.92	76.48	68.53	98.34	94.31	60.60	61.36	97.46	90.11	84.65	83.08	97.89	92.05	65.02	64.73
Co-teaching+ (Transformer)	99.02	97.15	81.60	71.86	98.22	94.71	65.83	62.50	96.93	90.36	89.22	85.55	97.20	91.25	67.50	69.20
JoCoR (Transformer)	98.42	98.29	75.42	70.37	99.32	98.41	70.64	69.06	96.88	97.26	82.96	60.85	98.08	97.83	72.55	62.09
Co-learning (Transformer)	99.20	97.52	86.68	74.93	98.10	93.71	69.40	67.06	97.53	90.76	94.55	87.38	97.60	92.22	80.38	77.21
Standard (Transformer)	88.27	64.20	30.26	8.81	83.25	80.83	42.07	11.45	96.17	50.64	51.87	8.37	90.86	62.27	46.46	9.67
TFCCL	98.84	98.66	91.19	89.02	98.98	98.83	91.60	92.18	99.28	98.85	92.82	92.99	99.01	98.38	89.23	86.09

TABLE III
OUTLIER DETECTION PERFORMANCE OF SELECTED APPROACHES ON WAAM DATA SET 3 WITH DIFFERENT NOISE RATIOS

Approaches	Accuracy				Precision				Recall				F1 Score			
	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
Co-teaching (CNN)	99.05	96.27	76.62	71.31	98.05	89.72	54.44	48.97	96.94	91.26	71.08	67.39	97.47	90.36	56.16	51.64
Co-teaching+ (CNN)	98.90	96.57	80.51	78.21	98.17	89.47	58.13	53.41	96.84	91.66	77.69	75.88	97.50	90.55	66.48	63.58
JoCoR (CNN)	99.14	98.99	81.32	15.16	98.40	98.65	50.54	15.09	97.02	96.00	94.09	77.91	97.70	97.30	65.57	25.27
Co-learning (CNN)	98.78	96.87	84.23	81.06	98.02	89.92	62.44	56.53	96.96	91.36	81.23	78.66	97.43	90.63	70.65	65.75
Co-teaching (Transformer)	99.13	96.56	85.88	82.39	98.64	91.51	61.73	53.58	96.74	90.31	72.93	62.38	97.67	90.86	65.55	57.55
Co-teaching+ (Transformer)	98.98	96.86	86.08	85.69	98.54	91.91	57.23	56.68	96.12	90.01	77.53	65.88	96.84	90.78	67.03	59.58
JoCoR (Transformer)	98.49	98.25	79.27	63.32	99.25	98.90	70.19	74.33	96.61	95.73	86.75	78.55	97.91	97.29	75.52	66.41
Co-learning (Transformer)	98.83	96.56	87.28	85.39	98.36	92.16	71.73	59.88	96.82	89.46	82.33	70.28	97.73	90.60	73.53	61.58
Standard (Transformer)	77.04	58.59	35.88	9.01	72.47	61.71	47.15	8.88	93.98	82.92	50.58	5.46	84.04	70.76	48.80	6.76
TFCCL	98.91	98.78	93.71	91.95	99.55	98.23	94.35	92.24	98.65	98.40	90.16	87.50	99.10	98.31	91.53	89.45

usually remains below 50%. To verify the outlier detection performance of the TFCCL framework under severe label noise, the WAAM data sets with a noise ratio of 60% are utilized in the experiments.

B. Evaluation of WAAM Outlier Detection

The results of WAAM outlier detection on five data sets with noisy labels using selected approaches and the TFCCL framework are illustrated in Tables I-V. According to the results, the TFCCL framework shows the leading performance on five data sets with five different noise ratios. The selected approaches using the Transformer backbone demonstrate better performance than those using the CNN backbone. A detailed evaluation of different noise ratios is presented below.

1) 30% and 40% Noisy Labels: It can be found that all selected approaches with different backbones as well as the TFCCL framework all achieve satisfactory performance on five data sets. The standard approach also shows acceptable performance under 30% noise ratio. As indicated by the results, the TFCCL framework outperforms the other approaches on most data sets in terms of accuracy, precision, recall, and F1 score. Notably, on data set 2 with a 30% noise ratio, Co-teaching, Co-teaching+ and Co-learning (with both backbones) all achieve higher accuracy than TFCCL. data sets 1 and 2 with noise ratios of 30%, Co-learning-Transformer has higher accuracy than TFCCL. On data sets 2, 3 and 5 with noise ratios of 30% and 40%, JoCoR-CNN has higher accuracy than TFCCL as well. Nevertheless, TFCCL attains the highest

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

TABLE IV
OUTLIER DETECTION PERFORMANCE OF SELECTED APPROACHES ON WAAM DATA SET 4 WITH DIFFERENT NOISE RATIOS

Approaches	Accuracy				Precision				Recall				F1 Score			
	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
Co-teaching (CNN)	96.17	90.82	61.97	60.01	81.41	66.76	35.72	31.74	98.07	83.91	75.61	72.90	88.86	73.58	44.60	39.87
Co-teaching+ (CNN)	96.32	90.52	67.02	62.33	81.29	67.16	38.47	37.80	97.92	84.21	80.11	75.93	88.89	74.71	53.12	50.17
JoCoR (CNN)	98.27	96.82	73.63	12.18	91.18	83.70	37.76	11.79	98.40	99.08	99.89	74.72	94.64	90.66	54.55	20.37
Co-learning (CNN)	96.20	90.82	70.97	65.01	81.17	67.41	43.22	39.54	98.03	83.91	84.31	78.40	88.71	75.01	56.92	53.27
Co-teaching (Transformer)	96.54	90.77	66.33	63.53	83.78	65.71	37.89	30.87	96.68	86.47	83.96	79.48	89.68	74.35	49.32	43.00
Co-teaching+ (Transformer)	96.39	91.07	71.83	67.03	83.96	65.31	42.09	33.87	96.56	86.97	88.76	81.98	89.82	74.60	57.10	47.94
JoCoR (Transformer)	96.41	96.55	71.65	62.72	96.42	94.09	61.44	53.52	94.91	96.85	84.22	96.12	95.64	95.40	68.03	63.52
Co-learning (Transformer)	96.20	91.37	77.03	71.13	83.80	64.91	47.39	37.57	96.45	87.37	93.56	85.38	89.65	74.90	62.70	52.14
Standard (Transformer)	74.32	61.58	15.75	6.09	71.92	64.56	31.52	15.11	75.78	41.60	23.94	9.26	83.67	58.75	27.21	11.48
TFCCL	99.03	98.45	95.48	93.98	99.02	98.57	93.14	89.62	99.51	98.00	94.15	94.37	99.26	98.27	93.47	91.48

TABLE V
OUTLIER DETECTION PERFORMANCE OF SELECTED APPROACHES ON WAAM DATA SET 5 WITH DIFFERENT NOISE RATIOS

Approaches	Accuracy				Precision				Recall				F1 Score			
	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
Co-teaching (CNN)	98.80	96.56	73.93	68.36	96.91	89.33	44.45	44.42	95.36	89.94	69.53	69.04	96.10	89.15	49.43	47.22
Co-teaching+ (CNN)	98.92	96.80	77.62	75.12	96.74	89.80	49.00	46.67	95.51	89.59	74.63	72.59	95.70	89.66	55.73	54.62
JoCoR (CNN)	99.04	98.78	82.40	12.54	96.06	96.75	47.45	12.21	97.86	95.37	95.59	77.20	96.93	96.04	63.04	21.08
Co-learning (CNN)	96.17	90.87	71.09	65.11	81.41	66.86	43.17	39.29	97.74	84.61	80.89	79.20	88.90	74.68	57.01	52.55
Co-teaching (Transformer)	98.83	97.21	79.10	76.99	98.76	92.46	44.61	39.89	97.65	89.97	69.90	70.27	98.20	90.94	52.18	51.10
Co-teaching+ (Transformer)	98.68	97.56	84.33	80.13	98.64	92.91	50.28	42.27	97.81	89.72	74.79	73.60	98.12	90.64	56.34	53.25
JoCoR (Transformer)	98.14	97.85	78.64	64.41	99.39	98.36	69.62	64.99	94.92	95.72	87.96	83.00	97.10	97.02	72.97	65.66
Co-learning (Transformer)	98.53	97.86	86.53	84.23	98.52	92.61	55.03	45.72	97.26	90.12	79.29	77.80	97.63	90.39	66.48	62.19
Standard (Transformer)	79.42	73.21	47.45	7.07	84.60	83.31	69.09	14.23	84.81	75.20	39.43	7.62	84.71	79.05	50.21	9.92
TFCCL	98.95	98.42	95.20	92.22	99.42	98.06	90.10	87.58	99.01	99.62	96.51	89.53	99.21	98.83	93.07	88.37

TABLE VI
COMPARISON OF DIFFERENT CLUSTERING ALGORITHMS WITHIN THE TFCCL FRAMEWORK ON WAAM DATA SETS UNDER VARYING NOISE RATIOS

Approaches	Metrics (%)	Data Set 1				Data Set 2				Data Set 3				Data Set 4				Data Set 5			
		30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
TFCCL-Kmeans	Accuracy	98.12	96.22	89.02	84.23	97.84	95.68	86.43	80.48	97.97	96.09	88.13	82.47	98.07	95.28	89.52	83.08	97.91	95.42	89.03	82.91
	Precision	98.29	96.42	85.08	80.11	97.72	95.01	85.59	78.52	98.28	95.01	88.31	80.88	97.81	95.11	86.48	78.59	98.12	94.48	83.28	74.01
	Recall	98.34	96.09	86.12	77.68	97.81	94.01	79.02	70.01	96.68	94.32	82.52	74.48	98.42	94.02	86.12	79.79	97.62	96.58	87.92	71.01
	F1 Score	98.30	96.26	85.59	78.84	97.76	94.48	82.12	73.83	97.48	94.62	85.31	77.41	98.11	94.53	86.29	79.26	97.84	95.52	85.48	72.31
TFCCL-Hierarchical	Accuracy	95.58	90.97	82.33	75.47	94.83	90.28	80.24	72.36	95.13	90.26	81.16	73.84	95.18	89.62	80.76	72.28	94.68	89.78	80.08	71.96
	Precision	95.76	91.04	78.86	72.84	94.42	89.27	77.47	69.87	95.33	89.32	80.07	71.94	94.57	88.84	77.94	68.57	94.64	87.83	75.08	65.96
	Recall	95.27	90.23	78.16	70.48	94.34	87.77	70.77	62.27	94.02	87.93	74.16	65.47	95.04	86.28	75.36	66.07	93.78	88.03	79.06	63.44
	F1 Score	95.51	90.58	78.48	71.57	94.37	88.53	73.97	65.58	94.61	88.55	76.97	68.63	94.76	87.57	76.58	67.18	94.18	87.46	76.87	64.52
TFCCL-FCM	Accuracy	99.10	98.95	93.88	93.45	98.84	98.66	91.19	89.02	98.91	98.78	93.71	91.95	99.03	98.45	95.48	93.98	98.95	98.42	95.20	92.22
	Precision	99.30	99.49	90.98	89.93	98.98	98.83	91.60	92.18	99.55	98.23	94.35	92.24	99.02	98.57	93.14	89.62	99.42	98.06	90.10	87.58
	Recall	99.28	98.85	92.82	92.99	99.04	97.94	87.65	82.10	98.65	98.40	90.16	87.50	99.51	98.00	94.15	94.37	99.01	99.62	96.51	89.53
	F1 Score	99.29	99.17	91.67	91.16	99.01	98.38	89.23	86.09	99.10	98.31	91.53	89.45	99.26	98.27	93.47	91.48	99.21	98.83	93.07	88.37

TABLE VII
COMPARISON OF DIFFERENT UFCM UPDATING STRATEGIES WITHIN THE TFCCL FRAMEWORK ON WAAM DATA SETS UNDER VARYING NOISE RATIOS

Approaches	Metrics (%)	Data Set 1				Data Set 2				Data Set 3				Data Set 4				Data Set 5			
		30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
TFCCL (Linear)	Accuracy	98.26	94.87	88.65	73.00	95.70	91.78	85.43	76.62	95.57	89.94	85.27	84.36	92.80	90.37	83.99	77.69	95.80	92.76	86.34	80.14
	Precision	97.78	92.50	90.02	71.98	98.23	91.67	82.73	78.79	94.07	93.50	84.98	88.31	92.71	94.08	89.13	90.84	99.34	98.81	77.78	87.80
	Recall	99.54	97.76	89.18	97.62	92.63	94.53	96.52	84.42	98.90	89.57	93.82	85.09	96.66	91.10	86.47	75.60	93.79	90.32	96.23	86.91
	F1 Score	98.67	96.10	89.97	83.27	96.17	93.07	89.58	82.02	96.44	91.50	88.81	86.40	94.66	92.57	88.22	87.25	96.77	94.36	88.09	87.64
TFCCL (Logarithmic)	Accuracy	96.93	92.56	86.20	70.64	94.47	89.13	83.14	74.54	94.06	87.58	81.70	82.28	90.94	88.17	81.00	75.41	94.15	90.48	78.07	77.53
	Precision	96.65	90.00	87.83	70.42	96.78	88.81	80.03	76.95	92.54	91.31	82.13	86.46	91.40	91.72	86.79	88.57	97.55	96.57	75.33	85.58
	Recall	98.03	95.67	86.28	95.30	90.78	92.13	93.64	82.83	97.55	87.41	91.35	83.37	95.62	88.37	84.15	73.93	92.57	87.93	94.20	85.29
	F1 Score	97.10	93.21	87.81	81.36	94.92	90.50	86.69	80.03	94.82	89.16	86.19	84.83	93.50	89.77	85.82	85.32	94.95	91.70	86.62	85.53
TFCCL (Dynamic)	Accuracy	99.10	98.95	93.88	93.45	98.84	98.66	91.19	89.02	98.91	98.78	93.71	91.95	99.03	98.45	95.48	93.98	98.95	98.42	95.20	92.22
	Precision	99.30	99.49	90.98	89.93	98.98	98.83	91.60	92.18	99.55	98.23	94.35	92.24	99.02	98.57	93.14	89.62	99.42	98.06	90.10	87.58
	Recall	99.28	98.85	92.82	92.99	99.04	97.94	87.65	82.10	98.65	98.40	90.16	87.50	99.51	98.00	94.15	94.37	99.01	99.62	96.51	89.53
	F1 Score	99.29	99.17	91.67	91.16	99.01	98.38	89.23	86.09	99.10	98.31	91.53	89.45	99.26	98.27	93.47	91.48	99.21	98.83	93.07	88.37

performance on the remaining metrics in these scenarios, which demonstrates the overall effectiveness of TFCCL in outlier detection with a relatively small number of labels.

2) 50% Noisy Labels: In the scenario with a 50% noise ratio, the TFCCL remains superior performance, whereas the performance of all selected approaches decreases significantly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

TABLE VIII
RESULTS OF ABLATION STUDY ON WAAM DATA SETS WITH DIFFERENT NOISE RATIOS

	Metrics (%)	Data Set 1				Data Set 2				Data Set 3				Data Set 4				Data Set 5			
		30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%	30%	40%	50%	60%
w/o Pre-training	Accuracy	95.17	93.96	80.26	71.22	95.75	94.02	77.57	63.44	96.75	94.52	86.50	76.94	95.48	94.93	86.85	80.80	95.70	93.98	88.49	82.74
	Precision	99.84	96.42	82.23	70.66	95.77	92.07	75.03	61.49	94.95	91.71	90.85	75.34	93.65	92.85	89.35	83.48	95.84	92.63	92.85	96.35
	Recall	92.38	93.97	87.85	93.43	97.00	98.19	92.24	99.85	99.94	99.97	86.35	91.92	99.90	99.87	90.85	88.29	97.85	98.91	89.77	74.31
	F1 Score	96.04	95.18	84.95	80.46	96.38	95.03	82.75	76.11	97.38	95.66	88.54	82.81	96.67	96.29	90.09	85.81	96.83	95.67	91.29	85.26
w/o FCCL	Accuracy	92.74	92.18	46.50	11.99	94.97	94.01	72.39	27.62	95.29	93.06	66.07	34.54	88.13	85.02	52.50	46.65	94.00	93.16	41.72	20.68
	Precision	95.46	97.20	55.76	24.68	93.48	93.43	61.42	34.59	97.34	96.21	80.14	41.76	84.71	81.45	65.41	39.04	94.29	95.44	51.29	38.66
	Recall	91.95	90.27	50.66	18.91	98.22	96.53	59.10	27.03	94.79	92.15	56.25	21.11	98.04	93.54	61.84	18.90	96.94	94.33	56.15	30.78
	F1 Score	93.68	93.60	52.98	21.41	95.79	94.95	64.10	30.34	96.05	94.13	65.42	28.04	91.72	89.78	63.53	31.79	95.59	94.88	52.80	34.27
w/o UMAP	Accuracy	98.85	95.61	64.85	57.82	96.79	93.59	69.91	56.75	96.29	88.66	76.97	64.64	93.47	90.97	67.91	65.31	96.36	93.40	65.49	59.80
	Precision	99.04	93.90	66.89	69.83	96.53	94.21	70.33	62.12	97.60	92.88	76.06	65.28	90.97	98.63	70.38	59.61	95.94	95.95	71.53	71.87
	Recall	97.80	99.54	88.26	58.97	98.02	94.84	83.76	66.27	96.23	84.19	84.01	88.63	96.45	87.49	88.42	69.73	98.76	94.15	80.81	66.01
	F1 Score	98.41	96.64	76.08	63.94	97.27	94.52	76.46	64.13	96.91	91.42	82.54	75.18	95.27	92.73	78.38	64.79	97.33	95.04	75.89	68.81
w/o Dynamic UFCM Updating Strategy	Accuracy	98.28	94.83	84.76	70.83	95.70	91.75	84.32	75.20	95.58	89.90	82.62	81.92	92.81	90.34	82.13	76.29	95.78	92.72	79.05	78.57
	Precision	97.79	92.48	88.03	69.64	98.22	91.63	81.46	76.59	94.08	93.48	83.68	86.47	92.71	94.05	87.39	88.82	99.95	98.78	76.23	83.53
	Recall	99.54	99.97	87.93	95.76	92.62	94.49	94.66	82.80	98.90	89.53	92.19	83.08	96.66	91.08	85.11	73.51	93.77	90.28	94.31	84.84
	F1 Score	98.66	96.08	87.98	80.64	96.17	93.04	87.57	79.57	96.43	91.46	87.28	84.74	94.65	92.54	86.23	84.73	96.76	94.34	86.51	84.18
TFCCL	Accuracy	99.10	98.95	93.88	93.45	98.84	98.66	91.19	89.02	98.91	98.78	93.71	91.95	99.03	98.45	95.48	93.98	98.95	98.42	95.20	92.22
	Precision	99.30	99.49	90.98	89.93	98.98	98.83	91.60	92.18	99.55	98.23	94.35	92.24	99.02	98.57	93.14	89.62	99.42	98.06	90.10	87.58
	Recall	99.28	98.85	92.82	92.99	99.04	97.94	87.65	82.10	98.65	98.40	90.16	87.50	99.51	98.00	94.15	94.37	99.01	99.62	96.51	89.53
	F1 Score	99.29	99.17	91.67	91.16	99.01	98.38	89.23	86.09	99.10	98.31	91.53	89.45	99.26	98.27	93.47	91.48	99.21	98.83	93.07	88.37

on all five data sets. Specifically, the accuracy and the F1 score of the selected approaches are no more than 88% and 78%, respectively. For TFCCL, the accuracy and the F1 score across five data sets exceed 91% and 88%, respectively, which shows the robustness of TFCCL against a moderate number of noisy labels.

3) 60% Noisy Labels: In situations with the noise ratio of 60%, the baseline approach and JoCoR-CNN are affected by severe label noise and are no longer capable of handling outlier detection tasks. The performance of the other selected approaches is also not effective enough. It should be noted that TFCCL still achieves satisfactory performance across five data sets, with each data set's accuracy above 89% and F1 score above 86%. The results indicate the competitive performance of the TFCCL in outlier detection with a large number of noisy labels.

C. Comparison Experiments

In this paper, two comparison experiments are conducted. Specifically, the first experiment aims to verify the effectiveness of the selection of the FCM algorithm. The results are illustrated in Table VI. According to the results, the FCM algorithm shows competitive performance across all five data sets under different noise ratios compared with other hard clustering algorithms, demonstrating its robustness and adaptability in the presence of label noise. The second experiment aims to verify the effectiveness of the proposed dynamic updating strategy for the UFCM algorithm. As shown in Table VII, the developed dynamic updating strategy achieves the best results compared with other updating strategies.

D. Ablation Study

In this paper, a detailed ablation study is carried out to verify the effectiveness of the main components in TFCCL under various noise ratios. To be specific, the effects of the pre-training, the FCCL strategy, the UMAP and the dynamic UFCM updating strategy in TFCCL are verified. The results of the ablation study are illustrated in Table VIII. Based on the results, detailed analyses are given: 1) The pre-training helps the Transformer encoder exploit the internal structure and capture useful feature representations. When trained directly for classification without pre-training, the average performance of the TFCCL framework declines with increasing noise ratios. 2) The FCCL strategy significantly improves the performance of outlier detection with noisy labels by utilizing the intrinsic features of data. Employing the UMAP is able to enhance the performance of the FCCL strategy when dealing with severe label noise. 3) The dynamic UFCM updating strategy effectively stabilizes the training process and the final performance. 4) Each of the main components plays a critical role in the TFCCL framework. The best performance of TFCCL can be achieved only if all components work together.

V. CONCLUSION

In this paper, a new TFCCL framework has been developed for learning with noisy labels (LNL) in outlier detection on industrial time series. Specifically, a novel FCCL strategy has been introduced to update the Transformer encoder, where a UFCM algorithm has been proposed for identifying positive and negative sample pairs. A dynamic two-stage training scheme has been designed to train the outlier detector, incorporating a joint learning strategy to enhance its robustness against noisy labels. Moreover, a dynamic updating strategy has been employed to update the UFCM algorithm throughout the training process. To further improve model reliability, a label-consistency regularization term has been introduced to minimize the discrepancy between the outputs of the outlier detector and the UFCM algorithm. Experimental results have demonstrated the effectiveness of the developed TFCCL framework. Future research directions can be summarized as follows: 1) utilizing optimization techniques for automatic hyperparameter selection [7], [16], [29], [35], [36]; 2) extending the developed TFCCL framework to multi-class outlier detection tasks [21], [22]; 3) modifying the Transformer encoder architecture to enhance feature extraction performance [3], [27]; and 4) refining the FCCL strategy by adjusting the contrastive loss and improving the clustering algorithm [24], [28].

Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

References

[1] M. Boyapati and R. Aygun, Semanformer: Semantics-aware embedding dimensionality reduction using transformer-based models, In: *Proceedings of the 2024 IEEE 18th International Conference on Semantic Computing (ICSC)*, Laguna Hills, USA, Feb. 2024, pp. 134-141.

[2] A. Castellani, S. Schmitt and B. Hammer, Estimating the electrical power output of industrial devices with end-to-end time-series classification in the presence of label noise, In: *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Bilbao, Spain, Sep. 2021, pp. 469-484.

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

[4] Y. Dai, Z. Mei, J. Li, Z. Li, K. Wei, M. Ding, S. Guo and W. Chen, Clustering-based contrastive learning for fault diagnosis with few labeled samples, *IEEE Transactions on Instrumentation and Measurement*, vol. 73, art. no. 2504913, 2023.

[5] J. Fang, Z. Wang, W. Liu, L. Chen and X. Liu, A new particle-swarm-optimization-assisted deep transfer learning framework with applications to outlier detection in additive manufacturing, *Engineering Applications of Artificial Intelligence*, vol. 131, art. no. 107700, 2024.

[6] J. Fang, Z. Wang, W. Liu, S. Lauria, N. Zeng, C. Prieto, F. Sikström and X. Liu, A new particle swarm optimization algorithm for outlier detection: Industrial data clustering in wire arc additive manufacturing, *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 2, pp. 1244-1257, 2022.

[7] W. Fang, B. Shen, A. Pan, L. Zou and B. Song, A cooperative stochastic configuration network based on differential evolutionary sparrow search algorithm for prediction, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2314481, 2024.

[8] M. Gong, Y. Zhao, H. Li, A. Qin, L. Xing, J. Li, Y. Liu and Y. Liu, Deep fuzzy variable C-means clustering incorporated with curriculum learning, *IEEE Transactions on Fuzzy Systems* , vol. 31, no. 12, pp. 4321-4335, 2023.

[9] E. Hallaji, R. Razavi-Far, M. Saif and E. Herrera-Viedma, Label noise analysis meets adversarial training: A defense against label poisoning in federated learning, *Knowledge-Based Systems*, vol. 266, art. no. 110384, 2023.

[10] A. Hamrani, F. Bouarab, A. Agarwal, K. Ju and H. Akbarzadeh, Advancements and applications of multiple wire processes in additive manufacturing: A comprehensive systematic review, *Virtual and Physical Prototyping*, vol. 18, no. 1, art. no. e2273303, 2023.

[11] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang and M. Sugiyama, Co-teaching: Robust training of deep neural networks with extremely noisy labels, In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montreal, Canada, Dec. 2018, vol. 31.

[12] F. He, L. Yuan, H. Mu, M. Ros, D. Ding, Z. Pan and H. Li, Research and application of artificial intelligence techniques for wire arc additive manufacturing: A state-of-the-art review, *Robotics and Computer-Integrated Manufacturing*, vol. 82, art. no. 102525, 2023.

[13] K. He, H. Fan, Y. Wu, S. Xie and R. Girshick, Momentum contrast for unsupervised visual representation learning, In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, Jun. 2020, pp. 9726-9735.

[14] Z. Huang, J. Zhang and H. Shan, Twin contrastive learning with noisy labels, In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, Jun. 2023, pp. 11661-11670.

[15] J. Li, R. Socher and S. C. H. Hoi, DivideMix: Learning with noisy labels as semi-supervised learning, *arXiv preprint arXiv:2002.07394*, 2020.

[16] Y. Liang, L. Tian, X. Zhang, X. Zhang and L. Bai, Multi-dimensional adaptive learning rate gradient descent optimization algorithm for network training in magneto-optical defect detection, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100016, 2024.

[17] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu and F. E. Alsaadi, A survey of deep neural network architectures and their applications, *Neurocomputing*, vol. 234, pp. 11-26, 2017.

[18] S. Lu, Z. Gao and Y. Liu, HFTL-KD: A new heterogeneous federated transfer learning approach for degradation trajectory prediction in large-scale decentralized systems, *Control Engineering Practice*, vol. 153, art. no. 106098, 2024.

[19] G. Mattera, J. Polden, A. Caggiano, L. Nele, Z. Pan and J. Norrish, Semi-supervised learning for real-time anomaly detection in pulsed transfer wire arc additive manufacturing, *Robotics and Computer-Integrated Manufacturing*, vol. 128, pp. 84-94, 2024.

[20] A. Oord, Y. Li and O. Vinyals, Representation learning with contrastive predictive coding, *arXiv preprint arXiv:1807.03748*, 2018.

[21] B. Qu, D. Peng, Y. Shen, L. Zou and B. Shen, A survey on recent advances on dynamic state estimation for power systems, *International Journal of Systems Science*, vol. 55, no. 16, pp. 3305-3321, 2024.

[22] B. Qu, Z. Wang, B. Shen and H. Dong, Decentralized dynamic state estimation for multi-machine power systems with non-Gaussian noises: Outlier detection and localization, *Automatica*, vol. 153, art. no. 111010, 2023.

[23] B. Song, S. Zhao, L. Dang, H. Wang and L. Xu, A survey on learning from data with label noise via deep neural networks, *Systems Science & Control Engineering*, vol. 13, no. 1, art. no. 2488120, 2025.

[24] Y. Sun, M. Chen, K. Peng, L. Wu and C. Liu, Finite-time adaptive optimal control of uncertain strict-feedback nonlinear systems based on fuzzy observer and reinforcement learning, *International Journal of Systems Science*, vol. 55, no. 8, pp. 1553-1570, 2024.

[25] C. Tan, J. Xia, L. Wu and S. Li, Co-learning: Learning from noisy labels with self-supervision, In: *Proceedings of the 29th ACM International Conference on Multimedia*, China, Oct. 2021, pp. 1405-1413.

[26] Y. Tu, B. Zhang, Y. Li, L. Liu, J. Li, J. Zhang, Y. Wang, C. Wang and C. R. Zhao, Learning with noisy labels via self-supervised adversarial noisy masking, In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, Jun. 2023, pp. 16186-16195.

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

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

[29] C. Wang, Z. Wang, Q. Liu, H. Dong and W. Sheng, Support-sample-assisted domain generalization via attacks and defenses: Concepts, algorithms and applications to pipeline fault diagnosis, *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6413-6423, 2024.

[30] H. Wang, C. Li, P. Ding, S. Li, T. Li, C. Liu, X. Zhang and Z. Hong, A novel transformer-based few-shot learning method for intelligent fault diagnosis with noisy labels under varying working conditions, *Reliability Engineering & System Safety*, vol. 251, art. no. 110400, 2024.

[31] R. Wang, X. Wu, T. Xu, C. Hu and J. Kittler, U-SPDNet: An SPD manifold learning-based neural network for visual classification, *Neural Networks*, vol. 161, pp. 382-396, 2023.

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

[33] H. Wei, L. Feng, X. Chen and B. An, Combating noisy labels by agreement: A joint training method with co-regularization, In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, Jun. 2020, pp. 13723-13732.

[34] J. Xia, H. Lin, Y. Xu, C. Tan, L. Wu, S. Li and S. Z. Li, Gnn cleaner: Label cleaner for graph structured data, *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 640-651, 2024.

[35] J. Xue and B. Shen, A survey on sparrow search algorithms and their applications, *International Journal of Systems Science*, vol. 55, no. 4, pp. 814-832, 2024.

[36] Y. Xue, M. Li, H. Arabnejad, D. Suleimenova, A. Jahani, B. C. Geiger, F. Boesjes, A. Anagnostou, S. J. E. Taylor, X. Liu and D. Groen, Many-objective simulation optimization for camp location problems in humanitarian logistics, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100017, 2024.

[37] J. Yan, L. Luo, C. Deng and H. Huang, Adaptive hierarchical similarity metric learning with noisy labels, *IEEE Transactions on Image Processing*, vol. 32, pp. 1245-1256, 2023.

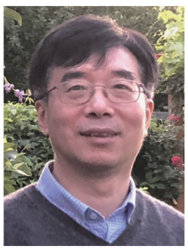
[38] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang and M. Sugiyama, How does disagreement help generalization against label corruption? In: *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, USA, Jun. 2019, pp. 7164-7173.

[39] Z. Yuan, C. Yao, X. Liu, Z. Gao and W. Zhang, Multiagent formation control and dynamic obstacle avoidance based on deep reinforcement learning, *IEEE Transactions on Industrial Informatics*, vol. 21, no. 6, pp. 4672-4682, 2025.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics

Jingzhong Fang (Member, IEEE) received his B.Eng. degree in automation from Shandong University of Science and Technology, Qingdao, China, in 2020, and the M.Sc. degree in data science and analytics from Brunel University of London, Uxbridge, U.K., in 2021. He is currently pursuing the Ph.D. degree in computer science at Brunel University of London, Uxbridge, U.K.

His research interests include intelligent data analysis and deep learning techniques. He is a very active reviewer for many international journals.



Zidong Wang (SM'03-F'14) received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China.

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

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals including IEEE Transactions on Automatic Control, IEEE Transactions on Control Systems Technology, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing, and IEEE Transactions on Systems, Man, and Cybernetics-Part C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.

Weibo Liu (Member, IEEE) received the B.Eng. degree in electrical engineering from the Department of Electrical Engineering & Electronics, University of Liverpool, Liverpool, U.K., in 2015, and the Ph.D. degree in artificial intelligence in 2020 from the Department of Computer Science, Brunel University of London, Uxbridge, U.K.

He is currently a Lecturer in the Department of Computer Science, Brunel University of London, Uxbridge, U.K. His research interests include intelligent data analysis, evolutionary computation, machine learning, deep learning and transfer learning. He serves as an Associate Editor for the Journal of Ambient Intelligence and Humanized Computing and the Journal of Cognitive Computation. He is a very active reviewer for many international journals and conferences.

Nianyin Zeng (Senior Member, IEEE) was born in Fujian, China, in 1986. He received the B.Eng. degree in electrical engineering and automation and the Ph.D. degree in electrical engineering from Fuzhou University, Fuzhou, China, in 2008 and 2013, respectively.

From October 2012 to March 2013, he was an RA with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. From September 2017 to August 2018, he was an ISEF Fellow funded by Korea Foundation for Advanced Studies, Seoul, South Korea, and also a Visiting Professor with Korea Advanced Institute of Science and Technology, Daejeon, South Korea. He is currently a Professor with the Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, China. He has authored or co-authored several technical articles. His current research interests include intelligent data analysis, computational intelligence, time-series modeling, and applications.

Dr. Zeng is currently serving as an Associate Editor for *Neurocomputing*, *Evolutionary Intelligence*, and *Frontiers in Medical Technology*, and also an Editorial Board Member for *Computers in Biology and Medicine*, *BioMedical Engineering OnLine*, and *Mathematical Problems in Engineering*. He is a very active reviewer of many international journals and conferences.

Yimeng He received the B.Eng. degree from the School of Automation, Central South University, Changsha, China, in 2020, and received a Ph.D. degree in control science and engineering at the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2025.

Her research interests include causal discovery, data-driven industrial modeling, machine learning, and deep learning.

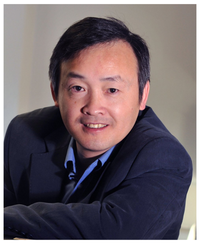
Yu Cao received his B.Eng. degree in computer science from Shandong Technology and Business University, Yantai, China, in 2019, and the M.Sc. degree in data science and analytics from Brunel University London, Uxbridge, U.K., in 2020. He is currently pursuing the Ph.D. degree in computer science at Brunel University London, Uxbridge, U.K.

His research interests include video anomaly detection and deep learning techniques.

Linwei Chen received her B.Eng. degree in electrical and electronic engineering from the University of Warwick, Coventry, U.K., in 2022. She is currently pursuing the Ph.D. degree in computer science at Brunel University London, Uxbridge, U.K.

Her research interests include transfer learning and optimization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2025.3616850, IEEE Transactions on Industrial Informatics



Xiaohui Liu received the B.Eng. degree in computing from Hohai University, Nanjing, China, in 1982 and the Ph.D. degree in computer science from Heriot-Watt University, Edinburg, U.K., in 1988.

He is currently a Professor of Computing at Brunel University London, Uxbridge, U.K., where he conducts research in artificial intelligence, data science, and optimization, with applications in diverse areas including biomedicine and engineering.