Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 15, Issue. 1, January 2026, pg.1 – 14

Cost-Based Resource Allocation in Computer Networks using Artificial Intelligence Techniques

Maryam Thajeel Hussein

Ministry of Higher Education and Scientific Research – Iraq - Sumer University
maream.997d@gmail.com

DOI: https://doi.org/10.47760/ijcsmc.2026.v15i01.001

Abstract: A data management process is an elaborate representation of a technical information system which will be with technical information and human resources across all business stages. Cloud automation software utilizes resource allocation methods. These methods are used to monitor the performance of applications and the capacity of data centers. This is done based on the number of requests users make. This study focuses on resource allocation in computer networks aimed at minimizing costs. To achieve this, the study will use artificial intelligence (machine learning and deep learning). The soaring popularity of distributed network (cloud computing) calls for an efficient approach to resource allocation to achieve high performance at low cost. Research proposes intelligent modelling of data supported with algorithms of autoencoder and blue whale feature selection algorithm and integrating this with any intelligent clustering. The model was tested on real-world data, IBM ILOG CPLEX data and SNDlib data. According to the findings, the suggested model outstands other models based on various parameters. In this research, we use deep neural networks to classify the data. The accuracy of proposed method is around 93%.

I. Introduction

Computers are interconnected with the help of networking software. These computers can be either personal computers or server computers. Further, they are attached to each other using various electronic devices like switches and routers. However, today many devices can connect to a computer network, including CCTV, IP-based phones, tablets, smartphones, and game consoles. The Internet of Things, more than all of these, provides the ability to allow anything connected to computer networks.

World's digital transformation is advancing rapidly across the globe. The application of information technology next to generation is growing at an incredible speed. In an essential manner, industry, administration systems, and society have gone through fundamental changes. The advancement of IoT technology provides new development opportunity and owing to its involvement into the digital economy has also become a powerful propulsive force for the development of digital economy further in the near future。

New technologies have emerged as a result of its effective integration in other settings. The popularization of IoT technologies was made possible through the use of networks and platforms.

A data management system is the overall representation of an enterprise-wide technical information system. The company product integrates the data, functions and human resources of a business at all stages of its life cycle while allowing all this to be done in an integrated business system. You can easily manage your company, project or technical facility with this system in a professional manner. It also assists in collaborating, communicating, and integrating designs internally, which will considerably lower costs, internal production expenses, and ensure the quality of on-time completion.

II. Looking for options at the lowest rates that can give maximum returns

Companies and governments earmark funds for investments that promise to maximize profits. Essentially, controlling the cost of construction is to achieve this. In order to make sustainable and competitive development, the assessment and costing is early on before the project initiation. This is very crucial for the cost estimation on time and in a proper manner. marginal costing is the method used by the managers in taking profit-maximised decisions on the pricing of the products.

Controlling the most important cost determinants, utilizes scientifically-based cost calculations is essential for success.

In recent years, many intelligent applications have become popular, such as facial recognition, AR/VR, and driverless driving, due to the rapid development of 5G technology for communication. These applications consume huge resources and exhibit high latency. The expansion and popularization of mobile devices, however, is limited primarily by low power consumption. A practical solution for wireless multimedia communications is mobile edge computing (MEC), which offers computation resources close to the end-users. MEC schemes are capable of reducing the processing load of smart applications in neighbouring servers, which get transmitted through the base stations. In comparison to cloud computing, the MEC has a critical edge in lessening the data transmission latency which enhances the services. Effective multiple tenant scheduling is crucial to meeting diverse service requirements. To ensure the best quality of service for every tenant, resource isolation needs to be maintained.

The Service Function Chains (SFCs) based on network slicing in 5G are designed to provide custom and commercial traffic management for applications with diverse Quality of Service (QoS) requirements like low latency, high reliability, high data rate, etc. Therefore, it is essential to virtualize the physical resources that are hardware, like machine and network components. 5G slicing allows tenants to create multiple slices, each definition of SLAs for different purposes. The delay and jitter-sensitive services cannot be handled in the Global Cloud Computing Centers (GCCs) owing to the distance between the centres. The European Telecommunications Standards Institute (ETSI) has put forth a plan to solve this crucial problem. ETSI defines Multi-Access Edge Computing (MEC) technology that incorporates processing, jitter control, and storage within the Radio Access Network (RAN) which can be deployed close to the User Device (UE). The end-to-end latency and variance are reduced by this method, allowing better resource allocation.

III. Problem Statement

Over the years, computers are increasingly designed as distributed networks instead of their traditional centralized models. Multiple data processors, each having non-volatile files on its own auxiliary storage media, make up these networks. Individuals who are directly connected to those processors as well as other users on the network will be able to access these files. At any node in the network, users can read data, update it, and execute files in others.

To achieve this infrastructure needs to consider connection reliability and latency across nodes. As any file requested can reside at more than one node in the network, there is always a tradeoff to be observed between cost, reliability and time. Cloud management and automation software typically performs resource management by deploying virtual machines (VMs) in cloud data centers to gather compute resources to meet the service requirements requested by users.

The above software will identify and allocate the appropriate physical resources needed for every requested virtual machine. These are based on accurate information about the resources and capabilities of the various data centres. Further, the resource allocation also depends on the current and future status of the various data centres. The user application performance, as well as the capability of a data centre to serve the maximum number of requests, is critically dependent on the resource allocation algorithms of cloud software.

The pronounced negative influence is more felt at the distributed cloud level, where algorithms may need to split a single request's resource between several relatively small data centres. When a user requests an application to run in a data center, the user specified the number of virtual machines needed and the communication requirement between them. There are several programs which are resource allocation programs. These programs select data centers to which these machines are allocated for the best performance with the least resources.

The two objectives in question – performance and reduced consumption – integrate over time. This may present challenges such as the lack of a rack with the necessary capacity to host all the required machines in one data centre. Hence, cloud automation software aims to reduce the communication volume as much as possible. Real-time analysis of the network is used to minimize allocation costs using intelligent cost-based algorithms. Cloud environments benefit from the effectiveness of machine learning and evolutionary algorithms for classifying and optimizing resource allocation behavior.

IV. Research Objectives

1. Enhancing the efficiency of resource allocation in distributed computing networks using AI.
2. Minimized communication costs in nodes through machine learning techniques based intelligent models.
3. Examine how the proposed model performs on CPLEX and SNDlib benchmark datasets.
4. By examining accuracy, sensitivity, and specificity, we compared the designed model with other models to find the better one.

V. Proposed Method

Advancement and improvement of the networks are not just concerned with the hardware portion of the networks but also dedicate developing efficiencies of the systems via software development. In the last few years, a lot of studies have been done to improve IoT systems' performance by offering software solutions. The Resource Allocation (RA) problem is the one of the main problems examined by these studies. The aim of this challenge is to reduce communication costs between the nodes in the network, which in turn increases the efficiency of connected systems such as RFID systems and 5G systems.

Two types of nodes mainly make up these systems. Source node is the first type node. It senses the environment and sends data to the entire network. The issue here is to allocate resources for each specific gateway in a case such that the cost of the communication is the least.

Before presenting the proposed algorithm, the resource allocation problem must first be defined. The model examined in this study consists of two node types. The first type consists of nodes that depict the resources allocated to service instances. The second type consists of gateways that will interface with these resources. Further, they will help to manage the requirement of data movement. Gateways connect various IoT components with each other. Each gateway manages multiple resources that can connect to several gateways. The connection cost of each gateway with the resource it is attached to is given.

A crucial issue is how to allocate resources among gateways to achieve the minimum overall cost of communication. Much of the cost of connectivity is that between the gateways, so it becomes important to design a communication structure between the gateways such that the number of links is as less as possible. This is commonly accomplished via the construction of bus-architecture or ring-architecture. As communication models undergo shifts, the costs of communication also evolve; thus, a goal of this model is to uncover resource allocation patterns that support the lowest cost.

The issues of file allocation and capacity allocation are problems computer networks face due to their complexities. The aim is to assign files to network nodes economically with consideration for the response time to request the files and their availability. To surmount this issue, a mathematical model was developed, a model which turned out to be a member of the integer nonlinear numerical programming problems which involve an intricate calculation even for a modest application.

Lately, the cost-based resource allocation through intelligent algorithms has gained popularity in lowering the running cost of applications. Deep learning and machine learning algorithm are popular software to control resource allocation cost. Creating an intelligent model that achieves optimal and viable outcomes necessitates careful methodological steps.

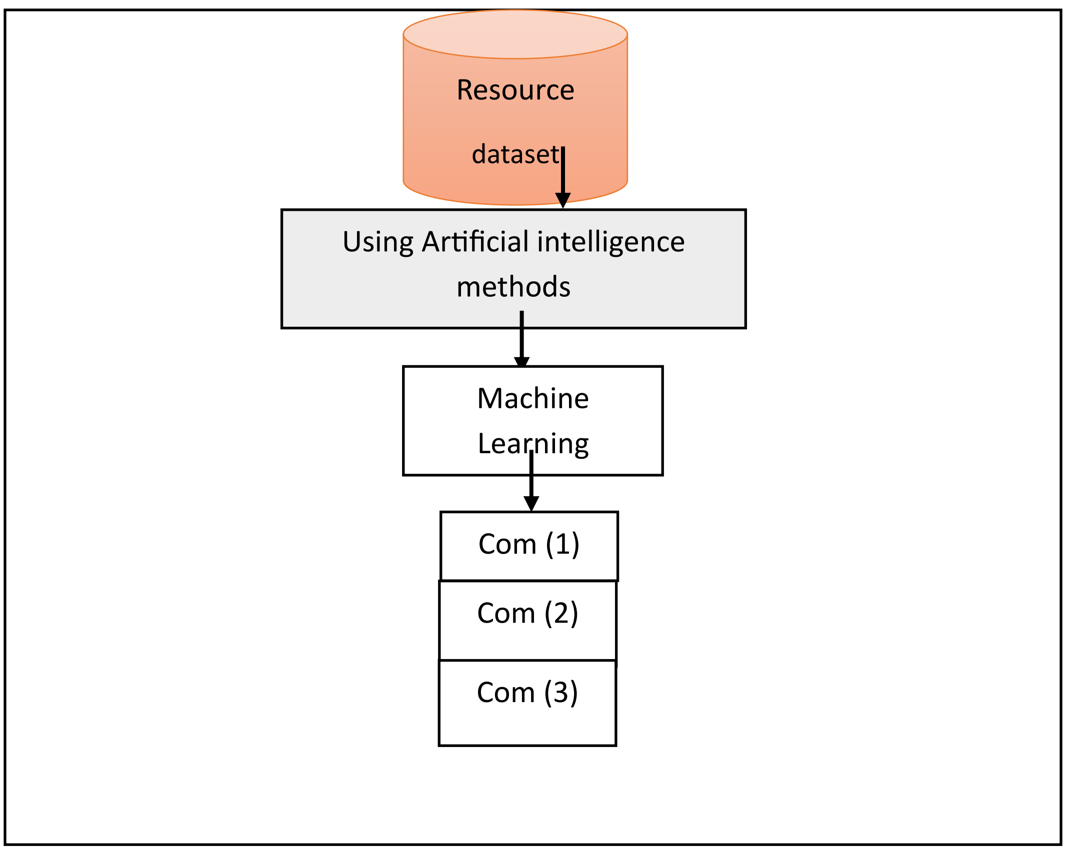

Figure 3-1. Main Research Method

When determining values, the one-way cost is selected as the reference sequence solution, and the correlation between different characteristic indices and the one-way cost is compared. A greater degree of correlation indicates a stronger association between the eigenvalue and the one-way cost.

VI. Feature Selection

In recent years the focus has shifted from hardware to software in network development. Further, there is a greater emphasis on improving system efficiency. Resource allocation is one of the most significant issues in this context that seeks to minimize communication cost between the nodes of the network. If the cost will go down, it will help in better performance of the connected systems like radio frequency identification (RFID) systems and 5G.

The first type includes source nodes, which sense the environment and transmit data. The second type consists of gateways that connect different parts and control the flow of data. These systems will be looked at in a closer way in hereafter. The question is how to link each group of resources to the suitable gateways so that overall communication cost can be minimized.

In order to present the algorithm offered, it is necessary to define the RA problem. In the present issue, resources for performing certain services are allocated from the first type of nodes, while the second type is represented by gateways connecting resources and regulating communications. The information concerning the cost of communication from any gateway to its set of resources is known beforehand. Therefore, how the resources can be allocated to the gateways in such a way that it incurs a minimum cost is a crucial challenge.

Gateway communication costs are also an important part of the overall cost. Due to its expensive, it is proposed to minimize connections between gateways whether bus or ring connection. Every time the communication pattern changes, the communication cost varies, necessitating a resource allocation pattern that minimizes the communication cost.

Conversely, there is the same issue of allocating files and storage capacity within networks. File allocation has the aim of minimizing the overall cost subject to network delay restrictions and the availability of files. A model has been developed to solve this problem and it is likely to belong to the family of nonlinear problems in integer programming which are difficult to solve using standard methods, especially when data is large.

In this sense, artificial intelligence algorithms have emerged as a promising means of alleviating the cost of resource allocation. Algorithms that employ deep learning techniques are able to tackle these challenges if they sequentially execute predefined actions leading to the construction of an appropriate and optimal intelligent model.

The three stages are important for feature selection in cost-dependent data.

1. The study aims to identify the features that are relevant to the cost of transmission and transfer projects. The features are clustered through the K-Means algorithm and correlation distances are derived.
2. The second method for feature selection is calculating the importance of each feature. This is done by replacing original features with random features and creating an alternative training set. Thus, we can find the most influential features.
3. The ineffective features are deleted using specific thresholds of the number of features & number of clusters. They depend on the real-time channel statistics & the network traffic information..

Cost management is one of the basic concepts of management. It aims to allocate the resources that improve efficiency. Overall efficiency is the ratio of minimum cost to actual cost and has two types technical efficiency and allocative efficiency. When something is technically inefficient it is not performing at the efficient frontier. When something is allocatively inefficient, it is not using the correct resource mix.

The optimal solution to this problem begins with the selection of the best features which is implemented with the help of Blue Whale Algorithm wherein a whale represent a possible solution for the resource allocation problem.

VII. Resource Classification

The autoencoder (AE), an unsupervised neural network, is what this research looks at. An AE is an unsupervised learning model which tries to reconstruct input data[4] and it allows learning different features and patterns of the dataset. The typical structure of this model has an input layer, one or more hidden layers and an output layer. The neurons in the input and output layers must be equal in number.

An autoencoder has two parts. The encoder converts original data to some hidden representation(codes). The decoder reconstructs this code to produce the original data. Denoising AE, cumulative AE, sparse AE, and variational AE are among the different kinds of AE available.

A sparse model, for example, enables the retrieval of hidden representations based on only a small number of active neurons. Essentially, these neurons activate only for specific characteristics of the data. Using added constraints known as the "sparsity penalty", the arithmetic mean of the neuronal activations is forced to be close to zero, thereby enhancing performance. A stacked AE model consists of multiple layers of sparse AE models placed on top of each other, which is typically trained using a progressive training method called greedy layer-wise training.

VIII. Result

Data Envelopment Analysis (DEA) is a nonparametric technique used to evaluate the relative efficiency of decision-making units (DMUs) using linear programming (LP) which involves multiple inputs/outputs. DEA models typically evaluate the performance DMUs independently with the input and output goals considered individually. As a result, each DMU is projected onto the efficient frontier in isolation. Nonetheless, in many practical circumstances, all DMUs act under the authority of a centralized decision maker (DM). For instance, when they belong to the same organization whether public or private which assigns the needed resources to the DMUs to produce given outputs. Many applications like bank branches, hospitals, university departments, retail chains, police stations, and more can implement a Centralized Structure. In these macro systems, the decision maker is concerned with optimally enhancing performance. The overall input usage (or total output) across all DMUs is minimized (or total output is maximized) and it is not done separately for each DMU.

IX. Dataset

Resource Allocation in Computer Networks Datasets: There are a number of public datasets that may be exploited to aid research and practice. In order to demonstrate the innovative features of the suggested approach and to prove the feasibility of its practical implementation, the following are some of the datasets:

a) Experimental data analysis of 25 data points produced in 2004 of which five data points are given. This data is used to describe a set of branches which are supervised by a central unit with the power to check the performance of each branch and collect the resources between the branches. The variables to be used are two output variables, sales and profits, and two input variables, operating hours and the overall size of the retail space of the supermarkets. Operating hours is the number of staffs working in a particular time, whereas size is the amount of retail areas. The assumption was to estimate efficiencies and input targets of each store with a traditional BCC type DEA model, using the displayed inputs of columns two, three and five.
b) CPLEX: It offers sample data sets, as well as a collection of setup models of computer network resource allocation.
c) SNDlib: A data library with the information about the design of survivable fixed telecommunications networks, both the network topology and resource allocation data, to be used in research. SNDlib will be intended to offer realistic telecommunications network data (including traffic data) to the research community, to aid in the development and analysis of advanced solutions in this area.

X. Evaluation Methods

Evaluating the classification model is among the most important issues in classification processes. This assessment may take place using training and test samples, where the classification achieved from the model is compared with the correct classification of the image or object in question. The analysis of classifications is based on four major values or variables, namely True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). These values can help identify the states of the classes and input data categories.
Total accuracy criterion denotes which percent of the data as a whole has been classified correctly. The accuracy criterion can be calculated using Equation 4-1.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

Sensitivity reflects how many cases that have the condition the test correctly identifies. Equation 4-2 implies that the low sensitivity of the test suggests that a certain number of sick people are ignored.

$$Sensitivity(Recall) = \frac{TP}{TP+FN}$$

Another measure is specificity, which refers to the proportion of negative cases that the test accurately identifies as negative. Having a low specificity means that healthy people are being classified as patients according to equation 4.3.

$$Specificity = \frac{TN}{TN+FP}$$

Evaluation of the Proposed Method Based on the Experimental Data Set

The experimental data set was tested on the proposed algorithm. In the first scenario, it is assumed that the central decision maker's objective is to minimize the total inputs consumed by all DMUs in the organization. When it comes to centralized railways, GCRAlBCC is the major model. The total man-hours and size results indicate that when carrying out the proposed centralized approach, so that the 16.19% are reduced, the total man-hours and size is dropped to 1593.599 and 86.289 respectively, as compared to the 1732.605 and 93.739, when a decentralized approach is carried out. The efficiency results of the proposed method based on some of the discovered sizes are presented in Table 4-2.

Table 4-2. Comparison of the proposed model on the first dataset

Number	Sale	Earnings	Working hours
1	137.67	5.08	78.11
2	77.80	0.16	52.20
3	73.13	2.08	53.21
4	135.76	4.13	77.92
5	58.30	1.30	41.8
	1535.60(0%)	**36.76(0%)**	**950.74(-18.85%)**

Results based on IBM ILOG CPLEX dataset:

The second dataset is IBM ILOG CPLEX. The second dataset was used for test and evaluation of the algorithm. With the help of different clustering techniques, the total resources managed to get distributed to majority requestors. The requested resource also remained available. Table 4-3 displays the outcomes of the techniques for the IBM ILOG CPLEX dataset.

Table 4-3. Comparison of the proposed model on the IBM ILOG CPLEX dataset

Method	Accuracy	Sensitivity(Recall)	Specificity
Graph-based clustering + deep learning	85.13	85.78	**86.75**
Nearest Neighbor Clustering + Deep Learning	88.46	88.45	**88.48**
Hierarchical Clustering + Deep Learning	83.28	83.17	**81.92**
Proposed method (feature selection with the help of blue whale + deep learning)	92.54	92.66	**92.47**

The accuracy of the proposed method is better than other methods (as seen in the above table). The proposed Method achieves a result of at least 5% better than the other Methods in accuracy, specificity and sensitivity. The graph that follows demonstrates this trend. The previous resource allocation methods considered a single level of benefit, which could be a project-level benefit, or portfolio level benefit. The proposed model considers both project-level and portfolio-level benefit in resource allocation in a joint manner. To make a project schedule evaluation from a project activity level outward. After that, the sensitivity analysis is done to study the effect on the model output due to change in parameters. The proposed model can provide improved allocation performance as shown in simulation results.

XI. Results based on the SNDlib dataset:

The final data set is SNDlib. The algorithm proposed here outperforms methods of the second dataset. In contrast to the initial dataset, the proposed method managed to have a vast distribution of resources with a 94.3% accuracy. The methods for the SNDlib data set gave us results as per the table 4-4.

Method	Accuracy	Sensitivity(Recall)	Specificity
Graph-based clustering + deep learning	86.56	87.45	**87.36**
Nearest neighbor clustering + deep learning	88.32	90.51	**90.30**
Hierarchical clustering + deep learning	84.53	84.27	**86.34**
Proposed method (Blue Whale-assisted feature selection + deep learning)	94.03	94.65	**93.60**

XII. Conclusion

Software for resource allocation is designed to choose the best data centres for virtual machines (VMs) requested by users. This is aimed at reducing the resource utilization of the software as much as possible. Also, it must ensure an optimal level performance. High performance is achievable at the same time as the two above goals. Whenever a user request is received, there may not be enough emptied shelves to store the requested VMs. A goal of cloud automation software is to reduce messaging.

In recent times, cost-based approaches of resource allocation became popular to reduce the cost of the application. Deep learning and other machine learning techniques can significantly lower resource allocation costs. A number of steps are needed to ensure the output of the intelligent model is relevant and of high quality in cost assessment, a single cost is used as a base value to study the relationship between other cost variables and the base value. A stronger correlation between cost and feature value indicates a more significant relationship between them.

Artificial intelligence techniques with a focus on deep learning and intelligent feature selection can efficiently address the cost of resource allocation in computer networks, as shown through studies. As compared to traditional methods, the suggested method significantly enhances the performance of the networks with a low total cost in communication among the nodes. Several non-standard procedures have enabled this, including the utilization of features in the Blue Whale algorithm, employing autoencoders for model formation, and engaging in intelligent clustering. Results of the simulation using three datasets obtained from real-world scenario, IBM ILOG CPLEX and SNDlib module shows that accuracy of model is very high as it is 94.3% as well as it is superior to other method in accuracy, sensitivity and specificity.

The application of modern evaluation techniques for performance such as Data Envelopment Analysis (DEA) facilitates the revelation of the differences between the units of decision making (DMU) and what must be done to be technically efficient. While there is reduced cost, it is not the only benefit of the new approach. It helps to plan and manage resources at the project and team level of network designs too. Today's state-of-the-art resource-allocation models use a new paradigm. They now rely not just on computing power but rather on intelligence and dynamic behavior. This helps users to make knowledgeable and accurate decisions according to the parameters and changing requirements of the network.

References

[1]. Lee, W., Jo, O., & Kim, M. (2020). Intelligent resource allocation in wireless communications systems. IEEE Communications Magazine, 58(1), 100-105.

[2]. Sun, C., Wu, X., Li, X., Fan, Q., Wen, J., & Leung, V. C. (2021). Cooperative computation offloading for multi-access edge computing in 6G mobile networks via soft actor critic. IEEE Transactions on Network Science and Engineering.

[3]. M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource management for 5GNR integrated access and backhaul: A semi-centralized approach," IEEE Transactions on Wireless Communications, vol. 21, no. 2, pp. 753– 767, 2022.

[4]. X. Zhang, M. Peng, S. Yan, and Y. Sun, "Deep-reinforcement-learningbased mode selection and resource allocation for cellular V2X communications," IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6380–6391, Jul. 2020.

[5]. W. Xia, T. Q. S. Quek, J. Zhang, S. Jin, and H. Zhu, “Programmable hierarchical C-RAN: From task scheduling to resource allocation,” IEEE Transactions on Wireless Communications, vol. 18, no. 3, pp. 2003–2016, Mar. 2019.

[6]. A. Rezaei, P. Azmi, N. Mokari, M. R. Javan, and H. Yanikomeroglu, “Robust resource allocation for cooperative MISO-NOMA-based heterogeneous networks,” IEEE Transactions on Communications, vol. 69, no. 6, pp. 3864–3878, Jun. 2021.

[7]. B. Khorsandi, F. Tonini, and C. Rafaelli, “Centralized vs. distributed algorithms for resilient 5G access networks,” Photonic Network Communications, vol. 37, no. 3, pp. 376–387, Jun. 2019.

[8]. Nouruzi, A., Rezaei, A., Khalili, A., Mokari, N., Javan, M. R., Jorswieck, E. A., & Yanikomeroglu, H. (2023). Smart Resource Allocation Model via Artificial Intelligence in Software Defined 6G Networks. arXiv preprint arXiv:2302.04655.

[9]. Zhang, X., Yang, L., Zhang, L., Ye, J., Zhang, C., & Zhao, D. Y. (2021). Research on intelligent evaluation and monitoring method for cost of transmission and transformation project. In IOP Conference Series: Earth and Environmental Science (Vol. 645, No. 1, p. 012062). IOP Publishing.

[10]. Alicherry, M., & Lakshman, T. V. (2012, March). Network aware resource allocation in distributed clouds. In 2012 Proceedings IEEE INFOCOM (pp. 963-971). IEEE.

[11]. Chen, L., Zhang, H., & Wang, F. (2022). Intelligent Evaluation Method of Engineering Cost Feasibility Model Based on Internet of Things. Journal of Sensors, 2022.

[12]. Moore, W. J., & Starr, A. G. (2006). An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities. Computers in industry, 57(6), 595-606.

[13]. Zhou, Y., Fadlullah, Z. M., Mao, B., & Kato, N. (2018). A deep-learning-based radio resource assignment technique for 5G ultra dense networks. IEEE Network, 32(6), 28-34.

