A Literature Review on Few-Shot Anomaly Detection in Time Series Data

AbstractTime series anomaly detection (TSAD) is a critical task in numerous domains, including finance, industrial monitoring, and healthcare. However, most traditional methods require large volumes of labeled data, particularly of normal behavior. In many real-world scenarios, labeled data is scarce, and anomalies, by definition, are rare. This makes it challenging to train robust models. **Few-shot anomaly detection** has emerged as a crucial sub-field to address this challenge, aiming to build models that can identify anomalies after seeing only a minimal number of labeled examples (the "shots"). This review synthesizes the current landscape of few-shot TSAD, categorizing existing work into four primary approaches: **meta-learning**, **generative and reconstruction models**, **contrastive learning**, and the emerging use of **large foundation models**. We discuss the foundational concepts of each methodology, survey key contributions, and conclude with the outstanding challenges and future research directions.

---

1. Introduction

A time series is a sequence of data points indexed in time order. Anomaly detection in this context involves identifying data points, sequences, or patterns that deviate significantly from what is defined as "normal" behavior. These deviations can signify critical events such as a machine failure, a health crisis, or fraudulent financial activity.

The primary challenge in TSAD is the **severe class imbalance**. Anomalies are rare, and it is often impractical or impossible to collect a comprehensive dataset of all possible anomalous events. Traditional unsupervised methods attempt to model the entirety of the normal data, but they can be sensitive to noise and struggle to define a tight boundary for "normality."

The **few-shot learning (FSL)** paradigm offers a promising solution. FSL aims to build models that can generalize from a very small number of examples, often just one or five (1-shot or 5-shot). In the context of anomaly detection, this typically means training a model to identify anomalies using only a few reference samples of either normal data, anomalous data, or both. This review explores the key strategies researchers are currently employing to tackle this problem.

---

2. Core Methodologies and Current Work

The research in few-shot TSAD can be broadly classified based on the underlying technical approach.

2.1 Meta-Learning (Learning to Learn)

Meta-learning, or "learning to learn," trains a model on a wide distribution of different learning tasks. The goal is to enable the model to *quickly adapt* to a new, unseen task using only a few examples.

- **Concept:** In the TSAD context, a "task" might be a specific time series from a particular sensor or dataset. The meta-model learns a generalized representation or an optimization strategy across many time series. When presented with a new time series and its few-shot support set (examples of normal data), it can rapidly create a task-specific anomaly detector.
- **Current Work:**
  - **MAML (Model-Agnostic Meta-Learning):** While not specific to TSAD, MAML is a foundational approach used in this domain. It finds a set of model parameters that can be fine-tuned to a new task with just a few gradient descent steps.
  - **Adaptive Sparse Coding:** Recent work has explored using meta-learning to train an adaptive sparse coding layer. The model learns to construct a "dictionary" of normal patterns from a few normal samples at inference time. New data points that cannot be reconstructed well from this dictionary are flagged as anomalous.

2.2 Generative and Reconstruction Models

This category of methods focuses on learning a robust model of the *normal* data distribution from the few available samples. Anomalies are then detected as low-probability outliers from this learned distribution.

- **Concept:** Given a few-shot support set of normal data, a generative model (like a Variational Autoencoder or VAE) or a reconstruction model (like an Autoencoder) is trained. New data points are passed through the model. If the model fails to reconstruct the data accurately (i.e., high reconstruction error), the data point is considered anomalous, as it does not fit the learned "normal" model.
- **Current Work:**
  - **Energy-Based Models (EBMs):** EBMs learn an energy function that assigns low energy scores to normal samples and high energy scores to anomalies. Few-shot learning can be used to quickly adapt this energy landscape to a new task.
  - **Generative Augmentation:** A key challenge in few-shot learning is the limited data. Some recent methods use generative models, such as **diffusion models**, to synthesize a larger, more diverse set of high-fidelity *normal* samples from the

initial few shots. This creates a more robust "normal" distribution to compare against.
    - **Reconstruction-Based Pipelines:** Classic autoencoder-based reconstruction remains a strong baseline. The model is trained on the few normal shots, and a threshold is set on the reconstruction error to detect anomalies.

2.3 Contrastive Learning

Contrastive learning has become a dominant paradigm in representation learning. It learns an embedding space where similar (positive) samples are pulled close together and dissimilar (negative) samples are pushed far apart.

- **Concept:** In the few-shot setting, the support set of normal samples is used to define "positive" pairs. The core challenge is defining "negative" (anomalous) samples.
- **Current Work:**
    - **Synthetic Anomaly Generation:** Methods like **CAROTS (Causality-Aware Contrastive Learning)** generate negative samples by creating "causality-disturbing" augmentations of the time series. The model is trained to distinguish between causality-preserving (normal) and causality-disturbing (anomalous) samples.
    - **Using Labeled Anomalies:** If the few-shot set contains rare *labeled anomalies* (a setting known as few-shot anomaly *classification*), these can be used as strong negative examples. Frameworks like **ANEMONE-FS** use a multi-scale contrastive learning approach where labeled anomalies are explicitly pushed away from normal samples in the embedding space.
    - **Triplet Loss & Reconstruction:** Other methods, like **CL-TAD**, combine contrastive learning with reconstruction. They may use a triplet loss where the "anchor" is the original sample, the "positive" is a reconstruction, and the "negative" is another sample from the batch.

---

3. Key Challenges and Future Directions

Despite significant progress, several challenges remain in few-shot time series anomaly detection.

1. **Defining "Anomalous":** The greatest challenge remains the lack of negative samples. If the few shots are *only* normal, the model must infer what is anomalous. This is why synthetic anomaly generation and causality-aware learning are critical research areas.
2. **Benchmark Scarcity:** There is a lack of standardized, high-quality public benchmarks specifically designed for the *few-shot* TSAD problem. Most existing benchmarks are for unsupervised or semi-supervised settings.
3. **Interpretability:** Most deep learning methods operate as "black boxes." It is not enough to know *that* an anomaly occurred; operators need to know *why*. The emerging

LMM-based methods that provide natural language justifications are a promising step toward solving this.

4. **Generalization and Transfer:** A model trained on tasks from one domain (e.g., industrial sensor data) may fail to generalize to tasks in a completely different domain (e.g., ECG data). Improving the cross-domain generalization of meta-learning models is an open problem.
5. **Computational Efficiency:** Foundation models and large generative models are computationally expensive, which may limit their use in real-time, resource-constrained environments (e.g., on-device IoT monitoring).