Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 4, No. 3, 2025

Scalable Blockchain Architecture for Efficient Task Scheduling and Storage Optimization

Kaelith Drayven
University of Delaware, Newark, USA
kaelith86@udel.edu

Abstract: With the rapid development of Internet technology and big data computing, distributed systems have been widely used in various fields. However, traditional distributed computing architectures still rely on centralized management models and face challenges such as uneven distribution of computing resources, high storage overhead, and trust issues between nodes. The decentralized, tamper-proof, and traceable characteristics of blockchain technology provide a new solution for distributed computing and storage. This study proposes an efficient distributed system based on blockchain, combining an improved task scheduling strategy, a game-theoretic computing resource allocation mechanism, and an optimized block storage method to improve the fairness and resource utilization of computing tasks while reducing the synchronization overhead of storage nodes. This paper constructs a decentralized computing framework that enables computing nodes to collaborate on tasks through smart contracts and reduces the waste of computing resources by optimizing the consensus mechanism. In addition, to solve the problem of blockchain storage redundancy, this paper proposes a hierarchical storage strategy that uses blockchain sharding technology to reduce the burden on storage nodes and improve system scalability. The results of this study not only help to improve the applicability of blockchain in large-scale computing tasks, but also provide theoretical support and technical paths for the optimization of future distributed computing architectures.

Keywords: Blockchain, distributed computing, task scheduling, storage optimization

1. Introduction

With the rapid development of Internet technology, the global demand for data generation and processing has been growing exponentially. Particularly in domains such as finance, logistics, healthcare, and intelligent manufacturing, distributed data storage and efficient computation have become pressing challenges for enterprises and research institutions[1]. While traditional centralized systems can partially meet the demands of large-scale data processing, their architectural limitations lead to issues such as poor scalability, insufficient fault recovery capabilities, and inefficient utilization of computational and storage resources[2]. Additionally, centralized systems are highly susceptible to cyberattacks—any compromise or failure of a core server could lead to a complete system collapse. Against this backdrop, distributed systems have emerged as a mainstream solution, leveraging multiple computing nodes to achieve efficient data storage and computation. However, traditional distributed systems rely on a centralized trust mechanism, requiring all participating nodes to be pre-authenticated and trusted, which is difficult to implement in an open Internet environment. To address this issue, blockchain technology has introduced a novel approach to building trustworthy distributed systems[3].

Blockchain technology, characterized by decentralization, immutability, and traceability, has become a crucial solution for addressing trust issues in distributed systems. Fundamentally, blockchain is a distributed ledger jointly maintained by multiple nodes, utilizing cryptographic algorithms and consensus mechanisms to ensure data security and consistency. Within a blockchain network, each node can independently participate in data storage and verification without relying on a centralized trust authority. This feature has enabled blockchain technology to be widely adopted in financial transactions, supply chain management, and smart contracts. However, conventional blockchain systems still face significant challenges when handling high-throughput computational tasks, including inefficient resource utilization, low transaction throughput, and high storage overhead. By integrating blockchain with distributed systems, not only can security be enhanced, but a highly efficient and cost-effective computational and storage infrastructure can also be established in open Internet environments, providing robust technical support for various application scenarios.

Both academia and industry have conducted extensive research on the integration of blockchain and distributed systems. Existing distributed computing frameworks, such as Hadoop, Spark, and Flink, offer high scalability in data processing but lack an effective decentralized trust mechanism, continuing to rely on centralized management. In contrast, blockchain technology, through smart contracts, distributed storage, and consensus algorithms, enables secure data interactions without the need for centralized control. Consequently, applying blockchain to distributed systems can not only enhance system robustness but also provide security guarantees for inter-node collaboration. While some studies have explored the application of blockchain in cloud computing and IoT data sharing, significant technical challenges remain in optimizing large-scale computational data storage, resource management, and task scheduling—holds significant theoretical and practical value[4].

This study aims to explore the construction of blockchain-based distributed systems and propose improvements to address the existing bottlenecks of blockchain in distributed computing. Specifically, we will investigate how blockchain's consensus mechanisms and distributed storage technologies can be leveraged to establish a secure and efficient computational network where multiple nodes collaborate without relying on a central trust authority. Additionally, to mitigate computational resource wastage in blockchain systems, we will design an adaptive task scheduling algorithm that dynamically allocates tasks based on nodes' computational capabilities, thereby improving resource utilization. Furthermore, to reduce the storage burden of blockchain systems, we will develop an efficient data storage and retrieval mechanism that ensures data security while enabling lightweight storage. These research outcomes will provide new perspectives for large-scale distributed computing and advance the practical application of blockchain technology in data-intensive computational tasks[5].

Overall, research on blockchain-based distributed systems holds significant real-world relevance and application potential. On one hand, it will provide a more secure and trustworthy operating environment for existing distributed computing architectures while reducing reliance on centralized management. On the other hand, by optimizing the computational and storage efficiency of blockchain, its application scope can be further extended to high-performance computing, intelligent manufacturing, and the Internet of Things. Moreover, this research will contribute to the innovation and development of blockchain technology, facilitating its deeper integration into distributed computing. In the future, as blockchain technology continues to evolve alongside advancements in computational hardware, blockchain-based distributed systems are expected to become a fundamental component of next-generation computing architectures, driving the growth of the global digital economy.

2. Related work

Blockchain-based distributed systems have garnered increasing attention due to their potential to address centralized trust issues and improve system robustness. Zhou et al. [6] applied Temporal Convolutional Networks (TCNs) in high-frequency trading signal prediction on blockchain, demonstrating how deep learning can enhance decision-making efficiency in blockchain environments. Li et al. [7] proposed a reinforcement learning-based adaptive resource scheduling method that dynamically adjusts to complex system environments, offering promising insights for blockchain-integrated distributed computing systems where adaptability and efficiency are essential. Duan [8] conducted a systematic analysis of user interface perception, which can benefit blockchain platform usability through enhanced human-computer interaction design. Li [9] enhanced ResNeXt50 architectures for complex datasets, supporting high-performance learning in resource-intensive blockchain environments. An et al. [10] developed a boundary-aware semantic segmentation framework, while Li et al. [11] proposed a contrastive learning method for fraud detection in e-commerce—both offering powerful insights into optimizing feature representation and trust evaluation in decentralized systems. Zhan [12] explored human activity recognition through spatiotemporal feature learning networks, relevant for decentralized data analytics across edge nodes. Zhang [13] investigated social network user profiling using graph neural networks, which can be adapted to node behavior modeling and anomaly detection in blockchain-based systems. Wang et al. [14] leveraged diffusion models for generative UI design, suggesting ways to automate visual interfaces for managing decentralized resource flows. Wang [15] addressed data imbalance through adaptive weighting in Markov networks, an idea that can be translated to equitable task distribution in heterogeneous computing environments. Xiang et al. [16] proposed a multi-scale fusion transformer for 3D medical image segmentation, while Li et al. [17] designed a CNN-Transformer framework for cross-modal classification—both highlighting the potential of hybrid deep models for handling diverse data types in blockchain applications. Liu et al. [18] presented a transformer-based rule mining approach, enabling dynamic decision-making strategies that could optimize blockchain task scheduling. Gao et al. [19] introduced a hybrid model for few-shot text classification using transfer and meta-learning, and Liao et al. [20] fine-tuned T5 using knowledge graphs to manage complex tasks, providing insights into task adaptability and efficient smart contract design. Liu et al. [21] tackled product description generation under few-shot learning settings, reinforcing the utility of meta-learning under limited data—an often-encountered issue in decentralized systems. Lastly, Cao et al. [22] proposed a rough set-enhanced therapy-assisting metaverse system, whose system optimization strategies and multi-agent coordination offer parallels to blockchain-based collaboration frameworks.

3. Method

In this study, we proposed an efficient distributed system framework based on blockchain, aiming to solve the problems of trust management, computing resource allocation and storage optimization in the existing distributed computing and storage system. The framework adopts an improved consensus mechanism to enable computing nodes to adaptively allocate computing tasks, while optimizing the block storage structure to reduce data redundancy and improve system throughput. The method in this paper mainly includes three parts: computing task scheduling strategy, dynamic resource allocation mechanism and blockchain storage optimization strategy. An example of blockchain network topology is shown in Figure 1.

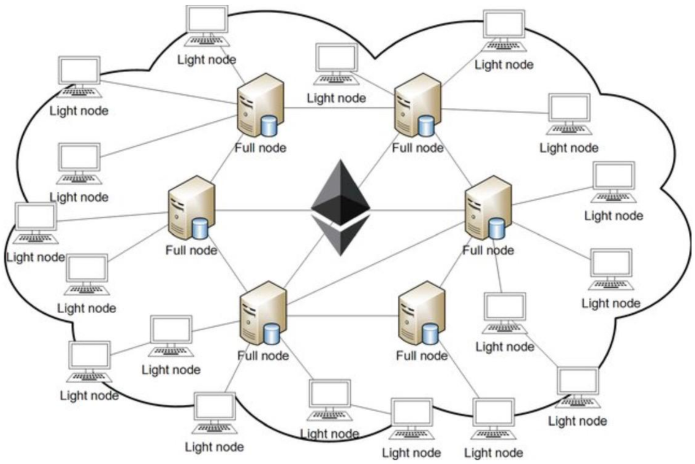

Figure 1. Blockchain network topology diagram

First, in order to improve the execution efficiency of distributed computing tasks, we define the distribution and execution model of computing tasks. Assume that there are N computing nodes in the blockchain network, and define the computing task set as $T = \{T_1, T_2, ..., T_M\}$, where each task T_i has different computing complexity $C(T_i)$ and data dependencies. The computing power of computing node N_j is defined as P_j, and the execution time of its computing task can be expressed as:

$$Texec(T_i, N_j) = \frac{C(T_i)}{P_j}$$

In order to optimize task allocation, we adopt a load balancing strategy based on computing power so that the task allocation matrix $A = [a_{ij}]$ satisfies:

$$a_{ij} = \begin{cases} 1, & \text{If task } T_i \text{ is assigned to node } N_j \\ 0, & \text{Otherwise} \end{cases}$$

The goal is to minimize the computation time of the entire system:

$$\min \max_j \sum_{i=1}^{M} a_{ij} Texec(T_i, N_j)$$

This optimization problem can be solved by an improved greedy algorithm, which allows nodes with high computing power to take on the computing load first while ensuring balanced distribution of tasks.

Secondly, we designed a dynamic allocation mechanism for computing resources based on game theory to ensure the fairness of computing tasks. Assume that each computing node N_j chooses whether to participate in the calculation with a certain probability p_j, and the computing income U_j of the node is determined by the computing reward R_j and the computing cost C_j:

$$U_j = p_j R_j - C_j$$

The optimal strategy for computing nodes is to maximize their own benefits, that is:

$$p_j^* = \arg\max_{p_j} U_j$$

In the Nash equilibrium state, the benefits of all computing nodes tend to be stable, that is:

$$\frac{\partial U_j}{\partial p_j} = 0, \quad \forall j \in N$$

By solving this optimization problem, we can obtain the optimal computing resource allocation strategy for computing nodes under different task loads, thereby ensuring the fairness and efficiency of computing tasks. Figure 2 shows a schematic diagram of the distributed computing task execution process.

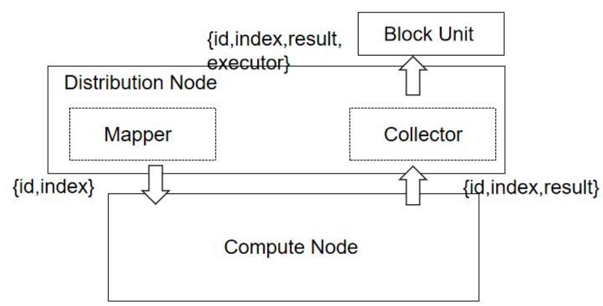

Figure 2. Schematic diagram of distributed computing task execution process

Finally, in order to reduce the blockchain storage overhead, we proposed an improved block storage compression method. The traditional blockchain storage model requires all nodes to store complete ledger data, which leads to excessive storage overhead[23]. We define the storage optimization objective function:

$$Stotal = \sum_{i=1}^{M} S(T_i) \cdot w_i$$

$S(T_i)$ represents the size of data generated by task T_i, and w_i represents the importance weight of the data. We introduce a hierarchical storage strategy, storing the calculation results on the main chain, and storing some intermediate calculation data on the auxiliary chain, and verifying the data integrity through the Merkle tree structure. Define the optimization goal of the storage node:

$$\min \sum_{i=1}^{M} w_i S(T_i), \qquad subject\ \text{to} H(S_{total}) \le H_{max}$$

$H(S_{total})$ represents the hash value of the stored data, which is used to ensure data integrity, and $H_{\max}$ is the maximum storage capacity constraint of the storage node. We dynamically adjust the value of w_i to give priority to the storage of important data, and use the sharding storage mechanism of the blockchain to reduce data redundancy and improve storage efficiency.

In summary, the method in this paper builds an efficient and decentralized distributed computing system through computing task optimization scheduling, game theory resource allocation, and block storage optimization. This system not only improves the utilization of computing resources, but also reduces storage overhead, providing a new technical solution for large-scale distributed computing.

4. Experiment

4.1 Datasets

In this study, two computers with identical hardware configurations were selected for testing, with a network bandwidth of 50MB. To simulate a wide-area network (WAN) environment, a decentralized and untrusted local area network (LAN) was established in the test environment. Within this network, no coordination nodes were present, and communication between nodes was conducted via the HTTPS protocol. Before processing received messages, each node assumed that the sender was a malicious entity and verified the received data accordingly.

The blockchain program developed for this study was implemented in Java, and blockchain data was visualized using eXtensible Markup Language (XML) files. XML files are well-suited for representing hierarchical data structures in computing environments, making them an ideal choice for storing and displaying structured data [24]. In XML files, data is encapsulated within tags that contain both start and end markers. In the blockchain system designed in this study, transactions stored in blockchain blocks were organized using a Merkle tree. This hierarchical tree structure establishes clear parent-child relationships between data elements, making XML a suitable format for data visualization. Additionally, each computation result block contained multiple computational results that maintained a parallel relationship with one another while being hierarchically nested within the corresponding block. This structured data representation further facilitates the efficient visualization of blockchain data using XML files.

4.2 Experimental Results

First, this paper conducts performance testing. The synchronization time of the new computing node is shown in Figure 3:

```
<log time="1559098110674" value="Connecting seed node,39.106.194.129:80"/>
<log time="1559098110777" value="The seed node is successfully connected."/>
<log time="1559098110777" value="Compute node synchronization network topology..."/>
<log time="1559098111125" value="Network topology synchronization completed"/>
<log time="1559098111125" value="Compute node synchronization custom algorithm..."/>
<log time="1559098111541" value="Custom algorithm synchronization completed"/>
```

Figure 3. Synchronization time of the new computing node in this test

The synchronization of the network topology took 451 ms, while the synchronization of the custom algorithm required 416 ms, resulting in a total synchronization time of 867 ms. By gradually increasing the number of nodes in the network, this method was used to analyze the relationship between the synchronization time of newly added computing nodes and the total number of nodes in the network, as illustrated in Figure 4.

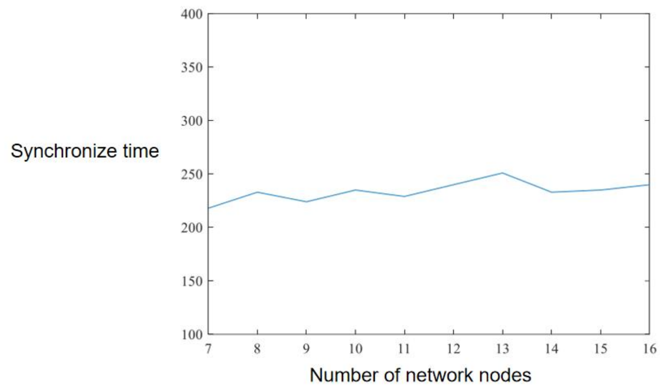

Figure 4. Relationship between the synchronization time of new computing nodes and the total number of network nodes

```
<log time="1559117474514" value="Connecting seed node,39.106.194.129:80"/>
<log time="1559117474614" value="The seed node is successfully connected."/>
<log time="1559117474614" value="Storage node synchronization network
      topology..."/>
<log time="1559117474902" value="Network topology synchronization completed"/>
<log time="1559117474902" value="Storage node synchronization custom
      algorithm..."/>
<log time="1559117474912" value="Custom algorithm synchronization completed"/>
<log time="1559117475303" value="Blocks synchronization completed"/>
```

Figure 5. The synchronization time of the new storage node in this test

In a blockchain network, when a new storage node joins, it needs to go through multiple stages of synchronization to ensure that it can correctly access the network and maintain complete data consistency. In the experiment, it takes 388ms to complete the synchronization of the network topology, 10ms to synchronize the custom Proof-of-Work (PoW) algorithm, and 391ms to synchronize the blockchain data. Therefore, the entire synchronization process of the new storage node takes a total of 789ms.

In the further experiment, we gradually increase the number of nodes in the network, and use the same method to measure the synchronization time of the new storage node, and analyze its relationship with the total number of nodes in the blockchain network, as shown in Figure 6. The experimental results show that, similar to the synchronization time characteristics of the computing node, the synchronization time of the new storage node is independent of the total number of nodes in the network. This phenomenon can be explained from the perspective of the decentralized architecture of the blockchain: in the process of the storage node joining the network, its main synchronization task is to obtain blockchain data, network topology information, and calculation rules related to the consensus mechanism, and these data are usually downloaded from a fixed set of storage nodes, rather than retrieved one by one from all network nodes[25]. Therefore, even though the total number of nodes in the network continues to increase, the synchronization time required for new storage nodes still depends mainly on the size of the blockchain data and is not significantly affected by the total number of nodes in the network.

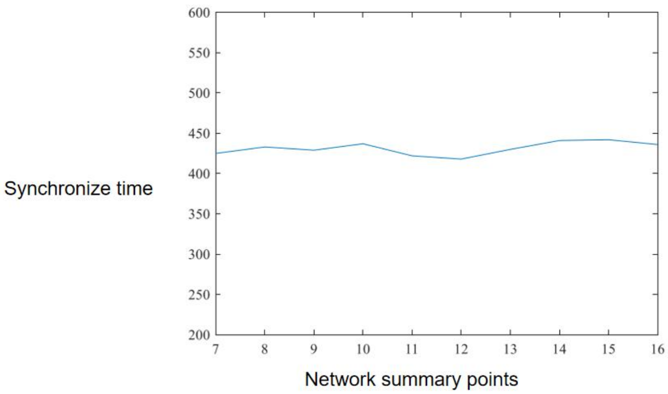

Figure 6. The synchronization time of the new storage node in this test

Finally, the relationship between the synchronization time of the new storage node and the size of the blockchain data is given, as shown in Figure 7.

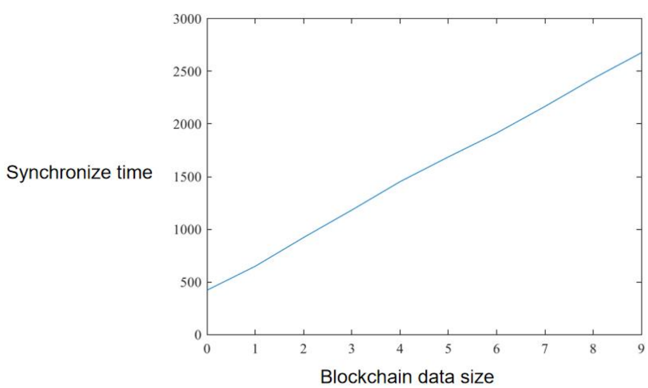

Figure 7. The relationship between the synchronization time of the new storage node and the size of the blockchain data

When a new storage node joins the blockchain network, its synchronization time is linearly related to the size of the blockchain data. Specifically, the larger the size of the blockchain data, the more data the newly added storage node needs to synchronize, which leads to an increase in synchronization time. This is because in the blockchain system, each storage node must maintain a complete copy of the ledger data to ensure data consistency and security across the entire network. Therefore, when a new storage node is connected, it needs to obtain blockchain historical data from existing nodes and gradually verify the hash value, transaction records, and related consensus information of each block to ensure that the received data is complete and has not been tampered with.

5. Conclusion

This study explores blockchain-based distributed systems and proposes an efficient storage and computing collaboration solution to address the challenges of current blockchain systems in computing resource allocation, data storage optimization, and node synchronization efficiency. By optimizing task scheduling strategies, introducing a computing resource allocation mechanism based on game theory, and improving the blockchain storage structure, we have implemented a decentralized, efficient, and scalable distributed computing system. Compared with existing distributed computing frameworks, our approach can not only make full use of the trust mechanism provided by blockchain technology, but also improve the overall efficiency of computing and storage while ensuring the decentralized nature. This research result not only provides new ideas for the application of blockchain technology in the field of distributed computing, but also lays the foundation for the design of future blockchain-based computing architectures.

Although the scheme proposed in this study has achieved optimization in many aspects, there are still some issues that deserve further study. For example, in the process of dynamic allocation of computing tasks, the computing power of computing nodes may be affected by factors such as hardware conditions and network delays, resulting in uncertainty in the efficiency of task execution. Future research can introduce more intelligent task scheduling algorithms, such as methods based on deep reinforcement learning, so that computing nodes can dynamically adjust their task acceptance strategies based on historical task data, thereby further optimizing the utilization of computing resources. In addition, the current storage optimization scheme is mainly based on hierarchical storage and data compression, and in the future, the application of blockchain sharding technology in storage nodes can be further explored to reduce the storage burden of a single storage node and improve the retrieval and synchronization efficiency of block data. In addition, in order to improve the fault tolerance of the system, decentralized storage protocols (such as IPFS) can be combined for data redundancy backup to ensure that the system can still operate stably under high concurrent access or partial node failure.

In terms of consensus mechanism optimization, although this study designed a consensus method based on a custom proof of work, this method still consumes a certain amount of computing resources. Future research can explore more efficient consensus algorithms, such as combining Proof of Stake (PoS) with Delegated Proof of Stake (DPoS) to reduce unnecessary computing overhead while improving the fairness and security of the consensus process. In addition, in order to further improve the throughput of the system, Layer 2 expansion solutions such as State Channels or Zero-Knowledge Proofs can be used to enable computing tasks to be quickly processed outside the main chain and recorded in the blockchain only when necessary, thereby reducing the burden on the blockchain main chain and improving the overall transaction throughput.

In general, blockchain technology has broad application prospects in the field of distributed computing and storage. With the continuous growth of computing needs and the continuous development of blockchain technology, future distributed computing systems need to be more intelligent, efficient, and able to adapt to the needs of different application scenarios. The results of this study provide new solutions for efficient distributed systems based on blockchain and provide research directions for future blockchain architecture optimization. In the future, we hope to further combine emerging technologies such as artificial intelligence and edge computing to provide blockchain-distributed computing systems with more flexible and efficient computing and storage capabilities to support a wider range of application scenarios, such as financial technology, healthcare, supply chain management, and the Internet of Things. Through continuous optimization and innovation, blockchain technology is expected to become an important part of the new generation of distributed computing infrastructure and promote the development of the global digital economy.

References

[1] Rao I S, Kiah M L M, Hameed M M, et al. Scalability of blockchain: a comprehensive review and future research direction[J]. Cluster Computing, 2024, 27(5): 5547-5570.

[2] Liu J, Chen C, Li Y, et al. Enhancing trust and privacy in distributed networks: a comprehensive survey on blockchain-based federated learning[J]. Knowledge and Information Systems, 2024, 66(8): 4377-4403.

[3] Ghani M A N U, She K, Rauf M A, et al. Enhancing Security and Privacy in Distributed Face Recognition Systems through Blockchain and GAN Technologies[J]. Computers, Materials & Continua, 2024.

[4] Khan A A, Laghari A A, Baqasah A M, et al. Blockchain-enabled infrastructural security solution for serverless consortium fog and edge computing[J]. PeerJ Computer Science, 2024, 10: e1933.

[5] Lakhan A, Mohammed M A, Abdulkareem K H, et al. A multi-objectives framework for secure blockchain in fog–cloud network of vehicle-to-infrastructure applications[J]. Knowledge-Based Systems, 2024, 290: 111576.

[6] Zhou, T., Xu, Z., & Du, J. (2025). Efficient Market Signal Prediction for Blockchain HFT with Temporal Convolutional Networks. Transactions on Computational and Scientific Methods, 5(2).

[7] Li, P., Xiao, Y., Yan, J., Li, X., & Wang, X. (2024, November). Reinforcement Learning for Adaptive Resource Scheduling in Complex System Environments. In 2024 5th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC) (pp. 92-98). IEEE.

[8] Duan, S. (2024). Systematic Analysis of User Perception for Interface Design Enhancement. Journal of Computer Science and Software Applications, 5(2).

[9] Li, P. (2025). Performance Boost in Deep Neural Networks: Improved ResNext50 for Complex Image Datasets. Transactions on Computational and Scientific Methods, 5(1).

[10] An, T., Huang, W., Xu, D., He, Q., Hu, J., & Lou, Y. (2025). A deep learning framework for boundary-aware semantic segmentation. arXiv preprint arXiv:2503.22050. https://arxiv.org/abs/2503.22050

[11] Li, X., Peng, Y., Sun, X., Duan, Y., Fang, Z., & Tang, T. (2025). Unsupervised Detection of Fraudulent Transactions in E-commerce Using Contrastive Learning. arXiv preprint arXiv:2503.18841.

[12] Zhan, J. (2025). Single-Device Human Activity Recognition Based on Spatiotemporal Feature Learning Networks. Transactions on Computational and Scientific Methods, 5(3).

[13] Zhang, Y. (2025). Social Network User Profiling for Anomaly Detection Based on Graph Neural Networks. arXiv preprint arXiv:2503.19380.

[14] Wang, S., Zhang, R., & Shi, X. (2025). Generative UI Design with Diffusion Models: Exploring Automated Interface Creation and Human-Computer Interaction. Transactions on Computational and Scientific Methods, 5(3).

[15] Wang, J. (2025). Markov network classification for imbalanced data with adaptive weighting. Journal of Computer Science and Software Applications, 5(1), 43-52.

[16] Xiang, Y., He, Q., Xu, T., Hao, R., Hu, J., & Zhang, H. (2025). Adaptive Transformer Attention and Multi-Scale Fusion for Spine 3D Segmentation. arXiv preprint arXiv:2503.12853.

[17] Li, M., Hao, R., Shi, S., Yu, Z., He, Q., & Zhan, J. A CNN-Transformer Approach for Image-Text Multimodal Classification with Cross-Modal Feature Fusion.

[18] Liu, J., Zhang, Y., Sheng, Y., Lou, Y., Wang, H., & Yang, B. (2025). Context-Aware Rule Mining Using a Dynamic Transformer-Based Framework. arXiv preprint arXiv:2503.11125.

[19] Gao, J., Lyu, S., Liu, G., Zhu, B., Zheng, H., & Liao, X. (2025). A Hybrid Model for Few-Shot Text Classification Using Transfer and Meta-Learning. arXiv preprint arXiv:2502.09086.

[20] Liao, X., Zhu, B., He, J., Liu, G., Zheng, H., & Gao, J. (2025). A Fine-Tuning Approach for T5 Using Knowledge Graphs to Address Complex Tasks. arXiv preprint arXiv:2502.16484.

[21] Liu, Z., Wu, M., Peng, B., Liu, Y., Peng, Q., & Zou, C. (2023, July). Calibration learning for few-shot novel product description. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1864-1868).

[22] Cao, J., Jiang, Y., Yu, C., Qin, F., & Jiang, Z. (2024, August). Rough set improved therapy-based metaverse assisting system. In 2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom) (pp. 358-364). IEEE.

[23] Agrawal R, Singhal S, Sharma A. Blockchain and fog computing model for secure data access control mechanisms for distributed data storage and authentication using hybrid encryption algorithm[J]. Cluster Computing, 2024, 27(6): 8015-8030.

[24] Akbar M, Waseem M M, Mehanoor S H, et al. Blockchain-based cyber-security trust model with multi-risk protection scheme for secure data transmission in cloud computing[J]. Cluster Computing, 2024, 27(7): 9091-9105.