Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100119
ISSN 2371-9524 Vol. 10 Num. 1

Cloud Based Error Correction and Big Data Mining for Computer Machine Learning Algorithms

Ziyi Huang[1,a,*]

[1]*School of Public Security Information Technology and Intelligence, Criminal Investigation Police University of China, Shenyang, 110035, Liaoning, China*
[a]*625188487@qq.com,*
[*]*Corresponding author*

Keywords: Machine Learning Algorithms, Cloud Based Error Correction System, Big Data Mining System, Data Preprocessing

Abstract: With the advent of the Big Data (BD) era, Machine Learning (ML) algorithms have been widely applied in various industries. Among them, ML algorithms are mainly trained on computers. Due to various problems that may arise during the training process, ML algorithms produce various errors and cannot achieve good results. This article analyzed a cloud based error correction system and BD mining system for computer ML algorithms. This system could effectively solve various problems that arose in ML algorithms, and its effectiveness was verified through experiments. By analyzing cloud error correction technology and BD mining technology, a cloud error correction system and BD mining system were constructed, and ML algorithms were used to verify the feasibility of the system in terms of average running time and accuracy. Through experimental research, it was found that the accuracy of using ML algorithms was 5.8% higher than using neural network algorithms. There were currently multiple ways to process BD, and using ML patterns to optimize BD processing was a relatively effective approach. Simulation experiments showed that the ML algorithm proposed in this paper had higher accuracy overall than neural network algorithms, thus making it an effective and practical optimization algorithm.

1. Introduction

With the rapid development of computer technology, there is an increasing amount of network data, and extracting useful information from the massive network has become particularly crucial. Data mining technology can solve the problem of adaptive demand for different types of data mining caused by the huge scale of data by integrating high virtualization and high availability characteristics, thereby improving the efficiency and accuracy of BD mining. On this basis, companies or operators can build their own internal data mining models according to their own needs, so as to obtain more effective business value.

The widespread application of BD has brought convenience to many fields, and many experts and scholars have conducted research on it. Shengdong Mu analyzed the hardware implementation method of a cloud based control system for BD in response to the current problems of large data

collection volume, large transmission calculation volume, and poor real-time scheduling control performance [1]. Wang Tian made it possible to synchronously collect large-scale data and information. However, in practical applications, the complex and variable perception conditions made it difficult to fully trust the collected data. The conventional data cleaning methods based on perception nodes were no longer able to cope with massive amounts of data, and BD computing brought new opportunities to them [2]. Faced with the increasing volume of data, BD security faces severe challenges. With the deepening of research on BD security, there are more and more issues related to infrastructure security, data privacy protection, data management, and data integrity. Currently, the processing, analysis, and storage of BD mainly rely on cryptographic technology, which cannot meet the security requirements of BD. For this reason, Narayanan Uma was researching a new solution to BD security issues [3]. Based on cloud image processing technology, Xu Zheng utilized BD analysis technology to conduct case analysis in the BD environment [4]. The above is the application of BD in various industries, but there is a lack of integration between BD and ML algorithms.

ML algorithms, as a widely used algorithm, have been combined by many scholars with BD. Guha Samayita explored the application of BD analysis in information systems, operational management, and healthcare. He analyzed the BD analysis methodology system of ML algorithms in order to provide new ideas for in-depth research and solutions of related problems [5]. Wang Shu Lin used ML algorithms combined with BD to provide better prevention and treatment plans for adolescents. Research found that factors such as the compatibility, complexity, and relative benefits of technology were closely related [6]. With the massive amount of information and the improvement of global network connectivity, utilizing ML algorithms has become a key technology to ensure economic and social development. Therefore, Zhong Wei was researching ML for BD [7]. Although the above research integrates and analyzes BD with ML algorithms, it lacks the application of ML algorithms in BD mining.

With the development of various industries, data processing is no longer just collected for data analysis purposes, but more for business operations. BD mining technology obtains important information for enterprise decision-making through deep analysis of massive data, enhances its competitiveness, and ultimately achieves maximum profits. This article analyzed BD mining systems and cloud error correction systems, and combined them with computer ML algorithms to address the current problems in BD mining. Based on ML, it analyzed Cloud error correction and BD mining systems for computer ML algorithms. The experimental results proved that the ML algorithm was effective and had high accuracy. The research in this article laid the foundation for further improving and perfecting the theory and methods of BD mining.

2. Cloud Based BD Evaluation

2.1 BD Collection and Storage

The characteristics of BD are its huge scale, diverse types, and rapid dynamic updates. On this basis, optimizing the excavation of existing data would become a task that modern information industry needs to do in the future development process. With the development of technology, unstructured data is rapidly increasing, which requires basic devices with good performance, high update rate, and large capacity to achieve the storage of BD.

BD collection technology can be applied to water volume data monitoring, which is stored in the cloud for analysis of flow management [8]. For decades, valuable data has been available in product surveys. The quality of data from different sources has improved. Through a review of relevant literature, it can be found that the current product development methods mainly based on theoretical modeling are gradually transforming into data-based product development methods. BD collection

technology and data mining are currently the most popular research directions [9]. BD collection and storage provide new capabilities for dynamic provisioning, monitoring, and management of resources, making it easy to expand services and implement new categories of existing applications [10]. For a long time, people have believed that improving these useful tools to manage the massive data contained in this particular measurement system is a difficult task. In this context, BD collection and storage technology provides it with efficient information processing capabilities [11].

2.2 BD Preprocessing

BD preprocessing achieves effective mining of non-standard BD through preprocessing. To obtain accurate and credible information, it is necessary to obtain accurate and credible conclusions. Due to the fact that only structurally consistent data can be used on BD platforms, it is necessary to preprocess many data before starting data mining. The higher the quality of the preprocessed data, the more effective and reliable it can be.

2.3 Presentation and Application of BD Technology

Data presentation and application technologies can efficiently extract hidden information and knowledge from massive amounts of data, thereby achieving the goal of improving the intensification of socio-economic development. Currently, the application of BD is mainly focused on three aspects, namely government decision-making, public services, and business intelligence.

With the emergence of BD technology, it has provided unprecedented opportunities for people. In order to analyze, process, accumulate, absorb, and manage massive, organized, and unorganized health information, multiple methods are needed [12]. How to establish an effective classification model is a very important issue for classification systems with large sample features. Therefore, automated classification is an important task that requires a training approach that utilizes an input behavior for learning categories to assign them to a data object. New units can be identified based on predefined categories. When applying BD technology to analyze and identify data behavior, many vulnerabilities often arise. To solve this problem, it is necessary to develop new parsing algorithms and establish corresponding monitoring systems [13].

ML based BD processing methods can effectively integrate and evaluate massive and complex health data. However, to apply this method to the field of health, there are still some limitations that need to be overcome, including ethical issues in clinical applications and health services. Compared to conventional biological data analysis, data processing based on data mining has strong adaptability and scalability, so it can be widely applied in fields such as disease grading, disease diagnosis and classification, and patient survival prediction. However, there are still many problems in introducing it into the field of ML in the field of health monitoring [14]. After analyzing the power grid issues, the methodology required for BD and ML cautiously recognizes the advantages of BD and the importance of dealing with power grid problems. The most crucial thing is that they are able to plan and operate traditional power grids. BD and ML are based on the following aspects and provide a detailed introduction to their applications in different fields, such as electricity and energy, health and life sciences, government, telecommunications, internet and digital media, retail, finance, e-commerce, and customer service [15].

3. Cloud Error Correction and Big Data Mining

3.1 Data Mining Technology

In the cloud environment, achieving parallelization of data mining algorithms is a very important

technology. The parallelization methods mainly include several aspects such as association, clustering, classification, and regression. To adapt to the cloud computing environment and solve key problems in BD analysis, this article analyzed an efficient parallel method for BD analysis applications and applied it to practical applications.

In traditional data mining, the sources of data are very limited. This is mainly based on data warehouses. The types of data are also very simple. This is mainly based on structural data. On this basis, this article conducted parallel processing on cloud based data mining methods, so as to optimize the data structure of data mining to the maximum extent and improve the application field of the model. Data mining technology, as an important research content in artificial intelligence technology, can be applied to teaching systems to analyze learners' learning behavior, and can serve as a powerful means to improve teaching efficiency and provide students with educational decisions, thus providing a reference method for achieving personalized teaching. The learning behavior BD analysis technology aims to evaluate and statistically analyze a large amount of data, extract correlations between data, and discover individual learning behavior characteristics.

3.2 Cloud Based Error Correction Technology

Cloud error correction technology has a higher demand for computer reliability analysis in fields such as banking processing, information services, and financial computing. Therefore, it is particularly important to improve its availability, reliability, and maintainability. In a computer, if an error occurs at a certain node, it can be detected in a timely manner and then transferred to another node for business conversion. In a malfunctioning computer system, a controversial solution is achieved through arbitration mechanisms to discover, diagnose, and reconfigure the system. In other words, when a computer experiences an error, whether the computer can be detected by the system and make correct judgments plays a crucial role in the availability of the computer. Traditional fault-tolerant computers usually use a heartbeat mechanism to detect the condition of opponents. If the local computer does not receive the opponent's heartbeat within the specified time, it would be considered as an error and the opponent's services would be swapped back. Although this one-on-one arbitration mechanism is easy to implement, if there are no problems with the computer, the entire system would become a mess, and it would also lead to unnecessary switching, thereby increasing the cost of switching and leading to a decrease in system availability.

3.2.1 Common Errors in Computer Systems

The system error of a computer not only comes from errors in software operation, but also from errors in hardware. In recent years, due to the work of researchers, continuous improvements have been made to the circuit board technology. The main source of the problem is the transient failure of the computer's processor.

3.2.2 Evaluation of Computer System Fault Tolerance Technology

Compared to software level issues, hardware level issues are more likely to occur. Moreover, it is extremely difficult to recover. One of the main reasons for computer hardware problems is the failure during teleportation, which is often caused by a single system. Most of the reasons for short-term system downtime are due to system overload. In this case, the computer is merged due to multiple components: One is permanent incorrect component damage caused by a single system, and the repair company only replaces the damaged part.

The hardware fault-tolerant approach of computers is to quickly restart calculations in the event of a malfunction by controlling the computer's temperature or adding a backup. This can backup the device and provide failure information for easy maintenance. The design of multiple systems can

lead to computer problems, which leads to incorrect message feedback. It is necessary to have a unified management device to find faults. Due to the enhanced duplication of information, in the event of a computer malfunction, it is possible to confirm the integrity of secure information transmission and redundantly process the data.

3.3 The Application of Machine Learning in Cloud Error Correction and Big Data

In the field of computer science, machine learning algorithms play an important role in cloud error correction and big data mining. Machine learning shows great potential in data error correction and mining by analyzing a large number of data patterns to learn and identify specific patterns and trends.

In practical applications, all data processing is presented in a uniformly dispersed form. With the continuous expansion of data storage scale, hardware storage capacity can no longer fully meet the needs of centralized storage. In terms of data analysis, the attributes of data can be divided into numerical attributes, binary attributes, classification attributes, and sequential attributes. Using numbers to represent the types of attributes, attributes are divided into two categories: continuous attributes and discrete attributes. Two known quantities with binary characteristics are called binary quantities, usually represented by 0 or 1. Classification attributes are usually further extensions of binary attributes. The value it determines must be greater than 2, with each value corresponding to a different type and no distinction in size. There must be two or more known values in the sequence type attribute, with different values representing different levels and known values having higher and lower differences.

Addressing data differences is a key issue in the BD mining process. The solving effect depends on the algorithm and solving method. After unifying the basic attributes of all data, conduct centralized comparative analysis on high-dimensional mixed data; On the basis of preliminary work, the focus is on studying the unified processing method of BD and applying it to the unified calculation method of BD.

3.3.1 ML Algorithms

ML algorithms are an inevitable product of the development of computer technology, allowing computers to autonomously learn based on data and algorithms, and discover hidden patterns in data. Therefore, ML based methods have become a popular technology. The most important method in using ML is to classify it based on the obtained data.

ML algorithms can be applied to simpler programs, making the calculation methods more consistent. A summary and analysis were conducted on the multidimensional dataset $A=\{x_1, x_2,..., x_i,..., x_n\}$ containing n points.

Numerical characteristics can be divided into two categories: One is continuous characteristics, and the other is discrete characteristics. For continuous characteristics, the calculation method for their dissimilarity is as follows:

$$d_{x_i} = \frac{\left|x_i - x_j\right|}{\max x_i - \min x_j} \tag{1}$$

Among them, d_{x_i} represents the degree of property difference between data points x_i and x_j; x_i and x_j represent the values of the basic properties of the data points; $\max x_i$ represents the maximum value that can be taken from the entire dataset; $\min x_j$ represents the minimum value taken from the entire dataset.

The attribute differences of discrete data were calculated using two methods: classification and sequence. In the ordinal type characteristic, assuming that the given value of the ordinal type characteristic is in a set of ordered sequences of 1, 2,..., M, the formula for calculating the dissimilarity of ordinal type data is as follows:

$$d_{x_i,x_j} = \frac{|x_i - x_j|}{M - 1} \tag{2}$$

Among them, M represents the maximum value.

Due to the fact that the attributes of binary and classification only represent the relationship between the two types of attributes, without distinction between large and small levels, their calculation method is as follows:

$$d_{x_i,x_j} = \begin{cases} 0, & x_i \neq x_j \\ 1, & x_i = x_j \end{cases} \tag{3}$$

When the attribute values between x_i and x_j are the same, the value is 1, otherwise it is 0.

By unifying the processing of data, not only can the problem of large number and decimal fusion caused by the use of Euclidean formulas be solved, but also the optimal algorithm for BD mining can efficiently process high-dimensional mixed data while maintaining the same dimensionality.

3.3.2 Specific Applications

In terms of cloud error correction, machine learning algorithms can be applied to automatic error correction of various text data. These algorithms are based on contextual information and language models to identify and correct typos, grammar, and spelling. Different machine learning methods, such as supervised learning, unsupervised learning, and deep learning, can be applied here to achieve automatic and effective data error correction.

In big data mining, machine learning algorithms have also been widely applied. Through deep analysis and pattern learning of a large amount of data, machine learning can help us extract valuable information from massive and complex data. For example, important information such as group behavior patterns and consumer preferences hidden in big data can be discovered through algorithms such as clustering analysis and association rule mining. Not only that, machine learning can also solve some problems that traditional data processing methods are difficult to solve. For example, for the processing of unstructured and semi-structured data, machine learning algorithms can automatically and effectively extract valuable information from it.

In addition, machine learning algorithms can also efficiently store and query large amounts of data, making the processing of big data more convenient and efficient. In the future, with the improvement of computing power and the continuous development of big data technology, machine learning algorithms will play a greater role in cloud error correction and big data mining. For example, using deep learning technology, we can more accurately identify and correct text errors, and even achieve multilingual and multimodal data error correction. Meanwhile, by applying new machine learning algorithms, we can better extract valuable information from big data, providing strong support for decision-making and socio-economic development.

4. Evaluation of Experimental Results of ML Algorithms

In order to gain a more intuitive and comprehensive understanding of the practical application effects of machine learning algorithms in cloud error correction and big data mining, this article

compares them with neural network algorithms and compares the computational capabilities of different algorithms in cloud error correction and big data mining. To ensure the accuracy of the experimental results, these two algorithms were tested for their performance in the same testing environment. On this basis, two big data mining and cloud error correction methods were tested and the results were compared.

In cloud error correction and big data mining experiments, a fault-tolerant error correction method is used to initialize the server, and the tested pages are placed on the server for verification. After the server starts running, the monitoring management module can be used to check the corresponding business processes. This article analyzes redundant and non redundant processes. Research has found that machine learning algorithms applied in cloud error correction have better external service capabilities than non error correction technologies. The availability test fault injection results of cloud error correction are shown in Table 1.

Table 1 Cloud error correction availability testing and fault injection results

Category	State	Switch or not
Redundant processes	Continuous	Correct
Redundant segmentation process	Continuous	Deny
Non redundant processes	Interrupt	Deny
Non redundant segmented processes	Interrupt	Deny

In order to achieve more realistic experiments, this article will test two different methods of cloud error correction and big data mining in a cloud computing environment, and compare the service execution efficiency under different error information processing modes using methods such as ML and neural networks. After fault tolerance testing, restart the computer once to ensure that all conditions are the same; We recorded experiments with different numbers of threads to reduce errors caused by initializers.

This article tested its performance using neural network algorithms and machine learning algorithms. Among them, KB/sec represented the amount of data received from the server per second. The non fault tolerant test results of the two algorithms are shown in Table 2.

Table 2 Test results of two algorithms in non fault-tolerant manner

Number of data points	Neural network algorithm(KB/sec)	Machine learning algorithms(KB/sec)
1000	4139.2	5234.3
2000	4147.4	5347.2
3000	5168.7	6124.9
4000	5152.5	6243.4
5000	5149.6	6089.6

From the data in Table 2, it could be seen that as the number of data points increased, the amount of data received from the server per second by the system side showed a trend of first increasing and then decreasing. According to the analysis of the non fault-tolerant methods of the two algorithms, it was found that the data volume using ML algorithms was higher, so using ML algorithms could better achieve data reception and transmission.

Due to the tendency of computers to malfunction, system error correction was necessary. Next, a fault-tolerant approach was used to test the system. The fault tolerance test results of neural network algorithm and ML algorithm are shown in Table 3.

Comparing the test results in Table 2 and Table 3, it could be seen that the test data run in a fault-tolerant manner using the neural network method was lower. The amount of data tested using ML algorithms was improved. Therefore, under fault-tolerant conditions, ML algorithms should be used for testing.

Table 3 Test results of fault tolerance methods for two algorithms

Number of data points	Neural network algorithm(KB/sec)	Machine learning algorithms(KB/sec)
1000	3245.3	6234.4
2000	3142.7	6782.5
3000	3562.4	6920.6
4000	3842.1	7245.6
5000	3621.3	7872.6

Cloud error correction and BD require a certain amount of time to run. Next, the main test indicator, that is, average response time would be further compared and analyzed. The true values of the system under normal response, as well as the average response time after using neural network algorithms and ML algorithms, were compared. The average response time for the three scenarios is shown in Figure 1.

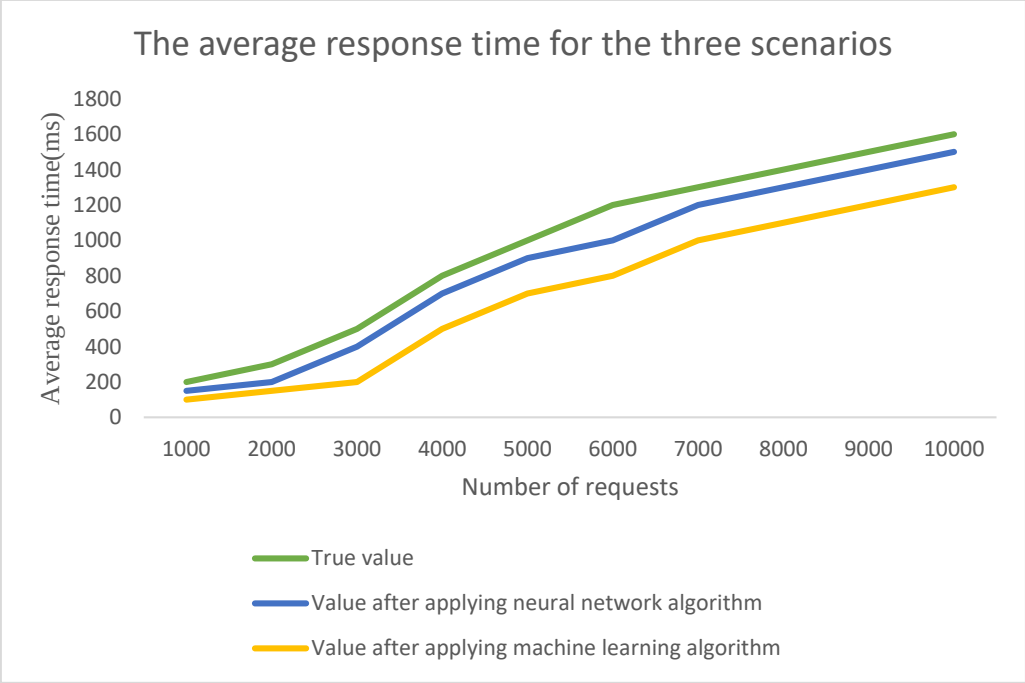

Figure 1 Average response time for three scenarios

From the data in Figure 1, it could be seen that by comparing the time taken by the two algorithms to perform data mining in different environments, the average running time of neural network algorithms and ML algorithms in Cloud error correction and BD mining systems was lower than the true value. As the number of requests increased, the running time of the system also continued to increase. However, in the same number of system requests, it was evident that ML algorithms took less time and were more efficient than neural network algorithms.

This article analyzed a cloud error correction system and BD mining system constructed using ML algorithms in a cloud computing environment. Experiments showed that the ML algorithm scheme had low performance loss, and could effectively improve system stability. It could avoid the complexity of hardware customization, and had the characteristic of being completely transparent to users and applications.

Next, 10 sets of data were selected. The accuracy of using neural network algorithms and ML algorithms in Cloud error correction and BD mining systems was compared. The accuracy of both operations in the system was analyzed. The operational accuracy of the two algorithms is shown in Figure 2.

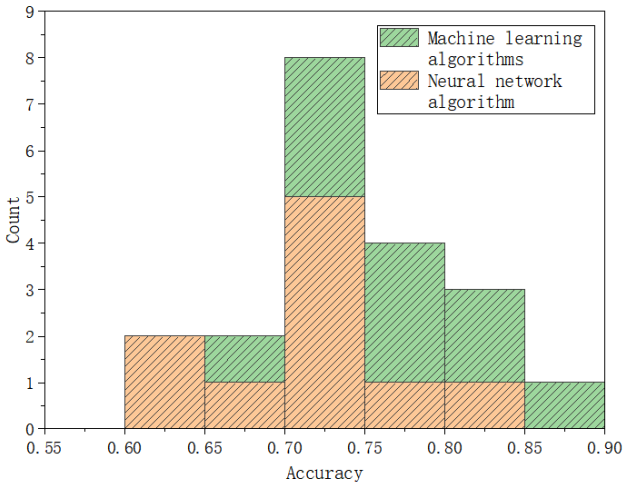

Figure 2 The running accuracy of two algorithms

As shown in Figure 2, the average accuracy of using neural network algorithms was 71.3%, while the average accuracy of using ML algorithms was 77.1%. The accuracy of using ML algorithms was 5.8% higher than using neural network algorithms. Therefore, according to the experimental results, it could be seen that the ML algorithm in this paper outperformed the neural network algorithm in terms of prediction accuracy, and the algorithm in this paper had high effectiveness and practicality.

5. Conclusions

With the increasing amount of massive data, existing storage systems cannot meet the increasing demand for applications. In the context of BD, with the new requirements of people for BD, technologies such as storage, networking, and computing are also rapidly developing. BD mining technology can quickly and efficiently mine and analyze a company's massive data, and extract useful information from it. This article analyzed a data mining system with high participation, low technical development requirements, and fast response rate. By introducing ML algorithms, this article analyzed cloud based error correction systems and BD mining systems for computer ML algorithms, and verified this system through experiments. The experimental results indicated that the ML algorithm had good feasibility. With the advent of the wireless internet era and users' desire for real-time information data, in the future era of cloud computing popularization and development, BD mining systems would fully demonstrate their charm and functionality. In the cloud environment, data mining technology has been proven to be a feasible and effective method that can effectively cope with the growth of massive data.

References

[1] Shengdong, Mu, Xiong Zhengxian, and Tian Yixiang. "Intelligent traffic control system based on cloud computing and big data mining." IEEE Transactions on Industrial Informatics 15.12 (2019): 6583-6592.

[2] Wang, Tian, Haoxiong Ke; Xi Zheng; Kun Wang; Arun Kumar Sangaiah; Anfeng Liu. "Big data cleaning based on mobile edge computing in industrial sensor-cloud." IEEE Transactions on Industrial Informatics 16.2 (2019): 1321-1329.

[3] Narayanan, Uma, Varghese Paul, and Shelbi Joseph. "A novel system architecture for secure authentication and data sharing in cloud enabled Big Data Environment." Journal of King Saud University-Computer and Information Sciences 34.6 (2022): 3121-3135.

[4] Xu, Zheng, Cheng Cheng, and Vijayan Sugumaran. "Big data analytics of crime prevention and control based on image processing upon cloud computing." Journal of Surveillance, Security and Safety 1.1 (2020): 16-33.

[5] Guha, Samayita, and Subodha Kumar. "Emergence of big data research in operations management, information systems, and healthcare: Past contributions and future roadmap." Production and Operations Management 27.9 (2018): 1724-1735.
[6] Wang, Shu Lin, and Hsin I. Lin. "Integrating TTF and IDT to evaluate user intention of big data analytics in mobile cloud healthcare system." Behaviour & Information Technology 38.9 (2019): 974-985.
[7] Zhong, Wei, Ning Yu, and Chunyu Ai. "Applying big data based deep learning system to intrusion detection." Big Data Mining and Analytics 3.3 (2020): 181-195.
[8] Smys, S., Abul Basar, and Haoxiang Wang. "CNN based flood management system with IoT sensors and cloud data." Journal of Artificial Intelligence 2.04 (2020): 194-200.
[9] Kuo, Yong-Hong, and Andrew Kusiak. "From data to big data in production research: the past and future trends." International Journal of Production Research 57.15-16 (2019): 4828-4853.
[10] Meryem, Amar, and Bouabid EL Ouahidi. "Hybrid intrusion detection system using machine learning." Network Security 2020.5 (2020): 8-19.
[11] Malik, Rahul, and Madaan Nishi. "RETRACTED ARTICLE: Flexible big data approach for geospatial analysis." Journal of Ambient Intelligence and Humanized Computing 13.2 (2022): 737-756.
[12] Rehman, Arshia, Saeeda Naz, and Imran Razzak. "Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities." Multimedia Systems 28.4 (2022): 1339-1371.
[13] Guezzaz, Azidine, Younes Asimi; Mourade Azrour; Ahmed Asimi. "Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection." Big Data Mining and Analytics 4.1 (2021): 18-24.
[14] Ngiam, Kee Yuan, and Wei Khor. "Big data and machine learning algorithms for health-care delivery." The Lancet Oncology 20.5 (2019): e262-e273.
[15] Salkuti, Surender Reddy. "A survey of big data and machine learning." International Journal of Electrical & Computer Engineering 10.1 (2020):2088-8708.