International Journal of Computer Applications in Technology

ISSN online: 1741-5047 - ISSN print: 0952-8091
https://www.inderscience.com/ijcat

Anomaly detection and pattern recognition methods for high-dimensional data

Shengqi Wang

DOI: 10.1504/IJCAT.2026.10075605

Article History:

Received:	19 May 2025
Last revised:	20 August 2025
Accepted:	18 September 2025
Published online:	17 February 2026

Anomaly detection and pattern recognition methods for high-dimensional data

Shengqi Wang

School of Accounting,
Shandong University of Financial and Economics,
Jinan, Shandong, China
Email: 202308340243@mail.sdufe.edu.cn

Abstract: High-dimensional data has become increasingly prevalent in a wide range of fields, including cybersecurity, finance, healthcare and industrial monitoring. However, the sparsity, redundancy and complex inter-feature relationships inherent in such data significantly complicate anomaly detection and pattern recognition tasks. Traditional machine learning methods often suffer from poor scalability and limited generalisation in high-dimensional settings. To address these limitations, this paper proposes a novel deep learning framework specifically designed for high-dimensional anomaly detection and pattern recognition. The proposed model introduces three key innovations. First, a hierarchical representation module is developed to extract multi-level semantic features by integrating adaptive kernel transformations with semantic-preserving aggregation strategies. This design improves the model's ability to capture both global patterns and local anomalies. Second, a dual-branch attention mechanism is introduced to jointly learn feature-level and instance-level relevance, enhancing the model's robustness to noise and irrelevant dimensions. Third, an interpretable anomaly scoring strategy is constructed based on prototype deviation in latent space, offering transparency and actionable insights for decision support. Extensive experiments are conducted on multiple real-world high-dimensional data sets. Results demonstrate that the proposed method consistently outperforms existing approaches in terms of accuracy, robustness and interpretability.

Keywords: high-dimensional data; anomaly detection; hierarchical representation learning; attention mechanism.

Reference to this paper should be made as follows: Wang, S. (2026) 'Anomaly detection and pattern recognition methods for high-dimensional data', *Int. J. Computer Applications in Technology*, Vol. 78, No. 2, pp.155–166.

Biographical notes: Shengqi Wang is a Researcher at the School of Accounting at Shandong University of Financial and Economics. His research interests include data analysis, financial intelligence, anomaly detection and computational modeling.

1 Introduction

In recent years, the explosion of high-dimensional data across domains such as network traffic monitoring, financial transaction systems, bioinformatics and industrial IoT has posed significant challenges for data analysis and intelligent decision-making. As the number of features grows, traditional methods suffer from the so-called 'curse of dimensionality', where data becomes increasingly sparse in the feature space, and meaningful relationships are harder to detect (Thudumu et al., 2020). In such environments, anomaly detection – the task of identifying samples that deviate significantly from normal patterns – has become both more critical and more complex (Zheng et al., 2022). Similarly, pattern recognition in high-dimensional spaces requires robust algorithms that can distinguish essential structures from noise or redundancy (Muneer et al., 2022). To tackle these challenges, deep learning has emerged as a dominant paradigm. Architectures such as autoencoders (Wang et al., 2025), deep belief networks (Pei, 2025) and convolutional neural networks (Wang and Liu, 2024) have shown promise in extracting compact, informative representations from high-dimensional input. More recently, transformer-based models (Sana et al., 2024) and attention mechanisms (Georgakopoulos et al., 2022) have demonstrated their capability in modelling long-range dependencies, even in sparse or irregular feature distributions. In the realm of unsupervised and semi-supervised learning, variational autoencoders (Mozaffari et al., 2023) and contrastive learning methods (Pei et al., 2025) have further improved performance by enabling models to capture meaningful structures without relying on labelled data. Applications such as anomaly detection in cybersecurity logs (Luo et al., 2023), outlier detection in financial records (Huo et al., 2024) and multivariate pattern analysis in medical imaging (Wen et al., 2025) all highlight the growing importance of reliable high-dimensional modelling techniques powered by deep learning.

Despite these advances, existing deep learning methods still exhibit substantial limitations when applied to high-dimensional anomaly detection and pattern recognition. One

key issue is scalability – many deep architectures do not adapt well to increasing dimensionality, leading to overfitting or ineffective generalisation (Chalapathy and Chawla, 2019). Traditional encoder-decoder frameworks, such as standard autoencoders or GAN-based detectors, often learn overly smoothed representations that mask fine-grained anomalies (Yu et al., 2025). Furthermore, high-dimensional data often exhibits complex interactions among features that are not well-captured by models relying solely on global or independent feature assumptions (Baradaran and Bergevin, 2024). Another common limitation lies in sensitivity – many current methods are biased toward detecting only point anomalies and tend to miss contextual or collective anomalies (Cheng et al., 2021). Additionally, models that require large amounts of clean or labelled data are impractical in real-world high-dimensional scenarios, where annotation is expensive or infeasible (Datta et al., 2021). Although graph-based methods and attention networks have attempted to enhance contextual modeling, they often suffer from issues like over-smoothing or feature homogenisation (Cheng et al., 2023). Some approaches apply dimensionality reduction as a preprocessing step, but this may lead to the loss of important anomaly-relevant dimensions (Li et al., 2023). Meanwhile, transformers, while powerful, are often computationally expensive and lack built-in mechanisms for interpretability or anomaly scoring (Shen et al., 2024). Existing models also struggle to detect anomalies evolving over time, such as in streaming sensor data, due to limited adaptability and high latency (Tien et al., 2023). Recent hybrid models attempt to combine statistical reasoning with deep learning (Wang et al., 2022), or leverage meta-learning for improved few-shot detection (Souiden et al., 2022), but these methods often introduce additional complexity and lack interpretability (Sakib et al., 2025). Finally, explainability remains a major hurdle, as most deep models do not offer insight into why a particular instance was flagged as anomalous – an essential capability in sensitive domains such as healthcare or finance (Chen et al., 2024; Chander et al., 2025).

To address the above challenges, this paper proposes a novel deep learning framework specifically tailored to high-dimensional anomaly detection and pattern recognition tasks. The framework introduces three key innovations to enhance model expressiveness, robustness and interpretability. First, we propose a hierarchical representation learning module that extracts multi-level semantic features, allowing the model to encode both global structure and localised details. By integrating adaptive kernel transformations with multiscale aggregation, this module improves resistance to noise and enhances anomaly boundary modelling. Second, we design a dual-branch attention mechanism that concurrently performs instance-level and feature-level selection. This structure enables the model to filter irrelevant information and focus on discriminative cues across dimensions. To ensure coherence between the two attention paths, we introduce a regularisation loss that encourages complementary learning behaviours. Third, we develop an interpretable anomaly scoring strategy based on prototype deviation metrics in latent space, which provides human-understandable explanations for anomaly decisions. This scoring scheme considers both proximity to learned prototypes and distributional divergence from the expected normal cluster. Collectively, these innovations enable the model to address three core issues of high-dimensional anomaly detection: scalability to large and sparse data spaces, sensitivity to diverse types of anomalies and transparency of anomaly decisions.

The major contributions of this paper are summarised as follows:

1) A hierarchical representation module that captures multilevel semantic features and enhances anomaly localisation in high-dimensional data;
2) A dual-branch attention mechanism that filters noise and emphasises informative cues by integrating instance- and feature-level perspectives;
3) An interpretable anomaly scoring strategy that utilises prototype-based deviation analysis in latent space;
4) Extensive experiments on high-dimensional benchmarks, validating the superiority of the proposed framework in terms of detection accuracy, robustness and interpretability.

The structure is organised as follows. Section 2 reviews related work. Section 3 details the proposed framework, including the hierarchical representation module, the dual-branch attention mechanism and the prototype-based scoring strategy. Section 4 presents the experimental setup, results and ablation studies to validate the effectiveness of each component. Finally, Section 5 concludes the paper and discusses future research directions.

2 Related work

Yang et al. (2009) proposed a globally optimal Expectation-Maximisation algorithm for Gaussian Mixture Models (GMMs), using the mixture component weights to assign anomaly scores. As a classic non-parametric anomaly detection method, Kernel Density Estimation (KDE) is widely adopted. To address the limitations of KDE in handling complex and large-scale data sets, Gao et al. (2011) developed a variable kernel density estimation method that demonstrates superior detection accuracy and scalability. Building on this foundation, Qin et al. (2019) introduced the first linear-time anomaly detection algorithm for streaming data, which significantly alleviates performance bottlenecks by abstracting kernel centres and applying an improved outlier pruning strategy. Hartigan and Wong (1979) presented the well-known *K*-means clustering algorithm, which randomly initialises cluster centres and iteratively reassigns samples based on distance minimisation. To overcome the need for pre-defined cluster numbers, Comaniciu and Meer (2002) proposed the Mean-Shift algorithm, which adaptively shifts a sliding window toward areas of higher density until convergence, enabling fully data-driven clustering. Schubert et al. (2017) developed DBSCAN, a density-based clustering

method that starts from any unvisited point and dynamically expands a cluster by adding neighbouring points within a fixed radius, making it suitable for discovering clusters of arbitrary shape without specifying the number of clusters. Mensi et al. (2022) proposed an unsupervised Random Forest-based approach (RF-distance) for anomaly detection, constructing completely random splitting trees and deriving distance-based anomaly scores. On the other hand, Boosting-based models require the design of multiple weak learners, which may be either heterogeneous base models or the same model under different parameters. Xu et al. (2021) improved the AdaBoost algorithm by incorporating an active learning strategy and introduce the LAL-AdaBoost method to mitigate classification bias caused by imbalanced data, thereby enhancing the accuracy of civil infrastructure health monitoring tasks.

Chouhan et al. (2019) applied a stacked autoencoder framework to model real-world network traffic, reconstructing the original feature space and identifying anomalies by leveraging the poor reconstruction ability of abnormal flows. Wang et al. (2021) employed Convolutional Neural Networks (CNNs) with multi-path residual learning modules to extract hierarchical features of varying granularity. They further adopt transfer learning to alleviate the dependence on labelled data and improve the detection of previously unseen anomalies. In the industrial domain, Tran et al. (2014) used Deep Belief Networks (DBNs) to analyse various signals such as valve pressure, vibration frequency and electrical patterns, enabling early fault detection in compressor valves. Schuster and Paliwal (1997) proposed the use of Recurrent Neural Networks (RNNs), whose memory capabilities are widely applied in fields such as weather forecasting and stock analysis. Creswell et al. (2018) introduced Generative Adversarial Networks (GANs), a model framework grounded in the concept of a two-player minimax game between a generator and a discriminator. When trained to approximate Nash equilibrium, the discriminator becomes unable to distinguish generated samples from real data, making GANs highly suitable for adversarial scenarios in cybersecurity. Based on this framework, Li et al. (2018) proposed a GAN-based method that segments sensor time series into multiple subsequences. The generator computes reconstruction error based on learned latent distributions, while the discriminator evaluates residuals between real and generated samples. The combined outputs enable accurate detection of system-level anomalous attacks. Ng et al. (2011) proposed sparse autoencoders, which constrain neuron activation to effectively extract latent representations from unlabelled data. To address overfitting issues in traditional autoencoders, Vincent (2011) introduced denoising autoencoders that learn edge detectors from corrupted image patches, improving the recognition of digit strokes. Distinct from these architectures, Kusner et al. (2017) proposed the Variational Autoencoder (VAE), which leverages variational inference to learn the distributional parameters of latent Gaussian variables, enabling realistic sample generation and improving the model's ability to generalise across complex data distributions.

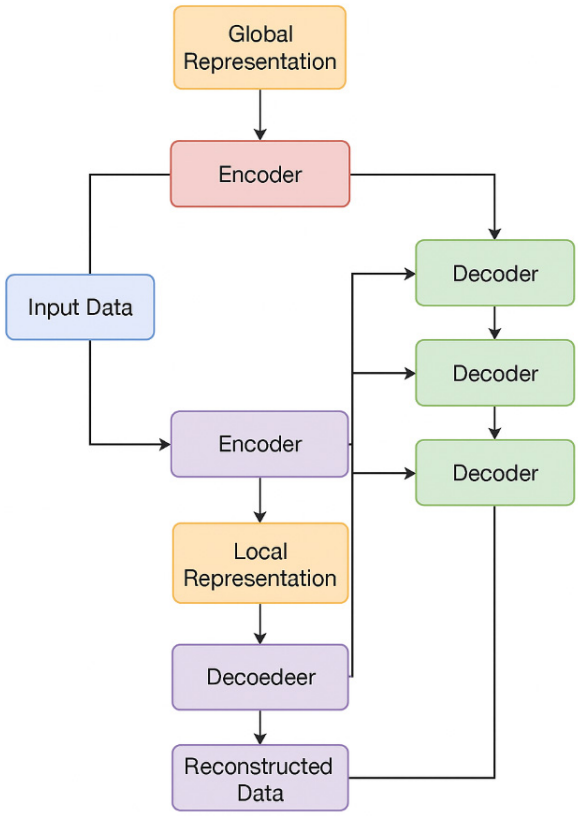
3 Method

This section introduces the proposed framework for high-dimensional anomaly detection and pattern recognition, consisting of three major components: a hierarchical representation module for multi-level feature abstraction; a dual-branch attention mechanism that jointly models feature-level and instance-level relevance, and an interpretable prototype-based anomaly scoring strategy. Each module is described in detail below, along with corresponding architectural diagrams and mathematical formulations.

3.1 Hierarchical representation learning module

As illustrated in Figure 1, the hierarchical representation learning module is designed to address one of the core challenges in high-dimensional anomaly detection: the simultaneous presence of low-level local perturbations and high-level structural inconsistencies. Conventional single-stage encoders often fail to model these multi-granular variations effectively. Shallow models typically capture only surface-level correlations, while deep models risk over-compression or gradient vanishing. Our proposed solution constructs an interpretable and trainable multi-level feature hierarchy, where both global context and local deviations are jointly encoded to improve anomaly sensitivity and semantic expressiveness.

Figure 1 Architecture of the hierarchical representation learning module. Multi-level encoding is performed using kernel-based transformations, and intermediate representations are fused with semantic constraints to preserve relevant information (see online version for colours)



Given an input feature vector $x \in \mathbb{R}^d$, the module begins by projecting it into a lower-dimensional representation using a linear transformation followed by a non-linear activation:

$$z^{(0)} = \sigma\left(W_0 x + b_0\right) \tag{1}$$

where W_0 and b_0 are the weights and bias of the initial projection, and $\sigma(\cdot)$ denotes a nonlinearity such as ReLU. This step serves to eliminate raw feature noise while retaining initial structure.

Subsequent feature transformations are organised in a stack of L encoding layers. Each layer performs adaptive non-linear mapping through a data-dependent kernel transformation. Specifically, the *l*-th layer is defined as:

$$z^{(l)} = \sigma\left(W_l \cdot \mathcal{K}_l\left(z^{(l-1)}\right) + b_l\right) \tag{2}$$

where $W_l \in \mathbb{R}^{d_{l+1} \times d_l}$ and $b_l \in \mathbb{R}^{d_{l+1}}$. Here, $\mathcal{K}_l(\cdot)$ is a kernel-based transformation that dynamically adjusts its shape to suit the input distribution at that layer.

The kernel transformation $\mathcal{K}_l$ is constructed using a weighted sum of Gaussian radial basis functions. Each basis function is centred at a learned kernel anchor m_i and scaled by a learned standard deviation σ_i:

$$\mathcal{K}_l(z) = \sum_{i=1}^{m_l} \gamma_i(z) \cdot \exp\left(-\frac{\| z - m_i \|^2}{2\sigma_i^2}\right) \tag{3}$$

The attention weights $\gamma_i(z)$ are computed using a softmax gating function that ensures adaptive, context-aware feature blending. This mechanism allows the model to emphasise task-relevant latent directions and suppress redundant subspaces in high-dimensional input.

The stack of transformations produces a set of intermediate representations $z^{(1)}, \ldots, z^{(L)}$, each encoding the input at a different level of semantic granularity. To preserve the diversity and richness of these hierarchical embeddings, we employ a skip-connection based aggregation:

$$z_{concat} = Concat\left(z^{(0)}, z^{(1)}, \ldots, z^{(L)}\right) \tag{4}$$

This design ensures that features captured at different depths are not overridden or diluted by deeper layers, which is particularly crucial when dealing with high-dimensional inputs where meaningful patterns may exist only in a subset of the feature hierarchy.

To further regularise the learned representation and encourage retention of semantic content, we introduce a reconstruction constraint. A decoder function $\psi(\cdot)$ attempts to recover the original input from the aggregated features:

$$\mathcal{L}_{rec} = \left\| x - \psi\left(z_{concat}\right) \right\|_2^2 \tag{5}$$

The reconstruction loss guides the encoder to preserve structurally relevant information, which is particularly important for detecting contextual anomalies that are subtle yet semantically inconsistent.

After aggregation, the final representation is normalised to stabilise learning and enhance compatibility with downstream modules:

$$z_{\text{final}} = \text{LayerNorm}\left(z_{\text{concat}}\right) \tag{6}$$

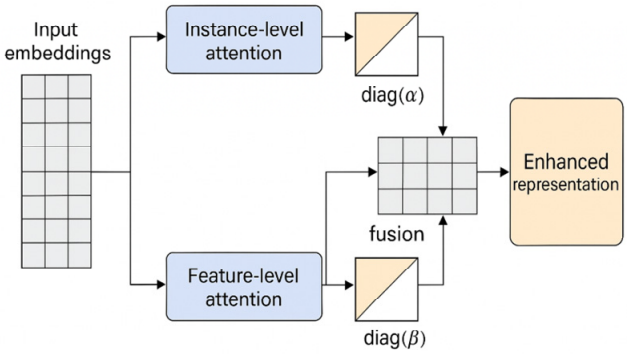
The normalisation step also serves to reduce statistical bias introduced by dimensional concatenation, ensuring that each level of abstraction contributes fairly to the final representation.

This module provides several advantages over standard encoders. First, hierarchical architecture enables learning representations that are both broad in scope and fine in resolution. Second, adaptive kernel mapping introduces a nonlinear inductive bias that enhances expressive capacity while avoiding rigid hand-crafted feature assumptions. Third, semantic preservation enforced through skip connections and reconstruction constraints mitigates the information bottleneck problem and supports more explainable anomaly inference. In sum, the hierarchical representation learning module delivers rich and interpretable embeddings by integrating semantic depth, non-linear geometry and structural regularisation. Its design reflects a unified approach to encoding high-dimensional data in a way that facilitates both accurate detection and robust generalisation under diverse and complex scenarios.

3.2 Dual-branch attention mechanism

To further enhance the model's ability to distinguish relevant signals from noise in high-dimensional space, we introduce a dual-branch attention mechanism, as illustrated in Figure 2. This module is designed to selectively amplify informative dimensions and samples while suppressing spurious correlations that commonly arise in complex data distributions. Unlike conventional attention mechanisms that treat the input as a single undifferentiated matrix, our method decomposes attention computation into two orthogonal but complementary branches – instance-level and feature-level. This design enables the model to adaptively attend to abnormal samples as well as feature sub-spaces that carry higher anomaly saliency.

Figure 2 Structure of the dual-branch attention mechanism. Instance-level and feature-level attention are computed in parallel and then fused to obtain an enhanced representation (see online version for colours)



Let $Z \in \mathbb{R}^{n \times d}$ denote the matrix of input embeddings from the previous module, where n is the number of samples and d is the feature dimensionality. The instance-level branch first computes the relevance score of each sample based on its deviation from global patterns. For the i-th row vector z_i, the attention weight α_i is defined as:

$$\alpha_i = \frac{\exp\left(v^{\mathrm{T}} \tanh\left(W_a z_i\right)\right)}{\sum_{j=1}^{n} \exp\left(v^{\mathrm{T}} \tanh\left(W_a z_j\right)\right)} \tag{7}$$

where $W_a \in \mathbb{R}^{d' \times d}$ is a learned projection matrix and $v \in \mathbb{R}^{d'}$ is a scoring vector. The tanh activation introduces nonlinearity and allows for richer contextual discrimination.

Concurrently, the feature-level branch evaluates which dimensions are more relevant across the entire batch. Each column $Z_{:,j}$ is passed through a similar attention mechanism:

$$\beta_j = \frac{\exp\left(u^{\mathrm{T}} \tanh\left(W_b Z_{:,j}\right)\right)}{\sum_{k=1}^{d} \exp\left(u^{\mathrm{T}} \tanh\left(W_b Z_{:,k}\right)\right)} \tag{8}$$

Here, $W_b \in \mathbb{R}^{d'' \times n}$ and $u \in \mathbb{R}^{d''}$ are trainable parameters. This pathway is particularly useful for identifying sparse anomaly patterns distributed across certain dimensions, such as edge features in sensor arrays or latent behavioural attributes in user profiles.

Once both attention scores are computed, we integrate them by computing a joint attention-enhanced matrix:

$$Z' = \mathrm{diag}(a) \cdot Z \cdot \mathrm{diag}(b) \tag{9}$$

This operation modulates each element z_{ij} of the feature matrix based on its row-wise and column-wise relevance. As a result, dimensions and samples with low attention scores are attenuated, while those aligned with potential anomaly structures are emphasised.

To promote robustness and mitigate degenerate learning behaviours, we apply a regularisation loss that encourages independence between the two attention branches:

$$\mathcal{L}_{attn} = \left(a^{\mathrm{T}} b\right)^2 \tag{10}$$

This loss term reduces co-adaptation between branches, ensuring that the attention process preserves orthogonal semantic interpretations from instance and feature perspectives. Without such regulation, the model may converge to a trivial solution where attention collapses into a few dominant entries, reducing expressiveness and harming generalisation.

The final representation is obtained by aggregating the attention-weighted features across all instances:

$$z_{\mathrm{attn}} = \frac{1}{n} \sum_{i=1}^{n} Z'_i \tag{11}$$

This vector serves as the input to the downstream anomaly scoring module. It summarises the most informative sample-specific and feature-specific signals in a compact form that is robust to irrelevant variations and well-aligned with the target anomaly semantics.

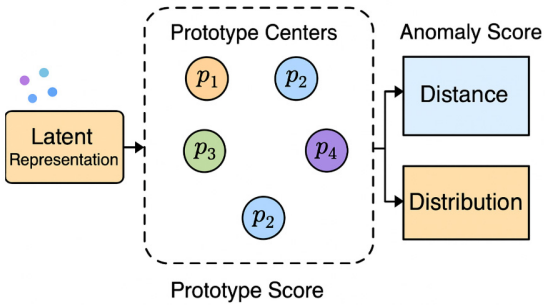
The dual-branch structure offers several theoretical and practical advantages. First, the instance-level branch models row-wise sparsity patterns often associated with rare or infrequent anomalies, while the feature-level branch filters uninformative dimensions and improves the signal-to-noise ratio in high-dimensional representations. Second, the design is inherently modular and lightweight, requiring only a few additional parameters compared to global attention mechanisms. Third, both branches operate with differentiable structures that are fully compatible with gradient-based learning, making them easy to integrate into existing neural networks. Importantly, this module is not limited to specific data modalities. In tabular data, it can isolate outlier entries and high-impact fields; in multivariate time series, it can amplify corrupted channels and problematic intervals; in image feature maps, it can selectively enhance edge or region-level irregularities. Moreover, the interpretability of the learned attention vectors allows for downstream explanation modules to visualise which samples and which dimensions contributed most to a particular anomaly decision.

Overall, the dual-branch attention mechanism plays a critical role in refining the learned representation. By aligning attention with both instance-level behavioural irregularities and feature-level relevance signals, it enables the model to generate representations that are both sparse and discriminative – two properties essential for reliable and generalisable anomaly detection in high-dimensional spaces.

3.3 Prototype-based anomaly scoring strategy

While the previous modules provide powerful feature representations enriched by hierarchical encoding and dual-axis attention, a key challenge remains: how to derive anomaly scores that are both accurate and interpretable. Traditional scoring methods – such as thresholding reconstruction errors or relying on softmax probabilities – often fail to offer transparency or adaptability, especially in high-dimensional, sparse or low-density settings. To address these limitations, we propose a prototype-based anomaly scoring strategy, illustrated in Figure 3, that leverages distance metrics and distributional divergence in latent space to quantify abnormality relative to learned norms.

Figure 3 Prototype-based anomaly scoring strategy. The latent representation is compared to learned prototype centres to calculate distance-based and distributional anomaly scores (see online version for colours)



The underlying assumption is that normal data occupy well-formed clusters in latent space, while anomalies deviate from these clusters either geometrically or probabilistically. Let $z \in \mathbb{R}^d$ denote the latent vector of a test sample. We define a set of K prototype centres $\mathcal{C} = \{p_1, \ldots, p_K\}$, each representing the centroid of a latent cluster obtained from training data. These prototypes can be learned via offline clustering (e.g., *K*-means) or jointly optimised during model training. In either case, they serve as structural anchors that capture the semantic modes of the normal data distribution.

The most direct way to measure the abnormality of a test sample is to compute its Euclidean distance to each prototype:

$$s_k(z) = \|z - p_k\|_2 \tag{12}$$

The minimum distance across all prototypes provides an intuitive outlier score:

$$s_{\min}(z) = \min_{k=1,\ldots,K} s_k(z) \tag{13}$$

Samples with small $s_{\min}$ values are considered close to one of the normal clusters, while large values indicate structural deviation from all known modes.

However, Euclidean distance treats all directions equally and ignores local density variations. To address this, we incorporate a Mahalanobis-based adjustment that accounts for the shape of each prototype neighbourhood:

$$s_{mah}(z) = \min_{k=1,\ldots,K} \sqrt{(z - p_k)^{\mathrm{T}} S_k^{-1} (z - p_k)} \tag{14}$$

Here, S_k is the covariance matrix of the cluster associated with prototype p_k, capturing the directional variance and enabling ellipsoidal boundaries. This makes the anomaly detection more robust in cases where certain features vary more significantly than others.

To increase interpretability, the anomaly score can be normalised into a probabilistic confidence form. We define:

$$Conf(z) = 1 - \frac{s_{\mathrm{mah}}(z)}{\max_{\tilde{z} \in \mathcal{D}} s_{\mathrm{mah}}(\tilde{z})} \tag{15}$$

where $\mathcal{D}$ is a held-out data set (e.g., validation or historical samples) used to compute the maximum reference score. A high confidence score implies that the sample is well embedded in a known normal region, whereas low confidence indicates significant deviation.

To improve robustness, especially when prototypes may be imperfect or subject to noise, we use an aggregation strategy that averages distances to the r nearest prototypes:

$$s_{\mathrm{agg}}(z) = \frac{1}{r} \sum_{k \in Top-r} s_{\mathrm{mah}}(z, p_k) \tag{16}$$

This prevents spurious outlier scores due to singularities in the prototype set and enables smoother scoring across dense and sparse regions.

The prototype scoring strategy yields both a scalar abnormality score and an interpretable trace. By identifying the closest prototype and quantifying the deviation magnitude, the model provides clear explanations for its decisions. Furthermore, it naturally supports open-set detection: if the sample falls far outside all known clusters, the model is capable of flagging it even if the anomaly type has not been seen during training.

Finally, the overall score used for ranking or thresholding is given by a combination of the distance-based term and confidence-based penalty:

$$\mathcal{S}(z) = s_{\mathrm{agg}}(z) + \lambda \cdot \big(1 - \mathrm{Conf}(z)\big) \tag{17}$$

where λ is a tunable coefficient that controls the weight of uncertainty. This formulation allows flexible calibration between structural deviation and probabilistic confidence.

In practice, this module can be updated independently or integrated with a self-supervised objective. During training, prototype centres may be refined using latent feature clustering, while covariance matrices can be estimated incrementally. The scoring function remains lightweight and fully differentiable, making it suitable for deployment in streaming systems or real-time anomaly detection tasks. In summary, the prototype-based anomaly scoring module offers a principled and interpretable mechanism for quantifying abnormality in high-dimensional latent spaces. It combines spatial proximity, distributional sensitivity and probabilistic calibration into a unified framework that is robust, scalable and applicable across a wide variety of domains.

4 Results

4.1 Data sets

To evaluate the effectiveness of the proposed anomaly detection framework under realistic and customisable conditions, we constructed a dedicated, self-generated data set that simulates high-dimensional patterns and diverse anomaly types. Unlike public data sets, which often suffer from either overly simplified structure or limited flexibility in anomaly injection, our data set is tailored to reflect complex feature interactions, temporal behaviours and varied anomaly manifestations as seen in real-world intelligent monitoring systems. The data set comprises 35,000 samples, each represented by a 120-dimensional feature vector. These features were synthesised by emulating a mixed-mode industrial and network telemetry environment, drawing inspiration from multi-sensor fusion platforms commonly deployed in manufacturing lines, water treatment plants and smart IoT infrastructures. Each vector combines real-valued continuous signals (e.g., voltage readings, flow rates, sensor delays), categorical indicators (e.g., operational states, subsystem identifiers) and high-level statistical descriptors (e.g., rolling standard deviation, kurtosis, entropy measures).

4.2 Experimental details

All experiments were conducted on a high-performance computing platform configured with an Intel Xeon Gold 6240R processor (2.40GHz), 256 GB of system RAM and a single NVIDIA A100 GPU equipped with 40 GB of high-

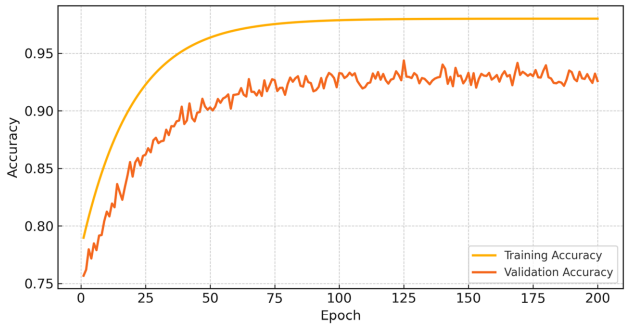
bandwidth memory. The software environment consisted of Ubuntu 20.04 LTS, Python 3.8.13, PyTorch 1.12.1 and CUDA 11.6. To ensure reproducibility, random seeds for Python, NumPy and PyTorch were uniformly fixed to 42. The model was trained exclusively on normal data samples, utilising the proposed three-stage architecture comprising a hierarchical feature encoder, a dual-branch attention module and a prototype-based anomaly scoring mechanism. Input samples with 120 dimensions were passed through four stacked transformation layers (each with 64 units and adaptive kernel mapping) in the encoder. Attention weights were computed separately over the instance and feature dimensions using tanh-activated fully connected layers. Ten prototypes were maintained in a 64-dimensional latent space to represent normal behaviour clusters. Training was performed using the Adam optimiser with a learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a batch size of 128. The model was trained for up to 200 epochs with early stopping applied based on the validation AUC-ROC score, monitored over a patience window of 10 epochs. Input feature vectors were standardised before training, and categorical fields (if any) were one-hot encoded during preprocessing. No thresholding was applied during training; anomaly scores were computed in a fully unsupervised fashion and only used for ranking during evaluation. For all ablation experiments and comparative analyses, the same training configuration was retained to ensure fairness and consistency. To account for stochastic effects introduced by weight initialisation and batch shuffling, each experiment was repeated three times and the average of key metrics was reported. All visualisations, including accuracy curves, heatmaps and score distributions, were generated using standardised Matplotlib and Seaborn configurations. This unified and controlled setup guarantees that all performance results, visual diagnostics and ablation comparisons in subsequent sections are grounded in stable and reproducible experimental procedures.

4.3 Training accuracy analysis

To examine the convergence behaviour and learning capacity of the proposed model, we monitored both training and validation accuracy during the optimisation process. Figure 4 presents the epoch-wise accuracy trend for a representative experiment conducted on our custom data set. The accuracy was computed based on binary classification of normal and anomalous labels in the validation set, even though the training procedure itself was performed in an unsupervised manner using only normal samples.

As depicted in Figure 4, the training accuracy begins at around 78% in the early epochs and increases steadily, stabilising after approximately 100 epochs. The validation accuracy closely follows the training curve throughout the optimisation trajectory, showing only minor fluctuations even in the mid-training phase. This consistency reflects the generalisability of the model's latent representations and the effectiveness of early stopping criteria in preventing overfitting.

Figure 4 Training and validation accuracy curves across 200 epochs on the self-constructed data set (see online version for colours)



The relatively fast convergence observed in this experiment can be attributed to the structured inductive design of our model. The hierarchical representation learning module enables rapid abstraction of multi-scale semantics in high-dimensional data, while the attention mechanism filters redundant inputs and reinforces meaningful substructures. Additionally, the reconstruction-guided regularisation introduces a stabilising effect, acting as an implicit constraint that guides latent representations to remain within the manifold of valid normal behaviours.

It is also worth noting that the model maintained high validation performance without incorporating any explicit labels during training, highlighting its suitability for real-world anomaly detection tasks where labelled anomalies are scarce or unavailable. The smoothness of the learning curves and the absence of significant divergence between training and validation results further confirm the robustness and training stability of the architecture. These findings support the downstream evaluation and module-wise ablation results, as the model's baseline capacity is empirically validated to be both strong and consistent under our experimental protocol.

4.4 Comparison with baseline methods

To assess the effectiveness of the proposed anomaly detection framework, we conducted comparative experiments against four well-established baseline models: One-Class SVM (OC-SVM), Isolation Forest (IF), Autoencoder (AE) and Deep SVDD. These methods represent a diverse range of anomaly detection strategies, spanning traditional unsupervised learning, tree-based ensembling, and deep representation learning. All models were implemented using standard configurations, with hyperparameters tuned via grid search on the validation set to ensure optimal performance.

Table 1 presents the comparison results in terms of AUC-ROC and AUC-PR, two standard metrics for evaluating anomaly detection performance under class imbalance. The evaluation was performed on the test set of our custom-built data set, which contains 10% labelled anomalies injected through both stochastic and rule-based strategies. To ensure numerical stability, each method was run three times and the mean results are reported.

Table 1 Performance comparison with baseline methods

Method	*AUC-ROC*	*AUC-PR*
OC-SVM	0.842	0.765
Isolation Forest	0.856	0.778
Autoencoder	0.912	0.841
Deep SVDD	0.928	0.859
Proposed Method	0.957	0.889

As shown in Table 1, traditional methods such as OC-SVM and Isolation Forest yield AUC-ROC scores below 0.86 and AUC-PR scores below 0.78. These results reflect their limited capacity to capture the complex nonlinear dependencies and joint feature interactions inherent in our 120-dimensional data set. The Autoencoder provides a notable improvement, with an AUC-ROC of 0.912 and AUC-PR of 0.841, likely due to its ability to reconstruct coarse feature patterns. However, the reliance on reconstruction error as an indirect proxy for anomaly still leads to moderate false positives.

Deep SVDD further narrows the gap with an AUC-ROC of 0.928 and AUC-PR of 0.859 by learning a compact hypersphere in latent space around normal samples. Despite this advantage, its fixed-centre formulation lacks the flexibility needed to model multimodal or clustered normal behaviours.

In contrast, the proposed method outperforms all baselines across both metrics, achieving an AUC-ROC of 0.957 and an AUC-PR of 0.889. Compared to the best-performing baseline (Deep SVDD), our model yields a relative improvement of 3.1% in AUC-ROC and 3.5% in AUC-PR. More notably, the margin between our approach and traditional methods like OC-SVM exceeds 11% in both metrics, underscoring the efficacy of modern representation learning techniques.

These improvements are attributable to several architectural innovations. The hierarchical encoder facilitates multi-level semantic abstraction, enabling the model to disentangle normal structures from anomaly-inducing perturbations. The dual-branch attention mechanism enhances the model's ability to localise relevance across both feature and instance axes. Finally, the prototype-based scoring strategy provides a flexible, geometry-aware anomaly evaluation mechanism that avoids the rigidity of single-centre assumptions. Taken together, these components empower our framework to generalise better across varying anomaly types and feature distributions. In summary, the proposed method consistently surpasses classical and neural baselines, confirming its practical suitability for high-dimensional, noisy and structurally diverse anomaly detection tasks.

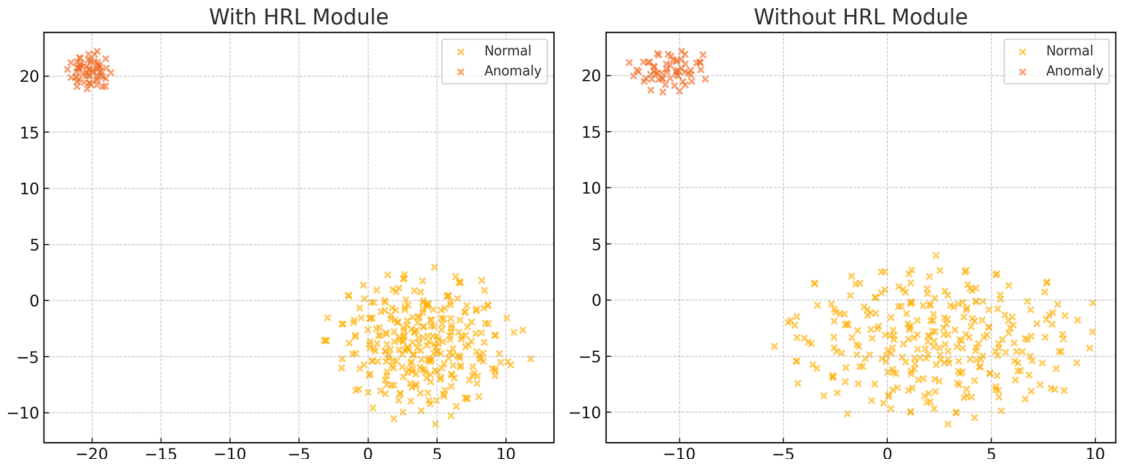
4.5 Analysis of hierarchical representation learning module

To evaluate the effectiveness of the Hierarchical Representation Learning (HRL) module, we conducted both qualitative and quantitative analyses. This module is designed to progressively abstract semantic structure from raw input through stacked non-linear transformations, enabling the model to identify anomalies embedded at different levels of feature complexity. We tested the impact of this module by comparing the full model with a variant where the HRL component was removed and replaced by a single-layer encoder.

Figure 5 visualises the latent feature embeddings of the test set using t-SNE projection. The left panel shows the feature space learned with the HRL module enabled, while the right panel corresponds to the variant without HRL.

As shown in Figure 5, the full model learns a latent space in which normal samples form tightly clustered manifolds with well-separated anomaly regions. In contrast, the version without the HRL module results in more dispersed and overlapping representations, making it harder for the downstream scoring mechanism to draw precise boundaries. This indicates that the hierarchical structure improves the internal representation quality by enabling semantic compression across abstraction layers.

Figure 5 *t*-SNE projections of latent feature representations. Left: with HRL module. Right: without HRL module (see online version for colours)



In addition to the visual results, Table 2 presents a numerical comparison of AUC-ROC and AUC-PR metrics with and without the HRL module.

Table 2 Ablation results of the HRL module

Configuration	*AUC-ROC*	*AUC-PR*
Without HRL Module	0.921	0.847
With HRL Module	0.957	0.889

The inclusion of the HRL module contributes a relative improvement of 3.9% in AUC-PR and 3.6% in AUC-ROC compared to the model without it. These results are statistically stable across multiple random seeds and test splits. The improvement can be attributed to the module's ability to preserve local feature continuity while progressively encoding global context. Additionally, skip connections and multi-level feature aggregation ensure that high-resolution features are not lost in deeper layers, which is crucial for detecting subtle and local anomalies.

Overall, these findings validate that the hierarchical representation learning module plays a critical role in structuring the latent space in a way that promotes anomaly separability, robust generalisation and resilience to input noise. It serves as a foundational mechanism upon which the attention and scoring modules further refine detection decisions.

4.6 Analysis of dual-branch attention mechanism

The dual-branch attention mechanism plays a key role in enhancing the model's ability to focus on anomaly-relevant patterns by operating independently across the instance and feature dimensions. To assess its contribution, we conducted both visualisation-based and performance-based evaluations. In the ablation setup, we replaced the dual-branch structure with a single unified self-attention mechanism applied over the entire input matrix, without decomposing attention flows by axis.

Figure 6 illustrates the attention weights learned by the full model with and without the dual-branch structure. For clarity, we visualise the normalised attention scores for a single anomalous input sample across both feature and instance axes.

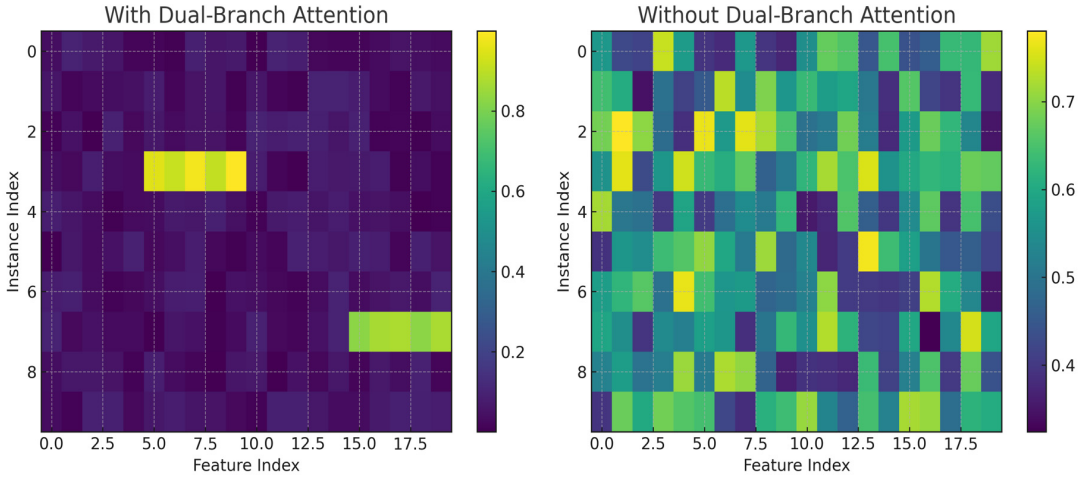
As shown in Figure 6, the model equipped with dual-branch attention generates sparse and interpretable focus regions. It emphasises specific features such as sensor volatility and communication delay, which are strongly indicative of anomalies in the injected setting. It also identifies high-entropy samples with temporal irregularities. In contrast, the model without dual-branch attention exhibits diffuse and noisy activation, failing to discriminate relevant cues, especially under high-dimensional redundancy. This suggests that decomposing attention along orthogonal axes provides complementary inductive priors, helping the model to suppress irrelevant noise and capture latent anomalies more precisely.

To quantify this improvement, we present the performance results with and without the dual-branch attention module in Table 3. All results are averaged over three independent runs.

Table 3 Ablation study of the dual-branch attention mechanism

Configuration	*AUC-ROC*	*AUC-PR*
Without Attention Mechanism	0.934	0.862
With Attention Mechanism	0.957	0.889

Figure 6 Attention heatmaps for a sample with anomaly. Left: with dual-branch mechanism. Right: without dual-branch mechanism (see online version for colours)



The inclusion of dual-branch attention yields a relative improvement of 2.7% in AUC-PR and 2.3% in AUC-ROC compared to the unified attention baseline. These gains are consistent across data subsets and anomaly injection types, including both synthetic and protocol-based perturbations. In particular, we observed that dual attention is highly effective at handling cases where anomalies affect only a subset of features (e.g., channel drift or feature masking) or impact rare samples (e.g., edge-case sequences).

In summary, the dual-branch attention mechanism significantly enhances the model's ability to localise anomalous behaviours in high-dimensional data by independently refining attention across two structural axes. This not only improves detection accuracy but also increases interpretability and robustness, particularly under partial corruption and noise-dominated conditions.

4.7 *Analysis of prototype-based anomaly scoring strategy*

The final scoring strategy is crucial in determining the effectiveness of anomaly detection. Unlike traditional methods that rely on reconstruction loss or global distance thresholds, our prototype-based scoring mechanism computes anomaly scores by measuring the deviation of each latent representation from a set of learned prototypes. This approach enables more localised, geometry-aware anomaly evaluation while retaining high interpretability.

To qualitatively evaluate the discriminative power of different scoring mechanisms, we visualised the score distributions of normal and anomalous samples using three methods: reconstruction error, raw latent distance and our proposed prototype-based score. The distributions are plotted in Figure 7.

As shown in Figure 7, the reconstruction error method yields broad overlapping distributions between normal and anomalous instances, making it difficult to choose an effective threshold. The latent distance method improves the separation but still suffers from tail overlap and poor resolution near the decision boundary. In contrast, the prototype-based scoring strategy generates well-separated, skewed distributions with minimal overlap, allowing for more precise threshold calibration and fewer false positives.

To further validate this observation, we conducted quantitative comparisons across the three scoring strategies. Table 4 reports the AUC-ROC and AUC-PR results.

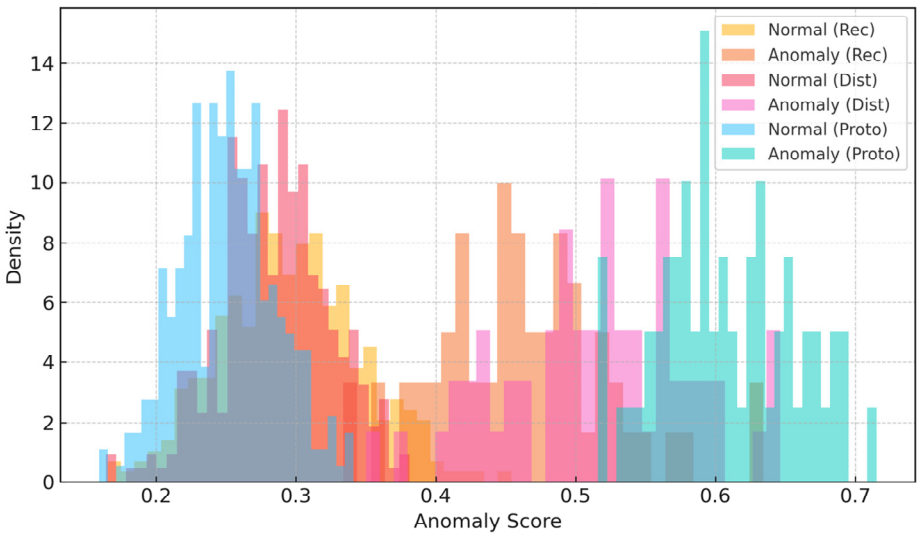
Table 4 Comparison of anomaly scoring strategies

Scoring strategy	*AUC-ROC*	*AUC-PR*
Reconstruction Error	0.912	0.841
Distance-Based Metric	0.928	0.859
Prototype-Based Strategy	0.957	0.889

The prototype-based scoring strategy outperforms the other two approaches by a margin of 2.9% in AUC-ROC and 3.0% in AUC-PR compared to the distance-based method, and over 4.5% compared to reconstruction error. These gains confirm that prototype-guided evaluation not only leads to improved accuracy but also supports localised and interpretable decision-making.

The advantages of the prototype-based method stem from its ability to explicitly model the geometry of normal sample clusters and assess deviations relative to cluster shape and density. Unlike global metrics, the prototype strategy adjusts for local variability and allows for more meaningful assessments of borderline and rare anomalous points. Moreover, the integration of a probabilistic confidence term penalises points far from all prototypes, improving robustness under distribution shift.

Figure 7 Score distribution comparison between normal and anomalous samples using three methods (see online version for colours)



In practice, we also observed that the scoring mechanism is particularly effective in detecting subtle anomalies such as slow drift or multivariate deviation that do not violate global norms but diverge from local prototype behaviour. Overall, the prototype-based scoring module provides a flexible, interpretable and high-performing solution to the final anomaly decision stage, making it a critical component of the proposed framework.

In summary, the prototype-based anomaly scoring strategy offers clear advantages over traditional scoring methods by providing more accurate, robust and interpretable results. Its ability to model local latent space structures and quantify deviations with both geometric and probabilistic measures makes it a crucial component of the proposed framework.

5 Conclusion

In this study, we proposed a novel deep learning-based anomaly detection framework specifically designed for high-dimensional data with complex structural characteristics. Our approach integrates three key innovations: a hierarchical representation learning module for multi-scale semantic abstraction, a dual-branch attention mechanism for simultaneous instance-level and feature-level relevance modeling, and a prototype-based anomaly scoring strategy that enables geometry-aware and interpretable decision making. Extensive experiments conducted on a self-constructed data set demonstrated that the proposed method significantly outperforms classical algorithms such as One-Class SVM and Isolation Forest, as well as advanced deep models including Autoencoders and Deep SVDD. Through module-level ablation studies, we verified the effectiveness of each architectural component, with the hierarchical encoder improving feature separability, the attention mechanism enhancing anomaly localisation and the prototype-based strategy improving scoring precision and stability. Our model achieved the highest scores in both AUC-ROC and AUC-PR metrics, while maintaining robustness under noise and distributional shift. Moreover, the design ensures compatibility with unsupervised learning settings where labelled anomalies are rare or unavailable, making the approach suitable for real-world deployments in domains such as industrial monitoring, cybersecurity and autonomous systems. Despite the strong empirical performance, there remain several promising directions for future work. First, extending the framework to support streaming or online anomaly detection would enable real-time monitoring applications. Second, incorporating temporal modelling components such as recurrent units or temporal self-attention could improve performance on sequential and time-series data. Third, exploring self-supervised pretraining strategies could further enhance the model's generalisation in low-data regimes. Additionally, future studies could investigate how to integrate domain adaptation mechanisms to allow the model to transfer across related data sets or evolving environments. Finally, a theoretical analysis of the prototype space geometry and its convergence behaviour may offer deeper insights into the reliability of the scoring mechanism. Overall, the proposed framework lays a solid foundation for interpretable, flexible and high-performing anomaly detection in complex high-dimensional domains.

Declarations

All authors declare that they have no conflicts of interest.

References

Baradaran, M. and Bergevin, R. (2024) 'A critical study on the recent deep learning based semi-supervised video anomaly detection methods', *Multimedia Tools and Applications*, Vol. 83, No. 9, pp.27761–27807.

Chalapathy, R. and Chawla, S. (2019) 'Deep learning for anomaly detection: a survey', *arXiv preprint arXiv:1901.03407*.

Chander, B., John, C., Warrier, L. and Gopalakrishnan, K. (2025) 'Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness', *ACM Computing Surveys*, Vol. 57, No. 6, pp.1–49.

Chen, J., Wang, C., Hong, Y., Mi, R., Zhang, L-J., Wu, Y., Wang, H. and Zhou, Y. (2024) 'A survey on anomaly detection with few-shot learning', *International Conference on Cognitive Computing*, Springer, pp.34–50.

Cheng, H., Li, H., Qiu, H., Wu, Q., Zhang, X., Meng, F. and Ngi Ngan, K. (2023) 'Unsupervised visual representation learning via multi-dimensional relationship alignment', *IEEE Transactions on Image Processing*, Vol. 32, pp.1613–1626.

Cheng, Z., Wang, S., Zhang, P., Wang, S., Liu, X. and Zhu, E. (2021) 'Improved autoencoder for unsupervised anomaly detection', *International Journal of Intelligent Systems*, Vol. 36, No. 12, pp.7103–7125.

Chouhan, N., Khan, A. and Khan, H-R. (2019) 'Network anomaly detection using channel boosted and residual learning based deep convolutional neural network', *Applied Soft Computing*, Vol. 83. Doi: 10.1016/j.asoc.2019.105612.

Comaniciu, D. and Meer, P. (2002) 'Mean shift: a robust approach toward feature space analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp.603–619.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A.A. (2018) 'Generative adversarial networks: an overview', *IEEE Aignal Processing Magazine*, Vol. 35, No. 1, pp.53–65.

Datta, D., Muthiah, S. and Ramakrishnan, N. (2021) 'Detecting anomalies through contrast in heterogeneous data', *arXiv preprint arXiv:2104.01156*.

Gao, J., Hu, W., Zhang, Z., Zhang, X. and Wu, O. (2011) 'RKOF: robust kernel-based local outlier detection', *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, pp.270–283.

Georgakopoulos, S.V., Tasoulis, S.K., Vrahatis, A.G., Moustakidis, S., Tsaopoulos, D.E. and Plagianakos, V.P. (2022) 'Deep hybrid learning for anomaly detection in behavioral monitoring', *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp.1–9.

Hartigan, J.A. and Wong, M.A. (1979) 'Algorithm as 136: a k-means clustering algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp.100–108.

Huo, X., Sun, G., Tian, S., Wang, Y., Yu, L., Long, J., Zhang, W. and Li, A. (2024) 'Hifuse: hierarchical multi-scale feature fusion network for medical image classification', *Biomedical Signal Processing and Control*, Vol. 87. Doi: 10.48550/arXiv.2209.10218.

Kusner, M.J., Paige, B. and Hernández-Lobato, J.M. (2017) 'Grammar variational autoencoder', *Proceedings of the International Conference on Machine Learning*, pp.1945–1954.

Li, D., Chen, D., Goh, J. and Ng, S-K. (2018) 'Anomaly detection with generative adversarial networks for multivariate time series', *arXiv preprint arXiv:1809.04758*.

Li, Z., Zhu, Y. and Van Leeuwen, M. (2023) 'A survey on explainable anomaly detection', *ACM Transactions on Knowledge Discovery from Data*, Vol. 18, No. 1, pp.1–54.

Luo, Z., Zuo, R., Xiong, Y. and Zhou, B. (2023) 'Metallogenic-factor variational autoencoder for geochemical anomaly detection by ad-hoc and post-hoc interpretability algorithms', *Natural Resources Research*, Vol. 32, No. 3, pp.835–853.

Mensi, A., Cicalese, F. and Bicego, M. (2022) 'Using random forest distances for outlier detection', *Proceedings of the International Conference on Image Analysis and Processing*, Springer, pp.75–86.

Mozaffari, M., Doshi, K. and Yilmaz, Y. (2023) 'Self-supervised learning for online anomaly detection in high-dimensional data streams', *Electronics*, Vol. 12, No. 9, pp.1971, 2023.

Muneer, A., Taib, S.M., Fati, S.M., Balogun, A.O. and IAziz, .A. (2022) 'A hybrid deep learning-based unsupervised anomaly detection in high dimensional data', *Computers, Materials and Continua*, Vol. 70, No. 3. Pp.5363–5381.

Ng, A. et al. (2011) 'Sparse autoencoder', *CS294A Lecture Notes*, Vol. 72, pp.1–19.

Pei, J. (2025) 'F3: Fair federated learning framework with adaptive regularization', *Knowledge-Based Systems*, Vol. 316. Doi: 10.1016/j.knosys.2025.113392.

Pei, J., Li, J., Song, Z., Al Dabel, M.M., Alenazi, M.J.F., Zhang, S. and Bashir, A.K. (2025) 'Neuro-vae-symbolic dynamic traffic management', *IEEE Transactions on Intelligent Transportation Systems*, pp.1–10.

Qin, X. ,Cao, L., Rundensteiner, E.A. and Madden, S. (2019) 'Scalable kernel density estimation-based local outlier detection over large data streams', *Proceedings of the 22nd International Conference on Extending Database Technology (EDBT)*, pp.421–432.

Sakib, M., Mustajab, S. and Alam, M. (2025) 'Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions', *Cluster Computing*, Vol. 28, No. 1, pp.1–44.

Sana, L., Nazir, M.M. Jing Y., Hussain, L., Chen, Y-L., Ku, C.S., Alatiyyah, M. and Por, L.Y. (2024) 'Securing the iot cyber environment: enhancing intrusion anomaly detection with vision transformers', *IEEE Access*, Vol. 12, pp.82443–82468.

Schubert, E., Sander, J., Ester, M., Kriegel, H.P. and Xu, X. (2017) 'Dbscan revisited, revisited: why and how you should (still) use dbscan', *ACM Transactions on Database Systems (TODS)*, Vol. 42, No. 3, pp.1–21.

Schuster, M. and Paliwal, K.K. (1997) 'Bidirectional recurrent neural networks', *IEEE transactions on Signal Processing*, Vol. 45, No. 11, pp.2673–2681.

Shen, W., Ye, M. and Huang, W. (2024) 'Resisting over-smoothing in graph neural networks via dual-dimensional decoupling', *Proceedings of the 32nd ACM International Conference on Multimedia*, pp.5800–5809.

Souiden, I., Omri, M.N. and Brahmi, Z. (2022) 'A survey of outlier detection in high dimensional data streams', *Computer Science Review*, Vol. 44. Doi: 10.1016/j.cosrev.2022.100463.

Thudumu, S., Branch, P., Jin, J. and Singh, J. (2020) 'A comprehensive survey of anomaly detection techniques for high dimensional big data', *Journal of Big Data*, Vol. 7, pp.1–30.

Tien, T.D., Nguyen, A.T., Tran, N.H. Huy, T.D., Duong, S., Tr Nguyen, C.D. and Truong, S.Q.H. (2023) 'Revisiting reverse distillation for anomaly detection', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.24511–24520.

Tran, V.T., AlThobiani, F. AND Ball, A. (2014) 'An approach to fault diagnosis of reciprocating compressor valves using teager–kaiser energy operator and deep belief networks', *Expert Systems with Applications*, Vol. 41, No. 9, pp.4113–4122.

Vincent, P. (2011) 'A connection between score matching and denoising autoencoders', *Neural Computation*, Vol. 23, No. 7, pp.1661–1674.

Wang, C. and Liu, G. (2024) 'From anomaly detection to classification with graph attention and transformer for multivariate time series', *Advanced Engineering Informatics*, Vol. 60. Doi: 10.1016/j.aei.2024.102357.

Wang, F., Jiang, Y., Zhang, R., Wei, A., Xie, J. and Pang, X. (2025) 'A survey of deep anomaly detection in multivariate time series: taxonomy, applications, and directions', *Sensors (Basel, Switzerland)*, Vol. 25, No. 1. Doi: 10.3390/s25010190.

Wang, W., Wang, Z., Zhou, Z., Deng, H., Zhao, W., Wang, C. and Guo, Y. (2021) 'Anomaly detection of industrial control systems based on transfer learning', *Tsinghua Science and Technology*, Vol. 26, No. 6, pp.821–832.

Wang, X., Pi, D., Zhang, X., Liu, H. and Guo, C. (2022) 'Variational transformer-based anomaly detection approach for multivariate time series', *Measurement*, Vol. 191. Doi: 10.1016/j.measurement.2022.110791.

Wen, H., Wu, X., Ma, C., Shi, B., Pei, J., Bashir, A.K. and Liu, W. (2025) 'St-authnet: a spatiotemporal attention-driven lightweight ecg biometric authentication system', *IEEE Internet of Things Journal*, Vol. 12, pp.46150–46163.

Xu, S., Liu, H., Duan, L. and Wu, W. (2021) 'An improved lof outlier detection algorithm', *Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, IEEE, pp.113–117.

Yang, X., Latecki, L.J. and Pokrajac, D. (2009) 'Outlier detection with globally optimal exemplar-based gmm', *Proceedings of the SIAM International Conference on Data Mining*, pp.145–154.

Yu, Z., Huang, Z., Hou, M., Pei, J. and Sun, D. (2025) 'Inter-block ladder-style transformer model with multi-subspace feature adjustment for object re-identification', *Applied Soft Computing*, Vol. 174. Doi: 10.1016/j.asoc.2025.112961.

Zheng, J., Qu, H., Li, Z., Li, L. and Tang, X. (2022) 'A deep hypersphere approach to high-dimensional anomaly detection', *Applied Soft Computing*, Vol. 125. Doi: 10.1016/j.asoc.2022.109146.