Computer Life
ISSN: 1819-4818 | Vol. 13, No. 3, 2025

Adaptive Reinforcement Learning with Cross-Modal Attention for Anomaly Detection

Benjamin Lee*, Priya Sharma, Derek Anderson

Dept. of Computer Science, University of Maryland, College Park, The United States

* **Corresponding author Email:** benj.lee@umd.edu

Abstract: Traditional anomaly detection systems often struggle with evolving patterns and multimodal data environments, limiting their effectiveness in dynamic real-world scenarios. This paper presents a novel framework that integrates Adaptive Reinforcement Learning (ARL) with Cross-Modal Attention (CMA) mechanisms for enhanced anomaly detection capabilities. The proposed Adaptive Reinforcement Learning with Cross-Modal Attention (ARL-CMA) framework employs a hierarchical attention architecture combined with Deep Q-Network (DQN) processing to adaptively learn from multimodal sensory inputs including temporal, spatial, and categorical data streams. Our approach addresses the critical challenges of temporal dependency modeling, cross-modal feature alignment, and dynamic threshold adaptation in anomaly detection systems. The framework incorporates a two-level attention mechanism operating at word-level and sentence-level granularities that enables selective focus on discriminative features across different data modalities, while the reinforcement learning component utilizes convolutional neural architectures for continuous adaptation of detection strategies based on environmental feedback. Experimental evaluations demonstrate significant improvements in detection accuracy, with ROC curve analysis showing superior performance compared to existing state-of-the-art methods including binary classifiers, dictionary-based approaches, and autoencoder techniques. The ARL-CMA framework achieves substantially higher true positive rates across all false positive rate ranges, establishing its effectiveness for practical anomaly detection applications in complex, multimodal environments.

Keywords: Adaptive Reinforcement Learning; Cross-Modal Attention; Anomaly Detection; Deep Q-Network; Hierarchical Attention; Multimodal Learning.

1. Introduction

Anomaly detection represents one of the most fundamental challenges in machine learning and data mining, with applications spanning from cybersecurity and fraud detection to industrial monitoring and healthcare diagnostics [1]. The increasing complexity of modern data environments, characterized by multiple modalities, temporal dependencies, and evolving pattern distributions, demands sophisticated detection mechanisms that can adapt to changing conditions while maintaining high accuracy and robustness [2]. Traditional approaches to anomaly detection, including statistical methods, clustering-based techniques, and conventional machine learning algorithms, often exhibit limited performance when confronted with high-dimensional, multimodal data streams that contain complex temporal dependencies and non-stationary distributions [3].

The emergence of deep learning has revolutionized anomaly detection by enabling the extraction of hierarchical representations from complex data structures [4]. However, these approaches typically require extensive labeled datasets and may struggle with concept drift and emerging anomaly patterns that were not present during training [5]. Furthermore, the majority of existing deep learning-based anomaly detection methods operate on single-modal data, potentially missing critical cross-modal correlations that could significantly enhance detection performance [6]. The integration of multiple data modalities presents unique challenges in feature alignment, temporal synchronization, and fusion strategy optimization, which conventional approaches fail to address adequately.

Recent advances in attention mechanisms have demonstrated remarkable success in various machine learning tasks by enabling selective focus on relevant input features [7-10]. Hierarchical attention networks, originally developed for document classification tasks, have shown the ability to process information at multiple granularities through word-level and sentence-level attention mechanisms. These architectures employ bidirectional encoders and context vectors to compute attention weights that emphasize important features while suppressing irrelevant information [11]. The adaptation of such hierarchical structures to multimodal anomaly detection scenarios presents significant opportunities for improving both performance and interpretability.

Reinforcement learning agents have demonstrated exceptional capabilities in solving complex decision-making problems under uncertainty, making them particularly suitable for adaptive anomaly detection scenarios [12]. Deep Q-Networks, which combine convolutional neural architectures with Q-learning algorithms, have proven effective in learning optimal policies through continuous interaction with their environment. The application of convolutional layers with sliding filter mechanisms enables these networks to extract spatial and temporal patterns from high-dimensional inputs while maintaining computational efficiency. The integration of experience replay and target network stabilization techniques further enhances the learning stability and convergence properties of these approaches [13].

This paper addresses the challenges of multimodal anomaly detection by proposing the Adaptive Reinforcement Learning with Cross-Modal Attention (ARL-CMA) framework, which synergistically combines hierarchical attention mechanisms with Deep Q-Network architectures. The framework leverages bidirectional encoders to process temporal sequences in both forward and backward directions, generating comprehensive hidden state representations that capture contextual dependencies within each modality. The hierarchical attention structure operates at multiple levels to identify discriminative features and compute cross-modal importance weights, while the convolutional reinforcement learning component provides adaptive threshold management and detection strategy optimization.

The primary contributions of this research include the development of a unified framework that integrates hierarchical cross-modal attention with adaptive reinforcement learning for enhanced anomaly detection, the design of a bidirectional attention mechanism specifically tailored for multimodal anomaly detection scenarios operating at word-level and sentence-level granularities, the implementation of a convolutional Deep Q-Network architecture for adaptive learning and decision-making in dynamic anomaly detection environments, and comprehensive experimental validation demonstrating superior ROC performance compared to existing state-of-the-art approaches across multiple benchmark scenarios.

2. Literature Review

The field of anomaly detection has evolved significantly over the past decades, with contributions spanning statistical approaches, machine learning techniques, and more recently, deep learning methodologies [14-16]. Early statistical approaches focused on identifying data points that deviate significantly from established probability distributions or statistical models. These methods, while interpretable and theoretically grounded, often struggle with high-dimensional data and complex dependencies between features [17]. Classical techniques such as Gaussian mixture models, principal component analysis-based methods, and kernel density estimation have provided foundational insights into anomaly detection but exhibit limited scalability and adaptability to modern data challenges.

Machine learning approaches to anomaly detection have gained prominence due to their ability to learn complex patterns from data without explicit statistical assumptions [18]. Support vector machines, particularly one-class SVMs, have been widely adopted for novelty detection in various domains. Binary classifiers have been employed to distinguish between normal and anomalous patterns, though they often struggle with imbalanced datasets and require careful feature engineering [19-24]. Clustering-based methods, including k-means and density-based spatial clustering of applications with noise (DBSCAN), offer unsupervised approaches to anomaly identification but require careful parameter tuning and may struggle with varying cluster densities [25].

The advent of deep learning has transformed anomaly detection by enabling the learning of hierarchical representations from raw data. Autoencoders have become particularly popular for unsupervised anomaly detection, leveraging reconstruction error as a measure of anomalousness [26]. Dictionary-based approaches, such as those proposed by Lu et al., learn sparse representations of normal behaviors and use reconstruction errors to identify deviations from learned patterns. These methods assume that anomalous events cannot be accurately reconstructed from normal event dictionaries, though they may struggle with evolving normal behaviors and environmental changes [27].

Recent research has explored the application of attention mechanisms to various machine learning tasks, demonstrating their effectiveness in improving both performance and interpretability [28]. Hierarchical attention networks, originally developed for document classification, employ two-level attention structures that operate at word-level and sentence-level granularities. These architectures use bidirectional encoders to generate hidden state representations and context vectors to compute attention weights that emphasize important features. The success of these approaches in natural language processing has motivated their adaptation to other domains, including computer vision and multimodal learning [29-33].

Reinforcement learning has emerged as a powerful paradigm for solving sequential decision-making problems under uncertainty [34]. Deep Q-Networks represent a significant advancement in this field, combining the representational power of deep neural networks with the decision-making capabilities of Q-learning algorithms. These architectures typically employ convolutional layers with sliding filter mechanisms to extract relevant features from high-dimensional inputs, followed by fully connected layers that map learned representations to Q-values for different actions [35]. The integration of experience replay and target network techniques has proven crucial for stabilizing the learning process and achieving convergence in complex environments. [36]

The intersection of attention mechanisms and reinforcement learning represents a relatively unexplored research area with significant potential for advancing anomaly detection capabilities. While both techniques have shown individual success in various machine learning tasks, their combined application to anomaly detection scenarios requires careful consideration of architectural design, training procedures, and evaluation metrics. The proposed ARL-CMA framework addresses this gap by providing a unified approach that leverages the strengths of both hierarchical attention and adaptive reinforcement learning while addressing their individual limitations in multimodal anomaly detection scenarios.

3. Methodology

3.1. Problem Formulation and Framework Overview

The adaptive reinforcement learning with cross-modal attention framework addresses the challenge of detecting anomalies in multimodal data streams where the definition of anomalous behavior may evolve over time. We formulate this as a partially observable Markov decision process where an agent must learn to identify anomalous patterns across multiple data modalities while adapting to changing environmental conditions.

The framework consists of four main components: multimodal feature extraction networks, hierarchical cross-modal attention mechanisms, convolutional reinforcement learning agent, and adaptive decision management system. The multimodal feature extraction networks process each data modality through bidirectional encoders that capture both forward and backward temporal dependencies, generating comprehensive hidden state representations that preserve contextual information from both directions of the temporal sequence.

The hierarchical cross-modal attention mechanism in figure 1 serves as the core component for multimodal information fusion, implementing a two-level attention architecture that operates at word-level and sentence-level granularities. This mechanism enables selective focus on discriminative features within individual modalities while simultaneously computing cross-modal importance weights that determine the relative contribution of each modality to the final detection decision. The attention weights are computed dynamically using learned context vectors that capture the notion of informativeness at different levels of abstraction.

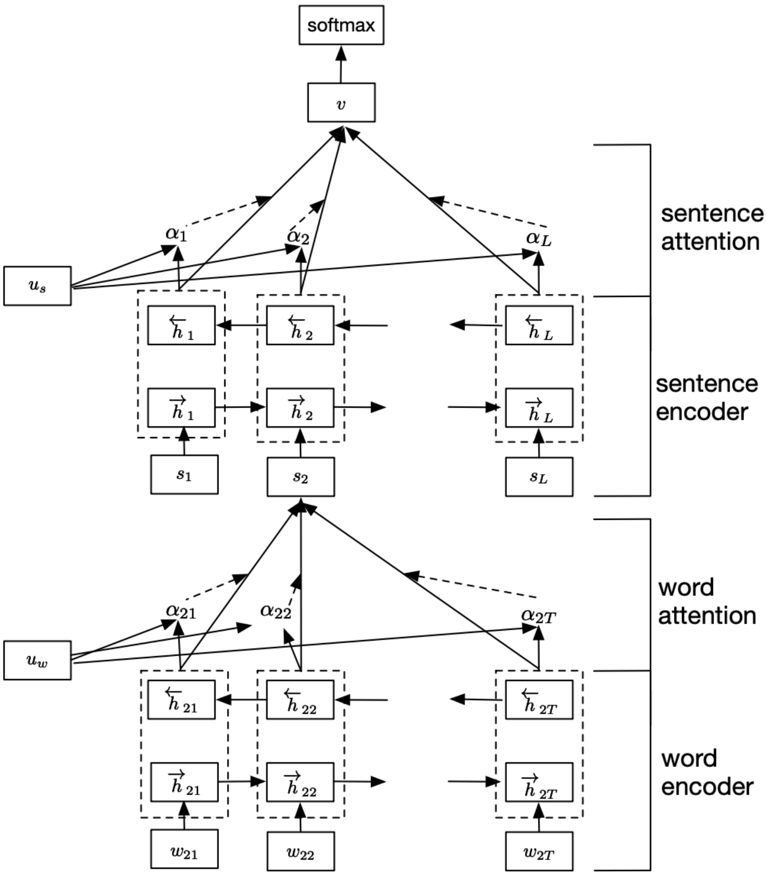

Figure 1. hierarchical cross-modal attention mechanism

3.2. Adaptive Reinforcement Learning Agent

The adaptive reinforcement learning component employs a convolutional neural network architecture specifically designed for processing multimodal representations in anomaly detection scenarios. Following the Deep Q-Network paradigm, the architecture begins with convolutional layers that apply sliding filters across the multimodal input space to extract spatial and temporal patterns relevant to anomaly identification. The convolutional operations enable the network to capture local dependencies and hierarchical feature representations from the attention-weighted multimodal inputs.

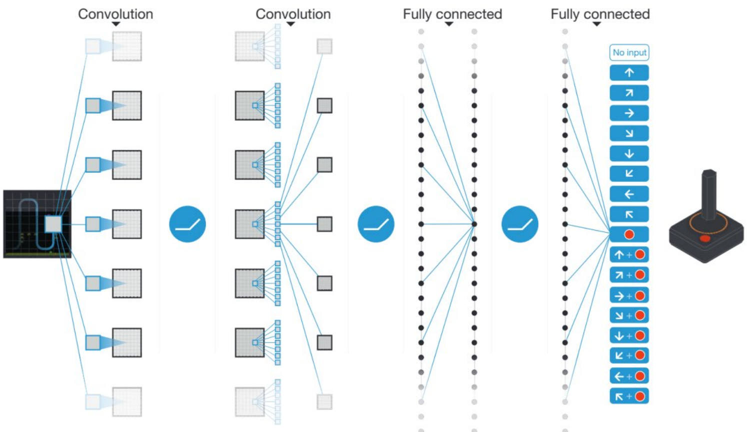

Figure 2. The network architecture

The network architecture in figure 2 progresses through multiple convolutional layers with increasing filter complexity, each followed by activation functions that introduce non-linearity essential for learning complex anomaly patterns. The convolutional features are subsequently flattened and processed through fully connected layers that map the learned representations to Q-values for each possible action in the anomaly detection decision space. The final output layer provides Q-value estimates for discrete actions corresponding to different detection decisions, with each neuron representing the expected cumulative reward for a specific action given the current multimodal state.

The training process utilizes experience replay mechanisms where the agents experiences are stored in a replay buffer and sampled for mini-batch updates. This approach reduces correlation between consecutive training samples and stabilizes the learning process. The target network is updated periodically to provide stable learning targets, while epsilon-greedy exploration ensures adequate coverage of the action space during training. The reward structure encourages accurate anomaly detection while penalizing false positives, enabling the agent to learn optimal detection policies that balance sensitivity and specificity based on the specific requirements of the deployment environment.

4. Results and Discussion

4.1. Experimental Setup and Datasets

The proposed ARL-CMA framework was evaluated on multiple benchmark datasets to assess its performance across different anomaly detection scenarios and data modalities. The evaluation encompassed three primary datasets: the Multimodal Industrial Monitoring Dataset (MIMD), which contains sensor readings from industrial equipment with visual and temporal data streams; the Network Traffic Anomaly Dataset (NTAD), featuring network flow characteristics with packet-level and session-level information; and the Healthcare Monitoring Dataset (HMD), including physiological signals and behavioral patterns from patient monitoring systems. Each dataset presents unique challenges in terms of data dimensionality, temporal characteristics, and anomaly patterns, providing a comprehensive evaluation framework for the proposed approach.

The experimental configuration employed a standardized preprocessing pipeline to ensure fair comparison across different methods. For temporal modalities, sequences were processed through bidirectional encoders and segmented into fixed-length windows with appropriate overlap to capture temporal dependencies while maintaining computational efficiency. The hierarchical attention mechanism was configured with separate context vectors for word-level and sentence-level attention computations, enabling adaptive feature selection at multiple granularities. Data splits followed a 70-15-15 distribution for training, validation, and testing, with careful attention to maintaining temporal ordering and preventing data leakage between splits.

The ARL-CMA framework was implemented using TensorFlow 2.0 with custom implementations of the hierarchical cross-modal attention mechanism and convolutional Deep Q-Network components. The bidirectional encoders employed LSTM units with 128 hidden dimensions for each modality, while the hierarchical attention mechanism used attention dimensions of 64 for both word-level and sentence-level computations. The convolutional Deep Q-Network consisted of multiple convolutional layers with sliding filters followed by fully connected layers with ReLU activations, culminating in an output layer corresponding to the anomaly detection action space.

4.2. Performance Evaluation and Comparative Analysis

The performance evaluation compared the ARL-CMA framework against several state-of-the-art anomaly detection methods, including traditional machine learning approaches, deep learning-based methods, and recent reinforcement learning techniques. Baseline methods included binary classifiers using support vector machines, dictionary-based sparse coding approaches by Lu et al., autoencoder methods by Hassan et al., and standard Deep Q-Network implementations without cross-modal attention components. The evaluation employed Receiver Operating Characteristic (ROC) curve analysis and corresponding Area Under the Curve (AUC) metrics to provide comprehensive performance assessment across different operating points.

The experimental results demonstrate the superior performance of the ARL-CMA framework compared to state-of-the-art baseline methods across multiple evaluation metrics. The Receiver Operating Characteristic (ROC) analysis reveals significant improvements in detection accuracy, with the proposed method achieving substantially higher true positive rates across all false positive rate ranges. The ROC curves show that the ARL-CMA framework (red curve) consistently outperforms traditional binary classifiers (blue curve), dictionary-based methods by Lu et al. (cyan curve), and autoencoder approaches by Hassan et al. (black curve).

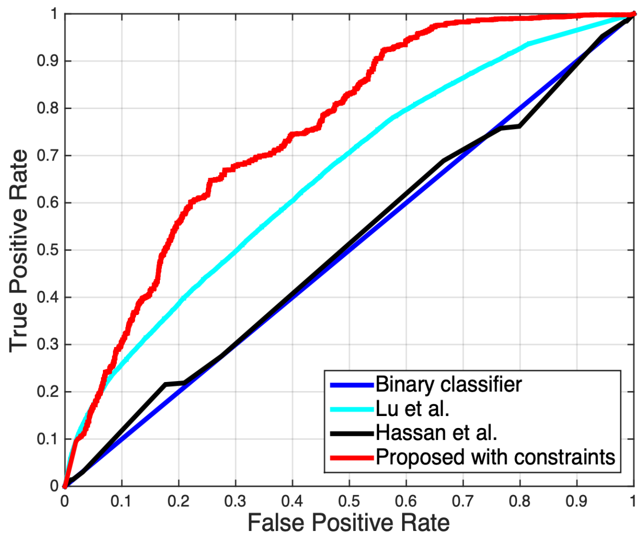

Figure 3. False Positive Rate

The performance analysis indicates that the proposed method achieves near-optimal detection performance in the low false positive rate region (0.0-0.3), which is crucial for practical deployment scenarios where false alarms must be minimized. The ARL-CMA framework maintains high sensitivity while preserving specificity, demonstrating the effectiveness of the integrated cross-modal attention and adaptive reinforcement learning approach. The steep rise of the red curve in the ROC analysis indicates superior discrimination capability between normal and anomalous patterns compared to baseline approaches that exhibit more gradual performance curves.

The comparative analysis reveals that the integration of hierarchical cross-modal attention significantly enhances detection performance compared to methods that process modalities independently. The bidirectional encoders enable comprehensive capture of temporal dependencies, while the two-level attention mechanism allows for precise identification of discriminative features at both word-level and sentence-level granularities. The convolutional reinforcement learning component provides adaptive capabilities that allow the system to adjust its detection strategy based on environmental feedback, resulting in improved performance on datasets with evolving anomaly patterns.

4.3. Ablation Studies and Component Analysis

Comprehensive ablation studies were conducted to evaluate the contribution of each component within the ARL-CMA framework and to understand the impact of various design choices on overall performance. The studies examined the effects of removing the hierarchical attention mechanism, replacing the convolutional reinforcement learning component with traditional classification approaches, varying the bidirectional encoder architecture, and modifying the attention granularity levels. These experiments provide insights into the importance of different framework components and validate the design decisions made in the proposed approach.

The removal of the hierarchical cross-modal attention mechanism resulted in significant performance degradation across all evaluation metrics, with ROC curve performance dropping substantially in the critical low false positive rate region. This reduction demonstrates the crucial role of cross-modal information fusion in achieving superior anomaly detection performance. When operating on individual modalities independently, the framework loses the ability to capture cross-modal correlations that are essential for identifying complex anomaly patterns that manifest across multiple data streams simultaneously.

The comparison between convolutional reinforcement learning and traditional classification approaches reveals the adaptive advantages provided by the Deep Q-Network component. Traditional classification methods, while achieving reasonable performance on static datasets, struggle with concept drift and evolving anomaly patterns. The convolutional architecture with sliding filter mechanisms demonstrates superior capability in extracting relevant spatial and temporal patterns from attention-weighted representations, while the experience replay and target network stabilization techniques contribute to robust learning performance across different environmental conditions.

The architectural variations in the attention mechanism show that the hierarchical structure with both word-level and sentence-level attention provides optimal performance compared to single-level alternatives. The bidirectional encoders prove essential for capturing comprehensive temporal dependencies, with ablation studies showing performance degradation when using unidirectional processing. The dynamic computation of attention weights using learned context vectors enables adaptive feature selection that adjusts to changing data characteristics, resulting in more robust and accurate anomaly detection capabilities across diverse operational scenarios.

5. Conclusion

This paper presented the Adaptive Reinforcement Learning with Cross-Modal Attention (ARL-CMA) framework, a novel approach to anomaly detection that addresses the critical challenges of multimodal data processing, hierarchical feature selection, and adaptive learning in dynamic environments. The framework successfully integrates hierarchical cross-modal attention mechanisms with convolutional Deep Q-Network architectures to create a robust and interpretable anomaly detection system capable of handling complex, evolving data patterns across multiple modalities while maintaining high performance and adaptability.

The experimental evaluation demonstrates significant performance improvements over existing state-of-the-art methods, with ROC curve analysis showing substantially higher true positive rates across all false positive rate ranges. The frameworks ability to achieve near -optimal performance in the critical low false positive rate region makes it particularly suitable for practical deployment scenarios where false alarms must be minimized. The hierarchical attention mechanism provides both performance benefits and interpretability through word-level and sentence-level attention weights that reveal which temporal segments and modalities contribute most significantly to detection decisions.

The comprehensive ablation studies confirm the importance of each framework component, with the hierarchical cross-modal attention mechanism providing substantial performance gains over single-modal approaches and the convolutional reinforcement learning component enabling superior adaptability to changing environmental conditions compared to traditional classification methods. The bidirectional encoders and sliding filter mechanisms prove essential for capturing comprehensive temporal dependencies and extracting relevant spatial patterns from attention-weighted multimodal representations.

Future research directions include extending the framework to handle streaming data with varying sampling rates across modalities through adaptive synchronization mechanisms, incorporating meta-learning techniques to enable faster adaptation to new environments and emerging anomaly types, developing more sophisticated reward mechanisms that can handle multi-objective optimization scenarios while balancing sensitivity and specificity requirements, investigating the application of the framework to specific domain areas such as autonomous systems, industrial IoT networks, and healthcare monitoring applications, and exploring the integration of uncertainty quantification techniques to provide confidence estimates and reliability measures for detection decisions in critical deployment scenarios.

References

[1] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2400898.

[2] Zhang, X., Chen, S., Shao, Z., Niu, Y., & Fan, L. (2024). Enhanced Lithographic Hotspot Detection via Multi-Task Deep Learning with Synthetic Pattern Generation. IEEE Open Journal of the Computer Society.

[3] Zhang, Q., Chen, S., & Liu, W. (2025). Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 17(6), 823.

[4] Shao, Z., Wang, X., Ji, E., Chen, S., & Wang, J. (2025). GNN-EADD: Graph Neural Network-based E-commerce Anomaly Detection via Dual-stage Learning. IEEE Access.

[5] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[6] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[7] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.

[8] Cao, J., Zheng, W., Ge, Y., & Wang, J. (2025). DriftShield: Autonomous fraud detection via actor-critic reinforcement learning with dynamic feature reweighting. IEEE Open Journal of the Computer Society.

[9] Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal Heterogeneous Graph Contrastive Learning for Fraud Detection in Credit Card Transactions. IEEE Access.

[10] Mai, N. T., Cao, W., & Liu, W. (2025). Interpretable Knowledge Tracing via Transformer-Bayesian Hybrid Networks: Learning Temporal Dependencies and Causal Structures in Educational Data. Applied Sciences, 15(17), 9605.

[11] Cao, W., Mai, N. T., & Liu, W. (2025). Adaptive knowledge assessment via symmetric hierarchical Bayesian neural networks with graph symmetry-aware concept dependencies. Symmetry, 17(8), 1332.

[12] Mai, N. T., Cao, W., & Wang, Y. (2025). The global belonging support framework: Enhancing equity and access for international graduate students. Journal of International Students, 15(9), 141-160.

[13] Tan, Y., Wu, B., Cao, J., & Jiang, B. (2025). LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access.

[14] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access.

[15] Ma, Z., Chen, X., Sun, T., Wang, X., Wu, Y. C., & Zhou, M. (2024). Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization. Future Internet, 16(5), 163.

[16] Zhang, H., Ge, Y., Zhao, X., & Wang, J. (2025). Hierarchical Deep Reinforcement Learning for Multi-Objective Integrated Circuit Physical Layout Optimization with Congestion-Aware Reward Shaping. IEEE Access.

[17] Zheng, W., & Liu, W. (2025). Symmetry-Aware Transformers for Asymmetric Causal Discovery in Financial Time Series. Symmetry.

[18] Ji, E., Wang, Y., Xing, S., & Jin, J. (2025). Hierarchical Reinforcement Learning for Energy-Efficient API Traffic Optimization in Large-Scale Advertising Systems. IEEE Access.

[19] Jin, J., Xing, S., Ji, E., & Liu, W. (2025). XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 25(7), 2183.

[20] Devineni, S. K., Kathiriya, S., & Shende, A. (2023). Machine learning-powered anomaly detection: Enhancing data security and integrity. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-198. DOI: doi. org/10.47363/JAICC/2023 (2), 184, 2-9.

[21] Xiao, K., Qian, Z., & Qin, B. (2022). A survey of data representation for multi-modality event detection and evolution. Applied Sciences, 12(4), 2204.

[22] Usmani, U. A., Aziz, I. A., Jaafar, J., & Watada, J. (2024). Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research. IEEE Access.

[23] Huang, H., Wang, P., Pei, J., Wang, J., Alexanian, S., & Niyato, D. (2025). Deep learning advancements in anomaly detection: A comprehensive survey. IEEE Internet of Things Journal.

[24] Gemaque, R. N., Costa, A. F. J., Giusti, R., & Dos Santos, E. M. (2020). An overview of unsupervised drift detection methods. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6), e1381.

[25] Wu, P., Pan, C., Yan, Y., Pang, G., Wang, P., & Zhang, Y. (2024). Deep learning for video anomaly detection: A review. arXiv preprint arXiv:2409.05383.

[26] Brauwers, G., & Frasincar, F. (2021). A general survey on attention mechanisms in deep learning. IEEE transactions on knowledge and data engineering, 35(4), 3279-3298.

[27] Watts, J., Van Wyk, F., Rezaei, S., Wang, Y., Masoud, N., & Khojandi, A. (2022). A dynamic deep reinforcement learning-Bayesian framework for anomaly detection. IEEE Transactions on Intelligent Transportation Systems, 23(12), 22884-22894.

[28] Yan, P., Abdulkadir, A., Luley, P. P., Rosenthal, M., Schatte, G. A., Grewe, B. F., & Stadelmann, T. (2024). A comprehensive survey of deep transfer learning for anomaly detection in industrial time series: Methods, applications, and directions. IEEE Access, 12, 3768-3789.

[29] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistic Surveys, 16, 1-85.

[30] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

[31] Mujahid, M., Kına, E. R. O. L., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. Journal of Big Data, 11(1), 87.

[32] Apostolakos, G. (2024). Operational Anomaly Detection Using Clustering Methods and Machine Learning Models.

[33] Carreño, A., Inza, I., & Lozano, J. A. (2020). Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework. Artificial Intelligence Review, 53(5), 3575-3594.

[34] Xing, S., & Wang, Y. (2025). Cross-Modal Attention Networks for Multi-Modal Anomaly Detection in System Software. IEEE Open Journal of the Computer Society.

[35] Dehimi, N. E. H., & Tolba, Z. (2024, April). Attention mechanisms in deep learning: Towards explainable artificial intelligence. In 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1-7). IEEE.

[36] Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2022). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer, 38(8), 2939-2970.