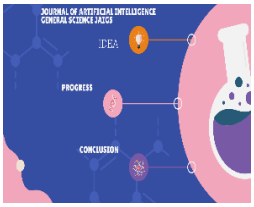
Vol.1, Issue 01, January 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS



Detection of Network Security Traffic Anomalies Based on Machine Learning KNN Method

Fanyi Zhao [1]*, Mingxuan Zhang[2], Shiji Zhou[3], Qi Lou[4]

[1]Computer Science, Stevens Institute of Technology, NJ, USA.

[2]Computer Science, University of California San Diego CA,USA.

[3]Computer Science, University of Southern California, CA,USA.

[4]Tian Yuan Law Firm, Hang Zhou, China.

*Corresponding author E-mail: rexcarry036@gmail.com

ARTICLEINFO
Article History:
Received: 01.01.2024
Accepted: 10.01.2024
Online: 22.01.2024
Keyword: Machine learning, network security, anomaly detection, network traffic

ABSTRACT

This paper discusses the application and advantages of machine learning in anomaly detection of network security traffic. By summarizing the existing methods and techniques of network anomaly detection, this paper focuses on the progress of clustering, classification, statistics, and information theory in research. In particular, innovations in data preprocessing, feature selection, and algorithm design, such as experimental validation based on an improved KNN algorithm, demonstrate the potential of machine learning in improving detection accuracy and efficiency. In the future, as the amount of data increases and algorithms are further optimized, these technologies are expected to drive further development in cybersecurity and address the challenges of increasingly complex cyber threats.

1. INTRODUCTION

Network anomaly detection plays a vital role in network security and is an indispensable part of network intrusion detection systems. Unlike feature code-based detection, anomaly detection can effectively detect unknown attack activities, including 0-day attacks, so it has significant practicability and value.

In the background of the long history and diverse network anomaly detection methods, researchers have adopted various methods, including clustering, classification, statistics, information theory, and digital signal processing. This paper realizes network anomaly detection based on machine learning. [1-3] Researchers have made remarkable progress Through continuous exploration and innovation. For example, Gaddam et al. combined the K-Means clustering algorithm and the ID3 algorithm to achieve a high anomaly detection accuracy, emphasizing the importance of defining a suitable similarity measurement method. On the other hand, Yan et al. successfully applied the SVM classifier, which labeled normal or abnormal data, to detect abnormal network traffic. However, the classification method relies on offline training data sets, which limits its application breadth in different network environments.

In addition, statistics-based methods are simple, fast, and interpretable for anomaly detection. However, these methods often need to meet specific constraints on the data, which are challenging to meet in the complex and changeable network environment. [4-5] Zempoaltecatl Piedras et al. proposed a new method based on entropy space, which has achieved good results in practical applications. Still, it is unsuitable for data scenarios with periodicity and large internal fluctuations.

Finally, applying machine learning technology in network anomaly detection improves detection accuracy and efficiency and brings more flexible and adaptable solutions. As the volume of data increases and algorithms improve, these technologies are expected further to push the forefront of cybersecurity in the future, addressing increasingly sophisticated and covert cyber threats.

2. RELATED WORK

2.1 Network traffic exception

According to a 2021 security report by Check Point Software, global cyberattacks against enterprises increased by 29 percent, with U.S. enterprises experiencing an average of 443 attacks per week and the Asia-Pacific region experiencing an average of 1,338 attacks per week. Among these frequent attacks, many use new attack tools and methods, which leads to the failure of detection and protection measures taken by many enterprises[6]. Detecting and blocking attacks from the network traffic side is currently the mainstream method of network security protection. The traditional traffic-side attack detection scheme relies on the static feature code extracted from the analysis of the attack traffic. [7]The detection speed of this scheme is breakneck, but the detection accuracy depends heavily on the signature database. If the signature code of an attack is not included in the signature database, the scheme cannot detect the attack. In addition, some attacks encrypt their traffic and cannot extract feature codes from the load without decryption. As a result, traditional feature code-based detection schemes cannot detect encrypted traffic.

The communication behavior of network traffic generated under normal conditions is obviously different from that of network traffic generated under attack, and this difference will not disappear because the load data is encrypted. On the other hand, if there is significantly different traffic from normal traffic on the network, does it indicate an attack on the network? [8]The answer is not necessarily. Because of the complex and changeable characteristics of the network, some normal behaviors will also cause network

traffic to change differently than usual. However, such cases are relatively rare, and if there is a good modeling of regular traffic, it can be controlled to a small range. [9-11]Therefore, when abnormal traffic is detected on the network, there is a high probability that an attack behavior generates it. You can further verify whether the abnormal traffic is an attack behavior by other means. Given this, network traffic anomaly detection is often used to detect unknown new attacks in unencrypted and encrypted traffic. It does not rely on the traditional static signature database and detects traffic significantly different from regular traffic through the analysis of network traffic communication behaviors so as to discover potential attack behaviors.

The concept of an exception has many different definitions. Kafadar et al. define it as a set of data that looks different from other data. Chandola et al. argue that anomalies are data patterns that deviate from clearly defined normal behavior. Lakhina et al. defined anomalies in the network as infrequent and dramatic changes at the level of network traffic. Through these definitions, it is evident that how to determine the normal state is a crucial step in anomaly detection.

Many security researchers have done research in the field of anomaly detection of network traffic and have put forward many feasible technical schemes. Through the study of these technical schemes, it can be found that all the schemes for anomaly detection from the network traffic side regard what kind of traffic data is collected and what kind of algorithm is used as the critical point. [12]In these schemes, traffic data and algorithms are not tightly coupled together; in fact, traffic data needs to be processed by a variety of different algorithms. Similarly, an algorithm can also process various types of traffic data, and technicians can freely combine traffic data and algorithms according to actual application scenarios to obtain the optimal detection scheme. To achieve this goal, this paper will review the relevant detection technologies from the two dimensions of network traffic data acquisition and anomaly detection algorithm model and predict the future research direction and possible challenges.

2.2 Network traffic data collection

What kind of traffic data to collect is an important anomaly detection problem. Different data depict network traffic from various angles and capabilities. You need to select the correct traffic data according to actual application scenarios and issues to be solved. [14-16]According to the description of features in the KDD '99 data set, traffic features in the network can be divided into three types: essential features of connections, content features of connections, and traffic statistical features. In addition, with the advent of deep learning technology, the raw load of packets can also be directly used for anomaly detection.

1. Basic characteristics of the connection

The essential characteristics of a connection are the basic property information of a connection established by different levels and different types of protocols in the Transmission Control Protocol/Internet Protocol [17](TCP/IP) family. This includes, but is not limited to, the duration of a connection, the type of protocol used, the amount of data sent, the amount of data received, etc. Jadidi et al. and Bartos et al. used the essential characteristics of Transmission Control Protocol (TCP) connections to train the model for anomaly detection.

Essential characteristics of the connection use some basic information and behavior information of the two parties in the communication process, but do not contain features related to the communication content. Therefore, the description of the connection is not comprehensive. However, in the scenario of encrypted traffic detection, the essential characteristics of the connection are mainly relied on.

2. Content characteristics of the connection[18]

The content feature of the connection is to analyze the load data in the connection and extract the characteristics that may reflect the intrusion behavior, including but not limited to the number of access to system-sensitive files and directories in a connection, the number of login failures, and the number of shell commands.

The content feature of the connection can describe the attack from the content with high accuracy[19]. Still, it requires experts in the field to analyze the attack and extract the related content, which has high labor costs and limited generalization ability and cannot be used in encrypted traffic scenarios.

3. Traffic statistics

Traffic statistics are extracted from a specified time window or number of connections to reflect the attack behavior instead of a single connection. You can collect statistics on the number of connections in the specified time window that have the same destination IP address as the current connection, the number of connections that have the same source IP[20] address as the current connection, and the number of connections that have the same service type as the current connection. You can also collect statistics on the number of connections whose destination IP addresses are the same as those of the current connection, the number of connections whose destination IP addresses are the same as those of the current connection, and the size of data sent to the same destination IP address, the number of new connections, and the number of sent and received packets in a specified time window. [21]Thottan et al., Fontugne et al., and Lim et al. counted the traffic characteristics between one host and other hosts in a specified time window from the perspective of hosts in the LAN and performed anomaly detection.

The advantage of data collection at the protocol level is that it can reflect the changes of each protocol and can discover the attack behavior closely related to a specific protocol[22]. Compared with connection-based features, traffic statistics can describe network traffic from a broader perspective, reflect the relationship between multiple connections, and can be used to detect complex and persistent attacks.

4. Original load

With the development of deep learning, the raw load in the packet can also be used as a training model for anomaly detection. According to the objective function, the deep learning model can automatically extract the features suitable for the target task from the original load through gradient descent and backpropagation techniques.

Using raw load features eliminates the need for feature engineering and reduces the dependence on domain experts, but this approach requires sample data and consumes a lot of computing resources. In addition, it is difficult for the model to converge to an ideal level when the load is encrypted[23].

2.3 The value of machine learning in cybersecurity

Traditional network security defense methods often rely on matching rules and patterns, but this method is challenging to cope with because of the increasingly complex means of network attack. By contrast, machine learning can better identify and predict various network anomalies and attacks by learning patterns and features from large amounts of data. Specifically, the application of machine learning in network security is mainly reflected in the following aspects:

0. Anomaly detection: By monitoring network traffic, system logs, and other data, the machine learning model can identify abnormal behaviors that are inconsistent with normal behaviors, such as abnormal data packets and abnormal login behaviors.

1. Threat Intelligence Analysis [24]: Machine learning can help analyze large amounts of threat intelligence data to identify potential threats and attackers and the means and targets they may take to attack.

2. Behavioral analysis: By analyzing the behavior patterns of users and devices, machine learning can identify abnormal behaviors, such as unauthorized access, abnormal data transmission, etc[25].

3. Malicious code detection: Machine learning can analyze the characteristics of software code, identify potentially malicious code, and prevent and remove it in time.

When it comes to the advantages of machine learning in detecting anomalies in network security traffic, there are two key aspects to consider:

1. Able to detect unknown attack patterns:

Traditional rule - or feature-based detection systems are often unable to deal effectively with new and unknown attacks (such as 0-day attacks) [26-28]because the features of these attacks are usually not in the set of known rules or features. By learning the patterns and behaviors of large amounts of data, machine learning models can identify and respond to the characteristics of novel attacks, thereby improving the coverage and accuracy of detection systems.

2. Adaptive and real-time:

The network security environment is changing dynamically, and the strategies and methods of attackers are constantly evolving. Without human intervention, machine learning models can automatically learn and adapt to changing threats. This adaptability enables the detection system to detect and respond to anomalies in real-time or near real-time, effectively reducing the loss and impact caused by the attack[29].

To sum up, the advantages of machine learning in network security traffic anomaly detection are mainly reflected in its ability to deal with unknown attacks and real-time dynamic environments effectively. These advantages make machine learning one of the indispensable technical means in modern network security defense.

3. Methodology

3.1 Anomaly detection algorithm design and data preprocessing

Abnormal traffic detection is an important research topic in network anomaly detection. Abnormal traffic is a situation that deviates significantly from normal traffic in the network, and regular traffic will change with the network environment and user behavior. Therefore, abnormal traffic must be compared with regular traffic running in the same network state to determine the abnormal expected behavior. [30]During traffic detection, the Bwd Packet Length Std characteristics of DOS attacks are pretty different from those of other types of traffic. Regular traffic and other types of attack traffic (except DOS) This feature value is mostly 0, while DOS attack traffic has a more significant value. Therefore, by learning the traffic's characteristics, normal and abnormal traffic can be correctly identified.

1. Algorithm design

The network anomaly detection model consists of a data collection and preprocessing module, a feature selection module, and a traffic classification module. The data collection and preprocessing module

preprocessed the KDD Cup99 data set, including transforming character features into numerical features, numerical standardization, and numerical normalization. The feature selection module uses a random forest algorithm to judge the importance of attributes, removes irrelevant features according to the extent of attributes, and weights them according to the importance of attributes. According to the weighted distance between the sample to be tested and the training sample, the traffic anomaly detection module classifies the traffic through the improved [32] KNN algorithm to determine whether it is attack traffic and attack type.

2. Data sets and preprocessing

The KDD Cup99 dataset was used in this experiment, with a total of 311029 data pieces, and the ratio of experimental training set to test set was 2:8.

The KDD Cup99 dataset consists of nine weeks of network connection data collected from a simulated U.S. Air Force Local Area Network, divided into labeled training data and unlabeled test data. The training data set contained one standard identifier type and 22 training attack types, with 14 attacks occurring only in the test data set. Each connection record in the KDD Cup99 training data set contains 41 fixed feature attributes and 1 class identifier, which indicates whether the connection record is regular or a specific attack type. Among the 41 fixed feature attributes, nine feature attributes are of discrete (symbolic) type, and the others are of continuous (continuous) type. These 41 attributes can be divided into four categories: essential TCP connection characteristics, TCP connection content characteristics, time-based network traffic statistics, and host-based network traffic statistics. The data set has four types of abnormal data: DOS, PROBE, U2R, and R2L.

3.2 Algorithm experiment and result analysis

1. Experimental design

In this section, the performance of the proposed algorithm was evaluated. All experiments were implemented in the windos10 operating system. Simulation of the CW-KNN anomaly detection algorithm was implemented using the KDD Cup99 data set, python3 compilation environment, and Pycharm editor.

2. Experimental evaluation indicators

Detection rate (DR) and false alarm rate (FAR) are essential indexes to determine the detection accuracy of anomaly detection algorithms. The detection rate refers to the percentage of correctly detected anomalies to the actual number of anomalies. The false positive rate refers to the percentage of errors detected and the number of anomalies detected. Based on the confusion matrix measurement equation as follows, where TP is the amount of abnormal data predicted as abnormal, TN is the amount of routine data predicted as usual, FP is the amount of routine data predicted as abnormal, FN is the amount of abnormal data predicted as usual, this paper uses the above indicators to verify the anomaly detection method proposed in this paper.

The detection rate and false positive rate are defined as follows:

$$DR = \frac{TP}{TP + FN} \times 100\%$$

$$FAR = \frac{FP}{FP + TN} \times 100\% \quad (1)$$

3. **Experimental analysis**

This section evaluates anomaly detection results using random forest dimensionality reduction and feature-weighted method combined with the KNN algorithm. To demonstrate the validity of the proposed model, we conducted many experiments using the KDD Cup99 dataset. The experimental results are compared with the original KNN and distance-weighted KNN (DIS-KNN) algorithm, and the experimental effectiveness of the CW-KNN algorithm is verified. The training and test data samples were selected from the original data set. To demonstrate the effectiveness of the proposed method, we conducted extensive experiments and determined the parameter range of K as 1,3,5,7,9,11, which changed the number of nearest neighbors to a certain extent and could effectively compare the experimental effect. The results are shown in Table 1.

Table 1. Comparison of detection rates of detection algorithms under different K values

	KNN	DIS-KNN	CW-KNN
K=1	88.93	94.52	96.76
K=3	89.94	95.75	97.38
K=5	96.53	97.95	98.40
K=7	96.13	96.36	98.38
K=9	93.62	96.78	97.39
K=11	92.23	95.19	96.40
Accuracy	92.90	96.10	97.45

As can be seen from Table 1, the detection accuracy of the three anomaly detection models also fluctuates with the change of the value of K. Compared with KNN and DIS-KNN (distance-weighted KNN algorithm), the CW-KNN algorithm has a higher detection rate of the anomaly detection model and can effectively reduce the false positive rate and false negative rate. As can be seen from the figure, CW-KNN has obtained a better detection rate and is more stable than the other two methods.

It can be analyzed that when the value of parameter K is between 3 and 5, the detection rate is greatly increased. When K is greater than 9, the detection rate gradually decreases for each algorithm. When parameter K is set to 5, CW-KNN can achieve the highest detection rate. On the whole, CW-KNN can achieve a higher detection rate.

Compared with the KNN algorithm, the CW-KNN algorithm has obvious advantages in terms of detection rate, which can verify the accuracy of the CW-KNN algorithm in anomaly detection of network data flow. Compared with the original KNN algorithm and the single distance-weighted method (DIS-KNN), the combined weighted KNN method has a higher detection rate. The detection rate of the four attack types is

also relatively ideal. Because U2r attack types take a small proportion in the test set, the detection rate is low, and the detection rate of Dos attacks can reach 100%, proving the algorithm's effectiveness.

4. Conclusion

By synthesizing the network traffic anomaly detection model based on KNN discussed in this paper, we can draw the following conclusions: First, through the combination of attribute weighting and distance weighting, this paper effectively optimizes the algorithm performance and significantly improves anomaly detection accuracy. Applying the random forest algorithm makes the importance ranking of attributes more accurate. Gaussian distance weighting effectively enhances the weight of abnormal data points in distance calculation to identify all kinds of attack behaviors more accurately.

Secondly, although the KNN algorithm has a significant computational load when processing high-dimensional data, which affects the detection efficiency, improving the method in this paper provides a new idea for future research. The following research direction will focus on introducing unsupervised learning techniques further to enhance the robustness and efficiency of the algorithm. By automatically learning the inherent structure and patterns of data, unsupervised learning methods are expected to better adapt to dynamic changes and complex network environments, thus driving the frontier of cybersecurity.

In summary, the network traffic anomaly detection model based on KNN proposed in this paper has significantly improved detection accuracy and provided necessary enlightenment for the future development direction of network security technology. As algorithms and data processing technologies evolve, we look forward to further research and innovation to achieve more efficient and intelligent network anomaly detection and protection strategies.

5. Acknowledgment

We want to express our sincere gratitude to Guo, L., Li, Z., Qian, K., Ding, W., and Chen, Z. for their insightful research on the Bank Credit Risk Early Warning Model Based on Machine Learning Decision Trees, published in the Journal of Economic Theory and Business Management. Their pioneering work has greatly inspired and assisted us in developing our research. We acknowledge their contributions as foundational to our understanding and approach to the topic.

We extend our sincere appreciation to Wu Y, Jin Z, Shi C, Liang P, and Zhan T for their pioneering research on the application of the Deep Learning-based BERT model in Sentiment Analysis, as detailed in their arXiv preprint (arXiv:2403.08217, March 13, 2024). Their innovative work has been instrumental in shaping our understanding and approach to sentiment analysis using advanced deep-learning techniques. We gratefully acknowledge their contributions, which have inspired and enriched our research endeavors in this field.

6. REFERENCES

[1] Li, S., Xu, H., Lu, T., Cao, G., & Zhang, X. (2024). Emerging Technologies in Finance: Revolutionizing Investment Strategies and Tax Management in the Digital Era. Management Journal for Advanced Research, 4(4), 35-49.
[2] Shi J, Shang F, Zhou S, et al. Applications of Quantum Machine Learning in Large-Scale E-commerce Recommendation Systems: Enhancing Efficiency and Accuracy[J]. Journal of Industrial Engineering and Applied Science, 2024, 2(4): 90-103.
[3] Wang, S., Zheng, H., Wen, X., & Fu, S. (2024). DISTRIBUTED HIGH-PERFORMANCE COMPUTING METHODS FOR ACCELERATING DEEP LEARNING TRAINING. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(3), 108-126.
[4] Zhang, M., Yuan, B., Li, H., & Xu, K. (2024). LLM-Cloud Complete: Leveraging Cloud Computing for Efficient Large Language Model-based Code Completion. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 295-326.
[5] Lei, H., Wang, B., Shui, Z., Yang, P., & Liang, P. (2024). Automated Lane Change Behavior Prediction and Environmental Perception Based on SLAM Technology. arXiv preprint arXiv:2404.04492.
[6] Wang, B., He, Y., Shui, Z., Xin, Q., & Lei, H. (2024). Predictive Optimization of DDoS Attack Mitigation in Distributed Systems using Machine Learning. Applied and Computational Engineering, 64, 95-100.
[7] Wang, B., Zheng, H., Qian, K., Zhan, X., & Wang, J. (2024). Edge computing and AI-driven intelligent traffic monitoring and optimization. Applied and Computational Engineering, 77, 225-230.
[8] Liu, Y., Xu, Y., & Song, R. (2024). Transforming User Experience (UX) through Artificial Intelligence (AI) in interactive media design. Engineering Science & Technology Journal, 5(7), 2273-2283.
[9] Zhang, P. (2024). A STUDY ON THE LOCATION SELECTION OF LOGISTICS DISTRIBUTION CENTERS BASED ON E-COMMERCE. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(3), 103-107.
[10] Zhang, P., & Gan, L. I. U. (2024). Optimization of Vehicle Scheduling for Joint Distribution in Logistics Park based on Priority. Journal of Industrial Engineering and Applied Science, 2(4), 116-121.
[11] Li, H., Wang, S. X., Shang, F., Niu, K., & Song, R. (2024). Applications of Large Language Models in Cloud Computing: An Empirical Study Using Real-world Data. International Journal of Innovative Research in Computer Science & Technology, 12(4), 59-69.
[12] Ping, G., Wang, S. X., Zhao, F., Wang, Z., & Zhang, X. (2024). Blockchain Based Reverse Logistics Data Tracking: An Innovative Approach to Enhance E-Waste Recycling Efficiency.
[13] Xu, H., Niu, K., Lu, T., & Li, S. (2024). Leveraging artificial intelligence for enhanced risk management in financial services: Current applications and future prospects. Engineering Science & Technology Journal, 5(8), 2402-2426.
[14] Shi, Y., Shang, F., Xu, Z., & Zhou, S. (2024). Emotion-Driven Deep Learning Recommendation Systems: Mining Preferences from User Reviews and Predicting Scores. Journal of Artificial Intelligence and Development, 3(1), 40-46.
[15] Wang, Shikai, Kangming Xu, and Zhipeng Ling. "Deep Learning-Based Chip Power Prediction and Optimization: An Intelligent EDA Approach." International Journal of Innovative Research in Computer Science & Technology 12.4 (2024): 77-87.
[16] Zhang, M., Yuan, B., Li, H., & Xu, K. (2024). LLM-Cloud Complete: Leveraging Cloud Computing for Efficient Large Language Model-based Code Completion. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 295-326.
[17] Wang, S., Xu, K., & Ling, Z. (2024). Deep Learning-Based Chip Power Prediction and Optimization: An Intelligent EDA Approach. International Journal of Innovative Research in Computer Science & Technology, 12(4), 77-87.

[18] Shi, Y., Shang, F., Xu, Z., & Zhou, S. (2024). Emotion-Driven Deep Learning Recommendation Systems: Mining Preferences from User Reviews and Predicting Scores. Journal of Artificial Intelligence and Development, 3(1), 40-46.
[19] Ping, G., Zhu, M., Ling, Z., & Niu, K. (2024). Research on Optimizing Logistics Transportation Routes Using AI Large Models. Applied Science and Engineering Journal for Advanced Research, 3(4), 14-27.
[20] Liu, Y., Xu, Y., & Song, R. (2024). Transforming User Experience (UX) through Artificial Intelligence (AI) in interactive media design. Engineering Science & Technology Journal, 5(7), 2273-2283.
[21] Ping, G., Wang, S. X., Zhao, F., Wang, Z., & Zhang, X. (2024). Blockchain Based Reverse Logistics Data Tracking: An Innovative Approach to Enhance E-Waste Recycling Efficiency.
[22] Shang, F., Shi, J., Shi, Y., & Zhou, S. (2024). Enhancing E-Commerce Recommendation Systems with Deep Learning-based Sentiment Analysis of User Reviews. International Journal of Engineering and Management Research, 14(4), 19-34.
[23] Choudhury, M., Li, G., Li, J., Zhao, K., Dong, M., & Harfoush, K. (2021, September). Power Efficiency in Communication Networks with Power-Proportional Devices. In 2021 IEEE Symposium on Computers and Communications (ISCC) (pp. 1-6). IEEE.
[24] Lin, Y., Li, A., Li, H., Shi, Y., & Zhan, X. (2024). GPU-Optimized Image Processing and Generation Based on Deep Learning and Computer Vision. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 39-49.
[25] Bao, Wenqing, et al. "The Challenges and Opportunities of Financial Technology Innovation to Bank Financing Business and Risk Management." Financial Engineering and Risk Management 7.2 (2024): 82-88.
[26] Shi, J., Shang, F., Zhou, S., Zhang, X., & Ping, G. (2024). Applications of Quantum Machine Learning in Large-Scale E-commerce Recommendation Systems: Enhancing Efficiency and Accuracy. Journal of Industrial Engineering and Applied Science, 2(4), 90-103.
[27] Wang, S., Zheng, H., Wen, X., & Fu, S. (2024). DISTRIBUTED HIGH-PERFORMANCE COMPUTING METHODS FOR ACCELERATING DEEP LEARNING TRAINING. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(3), 108-126.
[28] Zhang, M., Yuan, B., Li, H., & Xu, K. (2024). LLM-Cloud Complete: Leveraging Cloud Computing for Efficient Large Language Model-based Code Completion. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5(1), 295-326.
[29] Li, S., Xu, H., Lu, T., Cao, G., & Zhang, X. (2024). Emerging Technologies in Finance: Revolutionizing Investment Strategies and Tax Management in the Digital Era. Management Journal for Advanced Research, 4(4), 35-49.
[30] Xu, H., Li, S., Niu, K., & Ping, G. (2024). Utilizing Deep Learning to Detect Fraud in Financial Transactions and Tax Reporting. Journal of Economic Theory and Business Management, 1(4), 61-71.