THE JOURNAL OF APPLIED ENGINEERING AND TECHNOLOGIES
2026 Vol. 1, No. 3, ISSN (O): 3154-7877

Early Warning of Advanced Persistent Threats Through Enterprise Network Behavior Profiling

Mads Jensen[1], Sofie Nielsen[1], Lars Andersen[1], Astrid Petersen[1*]
[1]Department of Computer Science, University of Copenhagen, Copenhagen 2100, Denmark
Corresponding Author*: Astrid Petersen E-mail: a.petersen@dtu.dk

ARTICLE INFO

Keywords
APT detection
enterprise network security
campus network traffic
explainable AI
SHAP
ensemble learning
early-stage intrusion detection

Published:
01 September 2026

ABSTRACT

Advanced persistent threats often begin with low-volume and stealthy network activities, such as reconnaissance, command-and-control beaconing, credential probing, and slow data staging. These behaviors are difficult to detect because they may not produce obvious traffic spikes or signature-matching patterns. This study proposes a SHAP-guided ensemble learning approach for early-stage APT traffic detection in enterprise campus networks. The framework integrates CatBoost, Extra Trees, and Balanced Random Forest classifiers to identify weak abnormal signals from flow-level and temporal communication features. SHAP analysis is used to explain both global feature importance and individual alerts, allowing security analysts to understand why specific flows are classified as suspicious. Experiments are conducted on a campus-scale network traffic dataset collected from 14 subnets, 9,600 active endpoints, and 31 internal servers over 28 days. The dataset includes 8.73 million flow records, with injected APT-like scenarios involving beaconing, internal reconnaissance, credential spraying, and staged exfiltration. A total of 57 features are extracted, including periodic connection interval, low-volume destination repetition, DNS query irregularity, failed authentication-related flow ratio, rare external endpoint access, and off-hour communication frequency. The proposed method achieves 96.87% accuracy, 94.92% macro-F1, and 97.76% AUC under highly imbalanced traffic conditions. For early-stage command-and-control beaconing, the detection recall reaches 93.41%, which is 4.68% higher than the best standalone classifier. The false alarm rate is controlled at 2.36% after probability calibration. SHAP results show that periodic low-volume connections, rare external destination access, abnormal DNS behavior, and off-hour communication are key indicators of early APT activity. These findings show that interpretable ensemble learning can help identify stealthy enterprise network threats before large-scale compromise occurs.

1. Introduction

Advanced persistent threats (APTs) remain difficult to detect in enterprise campus networks because their early activities are usually quiet, slow, and deeply embedded in normal traffic. These activities may include reconnaissance, command-and-control beaconing, credential probing, low-frequency lateral access, and small-scale data staging. Unlike high-volume attacks, early APT traffic does not usually generate obvious traffic peaks, large packet bursts, or signatures that can be directly matched by traditional intrusion detection systems (Nguyen et al., 2026; Xue et al., 2025). Enterprise campus networks also contain heterogeneous endpoints, internal servers, user devices, authentication services, DNS systems, and cloud-connected applications, which makes weak malicious signals harder to separate from legitimate daily operations (Chowdhury, 2025; Yin et al.,

Citation: Jensen, M., Nielsen, S., Andersen, L., & Petersen, A. (2026). Early warning of advanced persistent threats through enterprise network behavior profiling. *The Journal of Applied Engineering and Technologies*, 1(3), 61-69.

https://doi.org/10.64744/tjaet.2026.286.

2026). Recent work on policy-driven vulnerability risk quantification for large-scale cloud infrastructure has shown that security risk assessment should not rely only on isolated technical indicators, but should also consider risk priority, exposure context, and operational policy constraints (Shao, 2026). This view is important for APT detection because early-stage attacks often appear as small deviations across multiple network layers rather than as single high-confidence events (Avgetidis et al., 2025; Yang, 2026).

Existing studies have explored several types of behavioral evidence for early threat detection, including flow records, temporal traffic changes, DNS behavior, abnormal access paths, and communication regularity (Deng & Yang, 2026; Subrahmanyam, 2025). These methods provide useful ways to capture stealthy activities that cannot be identified by signature-based rules alone. Flow-level features can describe connection duration, byte volume, packet count, service port, and communication direction. Temporal features can reveal off-hour access, repeated low-volume connections, or irregular access intervals. DNS-related indicators can help identify suspicious domain queries, rare domain usage, and possible command-and-control communication (Mahmood et al., 2025; Zhang, 2026). Access-path features can further expose unusual internal movement among endpoints and servers. However, these indicators are often weak when viewed separately, and their detection value depends heavily on the operational context of the campus network. Machine learning has become an important method for intrusion detection because it can learn complex and nonlinear relations among traffic features (Khayat et al., 2025; Li & Liu, 2026). Compared with rule-based detection, machine learning models can identify hidden patterns from large-scale traffic data and adapt to different combinations of abnormal behaviors (Wang et al., 2026). Deep learning models have also been applied to network intrusion detection and have reported strong performance in many experimental settings (Ahmed et al., 2025; Zhu et al., 2025). However, these models often require large amounts of labeled data, stable training distributions, and substantial computational resources. Their decision processes are also difficult for security analysts to interpret, especially when an alert must be verified quickly in daily operations. In practical campus networks, labeled APT samples are usually rare, traffic classes are highly imbalanced, and normal user behavior changes across time, departments, and service types (Farhan et al., 2025; Zhang et al., 2026). These conditions reduce the reliability of models that only pursue high accuracy without sufficient interpretability. Ensemble learning provides a practical alternative for detecting weak APT signals under unstable traffic conditions. By combining multiple classifiers, ensemble methods can reduce the dependence on a single model and improve robustness when feature distributions shift (Dornala & Senthilkumar, 2025; Qi & Qiao, 2026). Tree-based ensemble models are especially suitable for flow-level security data because they can handle nonlinear feature interactions, mixed numerical and categorical features, missing values, and imbalanced samples. Models such as CatBoost, Extra Trees, and Balanced Random Forest can capture different aspects of suspicious traffic behavior. CatBoost is effective for structured data and categorical feature handling, Extra Trees improves diversity through randomized tree construction, and Balanced Random Forest can reduce bias toward the majority class. Their combination is therefore suitable for campus-scale APT detection, where malicious samples are limited and normal traffic dominates the dataset. Interpretability is another key requirement for practical APT detection. Security analysts need to know not only whether a flow is classified as suspicious, but also why the alert is generated. Explainable methods such as SHAP can provide both global and local explanations by measuring the contribution of each feature to model decisions (Gao et al., 2026; Vasheghani & Sharifi, 2025). Global explanations can identify which traffic characteristics are most important across the whole dataset, while local explanations can help analysts examine individual alerts. This

is particularly useful for early-stage APT detection, where alerts may be triggered by subtle evidence such as rare external access, low-volume beaconing, abnormal DNS queries, unusual internal communication, or off-hour server access (Salih et al., 2025; Su & Wu, 2026). Nevertheless, many existing studies still use SHAP only as a post-analysis tool after model training, rather than using it to connect detection results with realistic APT behaviors and analyst verification needs.

To address these gaps, this study proposes a SHAP-guided ensemble learning method for early-stage APT traffic detection in enterprise campus networks. The proposed method combines CatBoost, Extra Trees, and Balanced Random Forest to detect weak abnormal signals from flow-level, temporal, access, and DNS-related features. SHAP is used to explain global feature importance and individual alert decisions, making the detection results easier to verify and interpret. Experiments are conducted on a campus-scale dataset containing 14 subnets, 9,600 endpoints, 31 internal servers, and 8.73 million flow records collected over 28 days. The purpose of this study is not only to improve early APT detection performance, but also to provide an interpretable workflow that can support practical security analysis. By linking ensemble learning with SHAP-based explanations, the study aims to help analysts identify low-volume and early-stage threat behaviors before they develop into larger intrusions, while keeping the model results transparent, traceable, and suitable for real enterprise campus environments.

2. Materials and Methods

2.1 Sample and Study Environment

This study used a campus-scale enterprise network dataset from a controlled security monitoring setting. The network contained 14 subnets, 9,600 active endpoints, and 31 internal servers. The endpoints included office computers, laboratory workstations, administrative terminals, and shared service devices. The servers included authentication, DNS, file, web, database, and application servers. Traffic records were collected for 28 consecutive days during normal teaching, research, administrative, and service activities. In total, 8.73 million flow records were obtained from gateway logs, internal switch mirrors, DNS logs, and authentication-related traffic summaries. To protect privacy, IP addresses, user identifiers, and host names were anonymized before analysis. Only flow-level metadata and derived behavior features were used. Packet payloads and personal content were not included.

2.2 Experimental Design and Control Settings

The experiment was designed to detect early APT-like traffic under imbalanced network conditions. The control group consisted of normal campus traffic collected during stable operation periods. It included web access, cloud service synchronization, software updates, DNS requests, file transfer, and internal server access. The experimental group contained injected APT-like scenarios. These scenarios included command-and-control beaconing, internal reconnaissance, credential spraying, rare external access, and staged data exfiltration. The attack traffic was added at a low volume to reflect the early stage of APT activity. To avoid a simple separation between normal and abnormal samples, attack behaviors were mixed with background traffic across different subnets and time periods. CatBoost, Extra Trees, and Balanced Random Forest were used as the main model components. Several standalone classifiers were used as baselines to test the effect of the ensemble structure on recall, false alarms, and model stability.

2.3 Measurement Methods and Quality Control

Flow records were generated from five-tuple traffic information. The fields included source

address, destination address, source port, destination port, protocol type, start time, end time, packet count, byte count, and connection state. Temporal features were calculated from repeated host-pair connections and destination access intervals. DNS features were measured using query frequency, domain rarity, failed resolution ratio, and repeated access to uncommon domains. Authentication-related features were derived from failed-login traffic patterns and repeated access attempts to protected internal services. Off-hour communication was measured by comparing host activity with its normal working-time baseline. Quality control was conducted before feature extraction. Duplicate records, incomplete logs, abnormal timestamps, and records with impossible duration values were removed. Timestamps from different log sources were converted to the same time zone. Extreme byte and packet values were checked by percentile rules. Label consistency was verified by matching injected attack windows with the corresponding flow records. The final feature set contained 57 flow-level and temporal communication features

2.4 Data Processing and Model Formulation

All records were cleaned, anonymized, and converted into host-pair flow sequences. Continuous features were normalized using statistics from the training set. Categorical fields, including protocol type and service category, were encoded before training. The dataset was divided into training, validation, and test sets in time order to reduce information leakage. Class imbalance was handled by balanced sampling and class-weight adjustment. The ensemble output was calculated from the calibrated probability scores of the three base classifiers. The final suspicious probability was defined as:

$$P(y=1|x)=\sum_{m=1}^{M} w_m P_m(y=1|x)$$

where $P_m(y=1|x)$ is the abnormal probability predicted by the m-th classifier, w_m is the weight based on validation performance, and M is the number of base classifiers. Model performance was measured by accuracy, precision, recall, macro-F1, AUC, and false alarm rate. The false alarm rate was calculated as:

$$FAR=\frac{FP}{FP+TN}$$

where FP denotes normal flows incorrectly labeled as suspicious, and TN denotes normal flows correctly labeled as normal. Probability calibration was performed on the validation set to reduce unstable alert scores and support threshold selection.

2.5 SHAP-Based Explanation and Evaluation Procedure

SHAP analysis was used to explain overall model behavior and single alert decisions. Global SHAP values were calculated to identify the main indicators of early APT traffic. Local SHAP explanations were used to show why a specific flow was classified as suspicious. Features with positive SHAP values increased the abnormal probability, while features with negative values supported a normal decision. The explanation results were checked together with the original flow context, including host-pair history, destination rarity, DNS behavior, connection periodicity, and access time. This process helped confirm whether the model used meaningful security signals instead of random traffic noise. The final evaluation compared the proposed SHAP-guided ensemble model with each standalone classifier. The evaluation focused on command-and-control beaconing recall, macro-F1 under class imbalance, and false alarm rate after probability calibration. The aim was to test whether the method could improve early APT detection while keeping the decision process clear for security analysts.

3. Results and Discussion

3.1 Overall Detection Performance

The proposed SHAP-guided ensemble model showed strong and stable performance on the campus-scale traffic dataset. It reached 96.87% accuracy, 94.92% macro-F1, and 97.76% AUC under highly imbalanced traffic conditions. These results show that the model could distinguish early APT-like flows from normal campus traffic without relying on large traffic bursts or fixed attack signatures. Compared with the single classifiers, the ensemble model achieved a better balance between recall and false alarm control. CatBoost handled mixed numerical and categorical traffic features well. Extra Trees reduced prediction variance in high-dimensional feature space. Balanced Random Forest improved the learning of minority attack samples. Their combination reduced the limitations of each single model. This result is consistent with recent ensemble-based intrusion detection studies, in which multi-model methods usually achieved higher F1 scores than individual classifiers on uneven network datasets (Jellison, 2026). For example, Xu S. et al. showed that ensemble models kept better F1 performance than several traditional models across different IoT anomaly datasets, especially when data distributions were imbalanced (Fig.1).

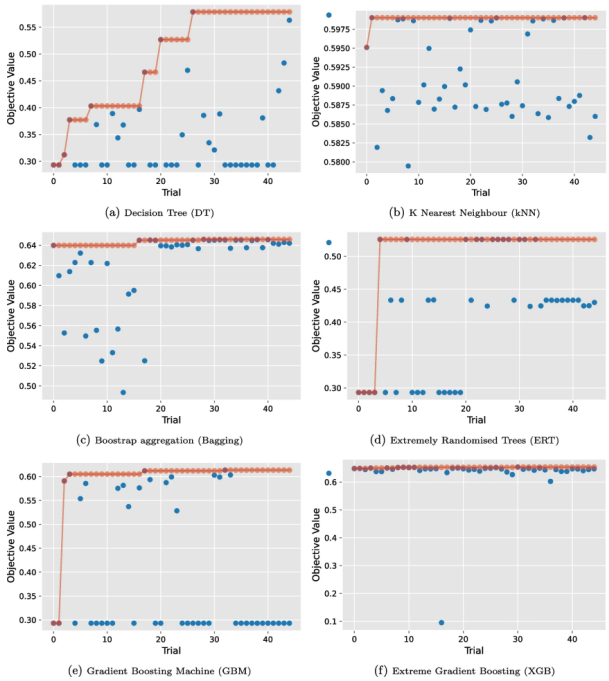

Fig. 1. Detection performance of the proposed ensemble model and baseline classifiers in early-stage APT traffic detection.

3.2 Early-Stage APT Detection Results

The model performed well in detecting early command-and-control beaconing. The recall for this class reached 93.41%, which was 4.68% higher than the best single classifier. This improvement is important because early beaconing traffic usually appears as repeated and low-volume communication, rather than clear attack traffic. The model also showed good sensitivity to internal reconnaissance and credential spraying. In these two scenarios, abnormal flows were sparse and mixed with normal service access. The results suggest that temporal and host-pair features are useful for detecting weak APT signals. In contrast, models that mainly used traffic volume and port information were less stable when the attack traffic was small. This finding is consistent with earlier intrusion detection studies, which reported that simple traffic-count features are not enough for stealthy attacks (Farzaan et al., 2025; Xu et al., 2026). The proposed model improved detection by

combining periodic connection interval, low-volume destination repetition, rare external endpoint access, failed authentication-related flow ratio, and off-hour communication frequency.

3.3 False Alarm Control and Comparison with Existing Studies

After probability calibration, the false alarm rate was reduced to 2.36%. This rate is acceptable for a campus network because thousands of endpoints can turn a small false positive rate into many daily alerts (Huang, 2026). Probability calibration reduced unstable alert scores and made threshold selection more reliable. Compared with recent ensemble IDS studies, this work paid more attention to early-stage detection and alert usability, rather than only reporting overall accuracy. Sufyan A. et al. proposed an ensemble IDS framework and compared several individual and ensemble models on public intrusion datasets. Their results showed that ensemble learning could improve generalization across different network intrusion datasets. However, their study mainly used benchmark data and did not focus on early APT traffic in campus networks (Fig.2). In contrast, the present study used a 28-day campus-scale dataset with 14 subnets, 9,600 endpoints, and injected APT-like scenarios. This setting better reflects the difficulty of detecting low-volume abnormal traffic in daily enterprise operations.

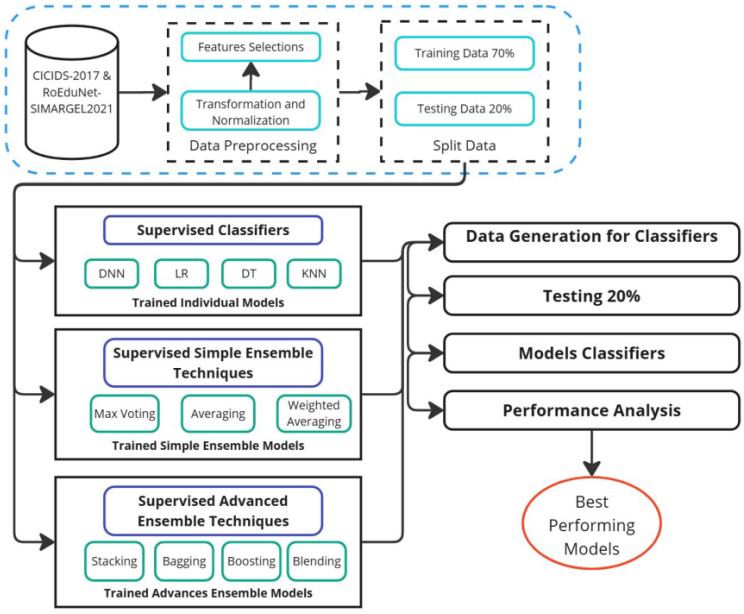

Fig. 2. SHAP feature contribution results for the main traffic indicators of early-stage APT activity.

3.4 SHAP Explanation and Security Meaning

The SHAP results showed that periodic low-volume connections, rare external destination access, abnormal DNS behavior, and off-hour communication were the main indicators of early APT activity. These features are consistent with common APT behavior. Command-and-control channels often use repeated connections with similar time intervals. Reconnaissance and credential probing may cause uncommon host-pair access and failed authentication-related flows. Staged exfiltration may appear as repeated access to rare external endpoints before a larger data transfer. Local SHAP results also explained single alerts by showing which features increased the suspicious score. This is useful for security analysts because an alert is hard to verify if no reason is given (Ma, 2026; Sufyan et al., 2026). Compared with black-box models, the proposed method provides both a detection score and an explanation. However, this study still has limitations. The APT-like scenarios were injected under controlled conditions, and real attackers may use more varied evasion

strategies. Future work should test the model over longer periods, across different campus networks, and with real incident records.

4. Conclusion

This study developed a SHAP-guided ensemble learning method for detecting early-stage APT traffic in enterprise campus networks. The model combined CatBoost, Extra Trees, and Balanced Random Forest to identify weak abnormal signals from flow-level and temporal communication features. Tests on a campus-scale dataset showed that the method achieved 96.87% accuracy, 94.92% macro-F1, and 97.76% AUC under highly imbalanced traffic conditions. For early command-and-control beaconing, the recall reached 93.41%, which was higher than that of the best single classifier. After probability calibration, the false alarm rate was reduced to 2.36%. These results indicate that the ensemble structure improved early detection and reduced unstable alerts. SHAP analysis showed that periodic low-volume connections, rare external destination access, abnormal DNS behavior, and off-hour communication were key signs of early APT activity. The main value of this study lies in combining accurate detection with clear model explanation. This design can help security analysts check suspicious flows before large-scale compromise occurs. The method may support enterprise security monitoring, campus network defense, and early warning systems. However, the APT-like scenarios were added in a controlled setting, and real attackers may use more complex evasion methods. Future work should test the method over longer periods, with real incident records, and in networks with different service structures.

Acknowledgments

We are grateful to all respondents who participated in this study.

References

1. Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E. H. M., Shahzad, T., & Khan, M. A. (2025). Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. Scientific Reports, 15(1), 1726.
2. Avgetidis, A., Faulkenberry, A., Adjibi, V., Galloway, T., Kintis, P., Alrawi, O., ... & Antonakakis, M. (2025, October). From Concealment to Exposure: Understanding the Lifecycle and Infrastructure of APT Domains. In 2025 28th International Symposium on Research in Attacks, Intrusions and Defenses (RAID) (pp. 488-505). IEEE.
3. Chowdhury, T. K. (2025). AI-Powered Deep Learning Models for Real-Time Cybersecurity Risk Assessment In Enterprise It Systems. ASRC Procedia: Global Perspectives in Science and Scholarship, 1(01), 675-704.
4. Deng, X., & Yang, J. (2026). Design of a Quantitative Futures Trading Model Incorporating Multimodal Feature Fusion and Tail Risk Control. Available at SSRN 6702398.
5. Dornala, S. K., & Senthilkumar, P. (2025). Real-Time threat identification and categorization in network traffic using deep learning behavioral analysis. International Journal on Smart Sensing and Intelligent Systems, 18(1).
6. Farhan, M., Waheed Ud Din, H., Ullah, S., Hussain, M. S., Khan, M. A., Mazhar, T., ... & Jaghdam, I. H. (2025). Network-based intrusion detection using deep learning technique. Scientific Reports, 15(1), 25550.
7. Farzaan, M. A. M., Ghanem, M. C., El-Hajjar, A., & Ratnayake, D. N. (2025). AI-powered system for an efficient and effective cyber incidents detection and response in cloud

environments. IEEE Transactions on Machine Learning in Communications and Networking, 3, 623-643.
8. Gao, G., Gao, R., Gao, R., Zhou, H., & Lu, C. (2026, March). Performance Study of Enterprise SMS Communication Scheduling Mechanisms in Triple-Play Convergence Environments. In 2026 IEEE 8th International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 82-86). IEEE.
9. Huang, R. (2026). ConfSched: Confidence-Threshold Routing for Efficient Edge-Cloud Activity Recognition. World Journal of Engineering and Technology, 14(2), 471-486.
10. Jellison, K. T. (2026). Women in Contemporary Japanese Hard Rock and Heavy Metal: Idol Images Outside, Rock Authority Inside (Doctoral dissertation, University of California, Irvine).
11. Khayat, M., Barka, E., Serhani, M. A., Sallabi, F., Shuaib, K., & Khater, H. M. (2025). Reinforcement learning with deep features: A dynamic approach for intrusion detection in IoT networks. IEEE Access.
12. Li, Y., & Liu, S. (2026, May). A Study on Dynamic Optimization of Alerting Policies and Multi-Agent Decision-Making Mechanisms in Cloud Environments. In 2026 7th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) (pp. 703-706). IEEE.
13. Ma, Y. (2026). Industrial Financial Risk Prediction Model Based on Graph Neural Network and Knowledge Graph Inference. Journal of Circuits, Systems and Computers, 35(16), 2650103.
14. Mahmood, A., Abbas, H., Amjad, F., Shahid, W. B., & Shah, S. Q. A. (2025). Real-Time Detection and Multi-Class Classification of DGAs with HybridBERT. IEEE Access.
15. Nguyen, T. H., Le, V. H., & Nguyen, T. D. (2026). A Smart Rule-Generation Approach for Network Intrusion Prevention Systems. KSII Transactions on Internet & Information Systems, 20(3), 1492.
16. Qi, C., & Qiao, X. (2026). Using AI to Monitor AI: Automated Operations Through Log-Driven Intelligence. Available at SSRN 6795840.
17. Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems, 7(1), 2400304.
18. Shao, W. (2026). Policy-Driven Vulnerability Risk Quantification framework for Large-Scale Cloud Infrastructure Data Security. arXiv preprint arXiv:2604.06252.
19. Su, D., & Wu, Y. (2026). Multilevel Growth and Social Network Modeling for Functional Communication Enhancement in Students with Intellectual Disabilities.
20. Subrahmanyam, S. (2025). Behavioral Analysis for Threat Detection. In Handbook of AI-Driven Threat Detection and Prevention (pp. 95-115). CRC Press.
21. Sufyan, A., Mujeeb-Ur-Rehman, M., Noreen, B., & Amin, S. (2026). Trends, capabilities, and challenges in modern cyber defense: A systematic review of detection and response technologies. Spectrum of Engineering Sciences, 4(1), 464-503.
22. Vasheghani, S., & Sharifi, S. (2025). Dynamic ensemble learning for robust image classification: A model-specific selection strategy. Available at SSRN 5215134.
23. Wang, Y., Chen, J., Arias, R., Wang, Y., & Yin, X. (2026). Development and Validation of a Patient-Friendly Digital Assessment Platform for Precision Screening of Oral Anti-Obesity Medications (AOMs).
24. Xu, S., Ma, Q., Liu, H., & Yue, L. (2026). Continuous Reorganization and Performance Preservation of Agent Memory Structure Under Distributed Change Environments.
25. Xue, Z., Zi, Y., Qi, N., Gong, M., & Zou, Y. (2025, August). Multi-level service performance forecasting via spatiotemporal graph neural networks. In 2025 5th International Conference on

Intelligent Communications and Computing (ICICC) (pp. 202-206). IEEE.
26. Yang, J. (2026). Stage-Coupled Computational Framework for Stratified Accessibility and Equity Analysis in Community-Based Elderly Care Services.
27. Yin, J., Huang, Y., & Rao, H. (2026). A Study on User Behavior Signal-Driven Predictive Models for Digital Platform Consumption Trends. Available at SSRN 6607298.
28. Zhang, Z. (2026). A Study on Return Optimization in E-commerce for Complex Consumer Goods Driven by Installation Information Quality. Available at SSRN 6734720.
29. Zhang, Z., Gao, Y., & Tong, Y. (2026). Semantic-Temporal Graph Learning for Demand Forecasting and Resource Allocation in Public Information Systems. Available at SSRN 6792279.
30. Zhu, W., Yao, Y., & Yang, J. (2025). Optimizing Financial Risk Control for Multinational Projects: A Joint Framework Based on CVaR-Robust Optimization and Panel Quantile Regression.