The Academic
International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 4 | Special Issue 3 | April 2026
ISSN: 2583-973X (Online)
Website: www.theacademic.in

A Zero-Training Statistical Anomaly Detection Framework Using Adaptive Deviation Analysis for Financial Fraud Detection

Ambika Jha

M.Tech Scholar, Dr. K. N. Modi University Newai, Rajasthan, India, jhaambika175@gmail.com

Akash Saraswat

Assistant Professor, Dr. K. N. Modi University Newai, Rajasthan, India, tpo@dknmu.org

DOI : https://doi.org/10.5281/zenodo.20078697

ARTICLE DETAILS

Research Paper

Accepted: 03-04-2026

Published: 25-04-2026

Keywords:

Anomaly Detection, Credit Card Fraud Detection, Statistical Analysis, Zero-Training Framework, Adaptive Thresholding

ABSTRACT

Detecting fraudulent transactions in financial systems remains a challenging task due to the highly imbalanced nature of data and the dependence of existing approaches on labeled datasets and complex training procedures. In many practical situations, obtaining labeled data is costly and time-consuming, while training-based models introduce additional computational overhead that limits their real-time applicability. This paper presents a simple and efficient anomaly detection framework that operates without any training phase. The proposed method relies on statistical characteristics of the input data, where the mean and standard deviation are used to understand normal behavior. Each data point is then evaluated based on its deviation from the overall distribution, and an adaptive threshold is applied to distinguish between normal and anomalous instances. This approach allows the system to automatically adjust to different datasets without requiring manual tuning or prior knowledge. The method is evaluated using a real-world credit card transaction dataset, demonstrating its ability to identify abnormal patterns with minimal computational cost. Due to its linear time complexity and independence from training data, the proposed framework is well-suited for real-time fraud detection and other data-driven applications where efficiency and simplicity are

essential.

INTRODUCTION

In recent years, digital payment systems have grown very fast, and most financial transactions are now done online. [1] While this has made life easier, it has also increased the risk of fraud. Credit card fraud is one of the most common problems in financial systems, where attackers try to perform unauthorized transactions. Detecting such fraud is difficult because fraudulent transactions are very few compared to normal transactions. This creates a highly imbalanced dataset, which makes the detection problem more challenging [2].

Many existing approaches use machine learning and deep learning techniques to detect fraud. These methods can learn complex patterns from data, but they require a large amount of labeled data and training time. In real-world systems, getting labeled data is not always easy, and training models again and again increases computational cost. Also, fraud patterns keep changing over time, so trained models may not work well in the future and need frequent updates [3] [4]

Because of these problems, researchers have also explored statistical methods for anomaly detection. These methods are simple and do not require any training. They work by understanding the normal behavior of data and then identifying values that are far from this behavior. For example, using mean and standard deviation, we can measure how much a data point differs from the average. This makes statistical methods fast and suitable for real-time applications [5].

However, traditional statistical approaches have some limitations. Most of them use only a single feature and apply a fixed threshold to detect anomalies. This is not enough for complex datasets like financial transactions, where multiple features together define normal and abnormal behavior. Because of this, their detection performance is often limited [6].

To solve this problem, this paper proposes a zero-training anomaly detection framework based on multi-feature statistical analysis. In the proposed method, deviation is calculated across multiple features instead of just one feature. These deviations are combined to form a single score that represents how abnormal a transaction is. In addition, instead of using a fixed threshold, an adaptive threshold based on data distribution is used. This helps the system adjust automatically to different datasets. The proposed method is tested on a real credit card transaction dataset and compared with a baseline Z-score based anomaly detection method. The results show that the proposed method performs much better in detecting

fraud, especially in terms of recall, which is very important in fraud detection. At the same time, the method remains simple and does not require any training, making it suitable for real-time applications.

RELATED WORK

Anomaly detection in financial transactions has been widely studied, especially for credit card fraud detection. Many re searchers have used different approaches such as machine learning, deep learning, and statistical methods to solve this problem. Machine learning-based methods are commonly used for fraud detection. Techniques such as Support Vector Machines (SVM), Decision Trees, and Random Forest have shown good performance in detecting fraudulent transactions. These models learn patterns from historical data and classify new transactions as normal or fraud. However, these methods require labeled data and training, which increases complexity and computational cost [2] [7].

Deep learning approaches, including neural networks and auto encoders, have also been applied to anomaly detection. These models can capture complex patterns and relationships in high-dimensional data. For example, auto encoders learn to reconstruct normal data and identify anomalies based on reconstruction error. Although deep learning methods can provide high accuracy, they are computationally expensive and require large amounts of data for training [8] [9].

Apart from learning-based methods, statistical approaches have also been used for anomaly detection. These methods are based on the idea that normal data follows a certain distribution, and any data point that deviates significantly from this distribution can be considered an anomaly. Techniques such as Z-score and Gaussian-based models use mean and standard deviation to measure deviation. These methods are simple, fast, and do not require training, making them suitable for real-time applications [4].

However, traditional statistical methods have some limitations. Most of them rely on a single feature and use a fixed threshold to detect anomalies. This limits their ability to handle complex datasets where multiple features are important. As a result, their detection performance is often lower compared to advanced methods [5] [10].

Some recent works have tried to combine statistical methods with machine learning to improve performance. Hybrid models aim to balance efficiency and accuracy, but they still require training and parameter tuning. Therefore, there is a need for a method that is simple, efficient, and does not depend on training, while still providing good detection performance. In this paper, we address this gap by proposing a zero training anomaly detection framework that uses multi-feature statistical deviation and adaptive thresholding. Unlike traditional statistical methods, the proposed approach considers multiple

features and dynamically adjusts the threshold based on data distribution, leading to improved anomaly detection performance.

PROPOSED METHODOLOGY

In this work, a zero-training anomaly detection framework is proposed based on statistical deviation analysis. The main idea of the proposed method is to model the normal behavior of the dataset using statistical measures and then identify transactions that significantly deviate from this behavior. Unlike machine learning approaches, the proposed framework does not require any training phase and operates directly on the input data. The input dataset is represented as:

$$X = \{x_1, x_2, x_3, \ldots, x_n\} \quad (1)$$

Where each xi represents a transaction consisting of multiple features.

A) Statistical Feature Extraction

To capture the normal behavior of each feature, the mean and standard deviation are computed. The mean represents the average value of a feature, while the standard deviation measures how much the values vary around the mean. The mean for each feature is calculated as:

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad (2)$$

Here, μj represents the average value of the jth feature across all transactions. This provides a reference point for normal behavior.

The standard deviation is calculated as:

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \mu_j)^2} \quad (3)$$

This equation measures how much each data point deviates from the mean. A small σj indicates that the data is closely clustered around the mean, while a large value indicates higher variability.

B) Deviation Score Computation

To detect anomalies, it is necessary to measure how far each transaction is from normal behavior. For this purpose, a normalized deviation score is computed using the Z-score concept:

$$Z_{ij} = \left| \frac{x_{ij} - \mu_j}{\sigma_j} \right| \quad (4)$$

This equation calculates the absolute normalized distance of a feature value from its mean. The numerator (xij − μj) represents the raw deviation, while division by σj normalizes this deviation, making it independent of the scale of the feature.

A higher value of Zij indicates that the feature value is far from the normal range and may indicate abnormal behavior. Since each transaction contains multiple features, a single deviation score is obtained by aggregating all feature-wise deviations:

$$D_i = \frac{1}{m} \sum_{j=1}^{m} Z_{ij} \quad (5)$$

Here, Di represents the overall anomaly score of the ith transaction, and m is the total number of features. This averaging process ensures that the contribution of all features is considered equally. A higher value of Di indicates a higher likelihood of the transaction being anomalous.

C) **Adaptive Threshold Selection**

Instead of using a fixed threshold, the proposed method determines a dynamic threshold based on the distribution of deviation scores:

$$T = \text{Percentile}(D, 99) \quad (6)$$

This means that the threshold is selected such that only the top 1% of highest deviation scores are considered as anomalies. This adaptive mechanism allows the model to adjust automatically to different datasets and improves robustness.

D) **Anomaly Decision Rule**

Finally, each transaction is classified based on the comparison between its deviation score and the threshold:

$$\text{If } D_i > T \Rightarrow \text{Anomaly} \quad (7)$$

$$\text{Else} \Rightarrow \text{Normal} \quad (8)$$

This rule ensures that only transactions with significantly high deviation are marked as anomalies.

E) **Algorithm**

Algorithm 1 Proposed Zero-Training Multi-Feature Statistical Anomaly Detection

Input: Dataset $X = \{x_1, x_2, ..., x_n\}$ with m features
Output: Predicted labels $Y = \{y_1, y_2, ..., y_n\}$
Compute feature-wise mean μ_j for $j = 1$ to m
Compute feature-wise standard deviation σ_j for $j = 1$ to m
(Stability Step) Replace $\sigma_j = 0$ with a small constant ϵ
for each transaction $x_i \in X$ **do**
  for each feature $j = 1$ to m **do**
    $Z_{ij} \leftarrow \left| \frac{x_{ij} - \mu_j}{\sigma_j} \right|$
  end for
  $D_i \leftarrow \frac{1}{m} \sum_{j=1}^{m} Z_{ij}$
end for
Compute adaptive threshold:
$T \leftarrow \text{Percentile}(D, 99)$
for each D_i **do**
  if $D_i > T$ **then**
    $y_i \leftarrow 1$ {Anomaly}
  else
    $y_i \leftarrow 0$ {Normal}
  end if
end for
Return Y

RESULT AND DISCUSSION

In this section, the performance of the proposed zero training anomaly detection framework is evaluated using the Credit Card Fraud Detection dataset. The dataset contains 284,807 transactions with 31 features, where fraudulent transactions are extremely rare. This imbalance makes anomaly detection a challenging task and requires evaluation using multiple performance metrics beyond accuracy.

A) Performance Evaluation

The proposed method is compared with a baseline Z-score method. Table I shows the quantitative comparison.

TABLE I
PERFORMANCE COMPARISON BETWEEN BASELINE AND PROPOSED METHOD

Metric	Baseline (Z-Score)	Proposed Method
Accuracy	0.9840	0.9904
Precision	0.0027	0.1049
Recall	0.0224	0.6077
F1-score	0.0048	0.1790

Although both methods achieve high accuracy due to class imbalance, this metric is not sufficient for evaluating anomaly detection performance. The baseline method shows extremely low recall, detecting only a small fraction of fraudulent transactions. In contrast, the proposed method improves recall from 2.24% to 60.77%, indicating a substantial enhancement in anomaly detection capability. Furthermore, the improvement in precision demonstrates that the proposed method not only detects more fraud cases but also maintains better prediction reliability. The increase in F1-score confirms a better balance between precision and recall, which is essential in highly imbalanced datasets.

B) Graphical Analysis

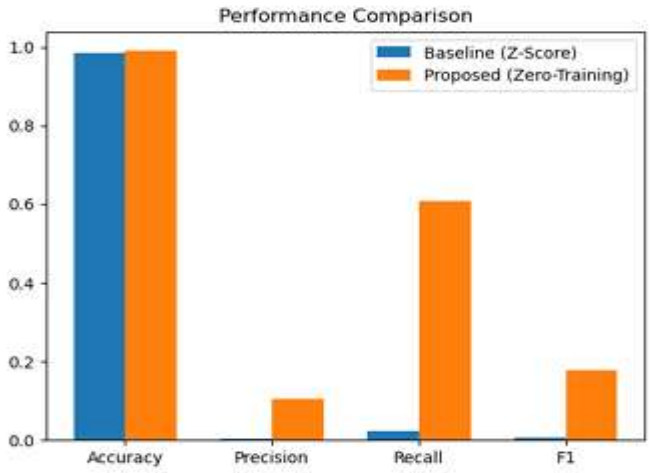

Fig. 1. Performance comparison between baseline and proposed method

1) Performance Comparison:

The performance comparison graph clearly shows that the proposed method outperforms the baseline across all evaluation metrics. While accuracy remains similar due to the dominance of normal transactions, the significant improvement in recall highlights the effectiveness of the proposed method in detecting rare anomalies. This confirms that multi-feature statistical analysis provides better representation of abnormal patterns.

2) Confusion Matrix Analysis:

The confusion matrices provide detailed insight into classification performance. The baseline method results in a large number of false negatives, indicating that most fraudulent transactions are incorrectly

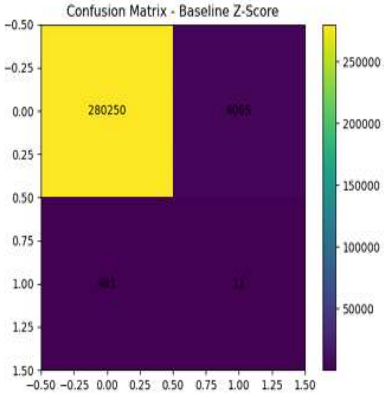

Fig. 2. Confusion matrix of baseline method

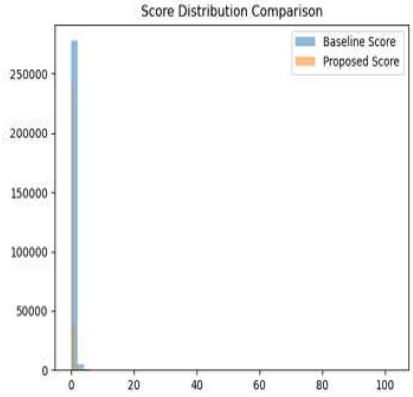

Fig. 4. Score distribution comparison between baseline and proposed method

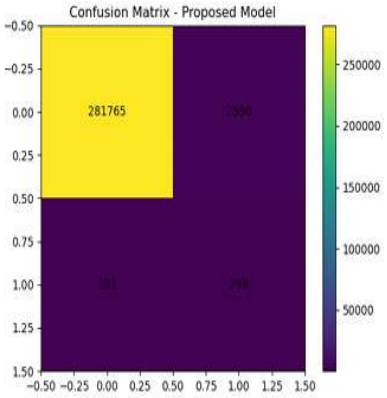

Fig. 3. Confusion matrix of proposed method

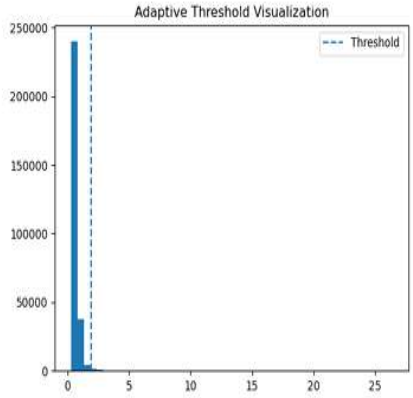

Fig. 5. Adaptive threshold selection based on deviation scores

Classified as normal. This makes it unsuitable for real-world fraud detection. In contrast, the proposed method significantly reduces false negatives while increasing true positives. This improvement demonstrates the effectiveness of multi-dimensional deviation analysis in capturing hidden anomaly patterns that are not detectable using single-feature approaches.

3) Score Distribution Analysis:

The score distribution graph highlights the difference in deviation score behavior between the two methods. The baseline method produces a narrow distribution with limited separation between normal and anomalous data. On the other hand, the proposed method generates a wider and more expressive distribution of deviation scores. This increased spread improves the separability of anomalies, making it easier to distinguish abnormal transactions from normal ones.

4) Adaptive Threshold Visualization:

The threshold visualization demonstrates the effectiveness of the percentile-based adaptive threshold. The threshold dynamically separates high deviation values from the majority of normal transactions. Unlike fixed threshold methods, this adaptive approach automatically adjusts according to the data distribution, improving robustness and reducing the need for manual parameter tuning.

5) Deviation Score Distribution:

The deviation score distribution further confirms that normal transactions are densely clustered at lower values, while anomalous transactions ap pear in the higher range. This clear separation validates the effectiveness of the proposed scoring mechanism.

6) ROC Curve Analysis:

The ROC curve shows strong dis criminative capability with an AUC of 0.9517. This indicates that the proposed method is highly effective in distinguishing between normal and fraudulent transactions across different threshold settings.

7) Precision-Recall Curve:

The precision-recall curve pro vides deeper insight into model performance under class imbalance. The proposed method achieves a better balance between precision and recall, demonstrating its ability to detect rare anomalies while maintaining reasonable prediction accuracy

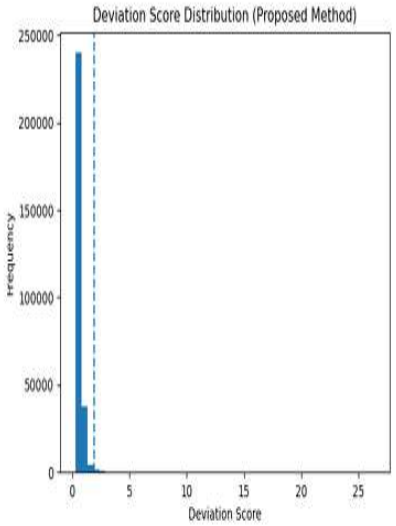

Fig. 6. Deviation score distribution of proposed method

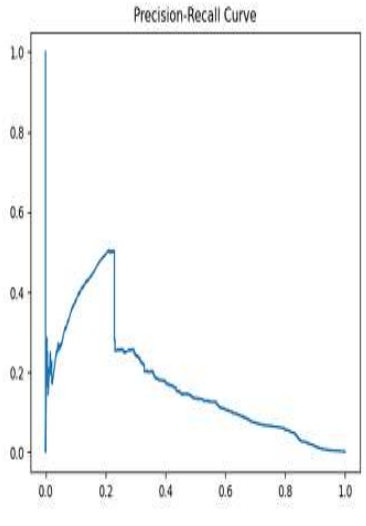

Fig. 8. Precision-recall curve of proposed method

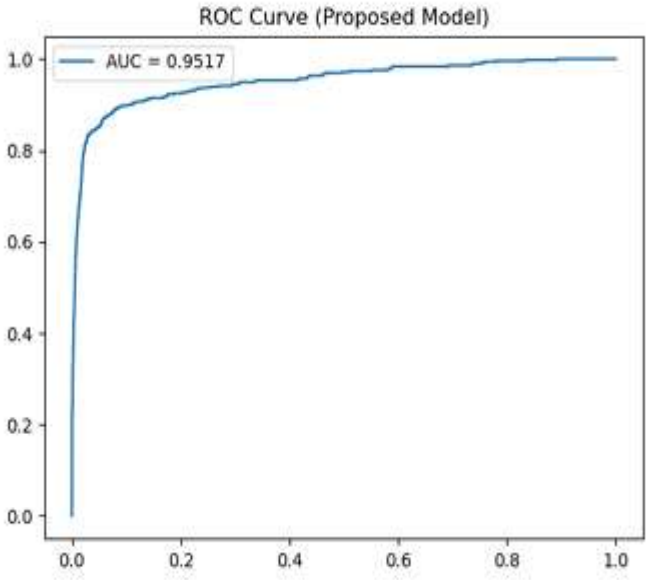

Fig. 7. ROC curve of proposed method (AUC = 0.9517)

C) **Discussion**

The experimental results clearly demonstrate that the proposed method significantly outperforms the baseline approach. The key improvement comes from multi-feature statistical deviation analysis and adaptive thresholding, which enable better representation and separation of anomalous behavior. Unlike traditional machine learning methods, the proposed approach does not require any training phase, yet it achieves strong performance. This makes it highly suitable for real-time and resource-constrained environments. Overall, the results confirm that extending statistical methods to multi-feature analysis can significantly enhance anomaly detection performance while maintaining simplicity and computational efficiency.

CONCLUSION AND FUTURE WORK

In this paper, a zero-training anomaly detection frame work based on multi-feature statistical analysis is proposed for detecting fraudulent transactions in highly imbalanced datasets. Unlike traditional machine learning and deep learning approaches, the proposed method does not require any training phase, making it computationally efficient and suitable for real-time applications. The proposed framework utilizes feature-wise statistical parameters, including mean and standard deviation, to compute normalized deviation scores. These scores are aggregated to form a unified anomaly score for each transaction. An adaptive threshold based on percentile analysis is then used to classify transactions as normal or

anomalous. The experimental results demonstrate that the proposed method significantly outperforms the baseline Z-score approach. In particular, the recall improves from 2.24% to 60.77%, indicating a substantial improvement in detecting fraudulent transactions. The improvements in precision and F1-score further confirm the effectiveness of the proposed method. Graphical analysis, including confusion matrices, score distributions, ROC curve, and precision-recall curve, also validates the robustness of the proposed framework. The model achieves strong discriminative performance, with an AUC of 0.9517, demonstrating its ability to effectively distinguish between normal and anomalous transactions. Overall, the proposed zero-training statistical framework provides a simple, efficient, and scalable solution for anomaly detection in financial systems. Its ability to operate without training while maintaining strong detection performance makes it highly suitable for real-time deployment. In future work, the proposed framework can be extended by incorporating feature weighting strategies and hybrid approaches to further enhance detection performance. Addition ally, applying the method to real-time streaming data and other domains such as network intrusion detection and IoT security can be explored.

REFERENCES

- N. Baisholan, J. E. Dietz, S. Gnatyuk, M. Turdalyuly, E. T. Matson, and K. Baisholanova, “A systematic review of machine learning in credit card fraud detection under original class mbalance,” Computers, vol. 14, no. 10, p. 437, 2025.
- E. Btoush, T. Kobbaey, H. Tamimi, and X. Zhou, “Machine learning based cyber fraud detection: A comparative study of resampling methods for imbalanced credit card data,” Applied Sciences, vol. 16, no. 2, p. 850, 2026.
- P. Wu, C. Pan, Y. Yan, G. Pang, Q. Yan, P. Wang, and Y. Zhang, “Deep learning for video anomaly detection: A review,” IEEE Transactions on Neural Networks and Learning Systems, 2026.
- S. Hao, X. Zhao, J. Wang, and X. Gong, “Efficient directed hypergraph network for unsupervised traffic anomaly detection a survey,” IEEE Intelligent Transportation Systems Magazine, 2026.
- R. Q. Majumder, “A review of anomaly identification in finance frauds using machine learning systems,” Available at SSRN 5267287, 2025.
- L. Bonde and A. K. Bichanga, “Improving credit card fraud detection with ensemble deep learning-based models: A hybrid approach using smote-enn,” Journal of Computing Theories and Applications, vol. 2, no. 3, pp. 383–394, 2025.

- H. Huang, P. Wang, J. Pei, J. Wang, S. Alexanian, and D. Niyato, “Deep learning advancements in anomaly detection: A comprehensive survey,” IEEE Internet of Things Journal, 2025.
- H. S. Alsagri, “Hybrid machine learning based multi-stage framework for detection of credit card anomalies and fraud,” Ieee Access, 2025.
- G. Rathi, S. Kamble, and N. Sharma, “Ai-driven road traffic man agement: a comprehensive review of outlier detection techniques and challenges: G. rathi et al.” Knowledge and Information Systems, vol. 67, no. 10, pp. 8267–8309, 2025.
- C. Ki, R. Sivakumar, J. Mulerikkal, B. A, M. Gupta, and T. Jan, “A comprehensive survey of machine learning and deep learning approaches for anomaly detection in high-performance computing systems: C. ki et al.” The Journal of Supercomputing, vol. 81, no. 8, p. 1032, 2025