International Journal of Multidisciplinary Conference Proceedings
editor@ijmcp.com
https://www.ijmcp.com

AI-Based Anomaly Detection in Cloud Computing Environments Using Ensemble Learning

1st Muhammad Zulqarnain Siddiqui, 2nd Kamran Khan
1st Associate Professor, Department of Computer Science, Iqra University, Karachi, Pakistan
2nd Associate Professor, Department of Computer Science, Iqra University, Karachi, Pakistan

KEYWORDS

Cloud Computing, Anomaly Detection, Ensemble Learning, Intrusion Detection, XGBoost, RUSBoost, UNSW-NB15, CICIDS2017

ARTICLE HISTORY

Date of Preparation: 05-10-2026

Conference Organizer(s)

Research Consultancy on Social & Management Development

&

Social Work Department, University of Punjab, Lahore

ABSTRACT

Cloud computing environments generate continuously changing network and performance data, making anomaly detection essential for identifying intrusions, service degradation, and abnormal resource behavior. This paper examines AI-based anomaly detection in cloud computing with emphasis on ensemble learning. A structured comparative evidence analysis integrates verified peer-reviewed results from intrusion-detection and cloud-monitoring studies using KDDCup99, NSL-KDD, UNSW-NB15, CICIDS2017, DApp monitoring data, Server Machine Dataset, and Vichalana data. The comparison focuses on accuracy, precision, recall, F1 score, class-imbalance handling, robustness, and short-horizon prediction. Published evidence shows that ensemble methods often outperform individual classifiers under the same experimental conditions. On UNSW-NB15, an ensemble machine-learning model achieved 97.06% binary-detection accuracy with 98.45% F1, exceeding the individual Decision Tree, XGBoost, Random Forest, and Gradient Boosting models evaluated in the same study. The evidence indicates that ensembles strengthen anomaly detection through classifier diversity, variance reduction, error-focused learning, and imbalance handling. However, benchmark scores remain dependent on dataset realism, preprocessing, class distribution, and evaluation protocol. Reliable cloud deployment therefore requires representative data, multi-metric evaluation, and continuous validation under changing traffic and workload conditions.

Corresponding Email | kamrankhaniu@gmail.com

Volume-Issue-Page Number | 3(1), 114

Citation | Siddiqui, M. Z., & Khan, K. (2026). AI-based anomaly detection in cloud computing environments using ensemble learning. *International Journal of Multidisciplinary Conference Proceedings, 3(1), 114.*

1. Introduction

Cloud computing has become a core delivery model for computation, storage, databases, and software services, but its scale and elasticity also expand the attack surface that security systems must monitor. Multi-tenant infrastructure, virtualized resources, distributed applications, and continuously changing traffic patterns generate large volumes of heterogeneous telemetry. Conventional signature-based intrusion detection remains useful for known attacks, yet it is less effective when the observed behavior differs from predefined signatures or when new attack patterns emerge. Anomaly detection therefore occupies an important role in cloud security because it focuses on deviations from expected behavior rather than relying only on previously catalogued threats (Chandola et al., 2009; Khraisat et al., 2019).

Machine learning has strengthened anomaly detection by learning decision boundaries directly from traffic and system measurements. Tree-based models, support vector methods, distance-based algorithms, and neural networks have all been applied to network intrusion data. Their performance, however, can vary sharply with class imbalance, feature distributions, dataset age, and attack composition. Ring et al. (2019) showed that intrusion-detection datasets differ substantially in recording environment, traffic characteristics, and evaluation suitability, which makes a single accuracy figure an incomplete basis for judging operational reliability.

Ensemble learning addresses part of this problem by combining multiple learners rather than depending on one classifier. Bagging reduces variance by aggregating models trained on different samples, boosting emphasizes difficult instances across sequential learners, and stacking or voting combines complementary predictions at a higher decision layer. Random Forest demonstrates the benefits of aggregating decorrelated decision trees, while XGBoost provides a scalable boosting architecture that has become influential in high-dimensional classification tasks (Breiman, 2001; Chen & Guestrin, 2016). Sagi and Rokach (2018) further explain that ensemble performance depends on the balance between the predictive strength and diversity of component models.

Cloud anomaly detection adds two distinct technical requirements. The first concerns security anomalies such as denial-of-service activity, exploits, reconnaissance, malware-related traffic, and unauthorized access. The second concerns performance anomalies such as abnormal latency, throughput degradation, resource saturation, or service failure. Both forms of anomaly can disrupt availability and trust, but they involve different signals and may require different learning strategies. Recent work consequently combines ensemble classifiers, deep models, feature reduction, and temporal modeling to improve detection accuracy, robustness, and responsiveness in cloud settings (Shahzad et al., 2022; Xin et al., 2023).

This paper examines verified empirical evidence on AI-based anomaly detection in cloud and cloud-relevant network environments, with specific attention to ensemble learning. The analysis addresses three questions: (1) how ensemble methods compare with individual classifiers on established intrusion-detection datasets; (2) which ensemble structures show consistent gains across security and performance anomalies; and (3) what limitations affect the transfer of benchmark accuracy to operational cloud environments. The objective is to provide a compact technical assessment based on traceable peer-reviewed results rather than combining incompatible datasets into a single artificial score.

2. Literature Review

2.1 Anomaly Detection in Cloud and Network Environments

Anomaly detection identifies observations or sequences that depart from an expected pattern. Chandola et al. (2009) distinguish anomaly-detection methods according to the assumptions used to separate normal and abnormal behavior, while modern intrusion-detection systems increasingly use machine learning to automate this separation. In cloud settings, the problem becomes more difficult because service traffic is dynamic, resources scale automatically, workloads vary by tenant, and legitimate behavior can change over time. Moustafa et al. (2017) addressed this context through a collaborative anomaly-detection framework for large-scale cloud data and evaluated it using UNSW-NB15, emphasizing the need for scalable processing and low false-positive behavior.

The choice of benchmark strongly influences reported performance. UNSW-NB15 was created to include contemporary normal traffic and modern attack activity that older datasets such as KDDCup99 and NSL-KDD do not fully represent (Moustafa & Slay, 2015). CICIDS2017 similarly provides labeled network flows covering benign behavior and multiple attack scenarios and is widely used for machine-learning evaluation (Sharafaldin et al., 2018). Dataset surveys caution that traffic capture, feature construction, attack balance, and temporal realism differ across benchmarks, so performance values are most meaningful when interpreted within the dataset and experimental protocol that produced them (Ring et al., 2019).

2.2 Ensemble Learning for Intrusion Detection

Ensemble learning combines predictions from multiple models to improve generalization and reduce dependence on a single decision rule. Random Forest uses bagging and random feature selection to construct diverse decision trees, whereas boosting methods sequentially emphasize examples that previous learners classify poorly. XGBoost extends gradient boosting through regularized objectives, sparsity-aware processing, and efficient tree construction (Breiman, 2001; Chen & Guestrin, 2016). These properties are relevant to intrusion detection because traffic datasets often contain nonlinear feature interactions, high dimensionality, rare attack classes, and substantial class imbalance.

Empirical security research shows that ensembles can outperform individual classifiers when model diversity and data preprocessing are handled carefully. Dhaliwal et al. (2018) reported 98.70% accuracy for XGBoost on NSL-KDD, exceeding several comparison classifiers in the same study. Bhati et al. (2021) reported 99.95% accuracy for an XGBoost-based ensemble intrusion-detection technique on KDDCup99. Although these values demonstrate the strength of boosted trees, they are not directly comparable because the datasets differ in age, traffic composition, redundancy, class distribution, and experimental design.

2.3 Cloud-Focused Ensemble Models

Shahzad et al. (2022) developed a cloud-based anomaly-detection architecture that uses an ensemble machine-learning layer for binary detection and a CNN-LSTM component for multiclass categorization. On UNSW-NB15, the ensemble layer achieved 97.06% accuracy, 98.39% precision, 98.45% recall, and an F1 score of 98.45% for binary anomaly detection. The study also reported 99.91% accuracy for multiclass anomaly categorization using CNN-LSTM. Importantly, the individual tree-based classifiers in the same evaluation produced lower binary accuracy, including 93.57% for Random Forest, 93.34% for XGBoost, and 94.40% for Gradient Boosting, which provides direct within-study evidence for the ensemble gain.

Al-Sharif and Bushnag (2024) examined boosted and bagged ensemble classifiers for cloud intrusion detection using CICIDS2017-derived subsets. RUSBoost produced the strongest average performance, with 99.821% accuracy across evaluated subsets, while bagging reached 99.997% accuracy on one subdataset. On the full dataset, RUSBoost achieved 99.736% accuracy. These results underline the value of imbalance-aware ensemble learning because attack classes in intrusion datasets are rarely distributed uniformly.

2.4 Performance Anomalies and Robustness

Cloud security is not limited to malicious network traffic; runtime performance anomalies can also indicate service degradation, misconfiguration, resource contention, or emerging failures. Xin et al. (2023) evaluated an ensemble learning-based detection framework across DApp monitoring data, Server Machine Dataset data, and Vichalana data. Individual unsupervised methods such as Isolation Forest, k-nearest neighbors, local outlier factor, and one-class SVM performed inconsistently across datasets, demonstrating that the best detector for one distribution may not remain best when the data pattern changes. Isolation Forest itself is designed to isolate anomalies through randomized partitioning and is efficient in high-dimensional settings (Liu et al., 2008).

The deep ensemble in Xin et al. (2023) improved both accuracy and robustness. It achieved an F1 score of 0.8152 on the Server Machine Dataset and 0.8438 on Vichalana, outperforming the base methods reported for those datasets. The same framework maintained F1 scores above 0.8 for anomaly prediction within the next four minutes on the DApp monitoring data and obtained the highest reported combined accuracy-robustness-prediction score. This evidence extends the ensemble argument beyond network attack classification to operational cloud monitoring.

A recurring theme across the literature is that high benchmark accuracy does not automatically imply reliable deployment. Intrusion detection faces concept drift, class imbalance, unseen attacks, encrypted traffic, changing service behavior, and operational constraints on latency and false alarms. Khraisat et al. (2019) emphasize that anomaly-based systems can detect novel behavior but often face false-positive challenges. More recent cloud-focused evidence therefore favors evaluation with multiple metrics and heterogeneous datasets rather than dependence on accuracy alone.

3. Methodology

The study uses a structured comparative evidence analysis of peer-reviewed anomaly-detection research. Studies were included when they met four conditions: they examined network or performance anomaly detection relevant to cloud computing; they used machine learning or ensemble learning; they reported an identifiable dataset and quantitative evaluation metric; and the bibliographic record could be verified through a DOI-bearing publisher or recognized research repository. Foundational machine-learning and dataset papers were included to establish the technical basis of the models and benchmarks.

The empirical comparison centers on KDDCup99, NSL-KDD, UNSW-NB15, CICIDS2017, DApp monitoring data, Server Machine Dataset, and Vichalana data. These datasets represent different generations and types of anomaly-detection problems. KDDCup99 and NSL-KDD are established intrusion-detection benchmarks, UNSW-NB15 and CICIDS2017 contain more contemporary traffic and attacks, while the DApp and server-monitoring datasets represent time-dependent performance behavior in cloud applications. The analysis does not treat their scores as directly interchangeable.

The ensemble techniques are grouped into four families: bagging, boosting, voting or stacking, and deep ensemble integration. Random Forest represents bagging-based tree aggregation. Gradient Boosting, XGBoost, AdaBoost, LPBoost, and RUSBoost represent sequential or reweighting-based

ensembles. Voting models combine predictions from heterogeneous classifiers, while deep ensemble structures learn nonlinear combinations of outputs from multiple detectors. This grouping permits technical comparison without assuming that all ensembles have the same training objective.

Performance is assessed through accuracy, precision, recall, F1 score, false-alarm behavior, robustness across datasets, and prediction horizon where these metrics are directly reported. Accuracy receives less interpretive weight in heavily imbalanced settings because a classifier can obtain a high value by favoring the dominant class. Precision and false-positive behavior indicate the operational cost of alarms, recall reflects missed anomalies, and F1 combines precision and recall into a balanced summary measure.

Only numerical findings traceable to the original study are reported in the results. No pooled meta-analytic estimate is calculated because the studies differ in datasets, class definitions, train-test protocols, feature preprocessing, model parameters, and target tasks. The analytical focus is therefore within-study comparison, consistency of ensemble advantages, and the conditions under which reported gains appear strongest. This design preserves the meaning of the original experiments while allowing a coherent cross-study interpretation.

4. Results

4.1 Security Anomaly Detection

The evidence shows a consistent advantage for ensemble or boosted models in security-focused anomaly detection. Dhaliwal et al. (2018) obtained 98.70% accuracy with XGBoost on NSL-KDD, compared with 96.12% for Random Forest among the referenced comparison techniques. Bhati et al. (2021) reported 99.95% accuracy for an XGBoost-based ensemble on KDDCup99. These findings confirm that boosted tree structures can fit complex nonlinear interactions in network features, but the older benchmark composition limits direct inference to current cloud traffic.

The strongest within-study comparison comes from Shahzad et al. (2022). On UNSW-NB15, Decision Tree, XGBoost, Random Forest, and Gradient Boosting achieved binary accuracies of 91.86%, 93.34%, 93.57%, and 94.40%, respectively. Their ensemble machine-learning model increased accuracy to 97.06% and reported 98.39% precision, 98.45% recall, and 98.45% F1. Because the classifiers were evaluated within the same experimental setting, the improvement provides more persuasive evidence than cross-paper accuracy comparisons.

4.2 Imbalance-Aware Ensembles

The CICIDS2017-based results of Al-Sharif and Bushnag (2024) highlight the importance of class imbalance. RUSBoost achieved the best average performance across the evaluated subsets with 99.821% accuracy and recorded 99.736% accuracy on the entire dataset. Bagging achieved 99.997% accuracy on one subdataset, while RUSBoost reached 99.936% on another. The variation across subsets shows that no single ensemble dominates every traffic distribution, even when the overall accuracy remains high.

RUSBoost is particularly relevant to cloud intrusion detection because it combines boosting with random undersampling of the majority class. This mechanism reduces the tendency of a classifier to concentrate on abundant benign traffic while overlooking rare attack instances. The published results therefore support an architectural principle rather than only a headline score: imbalance handling needs to be integrated into model training when attack classes are sparse or highly unequal.

4.3 Cloud Performance Anomalies

For cloud application monitoring, Xin et al. (2023) found that base detectors changed rank across datasets. On the DApp monitoring data, k-nearest neighbors produced the best base F1 score of 0.8033,

whereas Isolation Forest performed best on Server Machine Dataset data with an F1 score of 0.7515 and one-class SVM was strongest among the reported base methods on Vichalana with 0.6778. This instability supports the use of ensembles when deployment environments differ in distribution and noise characteristics and also shows why isolation-based, distance-based, density-based, and kernel-based detectors can complement one another (Liu et al., 2008).

The deep ensemble improved F1 to 0.8152 on the Server Machine Dataset and 0.8438 on Vichalana. It also maintained F1 scores above 0.8 for predictions up to four minutes ahead on DApp monitoring data and produced the highest ARP score, 5.1821, in the study. These results indicate that ensemble learning can add value in cloud operations beyond intrusion classification by improving robustness and short-horizon anomaly prediction.

5. Discussion

The comparative evidence supports ensemble learning as a strong design choice for AI-based anomaly detection in cloud environments, but the source of the advantage matters. Bagging improves stability when individual learners are sensitive to sampling variation. Boosting improves difficult decision boundaries by concentrating learning on errors. RUSBoost adds explicit class-imbalance handling, and deep ensembles combine heterogeneous anomaly scores through nonlinear representations. The common benefit is reduced dependence on the failure pattern of one model.

A second finding is that dataset realism is as important as classifier sophistication. Very high accuracy on KDDCup99 or NSL-KDD can demonstrate algorithmic capability, yet these datasets cannot represent the full diversity of modern cloud traffic. Moustafa and Slay (2015) introduced UNSW-NB15 partly to address limitations in older benchmarks, and later studies increasingly use datasets such as CICIDS2017 or operational monitoring traces. Ring et al. (2019) similarly caution that dataset properties determine how far an intrusion-detection result can be generalized.

The results also show why accuracy alone is insufficient. In a cloud environment, a false-positive alert can trigger unnecessary investigation or automated mitigation, while a false negative can leave an attack or service failure undetected. Precision, recall, F1 score, confusion matrices, and false-alarm rates therefore provide more operationally meaningful evidence. Shahzad et al. (2022) strengthened their binary result by reporting precision, recall, and F1 alongside accuracy, while Xin et al. (2023) explicitly incorporated robustness and prediction ability into evaluation.

Model diversity should be deliberate rather than cosmetic. Combining several highly correlated learners can reproduce the same errors and deliver little gain. Ensemble design is strongest when component models capture different properties of the data, such as tree partitioning, distance structure, local density, linear separation, or temporal dependencies. Sagi and Rokach (2018) identify classifier diversity as a central condition for effective ensemble learning, and the cloud evidence reviewed here is consistent with that principle.

Operational deployment also requires attention to computational cost. Cloud telemetry arrives continuously, so training complexity, inference latency, memory use, and model-update frequency matter alongside classification quality. Tree ensembles can be efficient at inference, but large boosted models or deep ensembles can increase training and maintenance costs. Xin et al. (2023) reported fast testing for their prediction component, yet their broader discussion also recognizes the computational trade-offs associated with machine-learning methods.

Another challenge is concept drift. Normal cloud behavior changes as software versions, workloads, network paths, user populations, and resource allocations change. Attack behavior also

evolves. A detector trained on a fixed benchmark can therefore lose calibration after deployment. Ensemble systems can mitigate this risk when components are retrained or reweighted over time, but static ensembles are not inherently drift-resistant. Continuous validation against recent traffic is necessary for sustained reliability.

The evidence has three limitations. First, the reviewed studies use different datasets and protocols, preventing a valid pooled accuracy ranking. Second, several benchmarks contain synthetic or laboratory-generated attacks rather than unconstrained production traffic. Third, published studies commonly emphasize classification performance more than deployment factors such as alert workload, retraining frequency, encrypted traffic, and adversarial adaptation. These limitations do not weaken the value of ensemble learning, but they define the boundary between benchmark success and operational cloud security.

6. Conclusion

AI-based anomaly detection is an important component of cloud security and reliability because cloud environments produce dynamic traffic, heterogeneous telemetry, and attack patterns that cannot be handled completely through fixed signatures. The reviewed evidence shows that ensemble learning frequently improves anomaly detection by combining complementary classifiers, reducing variance, concentrating on difficult cases, or addressing class imbalance. Verified studies report strong results for XGBoost-based intrusion detection, ensemble machine learning on UNSW-NB15, RUSBoost on CICIDS2017-derived data, and deep ensembles for cloud performance anomalies. The most credible gains appear in within-study comparisons where ensembles are evaluated against their component models under the same dataset and protocol. At the same time, benchmark scores require careful interpretation because datasets differ in age, traffic realism, class balance, and task definition. Effective cloud anomaly detection therefore depends on four elements working together: representative data, diverse ensemble components, evaluation beyond accuracy, and continuous validation under changing operational conditions.

References

Al-Sharif, M., & Bushnag, A. (2024). Enhancing cloud security: A study on ensemble learning-based intrusion detection systems. IET Communications, 18(16), 950-965. https://doi.org/10.1049/cmu2.12801

Bhati, B. S., Chugh, G., Al-Turjman, F., & Bhati, N. S. (2021). An improved ensemble based intrusion detection technique using XGBoost. Transactions on Emerging Telecommunications Technologies, 32(6), e4076. https://doi.org/10.1002/ett.4076

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15. https://doi.org/10.1145/1541880.1541882

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Dhaliwal, S. S., Nahid, A.-A., & Abbas, R. (2018). Effective intrusion detection system using XGBoost. Information, 9(7), Article 149. https://doi.org/10.3390/info9070149

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity, 2, Article 20. https://doi.org/10.1186/s42400-019-0038-7

Moustafa, N., Creech, G., Sitnikova, E., & Keshk, M. (2017). Collaborative anomaly detection framework for handling big data of cloud computing. In 2017 Military Communications and Information Systems Conference (MilCIS) (pp. 1-6). IEEE. https://doi.org/10.1109/MilCIS.2017.8190421

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In 2015 Military Communications and Information Systems Conference (MilCIS) (pp. 1-6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 86, 147-167. https://doi.org/10.1016/j.cose.2019.06.005

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery, 8(4), e1249. https://doi.org/10.1002/widm.1249

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413-422). IEEE. https://doi.org/10.1109/ICDM.2008.17

Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP) (pp. 108-116). SciTePress. https://doi.org/10.5220/0006639801080116

Xin, R., Liu, H., Chen, P., & Zhao, Z. (2023). Robust and accurate performance anomaly detection and prediction for cloud applications: A novel ensemble learning-based framework. Journal of Cloud Computing, 12, Article 7. https://doi.org/10.1186/s13677-022-00383-6

Shahzad, F., Mannan, A., Javed, A. R., Almadhor, A. S., Baker, T., & Al-Jumeily OBE, D. (2022). Cloud-based multiclass anomaly detection and categorization using ensemble learning. Journal of Cloud Computing, 11, Article 74. https://doi.org/10.1186/s13677-022-00329-y