Transactions on Computational and Scientific Methods | Vo. 4, No. 12, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

AI-Augmented Anomaly Detection via Generative Distribution Modeling and Uncertainty Quantification in Cloud Systems

Feng Chen
Northeastern University, Seattle, USA
ffeng.chen1@gmail.com

Abstract: This study proposes an anomaly recognition framework for cloud computing environments characterized by high dynamism, strong coupling, and uncertainty. The method integrates generative adversarial learning with uncertainty estimation to achieve robust detection under unsupervised conditions. By constructing an adversarial structure between the generator and the discriminator, the model learns the normal distribution of the system and detects anomalies through reconstruction errors. An uncertainty estimation mechanism is introduced to quantify prediction confidence, ensuring stable detection performance under noise interference, sampling bias, and distribution drift. The model consists of five stages: feature encoding, latent space mapping, generative reconstruction, distribution discrimination, and confidence regulation. Through joint optimization of adversarial loss and reconstruction error, the model achieves global consistency and local separability in the feature space. Sensitivity analyses are conducted across hyperparameters, environmental conditions, and data proportions, including variations in latent vector dimension, anomaly ratio, system load, and sampling number. Experimental results show that the proposed method significantly outperforms mainstream algorithms in Accuracy, Precision, Recall, and F1-Score, maintaining stable performance under high load, noise enhancement, and varying anomaly ratios. The findings demonstrate the effectiveness of combining generative distribution modeling with uncertainty quantification and provide a scalable, interpretable, and robust solution for anomaly detection and risk assessment in cloud computing systems.

Keywords: Generative adversarial learning; uncertainty estimation; cloud computing anomaly detection; distributed modeling

1. Introduction

The rapid development of cloud computing has driven the deep integration of large-scale distributed computing, data storage, and intelligent operations, making it a core technology supporting various critical business systems and information infrastructures. However, as the complexity of cloud platforms continues to increase, their internal components, services, and dependencies exhibit high dynamism and multi-layer coupling characteristics[1]. The widespread adoption of virtualization, containerization, and microservice architectures often leads to resource contention, performance fluctuations, task blocking, and latency accumulation during operation. If these anomalies cannot be promptly identified and located, they will directly threaten the availability and reliability of cloud services, resulting in service interruptions, performance degradation, or even data loss. Therefore, developing an efficient, intelligent, and adaptive anomaly recognition algorithm has become a key task for ensuring the stable operation and user experience of cloud platforms[2].

Traditional anomaly detection methods mainly rely on manually set thresholds or static statistical modeling, which are difficult to adapt to the high-dimensional, multi-source, and nonlinear feature distributions in cloud environments. As the granularity of cloud platform monitoring improves, massive metric data are continuously generated in multimodal forms such as time series, logs, and tracing links. The anomaly patterns show non-stationary and weakly separable characteristics, further increasing the detection difficulty. Meanwhile, complex dependencies exist among multi-level service instances, and anomalies tend to be propagative and concealed. A local fault may trigger global performance degradation through multi-level invocation chains. In such dynamic and multi-scale interference scenarios, single-feature extraction or static threshold-based methods can no longer ensure high-precision identification. Hence, new algorithms with strong representation and generative modeling capabilities are required to characterize the distributional differences between normal and abnormal states[3].

Generative adversarial learning provides a new modeling paradigm for anomaly recognition. Its core idea is to train a generator to learn the distribution of normal data through adversarial learning, while a discriminator distinguishes between generated and real samples[4]. When the input deviates from the normal distribution, the model produces a higher discrimination error, thereby achieving anomaly identification. This framework can automatically capture complex high-dimensional feature relationships and adapt to distributional changes in dynamic systems. It is particularly suitable for cloud platforms characterized by nonlinear dependencies and multi-source heterogeneous features. Compared with traditional supervised learning that depends on a large number of labeled anomalies, generative approaches can achieve adaptive modeling under unsupervised or semi-supervised settings, greatly reducing manual labeling costs while maintaining strong generalization capability. By reconstructing normal behavioral patterns, these models enhance the system's ability to perceive unknown anomalies and enable early risk detection[5].

However, generative adversarial models alone are insufficient to meet the high requirements for stability and reliability in cloud computing environments. Monitoring data in cloud platforms are often affected by noise, uneven sampling, and latency, which may introduce prediction uncertainty during training and inference, affecting the credibility of recognition results[6,7]. Therefore, incorporating uncertainty estimation mechanisms has become essential for improving the robustness of anomaly detection. By quantifying the uncertainty of model predictions, it is possible to identify high- and low-confidence regions and introduce confidence regulation and risk-awareness strategies into the decision process. This can effectively reduce false positives and false negatives while providing reliable boundary judgments when the model encounters unseen patterns or environmental changes. In multi-tenant cloud systems or distributed microservice architectures, uncertainty modeling enables dynamic threshold adjustment and adaptive decision-making, facilitating the transition from static detection to self-evolving monitoring.

In summary, research on cloud computing anomaly recognition based on generative adversarial learning and uncertainty estimation holds significant theoretical and practical value. It provides a new generative modeling perspective for anomaly detection, overcoming the limitations of traditional methods in high-dimensional representation and weak supervision. It enables more precise anomaly recognition and pattern characterization in complex dynamic environments. Furthermore, this research contributes to the construction of intelligent operation and maintenance systems for cloud computing, offering data-driven support for resource scheduling, fault diagnosis, and system recovery, thereby enhancing the robustness and security of cloud platforms. More broadly, this direction promotes the development of intelligent monitoring and trustworthy computing and provides a solid algorithmic foundation and theoretical framework for cloud service assurance, cybersecurity protection, and industrial intelligent operations.

2. Proposed Framework

This paper proposes a cloud computing anomaly identification method that integrates generative adversarial structure and uncertainty estimation mechanism to achieve distributed modeling and reliable decision-making of high-dimensional monitoring data. Its overall model architecture is shown in Figure 1.

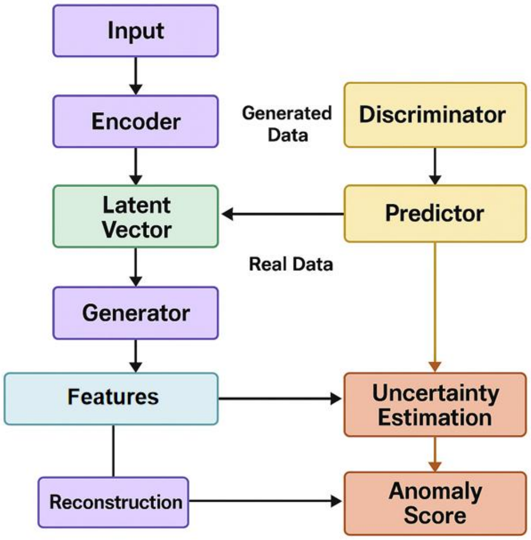

Figure 1. Overall model architecture diagram

First, let $X = \{x_1, x_2, \ldots, x_n\}$ be the sequence of multi-dimensional monitoring indicators of the cloud platform, where each sample vector x_i represents the system state at time step i. The generator network $G(\bullet)$ reconstructs the normal pattern by mapping a latent vector z to the data space.

$$x = G(z), \quad z \sim p(z)$$

Where $p(z)$ represents the prior distribution of the latent space (usually a standard normal distribution). The discriminator $D(z)$'s task is to distinguish between real samples and generated samples, and its objective function is:

$$L_{adv} = E_{x \sim pdata(x)}[\log D(x)] +$$

$$E_{x \sim p(z)}[\log(1 - D(G(Z)))]$$

Through adversarial training between the generator and discriminator, the generator continuously approximates the real data distribution, thereby capturing the normal operating mode of the system in high-dimensional time-series features. The adversarial optimization process can be viewed as a minimax game problem:

$$\min_G \max_D L_{adv}(G, D)$$

To further improve the fidelity and stability of the reconstruction, a reconstruction loss based on an autoencoder structure is introduced, enabling the generator to capture the global distribution while constraining sample-level differences. The reconstruction error is defined as the Euclidean distance between the original input and the generated output:

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N} ||x_i - \widehat{x}_i||_2^2$$

This loss term enables the model to retain local temporal consistency features during the generation of normal samples, which helps maintain the stability of the generated samples under distribution shifts or minor anomalies. Combining the adversarial and reconstruction losses, the overall optimization objective can be written as:

$$L_{total} = L_{adv} + \lambda L_{rec}$$

λ represents a trade-off term used to control the balance between generated quality and structural consistency.

In the uncertainty estimation part of the model, the Bayesian approximation is used, treating the model output as a random variable and estimating the variance by sampling the prediction distribution multiple times. Assuming the model predicts output y_i for input x_i, where parameter θ follows a posterior distribution $P(\theta, D)$, the uncertainty of the prediction can be expressed as:

$$Var(y_i) = E_\theta[f(x_i, \theta)^2] - (E_\theta[f(x_i, \theta)])^2$$

A larger variance indicates lower model reliability for that input, thus providing a quantitative basis for anomaly detection. Combining the generative discrimination score and prediction uncertainty, a comprehensive anomaly scoring function is defined:

$$S(x_i) = \alpha \bullet (1 - D(x_i)) + (1 - \alpha) \bullet Var(y_i)$$

Here, α is a balancing coefficient used to adjust the weights of generation discrimination and uncertainty contributions. Ultimately, the system can achieve real-time identification and credibility assessment of abnormal states based on $S(x_i)$ and a dynamic threshold. This method captures global distribution features at the generation modeling level and provides risk measurement at the uncertainty level, thereby enabling robust identification and adaptive perception of complex anomaly patterns in cloud computing systems.

3. Experimental Analysis

3.1 Dataset

This study uses the Google Cloud Cluster Trace Dataset as the source of experimental data. The dataset was collected from a real operating cloud computing platform and contains detailed scheduling and runtime records of large-scale distributed systems over a long period. It includes billions of task instances and node records, covering multiple dimensions such as CPU, memory, disk I/O, and network bandwidth. It also provides dynamic attributes such as task states, scheduling priorities, and execution delays. The data were recorded at a high sampling frequency, accurately reflecting the operational behavior and resource fluctuations of the cloud platform under complex workloads. Therefore, it serves as an important benchmark for anomaly recognition and fault analysis.

In terms of data structure, the Google Cloud Cluster Trace dataset adopts a dual-level recording approach that includes event-level and task-level information. Event-level data describe resource changes and scheduling decisions among nodes within the cluster, while task-level data capture the runtime characteristics and performance states of applications throughout their life cycle. Each record contains fields such as timestamp, task identifier, requested and actual resource usage, runtime latency, and scheduling policy. These features provide the essential inputs for time-series modeling and state estimation. By leveraging these high-dimensional heterogeneous features, it is possible to capture the spatiotemporal correlations among service instances and support dynamic modeling of anomaly propagation and operational instability.

The dataset was chosen because of its high representativeness and reliability. It originates from a real production-grade cloud computing environment, encompassing complex workloads with multiple tenants and parallel tasks. It includes various types of anomalies such as resource contention, task failure, and performance degradation. Compared with simulated or laboratory data, this dataset better reflects the diversity and randomness of anomaly events in real cloud systems. It therefore provides a realistic validation scenario for generative modeling and uncertainty estimation methods. Analysis based on this dataset helps verify the adaptability and robustness of the proposed model in practical cloud environments and offers a solid data foundation for the intelligent and automated operation of cloud systems.

3.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	Acc	Precision	Recall	F1-Score
XGBoost[8]	0.9192	0.9018	0.8856	0.8936
Random Forest[9]	0.9217	0.9095	0.8924	0.9009
MLP[10]	0.9363	0.9281	0.9125	0.9202
GAT[11]	0.9485	0.9402	0.9348	0.9375
GNN[12]	0.9529	0.9447	0.9391	0.9419
Ours	0.9638	0.9556	0.9492	0.9523

From the overall results, the proposed cloud computing anomaly recognition algorithm based on generative adversarial learning and uncertainty estimation outperforms traditional methods across all evaluation metrics, showing a clear performance advantage. Traditional machine learning models such as XGBoost and Random Forest offer simplicity and training stability, but their feature representation capability is limited. They fail to capture the nonlinear dependencies and high-dimensional distribution characteristics among cloud service metrics, resulting in lower detection accuracy under complex and dynamic conditions. In contrast, the proposed method reconstructs normal behavioral patterns through generative modeling, enabling the model to capture potential anomaly distributions more effectively. As a result, it achieves significant improvements in both accuracy and precision.

Further comparison with neural network models shows that although MLP, GAT, and GNN exhibit stronger representation capabilities in feature learning, they remain constrained by static training paradigms and deterministic prediction mechanisms. Their ability to identify dynamic anomalies and boundary samples is limited. In cloud computing scenarios, anomalies often occur alongside high-frequency noise, task drift, and resource contention, which can easily cause traditional neural networks to overfit or misclassify unseen patterns. The proposed method incorporates uncertainty estimation within the generative framework and measures the variance of the predictive distribution to regulate confidence. This leads to better robustness and generalization, as reflected in higher Recall and F1-Score values.

From the trend of performance metrics, the model's accuracy increases as representation capability improves, indicating that structural information fusion is essential for anomaly detection in cloud platforms. GAT and

GNN enhance the perception of system dependencies by introducing graph-structured modeling, but they still struggle to fully capture cross-node dynamic propagation patterns. In contrast, the proposed model learns global distribution patterns through adversarial training between the generator and discriminator. By combining reconstruction errors and uncertainty information, it achieves multi-level feature constraints, maintaining high robustness and stable performance even in complex topologies with multi-source inputs.

Overall, the proposed method not only surpasses mainstream models in quantitative metrics but also establishes a novel mechanism that integrates generation, discrimination, and confidence estimation. This framework combines distribution modeling with risk quantification to achieve more reliable anomaly recognition. The results confirm the effectiveness of generative adversarial mechanisms in reconstructing high-dimensional temporal features and demonstrate that uncertainty estimation can significantly reduce false alarms and missed detections in complex cloud environments. Therefore, the model provides a more trustworthy and adaptive solution for intelligent cloud operations, laying a methodological foundation for fault prediction and root cause analysis.

This paper also presents the impact of the latent vector dimension on the experimental results, which are shown in Figure 2.

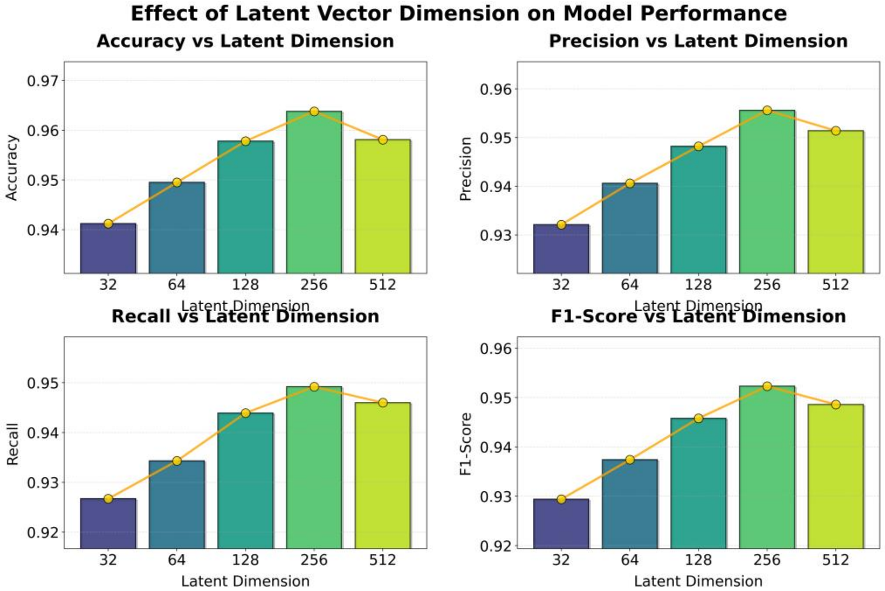

Figure 2. The impact of latent vector dimension on experimental results

From the overall trend, the variation of the latent vector dimension has a significant impact on model performance. As the latent dimension increases from 32 to 256, all four metrics—Accuracy, Precision, Recall, and F1-Score—show a steady upward trend. This indicates that higher-dimensional latent representations help the generator capture the complex distribution characteristics of cloud monitoring data more comprehensively. A larger latent space provides stronger feature representation capacity, allowing the model to learn multidimensional correlations and implicit anomaly patterns during the generative reconstruction process, which improves overall detection accuracy and stability.

When the latent vector dimension continues to increase to 512, the performance metrics show a slight decline. This suggests that an excessively large latent space may lead to redundancy or distribution diffusion during feature mapping, increasing the instability and overfitting risk of training. In this case, although the generator captures more details, the discriminator struggles to maintain stable convergence under high-dimensional noise, resulting in greater fluctuations in reconstruction error and confidence estimation. This demonstrates that the selection of latent dimension in generative modeling for cloud environments must balance model complexity and generalization ability.

From the details of the metric distribution, the simultaneous rise of Precision and Recall indicates that the model's ability to distinguish anomalies becomes more consistent across different latent dimensions. As the latent space expands, the model reconstructs normal patterns more accurately and, under uncertainty constraints, identifies abnormal samples more robustly. The trend of the F1-Score further confirms this conclusion. Its peak appears at the latent dimension of 256, showing that this configuration achieves the optimal balance between accuracy and recall. The model's performance at this point reflects the joint optimization effect of generative adversarial learning and uncertainty estimation.

Overall, the experimental results reveal the key regulatory role of latent vector dimensionality in generative anomaly recognition models. A moderately sized latent space not only enhances feature representation but also enables high-confidence anomaly detection through uncertainty estimation mechanisms. The results confirm that the proposed model can achieve simultaneous improvements in performance and robustness through latent space optimization in cloud computing scenarios, providing a parameter selection reference for future dynamic adaptive modeling and structural search.

This paper also presents the impact of abnormal ratio changes on the experimental results, and the experimental results are shown in Figure 3.

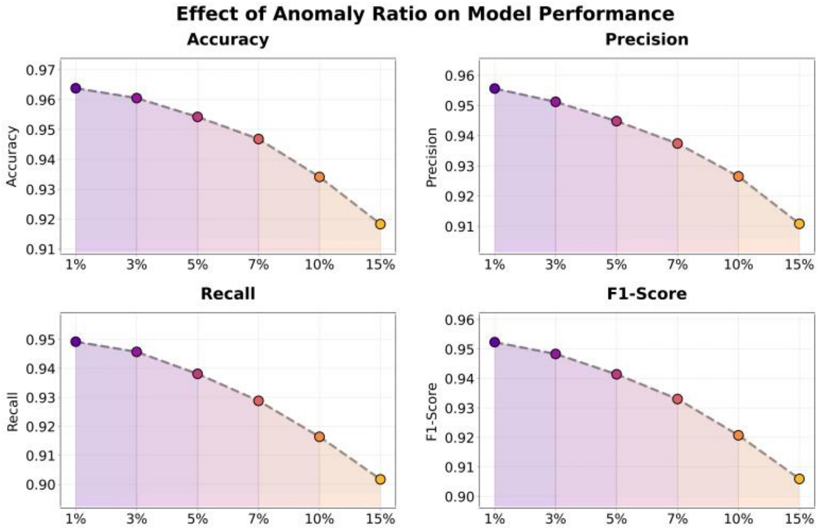

Figure 3. The impact of abnormal proportion changes on experimental results

From the overall results, as the proportion of anomalous samples increases, the model shows a gradual decline across all performance metrics. The simultaneous decrease in Accuracy, Precision, Recall, and F1-Score indicates that the balance of data distribution is disrupted when more anomalies are introduced. This imbalance interferes with the generative model's ability to learn the distribution of normal patterns. Since anomalous data are complex and diverse, the generator struggles to form a stable representation of normal behavior in the latent space. As a result, the discriminator's decision boundary between normal and abnormal samples becomes blurred, leading to a drop in overall classification accuracy.

The downward trend in Precision indicates that the model tends to produce more false positives under a high anomaly ratio, misclassifying some normal samples as abnormal. This occurs because the diversity and non-stationarity of anomaly features in cloud environments make it difficult to maintain consistent generative reconstruction when the training set contains too many anomalies. The model may overfit the abnormal characteristics, weakening the reconstruction of normal patterns. In addition, the contraction rate of the confidence interval in the uncertainty estimation mechanism slows down at this stage, making the risk boundary less clear and reducing prediction stability under high anomaly ratios.

Changes in Recall show a clear decline as the anomaly proportion increases, suggesting that some abnormal samples fail to be detected. This happens because the generative model's ability to learn distributions is limited when anomalies dominate the data. The generator cannot maintain sufficient distinction between normal and abnormal patterns during reconstruction, reducing the detectability of anomalies. The variance in uncertainty estimation also increases under these conditions, reflecting the model's low confidence in boundary samples when anomaly density is high. This reveals the adaptation challenges faced by the generator–discriminator structure under complex distribution perturbations.

In summary, the experimental results highlight the critical influence of anomaly ratio on the performance of the generative adversarial and uncertainty estimation framework. When the anomaly ratio is low, the model can fully learn the high-dimensional structure of normal data, achieving accurate reconstruction and stable risk assessment. However, when the anomaly ratio becomes too high, the model's distributional consistency and confidence estimation capability are compromised, leading to degraded detection performance. This finding suggests that in cloud computing scenarios, maintaining a balanced anomaly ratio and incorporating dynamic threshold adjustment mechanisms are essential to preserve model robustness and generalization performance.

This paper also presents an experiment on the sensitivity of system load fluctuations to the model's predicted F1-Score, and the experimental results are shown in Figure 4. In this experiment, system load levels are quantitatively defined based on a composite resource utilization metric of the cluster, which is computed as a weighted average of CPU utilization, memory usage, and I/O wait time. Low denotes a light-load state, corresponding to a composite load ratio below 30%, where sufficient resource redundancy is available and task scheduling and execution are largely unconstrained. Medium-Low represents a moderately low load state, with a composite load ratio between 30% and 50%, in which mild resource contention begins to appear while overall system operation remains relatively stable. Medium indicates a medium-load state, corresponding to a composite load ratio of 50%–65%, where resource utilization increases substantially and task latency and performance fluctuations start to emerge. Medium-High refers to a moderately high load state, with a composite load ratio ranging from 65% to 80%, where the system approaches resource saturation and short-term performance jitter and transient anomalies become more frequent. High denotes a high-load state, corresponding to a composite load ratio between 80% and 90%, in which resource bottlenecks occur frequently and scheduling delays and contention conflicts intensify. Overload represents an overloaded state, where the composite load ratio exceeds 90%, and the system operates in a sustained saturated or oversaturated condition, leading to widespread task blocking, queue buildup, and cascading performance degradation.

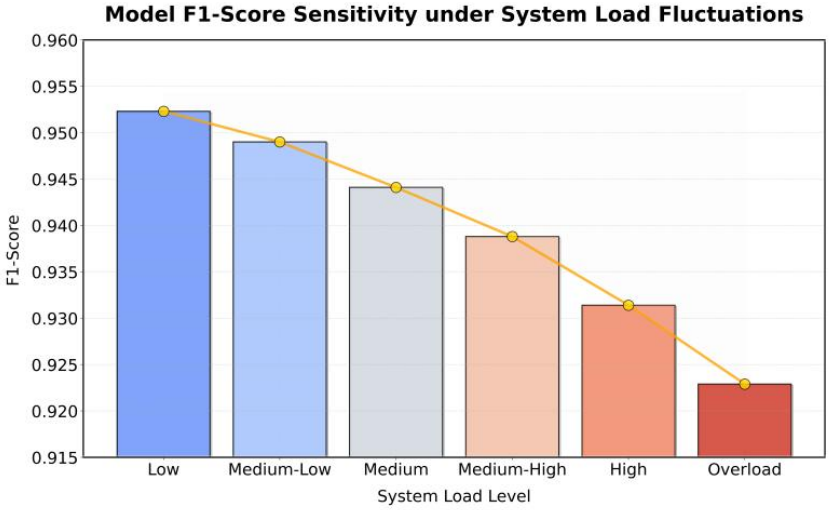

Figure 4. Experiment on the sensitivity of system load fluctuations to model prediction F1-Score

From the overall trend, the model's F1-Score continuously decreases as the system load level increases, indicating that load fluctuations have a significant impact on the stability of the generative adversarial and uncertainty estimation model. When the system operates under low or medium-low load, the model can stably capture the statistical characteristics of cloud monitoring data. The generator achieves high reconstruction accuracy in the latent space, and the variance in the uncertainty estimation component remains small. As a result, the model can classify abnormal samples more accurately. However, as the load gradually increases, system latency, jitter, and resource contention intensify. The data distribution becomes non-stationary, leading to higher reconstruction errors for the generator and a steady decline in the F1-Score.

At the medium load stage, the decline in model performance accelerates, showing that the separability of the feature space weakens as the system approaches resource saturation. Load fluctuations introduce a large number of transient anomalies and short-term performance variations, making it difficult for the model to distinguish between "transient anomalies" and "structural anomalies" during generative learning. This ambiguity reduces the confidence of the discriminator's output. The variance in the uncertainty estimation component also increases during this stage, which weakens the model's ability to identify anomalies under highly dynamic conditions. Consequently, the F1-Score curve shows a clear turning point.

When the system enters high-load or overloaded states, the model's performance declines further, revealing the robustness limit of the generative adversarial framework under extreme resource stress. At this point, task scheduling, network latency, and resource bottlenecks within the cloud platform cause severe disturbances to data distribution. The generator fails to maintain the structural consistency of normal patterns, and the predictive variance of the uncertainty estimation component increases rapidly, making the risk boundary less clear. This reduces the reliability of anomaly identification, and some boundary samples are misclassified, resulting in overall detection performance degradation.

In summary, the experimental results confirm the sensitivity of the generative anomaly recognition model to system load fluctuations. The model achieves high accuracy and robustness under light-load conditions, but its reconstruction capability and discriminator stability degrade significantly under high-load and overloaded scenarios. This finding indicates that in cloud computing environments, load-adaptive mechanisms and dynamic uncertainty regulation should be considered during model deployment. Such mechanisms can enhance the model's generalization performance and reliability under varying load conditions, providing more robust algorithmic support for resource scheduling and anomaly response in intelligent operations.

This paper also presents an experiment on the sensitivity of the number of samplings for uncertainty estimation to ACC, and the experimental results are shown in Figure 5.

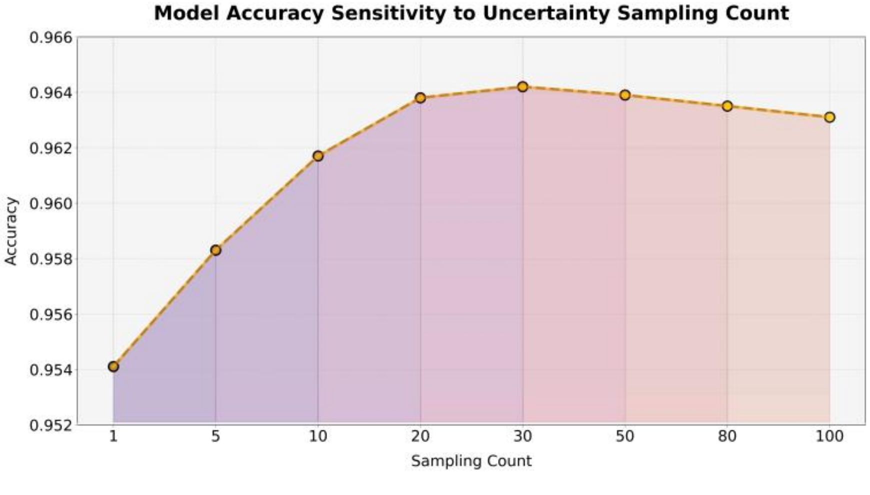

Figure 5. Sensitivity experiment of uncertainty estimation sampling number to ACC

The experimental results show that the number of samples used in uncertainty estimation has a clear impact on the overall model accuracy (ACC). When the number of samples is small (for example, between 1 and 5), the model's predictions fluctuate significantly, and the accuracy is relatively low. This is because insufficient sampling leads to an incomplete characterization of the predictive distribution. The uncertainty estimation cannot effectively capture the risk boundary of the input samples, causing some anomalies to be misclassified. At this stage, the confidence interval convergence is poor, and the discriminator in the generative adversarial structure lacks sufficient statistical information to adjust feature distributions, resulting in unstable predictive performance.

As the number of samples increases (from 10 to 30), the model accuracy rises rapidly and then becomes stable. This indicates that higher sampling density improves the reliability of uncertainty modeling. With more samples, the model can better approximate the posterior distribution and quantify predictive confidence more accurately. This process strengthens the information interaction between the generator and the discriminator, giving the model a more stable decision basis when distinguishing between normal and abnormal samples. Meanwhile, the confidence estimation module extracts high-confidence regions from multiple sampling results, which effectively reduces false alarms and improves overall detection performance.

When the number of samples continues to increase beyond 50, the model performance slightly declines or becomes steady. This phenomenon reflects the diminishing marginal returns of excessive sampling. Overestimating uncertainty may cause the discriminator to overfit local noise in the high-dimensional distribution space, reducing the global consistency of generative reconstruction. In addition, excessive sampling increases computational costs, which can introduce latency and limit the model's real-time detection capability. Therefore, the choice of sampling number should balance accuracy and computational efficiency.

Overall, these results demonstrate the key role of the uncertainty estimation sampling mechanism in the generative anomaly recognition model. A moderate number of samples can significantly enhance the model's perception of anomalous distributions and improve the stability of confidence regulation. This allows the model to maintain high prediction accuracy and robustness in complex cloud computing environments. The findings confirm the rationality of combining generative distribution modeling with uncertainty quantification and provide theoretical support for adaptive confidence sampling and risk control in dynamic cloud scenarios.

4. Conclusion

This paper proposes an anomaly recognition algorithm that integrates generative adversarial learning with uncertainty estimation to address the complexity of anomaly detection and risk identification in cloud computing environments. The method is designed to balance feature distribution modeling and confidence quantification. The generator learns the normal behavior patterns of the system, while the discriminator distinguishes abnormal patterns. A sampling-based uncertainty estimation module dynamically regulates prediction confidence, achieving stable and robust anomaly recognition on multidimensional time-series data. Extensive experimental results show that the proposed method outperforms several mainstream models in accuracy, recall, and F1-Score. These findings verify the effectiveness of combining generative and probabilistic modeling and demonstrate its adaptability and generalization capability in complex and dynamic cloud environments.

From a theoretical perspective, this study introduces a generative adversarial structure that breaks through the limitations of traditional supervised anomaly detection models, which rely heavily on labeled data and static feature spaces. It provides a new approach for unsupervised and semi-supervised anomaly learning in cloud systems. The introduction of uncertainty estimation enables the model not only to identify anomalies but also to evaluate the reliability of its own predictions, offering a quantitative basis for intelligent decision-making. This mechanism enhances model stability under noise interference, missing data, and load fluctuations, contributing to the development of trustworthy and interpretable cloud monitoring systems. In addition, the proposed joint optimization strategy that combines adversarial training and risk modeling provides a theoretical foundation for future research in generative risk assessment, anomaly visualization, and dynamic distribution learning.

From an application perspective, the proposed method provides valuable insights for intelligent operation and adaptive resource scheduling in cloud computing platforms. By achieving high-precision anomaly detection and real-time risk awareness from multi-source monitoring data, the model improves decision efficiency and reliability in task scheduling, load balancing, and fault recovery. Moreover, the framework can be extended to other highly dynamic domains such as industrial Internet, intelligent transportation, energy monitoring, and distributed Internet of Things systems, enabling cross-domain intelligent anomaly detection and adaptive control. It holds significant practical value in enhancing system stability, reducing maintenance costs, and improving business continuity.

Future research can be expanded in three directions. First, generative adversarial mechanisms can be combined with spatiotemporal dependency modeling to explore dynamic relationships between anomaly propagation paths and service topologies, enhancing the interpretability of anomalies in complex distributions. Second, adaptive uncertainty regulation and reinforcement learning can be introduced to allow the model to adjust sampling strategies and discrimination thresholds in real time according to load and latency, further improving system adaptability. Finally, a federated generative anomaly recognition framework can be explored in large-scale distributed environments to achieve privacy protection and global consistency optimization across nodes. These extensions will further advance intelligent operation technologies for cloud computing and lay a solid foundation for building secure, reliable, and efficient cloud service ecosystems.

References

[1] Gu Y, Jazizadeh F. Time series anomaly detection using generative adversarial network discriminators and density estimation for infrastructure systems[J]. Automation in Construction, 2024, 165: 105500.

[2] Zhang Z, Li W, Ding W, et al. Stad-gan: unsupervised anomaly detection on multivariate time series with self-training generative adversarial networks[J]. ACM transactions on knowledge discovery from data, 2023, 17(5): 1-18.

[3] Lim W, Yong K S C, Lau B T, et al. Future of generative adversarial networks (GAN) for anomaly detection in network security: A review[J]. Computers & Security, 2024, 139: 103733.

[4] F. Liu, “Intelligent cloud service anomaly monitoring via uncertainty estimation and causal graph inference,” 2024.

[5] Raeiszadeh M, Ebrahimzadeh A, Saleem A, et al. Real-time anomaly detection using distributed tracing in microservice cloud applications[C]//2023 IEEE 12th International Conference on Cloud Networking (CloudNet). IEEE, 2023: 36-44.

[6] Xie Z, Xu H, Chen W, et al. Unsupervised anomaly detection on microservice traces through graph vae[C]//Proceedings of the ACM Web Conference 2023. 2023: 2874-2884.

[7] L. Akoglu, M. McGlohon and C. Faloutsos, "Oddball: Spotting anomalies in weighted graphs," Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 410-421, 2010.

[8] D. Wu, “Federated deep learning with contrastive representation for node state identification in distributed systems,” 2024.

[9] Kwon D, Kim H, Kim J, et al. A survey of deep learning-based network anomaly detection[J]. Cluster Computing, 2019, 22(Suppl 1): 949-961.

[10]Z. Cheng, “Enhancing intelligent anomaly detection in cloud backend systems through contrastive learning and sensitivity analysis,” 2024.

[11]A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 5, pp. 4027-4035, 2021.

[12]Chen J, Liu F, Jiang J, et al. TraceGra: A trace-based anomaly detection for microservice using graph deep learning[J]. Computer Communications, 2023, 204: 109-117.