

Research Article

Big Data Analysis of Marketing User Intelligence Information Based on Deep Learning

Zhongqi Shen 

Business College, Jiangsu University of Science & Technology, Zhangjiagang, Jiangsu 215600, China

Correspondence should be addressed to Zhongqi Shen; 1412480418@st.usst.edu.cn

Received 7 May 2022; Revised 8 June 2022; Accepted 17 June 2022; Published 30 June 2022

Academic Editor: Muhammad Muzammal

Copyright © 2022 Zhongqi Shen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To solve the problem that the explosive growth of data cannot be effectively analyzed, a big data analysis and prediction system based on deep learning is proposed. The systems proposed in this paper optimize the design of circulatory neural networks using an optimization processing strategy in addition to circulatory neural networks based on a distributed parallel processing framework that studies and analyzes disrupted neural networks during the in-depth study. The results are as follows: using the DCNNPS parallel strategy can effectively improve the training speed of the convolutional neural network to a small extent, and the speed can be improved by more than 50%. Compared with the other two, the network optimized by DCNNPS is less sensitive to data growth in the scenario of a large amount of data and is more suitable for processing a large amount of data. Moreover, with the increase of computing nodes, the acceleration effect is more obvious, the acceleration ratio is higher, and the improvement energy is about 70%. It is proved that the system proposed in this paper can process massive data very effectively and make a great contribution to the development of the industry.

1. Introduction

Effective data analysis is integral to data processing calculations, machine learning, and other system support [1]. Traditional analytics systems are structured data-based online analytics processing and online transaction processing systems. The flow information processing mechanisms used by these traditional analytics tools are very effective in the data analysis process. Performance [2]. With the gradual development of deep learning, software tools applied to deep learning are also born and gradually popularized. Each software tool focuses on different emphases. These open-source deep learning tool systems provide good system support for deep learning, but for ordinary industry users, their use threshold is relatively high. They not only need a lot of time cost to learn relevant API and corresponding framework language syntax but also need a lot of hardware costs such as computing servers to build a deep learning framework. This difficulty can be well solved by building a deep learning system with the nature of cloud computing with the help of distributed computing

technology. Therefore, designing a big data analysis system based on deep learning, which can assist ordinary data analysts to carry out convenient, effective, and suitable deep learning analysis, has important practical significance for reducing the use threshold and improving the work efficiency of users [3, 4]. Figure 1 shows the big data analysis of marketing user intelligence information.

Establishing a large data analysis and forecasting system is a very convenient and popular solution to effectively improve the efficiency of data analysts and reduce operating costs. All parties design and implement big data analysis systems with different architectures for different objectives. PDminer has a highly modular and scalable 4-tier architecture, namely, workflow subsystem, user interface subsystem, data preprocessing subsystem, and data mining subsystem. It completes the analysis and mining tasks through flexible combination and efficient cooperation among algorithms [5]. Mhango et al. and others focus on the underlying basic computing functions, develop and integrate Python library and NumPy matrix calculation library suitable for deep learning, have strong algorithm scalability,

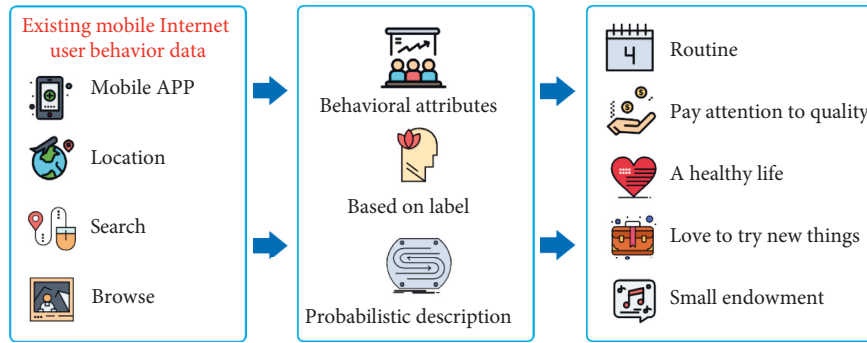


FIGURE 1: Big data analysis of marketing user intelligence information.

and provide GPU acceleration services [6]. Caffe, a deep learning architecture developed and constructed by Ma L. and others, focuses on different aspects of theano. It is designed and implemented with rich and complete basic modules to support the rapid construction of the neural network model and the high-speed training of the neural network model [7]. Liu et al. and others have shown that Caffe can process more than 40 million images per day using a single nvidia40 training depth convolution network (CNN), and the speed can reach 5000 FPS when training with Titan GPU [8]. The open-source big data collection and processing system developed by Gong B and others integrates the independently developed the Minos monitoring system, improves, and optimizes the cluster management, load balancing, job debugging, and other aspects of Hadoop and other frameworks and improves the overall computing performance of the system. The online workflow mechanism adopted by the ClowdFlows system shows strong usability and performance in distributed big data mining and analysis [9]. Liang et al., respectively, introduced deep learning into data processing scenarios in fields such as power load forecasting and military target image classification to improve system processing performance [10]. In addition, to better support the use of indepth learning training, Wang and Srikanth pointed out that academia and industry design and implement their own software systems to improve the efficiency of model training. Among them, the performance acceleration technology of each framework software is different, but it mainly includes GPU, ASIC, and FPGA [11]. BDA analysis engine takes Hadoop, spark, and other framework integration processing as the basic computing architecture and integrates heterogeneous resources, and the library of general algorithms for big data analysis provides a highly reusable distributed computing system, reducing the complexity of using big data analysis. The DI-X (Data Intelligence X) system developed by Tencent supports users to run the algorithm on the cloud GPU and CPU server through visualization, effectively reducing the use threshold of algorithm engineers and data scientists. Ali's data addition system provides many business functions that drastically reduce data-related workloads, such as modeling and ETL, and provides data services for various industries in combination with tools and engines. However, these systems are still insufficient in supporting deep learning, and the use of model structure is particularly complex, which cannot give

consideration to the convenience of use and the flexibility of the model. DLP (a parallel deep learning programming framework based on heterogeneous architecture) has developed an indepth training programming framework based on a hybrid heterogeneous architecture of the CPU and GPU [12]. The framework establishes a unified module library to realize the visual construction of deep learning model, so as to improve the convenience of programming. However, these systems and other systems with deep learning functions can achieve good results in specific application fields, but the improvement of model performance makes the model structure more complex and cannot maintain convenient use and flexible model construction at the same time. Moreover, when dealing with specific industry data, a single indepth learning model often cannot achieve the ideal effect. The combination of multiple algorithm models can face the changing needs and achieve the desired results, and the operation and use of these systems in this regard is more complex [13].

Based on the above research background and current situation, this paper designs and implements a big data analysis and prediction system based on deep learning, which aims to support the use of deep learning and reduce the user's use cost and operational complexity. Firstly, this paper investigates the existing processing framework and related technologies related to big data and indepth learning. Combined with the analysis of system design scenarios, functional requirements, and performance requirements, this paper designs and puts forward a five-tier overall architecture including data absorption, system framework, algorithm model, integration components, and service application, which lays the foundation for the specific implementation of the system.

2. Research Methods

2.1. Architecture Research and Design of Big Data Analysis and Prediction System Based on Deep Learning. The traditional general deep learning algorithm predictive analysis adopts a stand-alone operation mode, and the processing speed of big data is greatly limited. To improve the training efficiency of the deep learning algorithm model, the reasonable design of the deep learning algorithm model under a large amount of data, the use of parallel optimization method, and the combination of the deep learning and distributed computing

framework can effectively improve the processing speed of the original algorithm under large amount of data. Due to high requirements on users' programming ability for indepth learning prediction and analysis, it is difficult to get started. Combined with building system services in modular and visual forms, providing user operation services can reduce the difficulty of getting started [14, 15].

Therefore, to facilitate users to carry out rapid deep learning prediction analysis, we improve the support granularity of deep learning algorithm model and simplify the complexity of deep learning model construction. This topic breaks down big data hypothesis analysis tasks from data representation, data management, and task performance and builds and develops a big data analysis and prediction system based on an indepth study. The architecture includes a big data absorption module, data warehouse module, hybrid computing processing, algorithm model library, visual interface module, and other functional parts. The specific overall design of the architecture is seen in Figure 2, which includes five layers: data absorption, system framework, algorithm model, integration component, and service application.

2.2. Parallel Optimization Processing Strategy for the Convolutional Neural Network. In addition to the DCNNPS algorithm developed and implemented in this paper, the optimization development strategy divides the training servers into two types: computing server nodes and multiparameter server nodes and divides all model training into computational gradients. Update parameters by server node and parameter server node in two parts [16]. The calculation server node carries out gradient calculation based on the BP algorithm. In the process of forwarding propagation calculation, it adopts the idea of parallel batch SGD training to obtain the training data subset of the quantitative batch from the training data set for calculation. In the process of error back propagation, the original serial propagation is reconstructed, and then the parameter is updated. The hierarchical parameter update method is adopted, that is, the error back propagation of the previous level is completed, that is, the parameter change of this level is transmitted to the parameter server for parameter update [17, 18]. At the same time, the computing server node combines the cache mechanism of spark to reduce the loss of repeatedly reading disk data in the process of multiple rounds of iterative calculation.

Multiparameter server architecture is the most commonly used distributed scalable machine learning architecture. In the process of using the parameter server architecture, it is allowed to process multiple algorithms at the same time. As shown in Figure 3, the whole architecture is mainly divided into one parameter service group and multiple computing service groups. Each server node in the parameter service group retains a partition part of the globally shared parameters. Parameter server nodes can communicate with each other to transfer and migrate parameters to ensure the reliability and scalability of the service.

The computing service group is responsible for handling all computing tasks in model training. Each computing server is assigned partial partitions in the training data to perform local gradient computing tasks. Different from the parameter server node, the calculation server node only communicates with the parameter server node to pull parameters from the parameter server node and update parameters to the globally shared parameters of the parameter server node. Each computing service group has its own independent and isolated scheduling queue. The scheduling queue allocates computing tasks to the corresponding computing server nodes and monitors the operation status of each computing task at the same time. If the task fails, the scheduling queue will recall the unfinished task and assign it to other computing server nodes for execution [19].

2.3. Parallel Gradient Computing. Parallel microbatch SGD algorithm is the most commonly used optimization technology for CNN model training. The SGD computing process is executed in parallel in multiple computing server nodes. Each computing server node allocates and divides some partitions of all training data sets. After all the calculation tasks are completed, the weight results obtained from the calculation task will be aggregated into the parameter server, and the average evaluation and other operations will be carried out to update them to the shared parameters [20]. The process of calculating the parallel gradient of a computing server node is roughly divided into two stages: forward and reverse distribution. During the forward spread phase, the activation function is calculated using the input image and propagated in the forward direction. The calculation formula is

$$I_l = \sum_{i=0} W_l^i I_{l-1} + \theta_l. \quad (1)$$

Here, I_l and θ_l are the activation function and offset of layer l in CNN network, respectively, while W_l^i is the weight value of the i -th neuron in the eigenvector diagram between layer l and layers $l-1$ and W .

During the feedback phase, the error value is calculated based on the value of the activation function of the previous layer and the feedback. The calculation formula is shown in formula (2), where Err_l is the error value calculated by layer l and W_l^i is the weight value of the i -th neuron of layer l .

$$Err_l = \sum_{i=0} W_{l+1}^i Err_{l+1}. \quad (2)$$

The cost of communication between the computing server node and the parameter server node causes an update delay and the parameter aggregation becomes unstable.

2.4. Test and Analysis of Parallel Optimization Processing Strategy for Convolutional Neural Network. When testing the classification prediction accuracy and operation efficiency of the DCNNPS method, the classic Mnist, Cifar10, and Flowers data sets are used. The specific description is

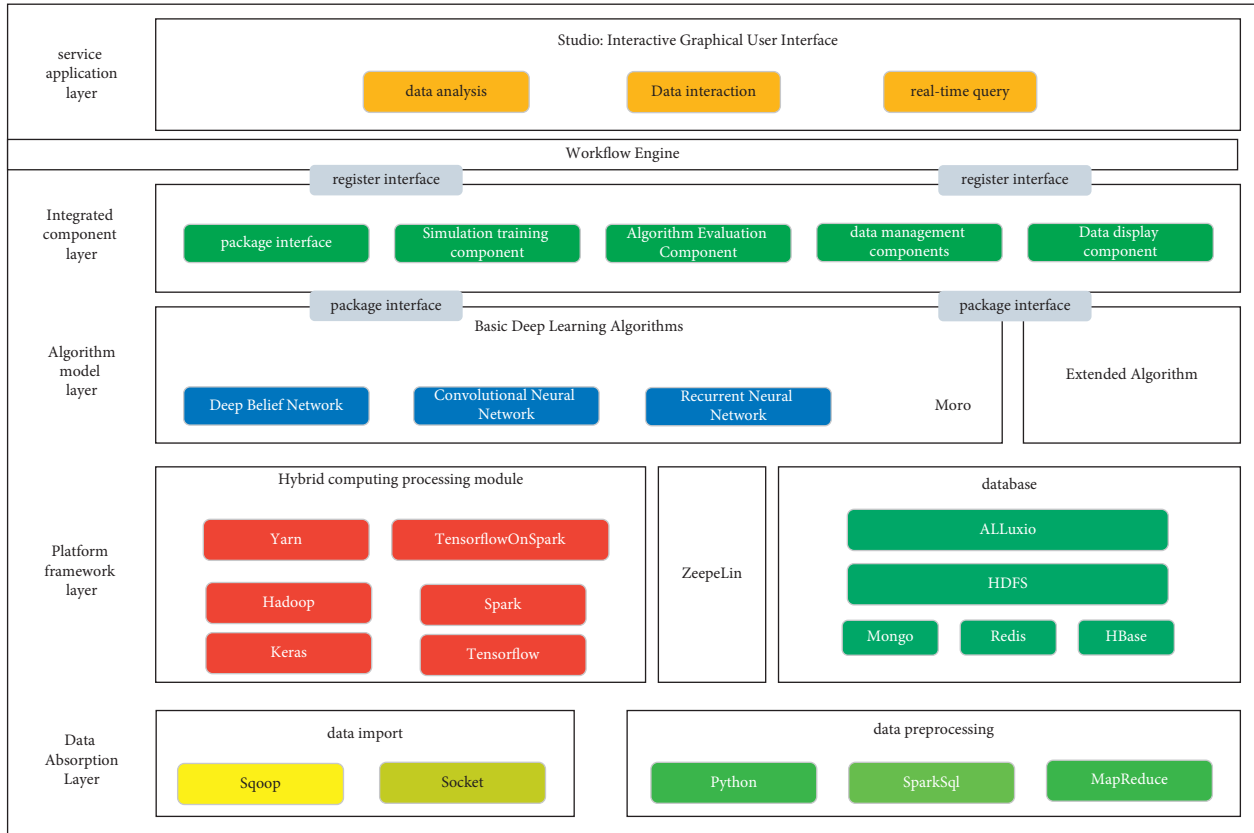


FIGURE 2: Architecture of big data analysis and prediction system based on deep learning.

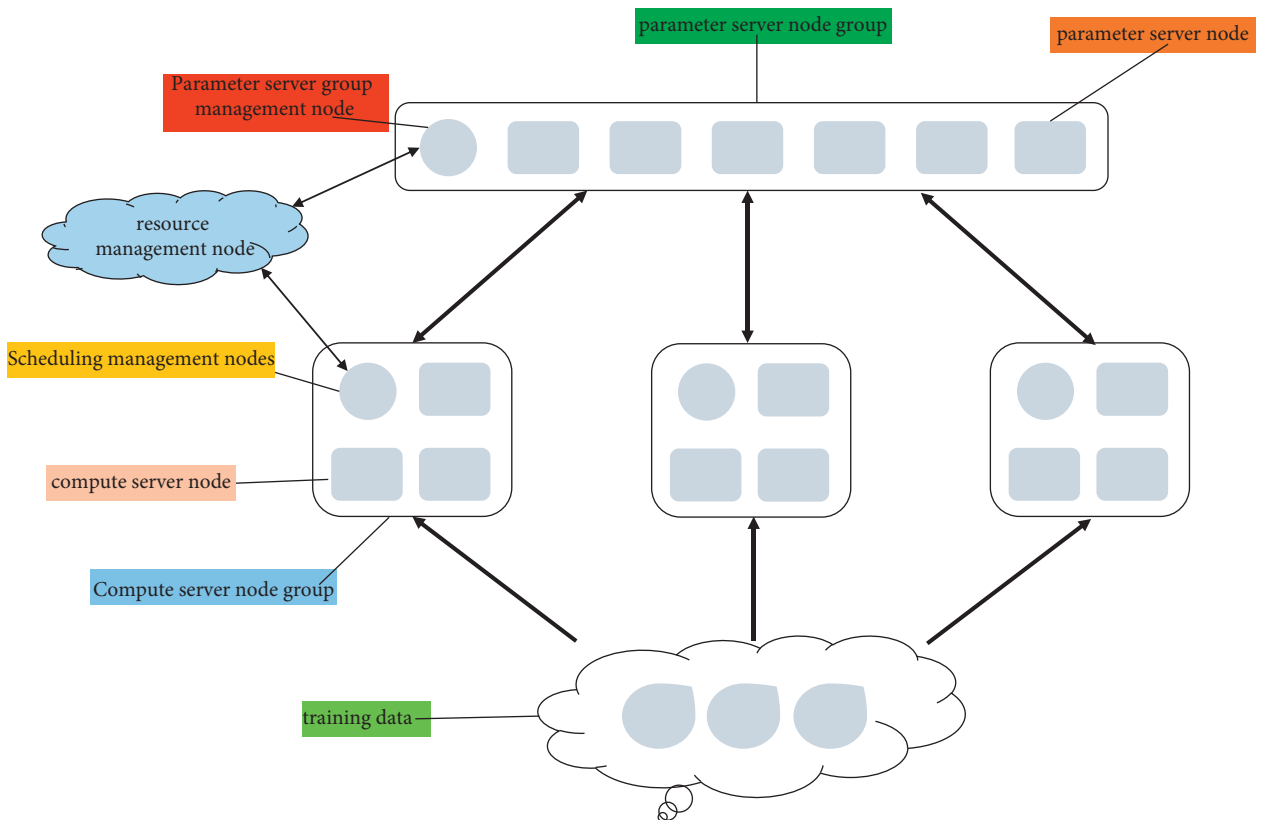


FIGURE 3: Architecture of the parameter server.

TABLE 1: Data set information.

Dataset name	Data volume	Picture size	Number of categories
Mnist	70000	28 * 28 * 1	10
Cifar10	70000	32 * 32 * 3	10
Flowers	3840	64 * 64 * 5	5

TABLE 2: Comparison of multiclassification results of different CNN network models.

	Mnist		Cifar10		Flowers	
	Time (s)	Accuracy	Time (s)	Accuracy	Time (s)	Accuracy
LeNet-5 in DCNNPS	18.525	0.973	17.934	0.802	15.528	0.821
Distributed LeNet-5	42.011	0.966	41.412	0.785	34.151	0.803
Single LeNet-5	57.329	0.986	54.917	0.831	9.964	0.840
ResNet-50 in DCNNPS	9038.475	0.967	8837.235	0.902	530.116	0.935
Distributed ResNet-50	25848.659	0.960	26276.250	0.882	1453.953	0.893

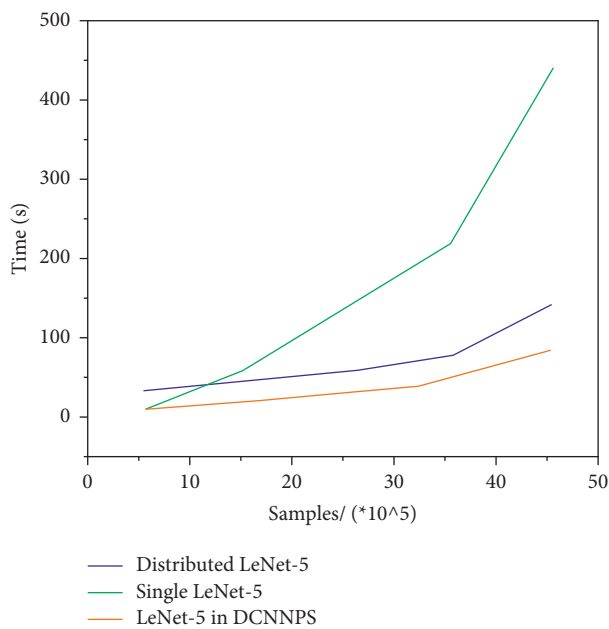


FIGURE 4: Training time experiment of different LeNet-5 networks under different data volumes.

shown in Table 1. Among them, the Mnist data set has 10 categories, including 70000 28 * 28 monochrome handwritten picture data. Each line of data is the pixel representation of a picture with a classification label; the Cifar10 data set has 10 categories, including 70000 32 * 32 color picture data. Each row of data is the pixel representation of one picture with a classification label. The Flowers dataset has five categories, including 3840 64 * 64 color picture data. Each row of data is the pixel representation of a picture with classification labels [21–23].

When testing the acceleration efficiency of DCNNPS method, Mnist data sets of 10000, 60000, 150000, 240000, and 480000 are used, respectively, and compared with the same algorithm implemented by single machine and distributed Tensorflow.

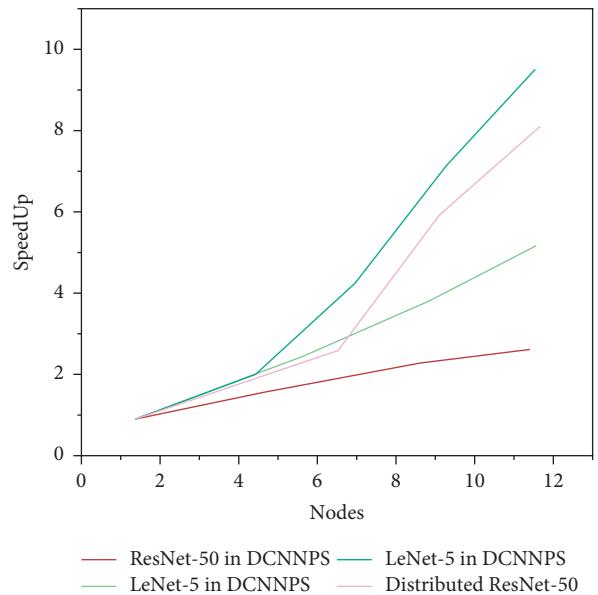


FIGURE 5: Acceleration ratio experiment of different CNN networks under different number of nodes.

3. Result Analysis

3.1. Classification Prediction Accuracy and Operation Efficiency Experiment. In this experiment, LeNet-5 in DCNNPS and ResNet-50 in DCNNPS represent the LeNet-5 and ResNet-50 networks optimized by DCNNPS, distributed ResNet-50 and distributed LeNet-5 represent the LeNet-5 and RESNET 50 networks realized by distributed Tensorflow, and Single LeNet-5 represents the LeNet-5 network realized by single Tensorflow. To train and test the above network model, this paper divides the training package and test package into 5:1 ratios, and the specific results are shown in Table 2. The results show that the DCNNPS parallel strategy can effectively improve the speed of convolutional neural network training with low-precision changes [24, 25].

3.2. Acceleration Efficiency Test. In this experiment, as shown in Figure 4, the training speeds of LeNet-5 in DCNNPS, single LeNet-5, and distributed LeNet-5 under different data sets are compared. In addition, the training acceleration efficiency of LeNet-5 in DCNNPS, ResNet-50 in DCNNPS, distributed ResNet-50, and distributed LeNet-5 in data sets of the same size, and different node numbers is compared. The results are shown in Figure 5. From the above results, it can be seen that the network optimized by DCNNPS is less sensitive to data growth and more suitable for processing a large amount of data than the other two in the scenario of a large amount of data. Moreover, with the increase of computing nodes, the acceleration effect is more obvious and the acceleration ratio is higher.

4. Conclusion

The analysis and prediction of a large amount of data are one of the important topics in the current era of data explosion. Nowadays, deep learning has gradually developed into an important method for a large amount of data analysis and prediction. However, unlike traditional data analysis and machine learning algorithms, the use of indepth learning to analyze and predict data requires not only the study and mastery of complex frameworks and complex programming models but also expensive equipment to build the running environment. Ordinary data analysts who need to use it need to pay high learning and equipment costs. At the same time, how to train the deep learning model more efficiently is also an important topic in the popularization and use of deep learning. Therefore, this paper fully analyzes these realistic demand scenarios, designs, and implements a high-performance big data analysis and prediction system based on deep learning. The specific work is as follows:

- (1) The in-depth study-based big data analysis and prediction system design requirements is defined in the design options, functional requirements, and performance requirements, as well as the five-story general architecture and functional modules of the big data analysis and prediction system studies and invents indepth learning.
- (2) In addition to advanced neural networks, an indepth optimization study is being conducted, and a strategy for the parallel optimization of convolutional neural networks based on a distributed processing system is being proposed. In addition to developing and using the system as a package, the optimized network is tested and analyzed.
- (3) Detailed testing and analysis of the operation of big data analysis and forecasting systems are based on indepth training. The results of the experiments show that the operation and performance of a big data analysis and forecasting systems based on the indepth learning meet design requirements and are practical in practical versions.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] H. Akta, T. Kzldeniz, and Z. Nal, "Classification of pistachios with deep learning and assessing the effect of various datasets on accuracy," *Journal of Food Measurement and Characterization*, vol. 16, no. 3, pp. 1983–1996, 2022.
- [2] G. Ciaparrone, L. Chiariglione, and R. Tagliaferri, "A comparison of deep learning models for end-to-end face-based video retrieval in unconstrained videos," *Neural Computing & Applications*, vol. 34, no. 10, pp. 7489–7506, 2022.
- [3] C. Rea, K. Montes, R. S. Granetz, R. Sweeney, and R. A. Tinguely, "Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks," *Nuclear Fusion*, vol. 61, no. 2, 2021.
- [4] B. Ramalingam, A. A. Hayat, M. R. Elara et al., "Deep learning based pavement inspection using self-reconfigurable robot," *Sensors*, vol. 21, no. 8, p. 2595, 2021.
- [5] A. Nasif, Z. A. Othman, and N. S. Sani, "The deep learning solutions on lossless compression methods for alleviating data load on iot nodes in smart cities," *Sensors*, vol. 21, no. 12, p. 4223, 2021.
- [6] J. K. Mhango, E. W. Harris, R. Green, and J. M. Monaghan, "Mapping potato plant density variation using aerial imagery and deep learning techniques for precision agriculture," *Remote Sensing*, vol. 13, no. 14, p. 2705, 2021.
- [7] L. Ma, Y. Li, J. Li, J. M. Junior, and M. A. Chapman, "Boundarynet: extraction and completion of road boundaries with deep learning using mobile laser scanning point clouds and satellite imagery," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 99, pp. 1–17, 2021.
- [8] S. Z. Liu, R. S. Sinha, and S. H. Hwang, "Clustering-based noise elimination scheme for data pre-processing for deep learning classifier in fingerprint indoor positioning system," *Sensors*, vol. 21, no. 13, p. 4349, 2021.
- [9] B. Gong, R. Cai, Z. Cai, Y. Ding, and M. Peng, "Selection of acoustic modeling unit for Tibetan speech recognition based on deep learning," *MATEC Web of Conferences*, vol. 336, no. 6, 2021.
- [10] F. Liang, C. Li, and X. Fu, "Evaluation of the effectiveness of artificial intelligence chest ct lung nodule detection based on deep learning," *Journal of Healthcare Engineering*, vol. 2021, no. 3, pp. 1–10, Article ID 9971325, 2021.
- [11] J. Wang and P. Srikantha, "Stealthy black-box attacks on deep learning non-intrusive load monitoring models," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3479–3492, 2021.
- [12] V. I. Ilie, C. O. Truica, E. S. Apostol, and A. Paschke, "Context-aware misinformation detection: a benchmark of deep learning architectures using word embeddings," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.
- [13] L. Jiang, X. Wang, W. Li, L. Wang, and L. Jia, "Hybrid multitask multi-information fusion deep learning for household short-term load forecasting," *IEEE Transactions on Smart Grid*, vol. 12, no. 99, p. 1, 2021.

- [14] Y. Li, P. Cui, C. Ye et al., "Accurate prediction of earthquake-induced landslides based on deep learning considering landslide source area," *Remote Sensing*, vol. 13, no. 17, p. 3436, 2021.
- [15] J. Wang and Y. Gao, "Suspect multifocus image fusion based on sparse denoising autoencoder neural network for police multimodal big data analysis," *Scientific Programming*, vol. 2021, no. 3, pp. 1–12, 2021.
- [16] L. Zhang, Y. Guan, and S. C. Jiang, "Investigations of soil autotrophic ammonia oxidizers in farmlands through genetics and big data analysis," *Science of the Total Environment*, vol. 777, no. 3, 2021.
- [17] W. Jaung, "Changes in human-nature relations during pandemic outbreaks: a big data analysis," *Science of the Total Environment*, vol. 768, no. 1, 2021.
- [18] Y. Qiu, X. Zhu, and J. Lu, "Fitness monitoring system based on internet of things and big data analysis," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.
- [19] T. Shang, Z. Zhao, X. Ren, and J. Liu, "Differential identifiability clustering algorithms for big data analysis," *Science China Information Sciences*, vol. 64, no. 5, pp. 1–18, 2021.
- [20] Y. Liang, D. Quan, F. Wang, X. Jia, and T. Li, "Financial big data analysis and early warning platform: a case study," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.
- [21] J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 113–126, 2022.
- [22] J. Guo, M. Zhang, Q. Shang, F. Liu, X. Wu, and X. Li, "River basin cyberinfrastructure in the big data era: an integrated observational data control system in the heihe river basin," *Sensors*, vol. 21, no. 16, p. 5429, 2021.
- [23] S. Arifuzzaman, M. Khan, and M. Marathe, "Fast parallel algorithms for counting and listing triangles in big graphs," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 1, 2020.
- [24] M. Guo, Z. Yu, Y. Xu, Y. Huang, C. Li, and J. G. Haro, "Menet: a deep convolutional neural network for extracting mangrove using sentinel-2a data," *Remote Sensing*, vol. 13, no. 7, pp. 1–24, 2021.
- [25] E. L. Ferguson, "Multitask convolutional neural network for acoustic localization of a transiting broadband source using a hydrophone array," *Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 248–256, 2021.