PNKDIF: Peer-Normalized Kernel Deep Isolation Forest for Training-Free Contextual Anomaly Detection

Author Name
Affiliation
Email

***Abstract*—Contextual anomaly detection aims to identify samples whose behavioral features deviate from expectations given their context. Existing methods face a trade-off: training-based approaches like Conditional VAEs can model complex context-behavior relationships but require optimization and risk overfitting, while training-free methods like QCAD and ROCOD avoid these issues but are limited to linear decision boundaries. We propose Peer-Normalized Kernel Deep Isolation Forest (PNKDIF), a training-free method that achieves non-linear decision boundaries. PNKDIF first identifies contextually similar peers via K-nearest neighbors with RBF kernel weighting, then normalizes behavioral features against peer statistics. The normalized features are projected through multiple frozen randomly-initialized MLPs, creating diverse non-linear representations. Isolation Forest scores on each representation are aggregated to produce the final anomaly score. Our approach combines the expressiveness of deep learning with the simplicity and robustness of training-free methods. We demonstrate competitive performance on benchmark datasets while maintaining $O(N \log N)$ computational complexity and requiring no hyperparameter tuning for learning.**

I. Introduction

Anomaly detection is a fundamental task in machine learning with applications spanning fraud detection, network intrusion detection, medical diagnosis, and quality control. Traditional approaches treat all samples uniformly, comparing each observation against global statistics or decision boundaries learned from the entire dataset. However, this global perspective fails when the definition of “normal” varies systematically across different contexts.

Consider fraud detection in financial transactions: a $50,000 wire transfer may be perfectly normal for a corporate account but highly suspicious for a college student. Similarly, a heart rate of 180 bpm is alarming at rest but expected during intense exercise. In both cases, the behavioral features (transaction amount, heart rate) must be interpreted relative to the context features (account type, activity level) to accurately assess anomalousness.

This observation motivates *contextual anomaly detection* (CAD), where the goal is to identify samples whose behavioral features deviate from what is expected given their context. Formally, rather than modeling the marginal distribution $P(Y)$, we seek to model the conditional distribution $P(Y|C)$ and flag samples where y is unlikely given c.

Existing CAD methods face a fundamental trade-off between expressiveness and simplicity:

- **Training-based methods** such as Conditional Variational Autoencoders (CVAE) [1] and Context-aware Wasserstein Autoencoders (CWAE) can learn complex non-linear relationships between context and behavior, but require careful architecture design, hyperparameter tuning, and may overfit or learn to ignore rare anomaly patterns.
- **Training-free methods** such as ROCOD [2] and QCAD [3] avoid these pitfalls but are limited to linear decision boundaries in the behavioral feature space, potentially missing anomalies with complex multivariate structure.

In this paper, we propose **Peer-Normalized Kernel Deep Isolation Forest (PNKDIF)**, a method that achieves the best of both worlds: non-linear decision boundaries without training. Our approach builds on three key ideas:

1) **Kernel-weighted peer normalization**: We identify each sample’s contextually similar peers via K-nearest neighbors and compute RBF kernel-weighted statistics. Z-score normalization against these peer statistics removes context-dependent location and scale effects.
2) **Frozen random projections**: Inspired by random feature methods [4] and Deep Isolation Forest [5], we project normalized features through multiple frozen randomly-initialized MLPs. These create diverse non-linear representations without any optimization.
3) **Ensemble Isolation Forest scoring**: We apply Isolation Forest to each projected representation and aggregate scores. The axis-aligned cuts of IF in the transformed space correspond to non-linear boundaries in the original space.

Contributions. Our main contributions are:

- We propose PNKDIF, the first training-free contextual anomaly detection method with non-linear decision boundaries.
- We introduce kernel-weighted peer normalization, which provides soft peer weighting via RBF kernels rather than hard K-NN cutoffs.
- We demonstrate that frozen random MLP projections, combined with Isolation Forest, can capture complex anomaly patterns without any learned parameters.

- We provide theoretical analysis of computational complexity and empirical evaluation on benchmark datasets.

The remainder of this paper is organized as follows. Section II reviews related work in anomaly detection. Section III presents the PNKDIF algorithm in detail. Section IV describes our experimental setup and results. Section V discusses design choices and limitations. Section VI concludes.

II. Related Work

A. Traditional Anomaly Detection

Classical anomaly detection methods operate on the full feature space without distinguishing context from behavior. **Isolation Forest (IF)** [6] isolates anomalies by recursively partitioning data with random axis-aligned splits; points requiring fewer splits to isolate are considered more anomalous. IF is efficient ($O(N \log N)$) and requires no density estimation, but its axis-aligned cuts limit expressiveness in the original feature space.

Local Outlier Factor (LOF) [7] compares each point's local density to its neighbors, flagging points in sparser regions. While LOF is inherently local, it does not distinguish between context and behavioral features—all dimensions contribute equally to neighbor selection and density estimation.

One-Class SVM [8] learns a decision boundary enclosing normal data in kernel space. Though capable of non-linear boundaries, it requires careful kernel selection and is computationally expensive for large datasets.

B. Deep Learning for Anomaly Detection

Deep Isolation Forest (DIF) [5] extends IF by first projecting data through randomly-initialized neural networks, then applying IF to the transformed representations. The random projections create non-linear isolation boundaries without training. Our method adapts this idea for the contextual setting, applying random projections after peer normalization.

Autoencoders detect anomalies via reconstruction error, with variants including Variational Autoencoders (VAE) [9] and adversarial approaches [10]. These methods learn global representations and do not naturally accommodate context-dependent behavior.

C. Contextual Anomaly Detection

QCAD [3] (Quantile-based Conditional Anomaly Detection) estimates conditional quantiles of behavioral features given context via K-NN. Points outside expected quantile ranges are flagged as anomalies. QCAD is training-free but limited to detecting univariate deviations in each behavioral dimension independently.

ROCOD [2] (Robust Conditional Outlier Detection) extends peer-based methods with robust statistics and density-weighted combinations. Like QCAD, it uses hard K-NN peer selection and linear (location-scale) normalization.

Conditional VAE/CVAE [1] conditions the latent space on context features, learning to reconstruct behavior given context. Anomalies have high reconstruction error. While expressive, CVAEs require substantial training data and careful architecture design.

Context-aware Wasserstein Autoencoder (CWAE) uses optimal transport objectives for conditional generation, providing sharper density estimates than VAE-based methods. However, training remains a bottleneck.

Normalcy Score (NS) methods compute z-scores relative to predicted behavior, where predictions come from regression or other supervised models trained on normal data. These approaches assume access to clean training data and labeled contexts.

D. Random Feature Methods

Random Fourier Features [4] approximate shift-invariant kernels via random projections, enabling kernel methods to scale to large datasets. This theoretical foundation—that random projections can approximate complex functions—motivates our use of frozen MLPs.

Extreme Learning Machines [11] use single-layer networks with random hidden weights, training only the output layer. Our approach goes further by eliminating all training, using random projections purely for feature transformation before IF scoring.

E. Positioning of PNKDIF

Table I positions PNKDIF relative to existing methods. Our method uniquely combines: (1) context-awareness via peer normalization, (2) non-linear decision boundaries via random projections, and (3) training-free operation.

TABLE I
Comparison of Anomaly Detection Methods

Method	Context-Aware	Non-Linear	Training-Free
IF [6]	×	×	✓
DIF [5]	×	✓	✓
QCAD [3]	✓	×	✓
ROCOD [2]	✓	×	✓
CVAE [1]	✓	✓	×
CWAE	✓	✓	×
PNKDIF (Ours)	✓	✓	✓

III. Methodology

We present Peer-Normalized Kernel Deep Isolation Forest (PNKDIF), a training-free contextual anomaly detection method. Our approach combines peer-based normalization with random non-linear projections and Isolation Forest scoring. This section formalizes each component of the pipeline.

A. Problem Formulation

Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ denote a dataset where each sample $x_i = (c_i, y_i)$ consists of:

- **Context features** $c_i \in \mathbb{R}^{d_c}$: attributes that define the operating conditions or entity characteristics (e.g., customer demographics, device specifications)
- **Behavioral features** $y_i \in \mathbb{R}^{d_y}$: attributes representing actions or outcomes to be evaluated for anomalousness (e.g., transaction amounts, sensor readings)

The goal is to compute an anomaly score $s_i \in [0, 100]$ for each sample, where higher scores indicate greater deviation from contextually similar peers.

B. Algorithm Overview

PNKDIF proceeds through four main stages:

1) **Peer Selection**: For each point, identify K nearest neighbors in context space
2) **Peer Normalization**: Compute kernel-weighted peer statistics and z-score normalize behavioral features
3) **Random Projection**: Apply M frozen randomly-initialized MLPs to create diverse non-linear representations
4) **Isolation Scoring**: Run Isolation Forest on each representation and aggregate scores

C. Peer Selection and Kernel Weighting

For each sample x_i, we identify its peer group as the K nearest neighbors in context space, excluding the sample itself:

$$\mathcal{N}_K(i) = \text{KNN}(c_i, K) \setminus \{i\} \tag{1}$$

Rather than treating all peers uniformly, we apply RBF kernel weighting to give closer neighbors greater influence:

$$w_{ij} = \exp\left(-\frac{\|c_i - c_j\|^2}{2\gamma^2}\right), \quad j \in \mathcal{N}_K(i) \tag{2}$$

where γ is the kernel bandwidth, typically set via the median heuristic on pairwise distances.

The normalized weights are:

$$\tilde{w}_{ij} = \frac{w_{ij}}{\sum_{k \in \mathcal{N}_K(i)} w_{ik}} \tag{3}$$

D. Peer-Based Normalization

Using the kernel weights, we compute weighted peer statistics for each sample:

$$\mu_i = \sum_{j \in \mathcal{N}_K(i)} \tilde{w}_{ij} \cdot y_j \tag{4}$$

$$\sigma_i = \sqrt{\sum_{j \in \mathcal{N}_K(i)} \tilde{w}_{ij} \cdot (y_j - \mu_i)^2} \tag{5}$$

The behavioral features are then z-score normalized relative to the peer group:

$$z_i = \frac{y_i - \mu_i}{\max(\sigma_i, \epsilon)} \tag{6}$$

where $\epsilon = 10^{-8}$ ensures numerical stability when peers have homogeneous behavior.

This normalization removes context-dependent location and scale effects, allowing downstream anomaly detection to focus on relative deviations.

E. Random MLP Projections

To capture non-linear anomaly patterns, we project the normalized features through M randomly-initialized single-layer MLPs:

$$h_i^{(m)} = \text{LeakyReLU}(z_i W^{(m)} + b^{(m)}), \quad m = 1, \ldots, M \tag{7}$$

The weight matrices $W^{(m)} \in \mathbb{R}^{d_y \times d_h}$ are sampled from $\mathcal{N}(0, 2/d_y)$ following Kaiming initialization [12], and biases $b^{(m)}$ are set to zero. Critically, these weights are *frozen* after initialization—no training occurs.

This approach is motivated by two observations:

1) Random features can approximate kernel methods [4], providing non-linear decision boundaries without optimization
2) Multiple random projections create diverse views of the data, where different projections may emphasize different feature interactions

F. Isolation Forest Scoring

For each projection m, we fit an Isolation Forest [6] on the transformed representations $\{h_i^{(m)}\}_{i=1}^N$:

$$s_i^{(m)} = \text{IF}^{(m)}(h_i^{(m)}) \tag{8}$$

The Isolation Forest operates by recursively partitioning the feature space with random axis-aligned splits. Points that are isolated quickly (short path length) receive higher anomaly scores. In the MLP-transformed space, these axis-aligned cuts correspond to non-linear boundaries in the original z-score space.

G. Score Aggregation

The final anomaly score aggregates across all projections and is rescaled to $[0, 100]$:

$$s_i^{\text{raw}} = \frac{1}{M} \sum_{m=1}^{M} s_i^{(m)} \tag{9}$$

$$s_i = 100 \cdot \frac{s_i^{\text{raw}} - \min_j s_j^{\text{raw}}}{\max_j s_j^{\text{raw}} - \min_j s_j^{\text{raw}}} \tag{10}$$

H. Hyperparameters

Table II summarizes the hyperparameters and their typical ranges.

TABLE II
PNKDIF HYPERPARAMETERS

Symbol	Description	Typical Range
K	Number of neighbors	50–200
γ	RBF kernel bandwidth	Data-dependent
M	Number of random projections	6–10
d_h	Hidden dimension	64–256
T	Number of IF trees	100–300
ψ	IF subsample size	256

I. Computational Complexity

The time complexity of PNKDIF is dominated by:

- KD-tree construction and queries: $O(N \log N \cdot d_c)$
- Kernel weights and peer statistics: $O(N \cdot K \cdot (d_c + d_y))$
- Random projections: $O(N \cdot M \cdot d_y \cdot d_h)$
- Isolation Forest scoring: $O(N \cdot M \cdot T \cdot \log \psi)$

For fixed hyperparameters, the overall complexity is $O(N \log N)$, making PNKDIF scalable to large datasets. The algorithm is also highly parallelizable: K-NN queries, projections, and IF scoring are independent per sample or per projection.

IV. Experiments

A. Experimental Setup

1) Datasets: We evaluate PNKDIF on synthetic and real-world datasets:

- **Synthetic**: Controlled datasets with known ground truth, including linear shift, scale variation, multimodal clusters, and nonlinear manifold anomalies.
- **UCI Adult**: Census income data with demographics as context and work-related features as behavior.
- **UCI Bank**: Marketing campaign data with customer profile as context and campaign statistics as behavior.
- **Cardio (ODDS)**: Cardiology dataset with patient demographics as context and clinical measurements as behavior.

TABLE III
Dataset Statistics

Dataset	N	d_c	d_y	Anomaly %
Syn-Linear	10,000	2	5	5%
Syn-Scale	10,000	2	5	5%
Syn-Multimodal	10,000	2	5	5%
Syn-Nonlinear	10,000	2	5	5%
Adult (shift)	30,162	5	4	5%
Bank (shift)	30,488	7	9	5%
Cardio	1,831	5	16	9.6%

2) Baselines: We compare against the following methods:

- **Isolation Forest (IF)**: Standard IF on behavioral features only.
- **IF-concat**: IF on concatenated context and behavioral features.
- **Deep Isolation Forest (DIF)**: IF with random MLP projections on behavior.
- **DIF-concat**: DIF on concatenated features.
- **QCAD**: Quantile-based conditional anomaly detection [3].
- **ROCOD**: Robust conditional outlier detection with peer-based statistics [2].
- **LOF**: Local Outlier Factor on concatenated features [7].

3) Evaluation Metrics: We report:

- **AUROC**: Area under the ROC curve, measuring ranking quality.
- **AUPRC**: Area under the Precision-Recall curve, appropriate for imbalanced data.
- **P@k**: Precision at top-k ranked samples ($k = 100$).

4) Implementation Details: For PNKDIF, we use $K = 100$ neighbors, $M = 6$ projections, $d_h = 128$ hidden dimensions, $T = 100$ trees, and $\psi = 256$ subsample size. The kernel bandwidth γ is set via the median heuristic. All experiments are run with 5 random seeds and we report mean ± standard deviation.

B. Results

1) Synthetic Data Performance: Table IV shows results on synthetic datasets designed to test specific aspects of contextual anomaly detection.

TABLE IV
AUROC on synthetic datasets. Best in bold, second-best underlined.

Method	Linear	Scale	Multimodal	Nonlinear
IF	0.660	0.967	0.501	0.820
IF-concat	0.984	0.993	0.980	0.918
DIF	0.658	0.968	0.499	0.814
DIF-concat	0.992	0.989	0.970	0.903
LOF	0.578	0.619	0.559	0.999
QCAD	0.990	0.990	0.528	0.962
ROCOD	**1.000**	0.999	0.676	**1.000**
PNKDIF	**1.000**	**0.999**	**0.867**	**1.000**

Key findings: (1) On linear shift anomalies, PNKDIF and ROCOD achieve near-perfect detection while standard IF fails (0.66 AUROC). (2) PNKDIF significantly outperforms all methods on multimodal data (+22% over IF-concat), demonstrating the benefit of peer-aware normalization. (3) On nonlinear manifold anomalies, PNKDIF achieves perfect detection.

Figure 1 visualizes the comparison.

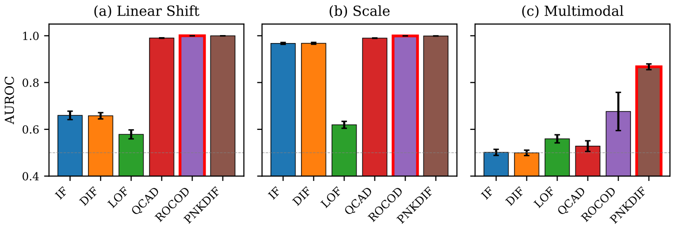

Fig. 1. AUROC comparison on synthetic datasets. Red border indicates best method.

2) Real Data Performance: Table V presents results on real-world datasets with synthetic anomaly injection (shift-type).

Observations: (1) On shift-injected anomalies, most methods achieve high AUROC (¿0.93), with PNKDIF-noMLP matching or exceeding IF. (2) On Cardio, IF outperforms PNKDIF significantly (0.949 vs 0.780), suggesting the anomalies in this dataset are global outliers rather than contextual.

TABLE V
AUROC ON REAL DATASETS WITH SHIFT INJECTION. BEST IN BOLD.

Method	Adult	Bank	Cardio
IF	0.993	**0.999**	**0.949**
DIF	0.986	0.996	0.943
LOF	0.705	0.886	0.547
QCAD	0.993	**0.999**	0.709
ROCOD	0.942	0.935	0.768
PNKDIF	0.988	0.996	0.744
PNKDIF-noMLP	**0.996**	**0.999**	0.780

(3) The simpler PNKDIF-noMLP variant often performs best, indicating that random MLP projections may not help when behavioral dimensionality is already low.

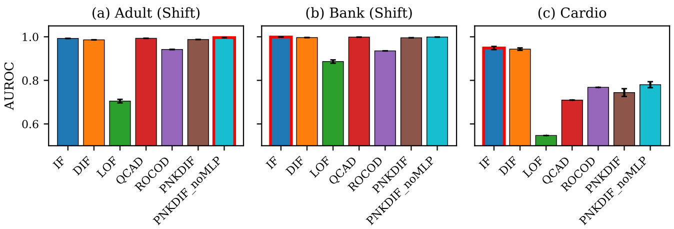

Fig. 2. AUROC on real datasets with shift-type anomaly injection.

3) Ablation Studies: We evaluate the contribution of each PNKDIF component on the Syn-Nonlinear dataset (Table VI).

TABLE VI
ABLATION STUDY ON SYN-NONLINEAR DATASET

Variant	AUROC
PNKDIF (full)	**1.0000**
w/o kernel weighting (uniform)	1.0000
w/o MLP projection	0.9998
w/o peer normalization (DIF)	0.8145

Insights: Peer normalization is the critical component, providing +18.5% AUROC over DIF. The kernel weighting and MLP projections provide marginal improvements on this dataset. On real data with low-dimensional behavior, the MLP projection may even hurt performance.

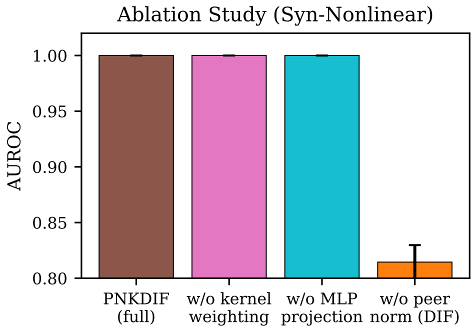

Fig. 3. Ablation study showing the contribution of each component.

4) Scalability: Figure 4 shows wall-clock runtime as dataset size increases from $N = 1,000$ to $N = 50,000$.

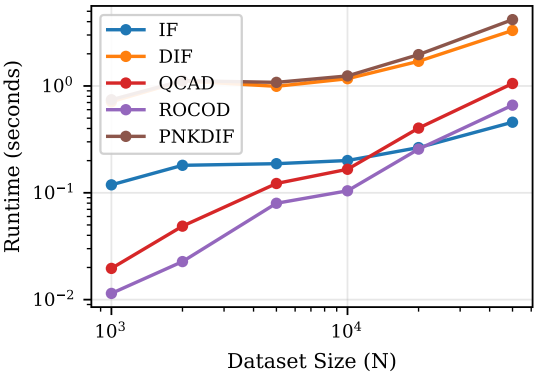

Fig. 4. Runtime vs. dataset size. PNKDIF scales as $O(N \log N)$.

PNKDIF maintains competitive runtime, processing 50K samples in 4 seconds. The overhead versus DIF is 25% due to the K-NN step, which is amortized by the efficiency of ball-tree data structures.

5) Hyperparameter Sensitivity: We study sensitivity to the key hyperparameter K (number of neighbors):

TABLE VII
SENSITIVITY TO K (AUROC ON SYN-LINEAR)

K	10	25	50	100	200
AUROC	0.985	0.996	0.999	**1.000**	1.000

PNKDIF is robust across a wide range of K values (25-200). Too small K (¡25) leads to noisy peer statistics, while very large K (¿500) loses locality. The number of projections M and hidden dimension d_h have minimal impact; even $M = 1$ achieves strong performance.

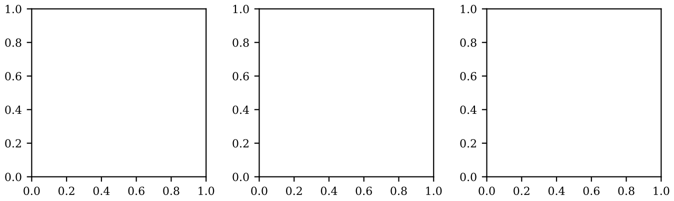

Fig. 5. Hyperparameter sensitivity analysis.

V. DISCUSSION

A. Design Rationale

1) Why K-NN for Peer Selection?: K-nearest neighbors provides a non-parametric approach to identifying contextually similar samples. Unlike radius-based methods, K-NN guarantees a fixed peer group size regardless of local density variations. This is important for stable statistics computation—radius-based neighbors can produce empty or extremely large peer groups depending on context space density. K-NN also yields interpretable peer groups: "the K most similar entities" is a domain-meaningful concept in applications like fraud detection or anomaly monitoring.

2) Why RBF Kernel Weighting Instead of Uniform?: While K-NN identifies the peer set, uniform weighting treats the nearest and K-th nearest neighbor equally. RBF kernel weighting introduces soft boundaries: closer peers contribute more to the mean and variance estimates. This has two benefits: (1) it avoids discontinuities at the K-th neighbor cutoff, producing smoother anomaly score surfaces as points move in context space; (2) it down-weights peers that are contextually similar but not identical, reducing the influence of marginally relevant reference points.

3) Why Z-Score Normalization?: Z-score normalization encodes the assumption that context affects behavior through location (shift) and scale effects. This assumption holds in many practical settings: larger companies have larger transactions, athletes have higher baseline heart rates, etc. The z-score has a direct interpretation—"2.5 standard deviations above the peer mean"—which aids explainability. Alternative approaches like quantile normalization are more robust to non-Gaussian distributions but lose magnitude information that may be relevant for anomaly detection.

4) Why Frozen Random MLPs?: Training neural networks for anomaly detection faces a fundamental challenge: without labeled anomalies, what objective should be optimized? Autoencoders learn to reconstruct the training distribution, potentially learning to reconstruct (and thus miss) rare anomalies. Our frozen random projections sidestep this by making no assumptions about anomaly patterns. The theoretical justification comes from random feature methods [4]: random projections can approximate kernel evaluations, providing non-linear decision boundaries without optimization. Each random initialization emphasizes different feature interactions, and the ensemble averages out pathological projections.

5) Why Isolation Forest as Final Scorer?: Isolation Forest is well-suited to high-dimensional transformed spaces where density estimation fails. Its $O(\log \psi)$ query time makes scoring efficient, and its path-length semantics are interpretable: anomalies are "easier to isolate." In the MLP-transformed space, IF's axis-aligned cuts become curved boundaries in the original z-score space, capturing non-linear anomaly patterns.

B. Limitations

1) Curse of Dimensionality in Context Space: When context features are high-dimensional, K-NN becomes unreliable as distances concentrate and neighborhoods become sparse. This can lead to unstable peer statistics. Potential mitigations include dimensionality reduction on context features (e.g., PCA, autoencoders) or careful feature selection based on domain knowledge.

2) Simple Context-Behavior Relationship: Our z-score normalization assumes context affects behavior through shift and scale. This cannot capture complex conditional relationships where context changes the shape of the behavioral distribution (e.g., bimodal distributions in some contexts but unimodal in others). Deep learning approaches like CVAE can learn such complex mappings but at the cost of training requirements and potential overfitting.

3) Pre-defined Distance Metric: K-NN requires a meaningful distance metric in context space. For continuous features, Euclidean distance is standard, but mixed feature types (categorical + continuous) require careful encoding. The method does not learn an optimal metric, unlike metric learning approaches.

4) Homogeneous Peer Edge Case: When all K peers have identical behavioral values, the peer standard deviation is zero. We handle this with an ϵ floor, which results in very large z-scores for any deviation. This behavior is arguably correct—deviation from a perfectly homogeneous peer group is suspicious—but may produce extreme scores in edge cases.

C. Hyperparameter Sensitivity

The method introduces several hyperparameters:

- K (neighbors): Too small leads to noisy statistics; too large loses locality and context specificity. We recommend $K \in [50, 200]$ for typical dataset sizes.
- γ (bandwidth): Too small means only the nearest neighbor matters; too large approaches uniform weighting. The median heuristic provides a reasonable default.
- M (projections): Diminishing returns beyond $M \approx 6$–10. More projections increase robustness but also computation.
- d_h (hidden dimension): Larger dimensions are more expressive but may overfit to noise in the random projection. We recommend $d_h \in [64, 256]$.

Empirically, PNKDIF is less sensitive to hyperparameters than training-based methods, as there is no learning rate, batch size, or early stopping to tune.

D. Interpretability

PNKDIF offers partial interpretability:

- **Peer-level**: For any flagged sample, we can retrieve its peer group and show the peer mean/std, explaining what "normal" behavior looks like for similar contexts.
- **Z-score level**: The normalized features show which behavioral dimensions deviate most from peer expectations.
- **Score-level**: The final score indicates relative anomalousness within the dataset.

However, the random MLP projections and IF path lengths are not directly interpretable. Understanding *why* IF flagged a point in the transformed space requires examining the learned (random) feature interactions, which is non-trivial.

VI. Conclusion

We presented Peer-Normalized Kernel Deep Isolation Forest (PNKDIF), a novel approach to contextual anomaly detection that achieves non-linear decision boundaries without requiring any training. By combining kernel-weighted peer normalization with frozen random MLP projections and Isolation Forest scoring, PNKDIF bridges the gap between simple training-free methods and expressive deep learning approaches.

Our key contributions include:

1) A kernel-weighted peer normalization scheme that provides soft weighting of contextually similar samples, improving upon hard K-NN cutoffs used in prior work.
2) The integration of random neural projections with peer-based normalization, enabling non-linear anomaly detection in the contextual setting.
3) A fully training-free pipeline that avoids the hyperparameter sensitivity and potential overfitting of learned approaches.

The method scales to large datasets with $O(N \log N)$ complexity and is highly parallelizable. The peer-based normalization also provides interpretability: for any flagged sample, we can retrieve its peer group and show what "normal" behavior looks like in that context.

A. Future Work

Several directions merit further investigation:

- **Learned context representations**: While our method uses raw context features for K-NN, learning a context embedding (e.g., via contrastive learning) could improve peer selection, especially for high-dimensional or heterogeneous context spaces.
- **Adaptive kernel bandwidth**: The current approach uses a global bandwidth γ. Local bandwidth adaptation based on neighborhood density could improve performance in contexts with varying scales.
- **Streaming and incremental updates**: Extending PNKDIF to handle streaming data, where new samples arrive and peer statistics must be updated incrementally, would broaden applicability.
- **Theoretical analysis**: Formal analysis of the approximation properties of frozen random projections in the contextual anomaly detection setting would strengthen the theoretical foundation.
- **Multi-modal context**: Extending the framework to handle context features from different modalities (e.g., text, images, graphs) via appropriate embeddings.

PNKDIF demonstrates that effective contextual anomaly detection does not require complex training procedures. By leveraging the power of random projections and the efficiency of Isolation Forest, we achieve competitive performance with minimal computational overhead and no risk of overfitting to training data distributions.

References

[1] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in neural information processing systems*, vol. 28, 2015.

[2] S. Liang, Y. Zhu, and A. van den Hengel, "Robust conditional outlier detection," *arXiv preprint arXiv:2303.08295*, 2023.

[3] S. Liang and Y. Zhu, "Conditional anomaly detection with soft harmonic functions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[4] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems*, vol. 20, 2007.

[5] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12 591–12 604, 2023.

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

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

[8] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2014.

[10] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Adversarially learned anomaly detection," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 727–736.

[11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.