DOI: 10.53469/jtpes.2024.04(01).07

Strategic Application of AI Intelligent Algorithm in Network Threat Detection and Defense

Kuo Wei[1,*], Hengyi Zang[2], Yiming Pan[3], Guanghui Wang[4], Zepeng Shen[5]

[1]Computer Science,Individual Contributor,Shenzhen, China
[2]Big Data and Business Intelligence,Independent research,Shanghai,China
[3]Computer science,colorado technical university,Austin, TX,USA
[4] Computer Science,Independent contributor,Shanghai
[5]Network Engineering, Shaanxi University of Technology, Shaanxi 723001, China
Correspondence Author, weikuo.ai@gmail.com

Abstract: ***With the rapid development of information technology, the network has become the main platform for important activities such as People's Daily life, business and government. However, the issue of network security has increasingly become the focus of attention. The rise of cyber intrusions has threatened the security of individuals, businesses and governments. Therefore, intrusion detection technology has become one of the important components of network security. In this paper, the algorithm technology combined with artificial intelligence is used to realize the application status of network security system intrusion detection and defense. The traditional K-means clustering algorithm has problems such as low efficiency, poor detection accuracy and passive processing in network intrusion behavior detection, such as preprocessing and k value determination. In order to solve the above problems, an improved k-means clustering algorithm network security detection model is adopted, and the detection model experiment is realized with the help of data sets.***

Keywords: Network security; Intrusion detection; Security defense; k-means clustering

1. INTRODUCTION

The continuous development of Internet technology has brought great convenience to People's Daily life and work, but the network application has gradually exposed some defects and loopholes in the design and security of the network. At the same time, the human invasion of the network is too hidden and destructive, and ordinary users cannot accurately judge the abnormal behavior in the network, resulting in a great impact on network security. For all kinds of information on the network, how to accurately extract and convey security text to users has become the primary problem. In recent years, researchers have improved the keyword extraction method and topic detection technology, and improved the relevant technology of topic tracking and detection. Before detecting and tracking network information security text topic, it is necessary to process information security text features. That is, text keyword extraction, text representation and feature weight calculation are carried out. After the text features are processed, text topics are detected and tracked by an improved hierarchical clustering algorithm and an adaptive topic tracking algorithm based on harmonic average similarity calculation.

Intrusion detection technology is a kind of technology that can monitor and discover network attack behavior, and can detect intrusion in time and respond and deal with it in time. Therefore, intrusion detection technology has very important application value in network security. Based on this, in order to effectively avoid the impact of network system security caused by network intrusion, the reasonable application of k-maeans clustering algorithm can effectively promote the improvement of network security detection ability. Therefore, this paper analyzes the application research of this clustering algorithm in network security detection.

2. RELATED WORK

2.1 Intrusion detection

In 1987, Dorothy Denning published the classic paper "Human assault detection Model" in the field of human assault detection. This xp system home document officially launched the research work in the field of human invasion detection, and is considered to be a pioneering achievement in the field of human invasion detection. The statistical analysis model proposed by Denning was well implemented in the early human intrusion detection

expert system (IDES). The IDES system mainly adopts the detection recommendations given in Anderson's technical report, but Denning's paper also includes other detection models.

The development process of intrusion detection technology is the competition process between intrusion and defense technology, and also the process of intrusion detection itself constantly surpassing and perfecting. At present, intrusion detection technology has developed from a simple event alarm to achieve a wide range of trend prediction and in-depth behavior analysis, and can achieve large-scale deployment, intrusion early warning, accurate positioning and supervision. Due to the advanced nature of the attack, the complexity of the environment, and the limitations of the user's understanding and application of intrusion detection, the development space of intrusion detection technology is still large.

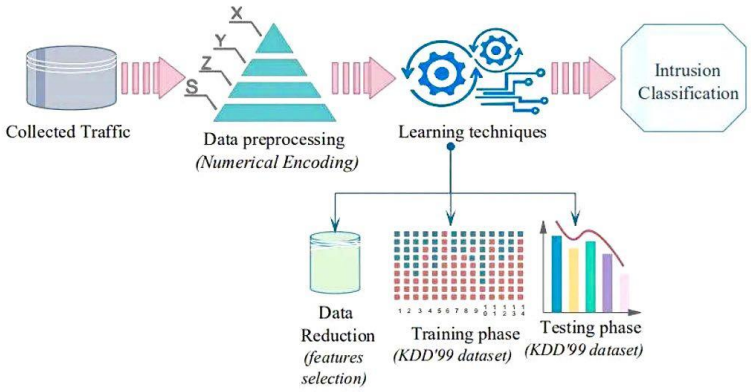

Figure 1: Intrusion detection technology principle model

The characteristics of intrusion detection technology.

(1) Intrusion detection technology is a technology that can detect and report illegal activities that have occurred or are in progress in a computer system or network. These illegal activities may be attacks on computer systems or network resources, unauthorized access, or other security incidents.

(2) The classification of intrusion detection technology is based on the monitoring and processing methods of intrusion detection technology.

Intrusion detection technology can be divided into the following categories: (1) feature-based intrusion detection technology. Feature-based intrusion detection technology refers to the technology that uses the known attack characteristics to judge whether an intrusion occurs. Behavior-based intrusion detection technology. Behavior-based intrusion detection technology is to learn and analyze the normal behavior in the computer system or network, so as to find the abnormal behavior, that is, the intrusion behavior. This technology mainly uses machine learning and data mining technology to achieve.

2.2 Intrusion detection in relation to network security

Intrusion detection technology has a wide range of applications in network security, including the following aspects:

(1) Attack detection. Intrusion detection technology can detect and report various types of attacks, such as port scanning, denial of service attacks, worm attacks, and virus attacks. Intrusion detection technology can detect these attacks in time, respond and deal with them accordingly, and prevent attackers from causing greater damage to the system and network.

(2) Anomaly detection. Intrusion detection technology can detect and report abnormal behavior in networks and computer systems, such as unauthorized access, malware, and insider threats. Intrusion detection technology can detect these abnormal behaviors, respond and deal with them in time, and prevent these behaviors from causing greater impact on the system and network.

(3) Security audit. Intrusion detection technology can record and store the behavior of network and computer system, including user access records, system logs and network traffic. This data can be used for security audits to assess and improve the security of networks and computer systems.

(4) Event response. Intrusion detection technology can detect intrusion events in time, and respond and deal with them accordingly, so as to prevent intruders from causing greater harm to the system and network.

(5) Malware detection. Intrusion detection technology can detect and report the presence of malware, including viruses, worms, and Trojans. Intrusion detection technology can detect these malicious software in time, and make corresponding response and processing, so as to prevent the malicious software from causing greater damage to the system and network.

2.3 Overview of k-means clustering

At present, cluster analysis algorithms have been widely used in various fields and fully meet the requirements of network data processing with more complex data and big data structures. Moreover,K-means clustering algorithm is also a data mining algorithm based on group analysis, which is divided into multiple subsets by combining data sets and related requirements.

At the same time, each subset data has a high similarity, but there are relatively obvious differences in attributes between subsets. In addition, the clustering algorithm pays more attention to the hierarchy, which can realize the classification of data, and also provides a guarantee for the similarity of each type of data, from which K clustering can be obtained. Specific k-means clustering algorithm workflow: First of all, in the detection of network security, because the amount of network data detected is large, and the data structure is more complex, it meets the requirements of big data. At this time, the data in the network can be divided into N data objects, and then K objects are randomly extracted from these objects as the initial clustering center. Then, according to the attributes of each object, it is classified into clusters.

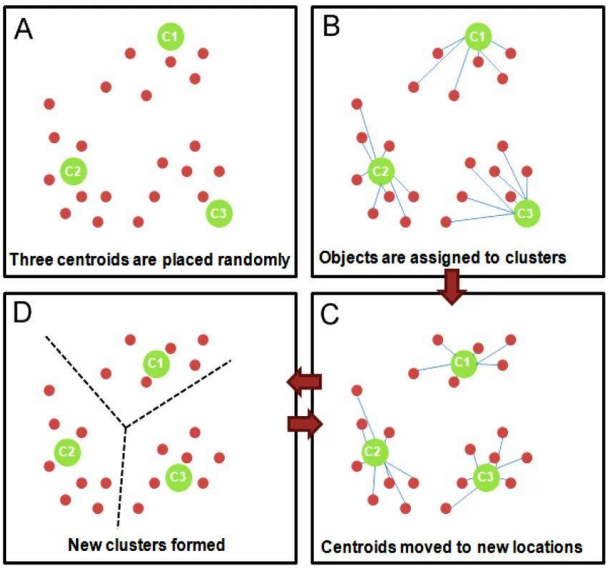

Figure 2: K-means clustering network intrusion principle model

In the process of cluster division, it is also necessary to take the distance of K objects as the standard of cluster division, so that K objects can be assigned to similar clusters respectively. Secondly, it is also possible to obtain the new cluster by using the computational method, which takes the new cluster as the center of all objects to achieve the mean value. Finally, with the help of repeated calculations, a new distance is formed, and the mean value of k is solved until the standard measure function marked in the mean square error gradually converges. In this way, the mean value of the best accuracy can be obtained. In the traditional k-means clustering algorithm, the mean square error is generally chosen as the function of standard measurement, and the intensive data is regarded as the clustering data object. Therefore, when there are obvious differences between all data objects, the traditional k-means algorithm can effectively meet the requirements of network security detection.

$$\arg\min_{\mathcal{C}} J(\mathcal{C}) = \sum_{k=1}^{K} \sum_{x^{(i)} \in \mathbb{C}_k} \|x^{(i)} - \mu^{(k)}\|_2^2 \tag{1}$$

The optimization algorithm steps are as follows:

1. Randomly select k samples as the initial cluster class center (k is a superparameter, representing the number of cluster classes). The value can be determined by prior knowledge and verification method);

2. Calculate the distance to k cluster centers for each sample in the data set, and assign it to the class corresponding to the cluster center with the smallest distance;

3. For each cluster class, recalculate its cluster center location;

4. Repeat the above 2 and 3 steps until a certain stop condition is reached (such as the number of iterations, the center position of the cluster class is unchanged, etc.).

Therefore, this paper proposes a k-means clustering algorithm network security detection model, and realizes the simulation experiment of the model with the help of data sets. Simulation results show that the proposed algorithm outperforms the traditional clustering algorithm in the accuracy and efficiency of network intrusion detection, and further reduces the false positive rate of network anomaly detection.

3. METHODOLOGY

With the continuous application of computer technology and network information technology, the data content in the network space has been increased exponentially. In the early stage, data mining technology was introduced into the network security intrusion detection system to realize the identification of intrusion behavior in the network. However, when adjusting parameters, over-reliance on manual methods leads to problems such as local minima, long calculation time and low accuracy of abnormal behavior detection, resulting in low overall detection efficiency and quality of network security.

3.1 Data collection and preprocessing

k-means is a widely used clustering algorithm. It creates k data sets with similar properties. Data instances that do not belong to these groups may be flagged as exceptions. Before we start k-means clustering, we use the elbow method to determine the optimal number of clusters.

Each row in the data set represents a network request that describes all the information for that request in the decision tree algorithm to predict forest cover

We know that features are divided into numeric and categorical types, and this dataset contains both

The final feature of each row of data is the target feature, such as most requests being marked as normal to indicate normal access

There are various other exception markers

As discussed earlier, we can completely train the model to make predictions using supervised learning techniques in the form of feature vectors + target features

But if there is an exception request that is not in all the target categories, that is not bad

So in order to find "unknown attacks", we do not use these target features in the algorithm:

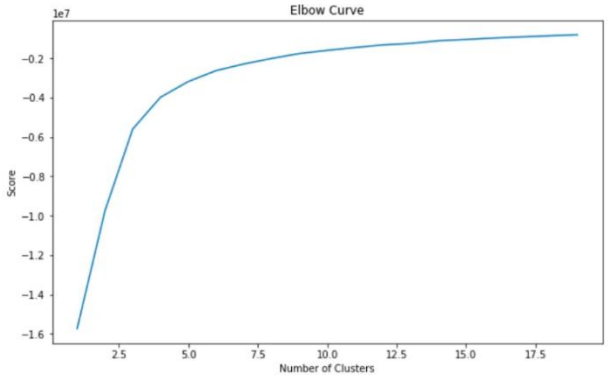

Figure 3: Abnormal detection data result

In order to find a reasonable number of distance centers, we try to have as many clustering centers as possible (from 1 to 20 clustering centers), and then we draw the Elbow curve. By looking at the Elbow curve, we find that the Elbow curve tends to converge when we increase the number of clustering centers to more than 10. Therefore, we can roughly set the number of clustering centers as 10.

Next we set the n_clusters for the K-means algorithm to 10, and then we visualize the data in 3D.

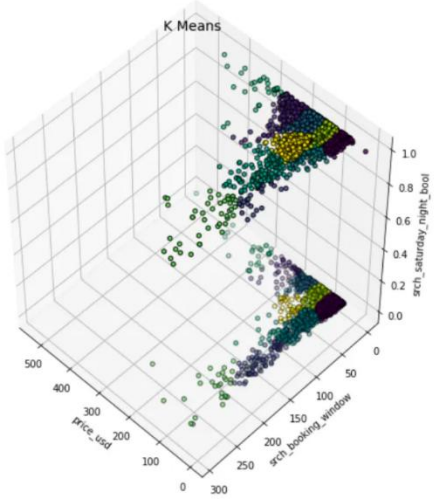

Figure 4: 3D visualization of data

3.2 Feature selection and extraction

We first carried out StandardScaler processing on the data, and then calculated the covariance matrix between the feature variables, which reflects the correlation between the feature variables. If the covariance between the two feature variables is regular, it means that they are positively correlated; if it is negative, it means that they are negatively correlated. If it is 0, it means that the feature variables are independent of each other and there is no correlation relationship (sometimes we will also use the correlation coefficient matrix instead of the covariance matrix).

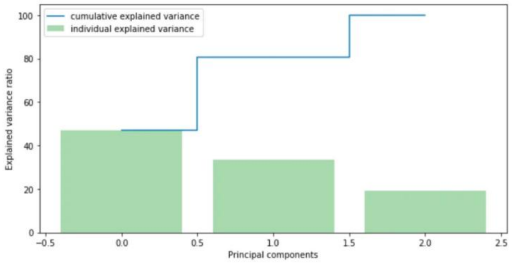

Figure 5: Data are standardized

Secondly, the eigenvalues and eigenvectors of the covariance matrix are calculated on the basis of the covariance matrix, and the explanatory variances and cumulative explanatory variances of each principal component (feature) are calculated according to the eigenvalues. The purpose of doing this is to select the principal components in the feature variables for the next step of principal component analysis (PCA). We chose the first two principal components because they have a cumulative explanatory variance of 80%.

3.3 Build data clustering

The assumption in clustering based anomaly detection is that if we cluster the data, the normal data will belong to the cluster and the anomaly will not belong to any cluster or belong to a small cluster. We use the following steps to find and visualize outliers.
Calculate the distance between each data point and its nearest cluster center. The maximum distance is considered abnormal.

Table1: Reduces dimension and sets parameter data n_components=2

	date_time	price_usd	srch_booking_window	srch_saturday_night_bool	cluster	principal_feature1	principal_feature2
0	2012-11-01 02:48:30	84.0	19	0	2	-0.889864	-0.521900
1	2012-11-01 03:06:43	78.0	16	1	0	0.230566	-0.272218
2	2012-11-01 09:04:18	114.0	56	1	3	0.567439	0.546642
3	2012-11-01 09:11:03	76.0	56	1	3	0.155659	0.581776

Set the proportion of outliers_fraction to 1%, because in the case of standard plus distribution (N(0,1)), we generally consider data other than 3 standard deviations as outliers, and data within 3 standard deviations contains more than 99% of the data in the dataset, so the remaining 1% of the data can be regarded as anomalies Be worth.

The number of outliers is calculated based on the proportion outliers_fraction. number_of_outliers

Set a threshold for determining outliers

The threshold is used to determine whether the data is an outlier

Visualization of data (both normal and abnormal).

3.4 Analysis of clustering results

It can be seen from the above experiments that the average distance decreases with the increase of k

There is no doubt about this, because as the number of population points increases, the data points must be closer to the nearest center of mass, and when the population points are equal to the number of data points, the average distance is 0, and each data point is its own center of mass

And a very strange phenomenon is that the distance is actually larger when k=35 than when k=30

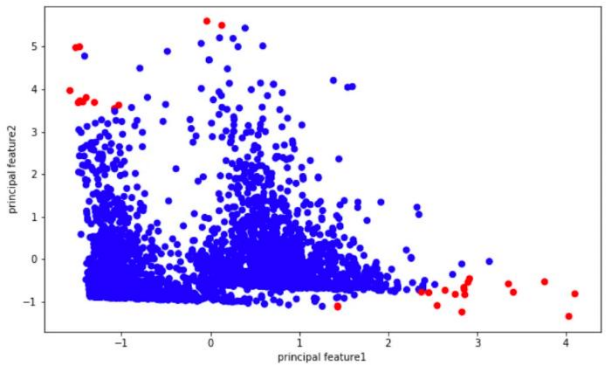

Figure 5: Clustering results scatter graph

This is because the iterative process of KMeans starts at a random point and therefore may converge to a local minimum

The case of k=35 May be due to a random initial center of mass, or it may be due to the algorithm ending before reaching a local minimum

To solve this possible problem, we can cluster a fixed k value multiple times, each time randomly with a different initial center of mass, and then choose the best among them.

From the above figure, it can be seen that the prices of outliers calculated by PCA and KMeans are mostly located at the highest and lowest points of the price range, which should be reasonable.

4. CONCLUSION

This paper explores the strategic application of AI intelligent algorithms in network threat detection and defense. With the escalating risks of cyber intrusions threatening individuals, businesses, and governments, intrusion detection technology has become pivotal for network security. The study delves into the limitations of traditional K-means clustering algorithms, such as low efficiency and poor detection accuracy in network intrusion behavior detection. To address these challenges, an enhanced K-means clustering algorithm is proposed for network security detection, and its efficacy is demonstrated through simulation experiments using relevant datasets.

The rapid growth of information technology has made the network a vital platform for daily life, business, and government activities. However, the escalating concern for network security, particularly due to cyber intrusions, necessitates advanced intrusion detection technology. This paper introduces an improved K-means clustering algorithm to overcome the inefficiencies of traditional methods in network intrusion detection. By strategically applying artificial intelligence, the proposed algorithm demonstrates superior accuracy and efficiency in detecting network intrusions, effectively reducing false positives in anomaly detection. The study underscores the significance of such advancements in bolstering network security capabilities and mitigating potential threats.

REFERENCES

[1] Li Junyi. Research on Network security detection application of K-means clustering algorithm based on big data [J]. Mechanical Design and Manufacturing Engineering, 2019,50(9):115-118.
[2] Zhou, H., Lou, Y., Xiong, J., Wang, Y., & Liu, Y. (2023). Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery. Frontiers in Computing and Intelligent Systems, 6(1), 103-106.6
[3] Implementation of an AI-based MRD Evaluation and Prediction Model for Multiple Myeloma. (2024). Frontiers in Computing and Intelligent Systems, 6(3), 127-131. https://doi.org/10.54097/zJ4MnbWW.
[4] Zhang, Q., Cai, G., Cai, M., Qian, J., & Song, T. (2023). Deep Learning Model Aids Breast Cancer Detection. Frontiers in Computing and Intelligent Systems, 6(1), 99-102.3
[5] Xu, J., Pan, L., Zeng, Q., Sun, W., & Wan, W. (2023). Based on TPUGRAPHS Predicting Model Runtimes Using Graph Neural Networks. Frontiers in Computing and Intelligent Systems, 6(1), 66-69.7
[6] Wan,Weixiang, et al. "Development and Evaluation of Intelligent Medical Decision Support Systems." Academic Journal of Science and Technology 8.2 (2023): 22-25.
[7] Dai Wei. Application of Intrusion detection technology in Network security [J]. Journal of Chongqing University of Technology (Natural Science),2018,32(04): 156-160+185.
[8] Xinyu Zhao, et al. “Effective Combination of 3D-DenseNet’s Artificial Intelligence Technology and Gallbladder Cancer Diagnosis Model”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 81-84, https://doi.org/10.54097/iMKyFavE.
[9] Shulin Li, et al. “Application Analysis of AI Technology Combined With Spiral CT Scanning in Early Lung Cancer Screening”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 52-55, https://doi.org/10.54097/LAwfJzEA.
[10] [1] Liu, Bo & Zhao, Xinyu & Hu, Hao & Lin, Qunwei & Huang, Jiaxin. (2023). Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. Journal of Theory and Practice of Engineering Science. 3. 36-42. 10.53469/jtpes.2023.03(12).06.

[11] Yu, Liqiang, et al. “Research on Machine Learning With Algorithms and Development”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 7-14, doi:10.53469/jtpes.2023.03(12).02.
[12] Xin, Q., He, Y., Pan, Y., Wang, Y., & Du, S. (2023). The implementation of an AI-driven advertising push system based on a NLP algorithm. International Journal of Computer Science and Information Technology, 1(1), 30-37.0
[13] Tian, M., Shen, Z., Wu, X., Wei, K., & Liu, Y. (2023). The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier. Academic Journal of Science and Technology, 8(2), 57-61.7
[14] Shen, Z., Wei, K., Zang, H., Li, L., & Wang, G. (2023). The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data. Academic Journal of Science and Technology, 8(3), 132-135.2
[15] Zheng He, et al. “The Importance of AI Algorithm Combined With Tunable LCST Smart Polymers in Biomedical Applications”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 92-95, https://doi.org/10.54097/d30EoLHw.
[16] Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model based on Machine Learning. (2024). Frontiers in Computing and Intelligent Systems, 6(3), 132-136. https://doi.org/10.54097/ObMPjw5n
[17] Liu, Y., Duan, S., Shen, Z., He, Z., & Li, L. (2023). Grasp and Inspection of Mechanical Parts based on Visual Image Recognition Technology. Journal of Theory and Practice of Engineering Science, 3(12), 22-28.1
[18] Xinyu Zhao, et al. “Effective Combination of 3D-DenseNet’s Artificial Intelligence Technology and Gallbladder Cancer Diagnosis Model”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 81-84, https://doi.org/10.54097/iMKyFavE.
[19] Liu, B. (2023). Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. International Journal of Computer Science and Information Technology, 1(1), 17-23.
[20] Pan, Linying, et al. “Research Progress of Diabetic Disease Prediction Model in Deep Learning”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 15-21, doi:10.53469/jtpes.2023.03(12).03.
[21] K. Tan and W. Li, "Imaging and Parameter Estimating for Fast Moving Targets in Airborne SAR," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 126-140, March 2017, doi: 10.1109/TCI.2016.2634421.
[22] Zhou, H., Lou, Y., Xiong, J., Wang, Y., & Liu, Y. (2023). Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery. Frontiers in Computing and Intelligent Systems, 6(1), 103-106.
[23] Liu, S., Wu, K., Jiang, C., Huang, B., & Ma, D. (2023). Financial Time-Series Forecasting: Towards Synergizing Performance And Interpretability Within a Hybrid Machine Learning Approach. arXiv preprint arXiv:2401.00534.
[24] Su, J., Nair, S., & Popokh, L. (2023, February). EdgeGym: A Reinforcement Learning Environment for Constraint-Aware NFV Resource Allocation. 2023 IEEE 2nd International Conference on AI in Cybersecurity (ICAIC), 1–7. doi:10.1109/ICAIC57335.2023.10044182
[25] Su, J., Nair, S., & Popokh, L. (2022, November). Optimal Resource Allocation in SDN/NFV-Enabled Networks via Deep Reinforcement Learning. 2022 IEEE Ninth International Conference on Communications and Networking (ComNet), 1–7. doi:10.1109/ComNet55492.2022.9998475
[26] Popokh, L., Su, J., Nair, S., & Olinick, E. (2021, September). IllumiCore: Optimization Modeling and Implementation for Efficient VNF Placement. 2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 1–7. doi:10.23919/SoftCOM52868.2021.9559076
[27] Bao, W., Che, H., & Zhang, J. (2020, December). Will_Go at SemEval-2020 Task 3: An Accurate Model for Predicting the (Graded) Effect of Context in Word Similarity Based on BERT. In A. Her belot, X. Zhu, A. Palmer, N. Schneider, J. May, & E. Shutova (Eds.), Proceedings of the Fourteenth Workshop on Semantic Evaluation (pp. 301–306). doi:10.18653/v1/2020.semeval-1.
[28] Moghaddam M, Charmi M, Hassanpoor H. Jointly humansemantic parsing and attribute recognition with feature pyramid structure in Efficient Nets . lET lmage Processing,2021，15(10):2281-2291.
[29] Du, Shuqian, et al. “Application of HPV-16 in Liquid-Based Thin Layer Cytology of Host Genetic Lesions Based on AI Diagnostic Technology Presentation of Liquid”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 1-6, doi:10.53469/jtpes.2023.03(12).01.
[30] Pan, Yiming, et al. “Application of Three-Dimensional Coding Network in Screening and Diagnosis of Cervical Precancerous Lesions”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 61-64, https://doi.org/10.54097/mi3VM0yB.

[31] Song, Tianbo, et al. “Development of Machine Learning and Artificial Intelligence in Toxic Pathology”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 137-41, https://doi.org/10.54097/Be1ExjZa.

[32] Qian, Jili, et al. “Analysis and Diagnosis of Hemolytic Specimens by AU5800 Biochemical Analyzer Combined With AI Technology”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 100-3, https://doi.org/10.54097/qoseeQ5N.

[33] Zheng, Jiajian, et al. “The Credit Card Anti-Fraud Detection Model in the Context of Dynamic Integration Selection Algorithm”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 119-22, https://doi.org/10.54097/a5jafgdv.