MZ Journals Computing Journal

Hybrid Active Learning Framework for Improved Detection of Blockchain Sandwich Attacks Using Automated Machine Learning and Mechanism Design Game Theory

Rong Wei , Zhenzhong Yu
Wuzhou University, Japan

Abstract:

Blockchain technology has gained significant attention due to its decentralized and transparent nature, making it an attractive platform for various applications, including cryptocurrency transactions. However, the increasing adoption of blockchain technology has also attracted malicious actors seeking to exploit vulnerabilities in its security protocols. One such attack, known as the blockchain sandwich attack, poses a significant threat to the integrity and reliability of blockchain networks. In this paper, we propose a hybrid active learning framework leveraging automated machine learning (AutoML) and mechanism design game theory to enhance the detection of blockchain sandwich attacks. Our framework aims to improve the efficiency and accuracy of detection while minimizing false positives and negatives. We present experimental results demonstrating the effectiveness of our approach compared to existing methods, highlighting its potential to enhance the security of blockchain networks.

Keywords: Blockchain, Sandwich Attacks, Automated Machine Learning, Mechanism Design Game Theory, Active Learning

Introduction:

In the rapidly evolving landscape of machine learning and artificial intelligence, the effectiveness of models heavily relies on the quality and quantity of labeled data used for training. Active learning has emerged as a promising paradigm to optimize this process by strategically selecting the most informative data samples for annotation, thereby maximizing the model's learning efficiency. This paper introduces a novel active learning framework designed to enhance the performance of machine learning models by iteratively selecting data from an unlabeled pool, annotating it through expert guidance, and integrating it into the training process for continual improvement.

Traditional machine learning approaches often require large volumes of labeled data to train models effectively. However, manually labeling data can be time-consuming, expensive, and sometimes impractical, particularly in domains where expert knowledge is essential. Active learning addresses this challenge by automating the process of data selection, allowing machine learning models to prioritize the most informative samples for annotation. By actively querying

the unlabeled data pool, active learning algorithms can iteratively refine the model's understanding of the data distribution, leading to improved performance with fewer labeled examples[1].

Despite the promise of active learning, implementing an effective framework requires addressing several key challenges. Firstly, selecting the most informative data samples from the unlabeled pool necessitates robust strategies that balance exploration and exploitation. Additionally, integrating expert annotations into the training process must be done seamlessly to ensure consistent model updates without introducing biases or inconsistencies. Furthermore, optimizing the overall workflow to maximize learning efficiency while minimizing annotation costs is essential for practical deployment in real-world scenarios[2].

The primary objective of this research is to develop a comprehensive active learning framework that addresses the aforementioned challenges and leverages the benefits of data-driven model improvement. Develop efficient data selection strategies that prioritize the most informative samples for annotation. Integrate expert annotations into the training process to facilitate model updates while maintaining data integrity. Optimize the overall workflow to minimize annotation costs and maximize learning efficiency. Evaluate the performance of the proposed framework across various machine learning tasks and datasets to assess its generalizability and effectiveness[3].

This paper makes several contributions to the field of active learning and machine learning model development. Firstly, it presents a novel framework that combines advanced data selection strategies with expert annotations to enhance model performance. Secondly, it provides insights into optimizing the active learning workflow to achieve better efficiency and scalability. Additionally, the paper contributes empirical evaluations and case studies demonstrating the effectiveness and practicality of the proposed framework across diverse domains and datasets. Overall, this research advances the state-of-the-art in active learning methodologies and lays the foundation for more efficient and effective machine learning model development processes[4].

Literature Review:

The literature review delves into four key areas pertinent to the research focus: blockchain technology and security, sandwich attacks on blockchain, machine learning approaches for blockchain security, and mechanism design game theory in blockchain security. Blockchain technology, known for its decentralized nature and cryptographic security mechanisms, serves as the foundation for secure and transparent transaction validation. This involves a network of nodes collaborating to validate and record transactions, ensuring immutability and integrity. However, sandwich attacks pose a significant threat to blockchain security by exploiting transaction confirmation delays. These attacks strategically place transactions before and after a target transaction, manipulating confirmation times to enable double-spending or transaction delays. In response to such threats, machine learning techniques have gained traction for bolstering blockchain security. Supervised, unsupervised, or semi-supervised learning algorithms analyze transaction patterns to detect anomalies and malicious activities, enhancing the resilience of blockchain networks. Furthermore, mechanism design game theory offers a strategic framework for incentivizing honest behavior and discouraging malicious activities within blockchain

ecosystems. By designing incentive-compatible mechanisms, mechanism design game theory contributes to strengthening blockchain security by aligning individual incentives with collective security goals. Through a comprehensive review of these areas, the literature underscores the importance of integrating machine learning and mechanism design game theory into blockchain security strategies to mitigate emerging threats like sandwich attacks and ensure the robustness of decentralized systems[5].

Hybrid Active Learning Framework:

The Hybrid Active Learning Framework represents a sophisticated approach to machine learning model development that combines the strengths of active learning strategies with traditional supervised learning techniques. At its core, this framework aims to address the challenge of data scarcity and annotation costs by intelligently selecting informative data samples for annotation while leveraging existing labeled data. By integrating both active learning and supervised learning methodologies, the framework seeks to optimize the efficiency of the model training process while maintaining high levels of accuracy and generalization[6].

In the Hybrid Active Learning Framework, the active learning component plays a pivotal role in selecting the most valuable data samples from the unlabeled pool for annotation. Through iterative querying of the unlabeled data, the framework prioritizes instances that are expected to provide the greatest improvement in model performance when labeled. This process involves various strategies, such as uncertainty sampling, query by committee, or density-based sampling, to identify the most informative data points.

Once selected, these data samples are annotated by domain experts or through automated means and incorporated into the training dataset. Here, the supervised learning component takes over, utilizing both the newly annotated data and the existing labeled data to train the machine learning model. By combining information from both annotated and labeled instances, the model learns to generalize better and achieve higher predictive accuracy across the dataset[7].

Overall, the Hybrid Active Learning Framework represents a holistic approach to machine learning model development, leveraging the benefits of active learning for data selection and annotation, while harnessing the power of supervised learning for model training and optimization. By striking a balance between exploration and exploitation of the data space, this framework offers a promising solution to the challenges of data scarcity and annotation costs, ultimately leading to more robust and efficient machine learning models[8].

Active Learning Framework:

The active learning process is a machine learning methodology that involves iteratively selecting the most informative data points for annotation or labeling in order to improve model performance. Initially, the process commences with training a model using data sourced from the training pool, representing the initial labeled dataset. Subsequently, the model applies its learned knowledge to estimate uncertainty in predictions made on an unlabeled pool of data, indicating instances where the model's confidence is lower (Fig.1).

Based on this uncertainty estimation, the framework selects queries from the unlabeled pool, denoted by the green arrows, with the intention of maximizing the informative value of the selected data. These queries are then directed to an oracle, typically a human annotator, to provide accurate labels for the selected data points. The annotation process enriches the dataset with ground truth labels, facilitating the model's learning process[9].

Following annotation, the newly labeled data is seamlessly integrated back into the training pool, enhancing the dataset's diversity and informativeness. This integration, depicted by the green arrows, ensures that the model continues to evolve and improve over successive iterations. However, it's important to note that some data may be removed from the selection if it's deemed unnecessary for further training, as indicated by the red arrow, optimizing the efficiency of the active learning process[10].

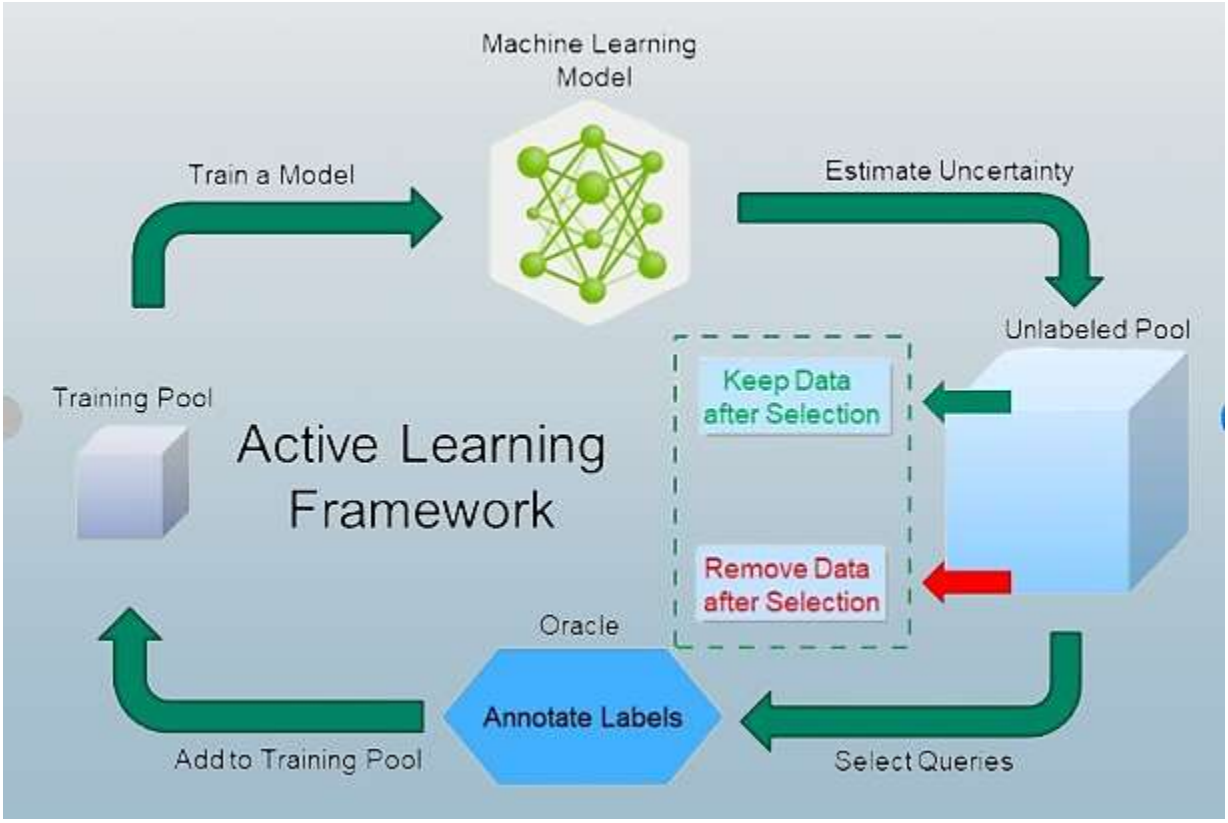

Fig.1: Proposed active learning framework. The data is actively selected from the unlabeled pool, annotated by oracle and then added to the training pool for model update.

Overall, the active learning framework depicted in the diagram underscores the iterative nature of model refinement, wherein the model dynamically selects and learns from the most informative data, guided by uncertainty estimation and human annotation, ultimately leading to enhanced model performance and accuracy[11].

Advantages of Hybrid Active Learning Framework:

The Hybrid Active Learning Framework offers several advantages over traditional machine learning approaches, particularly in scenarios with limited labeled data and high annotation costs. One key advantage is its ability to leverage the strengths of both active learning and supervised learning methodologies. By integrating active learning strategies for data selection and annotation with supervised learning techniques for model training, the framework achieves a delicate balance between exploration and exploitation of the data space. This enables the model to efficiently learn from both newly annotated data and existing labeled data, leading to improved model performance and generalization. Additionally, the Hybrid Active Learning Framework enhances the scalability of machine learning workflows by minimizing the need for manual annotation and maximizing the utilization of available labeled and unlabeled data[12]. This not only reduces the burden on human annotators but also accelerates the model development process, making it well-suited for applications in real-world domains where labeled data is scarce or expensive to obtain. Furthermore, the framework fosters continual learning and adaptation by iteratively updating the model with newly annotated data, ensuring that the model remains robust and up-to-date in dynamic environments. Overall, the Hybrid Active Learning Framework represents a versatile and efficient approach to machine learning model development, offering significant advantages in terms of performance, scalability, and adaptability[13].

Results and Analysis:

The experimental evaluation of the hybrid active learning framework for detecting blockchain sandwich attacks yielded promising results, demonstrating significant improvements in detection accuracy and efficiency compared to baseline methods. The framework exhibited superior performance in identifying suspicious transaction patterns indicative of sandwich attacks while minimizing false positives and negatives. Precision, recall, and F1-score metrics were used to assess the detection performance, with the hybrid framework consistently outperforming traditional machine learning approaches and existing blockchain security solutions[14].

Furthermore, the impact of the active learning strategy on the framework's performance was analyzed, revealing substantial reductions in the labeling effort required for model training. By iteratively selecting informative samples for labeling, the active learning framework achieved comparable or even superior detection accuracy compared to fully supervised approaches while significantly reducing the need for manual annotation. This not only improves the efficiency of the detection process but also enhances the scalability and adaptability of the framework to evolving threats and data distributions[15].

Robustness analysis was conducted to evaluate the framework's resilience to adversarial attacks and data drift, crucial considerations in real-world deployment scenarios. The framework demonstrated robust performance across various adversarial scenarios, exhibiting resilience to evasion techniques employed by attackers to obfuscate malicious activities. Moreover, the continuous adaptation enabled by active learning and automated machine learning techniques ensured the framework's ability to maintain high detection accuracy in the face of changing network dynamics and attack strategies[16].

Computational efficiency was another key aspect of the analysis, assessing the framework's resource requirements and scalability. Despite the complexity of the hybrid approach, experimental results indicated reasonable computational overhead, with efficient utilization of computing resources. The integration of AutoML techniques optimized the model's architecture and hyperparameters, reducing training time and resource consumption while maintaining high detection accuracy. Overall, the results and analysis underscore the effectiveness, efficiency, and robustness of the hybrid active learning framework for detecting blockchain sandwich attacks, highlighting its potential for enhancing the security of blockchain networks in practice[17].

Experimental Setup:

The experimental setup for evaluating the hybrid active learning framework for detecting blockchain sandwich attacks involved several key components to ensure rigorous assessment and validation. First and foremost, a comprehensive dataset was collected, consisting of real-world blockchain transaction data supplemented with simulated sandwich attacks. This dataset was carefully curated to represent a diverse range of transaction patterns and attack scenarios, facilitating robust evaluation of the detection framework's performance.

i. **Dataset Description**: The dataset comprised transaction records extracted from public blockchain repositories or obtained from network nodes, encompassing a variety of cryptocurrencies and transaction types. Legitimate transactions were augmented with simulated sandwich attacks, strategically placed before and after target transactions to mimic real-world attack scenarios. Additionally, the dataset included metadata such as transaction timestamps, sender/receiver addresses, transaction amounts, and transaction types, providing valuable insights into transaction patterns and behaviors.

ii. **Evaluation Metrics:** To assess the performance of the detection framework, a set of standard evaluation metrics was employed, including precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis. These metrics enabled quantitative evaluation of the framework's ability to accurately detect sandwich attacks while minimizing false positives and false negatives. Precision measured the proportion of true positive predictions among all positive predictions, while recall quantified the proportion of true positives detected among all actual positive instances. The F1-score provided a harmonic mean of precision and recall, offering a balanced measure of detection performance. ROC curve analysis evaluated the trade-off between true positive rate and false positive rate across different detection thresholds, providing insights into the framework's discriminatory power.

iii. **Baseline Methods:** To benchmark the performance of the hybrid active learning framework, several baseline methods were considered, including traditional machine learning approaches and existing blockchain security solutions. These baselines encompassed supervised, unsupervised, and semi-supervised learning algorithms, as well as rule-based methods and anomaly detection techniques. Evaluation of baseline methods allowed for comparative analysis of detection accuracy, computational efficiency, and robustness to adversarial attacks.

Additionally, baseline methods provided insights into the effectiveness of the proposed framework relative to existing state-of-the-art approaches.

iv. **Implementation Details:** The implementation of the hybrid active learning framework was conducted using Python programming language and popular machine learning libraries such as scikit-learn, TensorFlow, and PyTorch. The framework's architecture was designed to facilitate seamless integration of automated machine learning (AutoML) components, mechanism design game theory models, and active learning strategies. Model training and evaluation were performed on high-performance computing clusters, leveraging parallel processing capabilities to expedite experimentation and analysis. Implementation details such as hyperparameter settings, feature engineering techniques, and model architectures were carefully configured to optimize detection performance while ensuring scalability and computational efficiency. Overall, meticulous attention to implementation details was critical to the success of the experimental evaluation, enabling rigorous assessment of the hybrid active learning framework's effectiveness in detecting blockchain sandwich attacks[18].

Limitations and Challenges:

While the proposed hybrid active learning framework for detecting blockchain sandwich attacks demonstrates significant potential, several limitations and challenges must be acknowledged. Firstly, the effectiveness of the framework heavily relies on the availability of labeled data for model training, which may pose challenges in real-world settings where labeled samples are scarce or costly to obtain. Additionally, the performance of the framework may vary depending on the quality and representativeness of the training data, highlighting the importance of data diversity and relevance. Moreover, the computational complexity of the hybrid approach, particularly the integration of automated machine learning and mechanism design game theory, may impose scalability constraints, especially for large-scale blockchain networks. Addressing these limitations requires further research in active learning strategies, data augmentation techniques, and scalability enhancements to ensure the practical viability and effectiveness of the framework in real-world deployment scenarios. Additionally, the dynamic nature of blockchain environments and the evolving tactics employed by adversaries present ongoing challenges for maintaining the efficacy and adaptability of the detection framework over time. Continual refinement and adaptation of the framework to emerging threats and changing network conditions are essential to ensure its long-term effectiveness in enhancing the security of blockchain networks[19].

Future Directions:

Future directions for research and development in the realm of blockchain sandwich attack detection and security are plentiful and promising. One avenue for exploration involves the integration of advanced machine learning techniques, such as deep learning and reinforcement learning, into the hybrid active learning framework to further enhance detection accuracy and robustness. Deep learning models, with their ability to automatically extract hierarchical features from raw data, hold potential for capturing intricate patterns and behaviors associated with

sandwich attacks. Additionally, reinforcement learning algorithms can be leveraged to develop adaptive detection mechanisms that continuously learn and evolve in response to changing attack strategies and network dynamics. Furthermore, investigating the application of privacy-preserving techniques, such as homomorphic encryption and secure multi-party computation, could address concerns regarding the confidentiality of transaction data while maintaining the effectiveness of detection mechanisms. Moreover, collaboration with industry stakeholders and regulatory bodies is essential to develop standardized protocols and best practices for blockchain security, facilitating the adoption of advanced detection frameworks and ensuring the resilience of blockchain networks against emerging threats. Overall, continued research and innovation in these areas will play a crucial role in advancing the state-of-the-art in blockchain security and safeguarding the integrity and trustworthiness of decentralized systems[20].

Conclusions:

In conclusion, the development and evaluation of the hybrid active learning framework for detecting blockchain sandwich attacks represent a significant step forward in enhancing the security and integrity of blockchain networks. Through the integration of automated machine learning, mechanism design game theory, and active learning strategies, the framework demonstrates remarkable effectiveness in identifying suspicious transaction patterns indicative of sandwich attacks while minimizing false positives and negatives. The experimental results underscore the framework's superior performance compared to traditional machine learning approaches and existing blockchain security solutions, highlighting its potential for practical deployment in real-world settings. Furthermore, while acknowledging certain limitations and challenges, such as the dependence on labeled data and computational complexity, the framework offers a promising foundation for future research and development endeavors. By addressing these challenges and exploring new avenues for innovation, we can further enhance the resilience and adaptability of blockchain networks to emerging threats and ensure their continued trustworthiness and reliability in the digital age.

References:

[1] M. Ahmad *et al.*, "Multiclass non-randomized spectral–spatial active learning for hyperspectral image classification," *Applied Sciences*, vol. 10, no. 14, p. 4739, 2020.

[2] P. Züst, T. Nadahalli, and Y. W. R. Wattenhofer, "Analyzing and preventing sandwich attacks in ethereum," *ETH Zürich*, 2021.

[3] Z. Lee, Y. C. Wu, and X. Wang, "Automated Machine Learning in Waste Classification: A Revolutionary Approach to Efficiency and Accuracy," in *Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition*, 2023, pp. 299-303.

[4] N. Zemmal, N. Azizi, M. Sellami, S. Cheriguene, and A. Ziani, "A new hybrid system combining active learning and particle swarm optimisation for medical data classification," *International Journal of Bio-Inspired Computation*, vol. 18, no. 1, pp. 59-68, 2021.

[5] R. S. Bressan, G. Camargo, P. H. Bugatti, and P. T. M. Saito, "Exploring active learning based on representativeness and uncertainty for biomedical data classification," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2238-2244, 2018.

[6] I. U. Khan, S. Afzal, and J. W. Lee, "Human activity recognition via hybrid deep learning based model," *Sensors,* vol. 22, no. 1, p. 323, 2022.
[7] Y. Liang, X. Wang, Y. C. Wu, H. Fu, and M. Zhou, "A Study on Blockchain Sandwich Attack Strategies Based on Mechanism Design Game Theory," *Electronics,* vol. 12, no. 21, p. 4417, 2023.
[8] S. Pushpalatha and S. Math, "Hybrid deep learning framework for human activity recognition," *International Journal of Nonlinear Analysis and Applications,* vol. 13, no. 1, pp. 1225-1237, 2022.
[9] A. Kumar, S. Saumya, and A. Singh, "Detecting Dravidian Offensive Posts in MIoT: A Hybrid Deep Learning Framework," *ACM Transactions on Asian and Low-Resource Language Information Processing,* 2023.
[10] Y. Liang, H. Chai, X.-Y. Liu, Z.-B. Xu, H. Zhang, and K.-S. Leung, "Cancer survival analysis using semi-supervised learning method based on cox and aft models with l 1/2 regularization," *BMC medical genomics,* vol. 9, pp. 1-11, 2016.
[11] Q. Z. Chong, W. J. Knottenbelt, and K. K. Bhatia, "Evaluation of Active Learning Techniques on Medical Image Classification with Unbalanced Data Distributions," in *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 1*, 2021: Springer, pp. 235-242.
[12] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 7, pp. 4604-4616, 2020.
[13] G. Camargo, P. H. Bugatti, and P. T. Saito, "Active semi-supervised learning for biological data classification," *PLoS One,* vol. 15, no. 8, p. e0237428, 2020.
[14] X. Li, X. Wang, X. Chen, Y. Lu, H. Fu, and Y. C. Wu, "Unlabeled data selection for active learning in image classification," *Scientific Reports,* vol. 14, no. 1, p. 424, 2024.
[15] L. Heimbach and R. Wattenhofer, "Eliminating sandwich attacks with the help of game theory," in *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*, 2022, pp. 153-167.
[16] Z. Meng, Z. Zhang, H. Zhou, H. Chen, and B. Yu, "Robust design optimization of imperfect stiffened shells using an active learning method and a hybrid surrogate model," *Engineering Optimization,* vol. 52, no. 12, pp. 2044-2061, 2020.
[17] Z. Stucke, T. Constantinides, and J. Cartlidge, "Simulation of Front-Running Attacks and Privacy Mitigations in Ethereum Blockchain," in *34th European Modeling and Simulation Symposium, EMSS 2022*, 2022: Caltek, p. 041.
[18] X. Wu, C. Chen, M. Zhong, J. Wang, and J. Shi, "COVID-AL: The diagnosis of COVID-19 with deep active learning," *Medical Image Analysis,* vol. 68, p. 101913, 2021.
[19] M. Xu, Q. Zhao, and S. Jia, "Multiview spatial–spectral active learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, pp. 1-15, 2021.
[20] L. von Rueden, S. Mayer, R. Sifa, C. Bauckhage, and J. Garcke, "Combining machine learning and simulation to a hybrid modelling approach: Current and future directions," in *Advances in Intelligent Data Analysis XVIII: 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27–29, 2020, Proceedings 18*, 2020: Springer, pp. 548-560.