Research Article

Lei Tian and Ance Zhao*

Mechanical product design and manufacturing system based on CNN and server optimization algorithm

https://doi.org/10.1515/nleng-2024-0057
received August 6, 2024; accepted November 7, 2024

Abstract: Mechanical optimization refers to the use of advanced mechanical equipment and physical processes to optimize raw materials, products, *etc.*, in production, reduce energy consumption, and improve production efficiency and product quality. In the cloud manufacturing mode of manufacturing enterprises, the design of mechanical products requires server optimization based on the demand side. The existing server optimization algorithms are not intelligent enough for server discovery and optimization in this mode. A mechanical product design service and manufacturing method based on convolutional neural network and server discovery and optimization architecture is proposed, which combines deep learning and optimal path algorithm to select better mechanical product services. Research outcomes showed that the training accuracy and testing accuracy of the research network were 98.89 and 96.88%, respectively, with training loss and testing loss of 0.04619 and 0.08921, respectively. In contrast, the training accuracy and testing accuracy of the comparative network were 93.95 and 85.31%, respectively, with training loss and testing loss of 0.3556 and 0.4872. The running time of the two neural networks is 105 and 73 min, respectively. Overall, the accuracy of the research network is high, and the loss is small. The training accuracy and loss of the new activation function proposed in the study have always been superior to other functions. The path and running time of the server optimization path algorithm have better performance. It can be seen that the research proposed methods have good performance and advantages in mechanical product design services and optimization, which can provide technical references and directions for the development of manufacturing service-oriented enterprises in cloud manufacturing mode.

Keywords: convolutional neural network, server optimization, mechanical products, manufacturing system, product design

1 Introduction

With the rapid growth of information technology and the deepening of globalization, the manufacturing industry is facing unprecedented challenges and opportunities. The traditional manufacturing model is no longer capable of adapting to the rapid changes in market demands and the increasingly complex product production processes that are characteristic of the contemporary era. In contrast, the adoption of novel information technologies, including cloud computing, big data, and the Internet of Things (IoT), offers a promising avenue for the advancement of manufacturing industry innovation [1,2]. In this context, cloud manufacturing, as an innovative service-oriented networked manufacturing model, is gradually emerging and becoming a significant driver of transformation and upgrading in the manufacturing industry. Cloud manufacturing represents a convergence of advanced information technology, manufacturing technology, and emerging IoT technologies, exemplifying the concept of manufacturing as a service [3]. It employs state-of-the-art information technology concepts, including cloud computing, to facilitate the provision of high-value, cost-effective, and globally distributed manufacturing services for a diverse range of products in various network resource environments. Many researchers have put forward their insights on product manufacturing and design. IoT devices cannot address the issues of high-dimensional machine detection and data imbalance. Murugiah *et al.* proposed a new Industry 4.0 predictive manufacturing system for inspecting machineries. In the initial phase of the study, data were collected from sensors utilized in the IoT industry. Following the

* **Corresponding author: Ance Zhao,** Department of Students' Affairs, North China Electric Power University (Baoding), Baoding, 071003, China, e-mail: zac17325376537@126.com
Lei Tian: Department of Mechanical Engineering, North China Electric Power University (Baoding), Baoding, 071003, China

extensive cleaning of the data set, deep features were extracted using the multi-scale dilation attention convolutional neural network. A hybrid algorithm, namely probabilistic beetle swarm-butterfly optimization, was employed to extract deep weighted features and optimize the weights. Subsequently, the weighted features were allocated to the optimized hybrid fault detection conducted by the deep neural network and deep belief network [4]. The efficacy of the proposed model was evaluated based on a set of actual measurement values in the context of Industry 4.0. In response to the growing demand from consumers for products that are environmentally responsible, there has been a notable increase in the focus on sustainable product design in recent years. However, because of the lack and ambiguity of knowledge among clients and experts, including uncertainty in understanding, the design is not feasible. To address this gap, Zhou proposed a novel multi-objective optimization (MOO) ranking technique that integrates the technique for order preference by similarity to ideal solution (TOPSIS) method. Additionally, Zhou designed a fuzzy Markov Taguchi system method to mitigate cognitive uncertainty associated with customer preferences regarding optimization objectives. The implementation of the novel TOPSIS method yielded the optimal design solution, as evidenced by experimental results. This method provided an exemplar of sustainable substrate design and enabled sensitivity analysis scenarios and comparative studies to be conducted, thereby demonstrating the efficacy of the proposed method [5]. In the context of the contemporary industrial revolution, additive manufacturing represents a promising technological advancement with the potential to enhance the efficacy, flexibility, and competitiveness of the supply chain. Brito *et al.* proposed a new design method for additive manufacturing–supply chains using optimization techniques. The research results indicated that this method was evaluated in real use cases of elevator maintenance service providers, and the proposed design method had good capabilities in addressing the challenges brought by the widespread use of 3D printers in SC manufacturing [6]. High-resolution laser additive manufacturing markedly expands design autonomy, propels the evolution of topology optimization (TO), and propounds advanced structural design methodologies. To fully capitalize on the benefits of voxel-based forming techniques, Li *et al.* proposed a novel artificial intelligence-assisted TO approach for laser-based parallel optimization design, which aims to encompass the relationship between process performance and design variables. The method employed heuristic and gradient-based algorithms to concurrently design the micro-material properties and macrostructural topology of 3D components. In comparison to classical optimization techniques, numerical simulations demonstrated that this approach could markedly enhance the macroscopic structural mechanical performance. This collaborative design methodology could be extensively utilized in the design and optimization of intricate functional components, as well as case studies of artificial intelligence-assisted product evaluation [7].

The product service systems (PSSs) have precipitated a rapid evolution in the manner by which value is created within the manufacturing industry. The existing PSS configuration solutions afford customers the opportunity to select preferred product and service modules, characterized by a high degree of granularity. Zhang *et al.* addressed the product-oriented PSS configuration optimization problem from a fine-grained perspective. A multi-layer network comprising the product layer, service layer, and resource layer was constructed to represent the elements and relationships inherent to the PSS. It should jointly consider service activity selection and resource allocation, construct a mathematical model for optimizing PSS configuration, and make the calculation of optimization objectives under constraint conditions closer to the actual execution situation. The findings of the research indicated that the significance of service activities was perceived to enhance the performance of service activities with greater importance. The corresponding algorithm was enhanced and implemented to determine the optimal solution, and case studies in the automotive industry demonstrated the diverse advantages of the proposed method [8]. In the context of the Internet and big data, a novel production and operational organizational model for the intelligent automobile industry will facilitate its transformation in several key areas, including product planning, design, production, marketing, operation, and maintenance. Moreover, this model has the potential to facilitate service innovation oriented to the full life cycle of products. Yang *et al.* employed multi-attribute decision-making in the context of automotive manufacturing and service industries, utilizing a total of 20 attributes, comprising 14 attributes from the manufacturing industry and 6 attributes from the automotive service industry, as inputs. A BP neural network was subsequently trained. Finally, the effectiveness of the proposed methodology was validated through a set of numerical examples, demonstrating an average accuracy of 93.19% [9]. The design, machining, and assembly processes employed in mold fabrication are inherently knowledge-intensive, costly, and time-consuming operations that demand a significant investment of human resources. Lee and Ryu employed a methodology whereby product features were extracted from product images, and the degree of similarity between the features of new products and those of products previously manufactured was evaluated. The design features were extracted from the design data, and a similarity retrieval model framework was put forward based on the outcomes of both product and design feature extraction. A model was developed that encompassed collaborative processes, products, processes, and resources. The results demonstrated that this process of collaborative modeling facilitated the visualization and analysis of complex processes intrinsic to the creation of innovative products [10]. The current circular economy (CE) design framework and evaluation have failed to address two key areas where computational optimization can add value: determining the best-performing solution among many options and managing multiple potential conflicting or redundant CE goals. Ortner *et al.* proposed and tested a computational MOO method for CE product design, which improved the results of manual design processes and reduced waste and environmental impact throughout the product lifecycle. The results indicated that MOO could bring improvements to CE product design and provide a basis for future work discussions [11].

Ma *et al.* proposed four new optimization methods for multi-layer perceptron neural networks to predict this key parameter. The dataset used consists of 103 rows of information, with 7 influential parameters. In this work, the optimal fitting complexity for each set is determined through population-based sensitivity analysis. The grasshopper optimization algorithm stands out as the second most efficient optimizer with minimal complexity (overall size = 50) [12]. Gao *et al.* proposed a new method for evaluating the bond strength of fiber-reinforced polymers using an artificial intelligence-based model. The comparison of results shows that the new model can provide better performance. In conclusion, the proposed hybrid model can serve as a suitable alternative to empirical models in this research field [13]. Many advanced machine learning and deep learning techniques developed for big data have inadvertently provided solutions to small data problems. Dou *et al.* summarized and analyzed several emerging potential solutions for small data challenges in molecular science (including chemistry and biology), reviewed basic machine learning algorithms, as well as more advanced technologies, and discussed promising trends for small data challenges in molecular science [14]. Tajziehchi *et al.* studied the use of genetic algorithms for earthquake control and optimization and proposed a new method that uses binary genetic algorithms and natural numbers to select the optimal acceleration map and scale it for dynamic time history analysis, thereby obtaining an average response spectrum that matches the target spectrum appropriately and has a short distance, and indicating the expected earthquake of the site. The results show that this research can provide users with a series of optimal coefficients, crossover values, and mutations for each chromosome [15].

In summary, researchers have been involved and studied in data recognition, service cycle analysis, product design feature extraction, and algorithm classification for the optimization algorithms of mechanical products and servers. However, the application of service selection and design optimization for mechanical products is still insufficient, and there is no better segmentation and detailed analysis of product services to comprehensively analyze the service process. Therefore, this research proposes an intelligent service discovery algorithm and optimization architecture method for mechanical products based on convolutional neural networks (CNNs). Taking the optimization server as an example, this article innovatively studies the optimal service selection algorithm in cloud manufacturing mode based on directed graph and Dijkstra's algorithm, matches and optimizes the mechanical product features and demand side requirements, and provides a technical basis for enterprise mechanical product design and optimization.

2 Methods and materials

To implement the intelligent service discovery model in the intelligent service discovery scheme, research was conducted on the intelligent service discovery algorithm based on CNN. In recent years, CNNs have achieved remarkable results in image intelligence recognition in many industries (such as facial recognition, unmanned driving environment perception, *etc.*), but there is currently no research on their application in the field of mechanical products. To improve the accuracy of product image recognition, the advantages and disadvantages of existing neural network architectures and activation functions are studied, and a neural network architecture and activation function for mechanical product data are proposed. First, the intelligent service and optimization of mechanical products were studied and analyzed. Then, product service decomposition was carried out, and a CNN-based intelligent service discovery algorithm for mechanical products was proposed. Second, research was conducted on the optimal service selection algorithm in cloud manufacturing mode based on directed graphs and Dijkstra's algorithm.

2.1 Intelligent service discovery algorithm and optimal architecture for mechanical products based on CNN

In recent years, a new mechanical product service model has utilized cloud computing and artificial intelligence [16]. It achieves intelligent matching, recommendation, and optimization of mechanical product-related services through intelligent algorithms and data analysis to meet the diverse and personalized needs of users. Research is conducted on intelligent service discovery algorithms and optimization architectures, combining various services, as shown in Figure 1.

For the study of server optimization indicators, manufacturing capability and overall capability are selected. The former is for service quality evaluation, while the latter is for the overall evaluation of service time, cost, service praise, service needs, and other aspects. The evaluation indicators for overall capability are shown in the following equation:

$$\begin{cases} S = \dfrac{w_1}{T} + \dfrac{w_2}{C} + w_3 Q + w_4(N \times R) \\ \text{S.T.} \begin{cases} 0 \le w_i \le 1 \\ \sum\limits_{i=1}^{4} w_i = 1. \end{cases} \end{cases} \tag{1}$$

In Eq. (1), w_i represents the weight of each parameter. CNN has strong feature learning capabilities and can extract fine features from images, which is crucial for improving the accuracy of product image recognition and helping to more accurately identify user needs and product features in intelligent service discovery solutions. Traditional image recognition methods such as edge detection and threshold segmentation have complex pre-processing steps and are difficult to obtain the closest natural expression of image features. In contrast, CNN can automatically learn image features, avoiding complex preprocessing steps and improving recognition accuracy and efficiency. The study uses CNNs to analyze intelligent service discovery and designs a neural network architecture for mechanical product design based on CNN. CNN originates from fully connected neural networks, in which neurons (or nodes) are organized into multiple layers, and each layer's neurons are connected to the neurons of the previous and next layers, but there are no connections between neurons within the same layer [17]. In a fully connected neural network, each connection is associated with a weight, which is learned through the application of training data. The weighted sum of all inputs is calculated by each neuron, which then obtains its output through an activation function, such as the rectified linear unit (ReLU), the sigmoid function, or the tanh function [18]. The description of the n-th layer of a fully connected neural network is shown in the following equation:

$$e_{n,j} = \sum_j (W_{n,i,j} e_{(n-1),j} + b_{(n-1),i}). \tag{2}$$

In Eq. (2), $b_{(n-1),i}$ represents the bias voltage of the i-th neurons in the $n-1$-th layer network, $e_{n,j}$ represents the neurons in each layer, n represents the n-th layer neural network, $n = (0,\ 1,\ 2,\ \ldots,\ n)$, j represents the j-th neurons, f represents the activation function, $w_{n,i,j}$ represents the weight of neurons from the $n-1$-th layer neural network to the nth layer network, and $e_{0,j}$ represents the n-th of the input layer, denoted as 0.

The performance of a neural network is determined by its architecture and activation function [19]. Therefore, a new type of attention network, the novel CNN architecture, is proposed to improve the accuracy of attention networks [20]. There are limitations in the recognition accuracy of existing neural network architectures and the generated attention networks.

The attention map is a method used to represent the different weights assigned by deep learning models to information at different positions or time steps when processing input information [21]. First, the generation principle of attention networks is studied. Some attention networks use the key points of recognized objects as the centers of behavioral units to generate corresponding attention maps, with the center having the highest weight. The generated expression is shown in the following equation:

$$w = 1 - 0.095 d_m. \tag{3}$$

In Eq. (3), d_m means the distance from the pixel point to the center of the behavioral unit, and the highest weight is set to 1.0, which is inversely proportional to the distance.

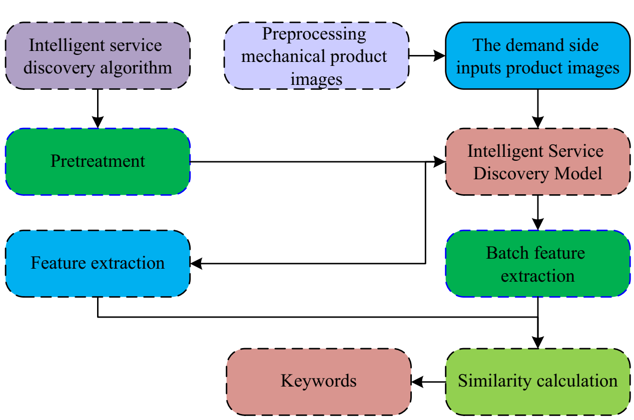

Figure 1: Intelligent service discovery solution.

However, the product image features generated by this method are different, which is not conducive to the subsequent construction of neural network architecture. Therefore, research has proposed a more compatible method that can generate enhancement layers for neural network architecture. In addition, an improved local binary pattern is used for image feature extraction, which can represent image features by setting numerical values as the attention map for generating images. The generation steps are shown in Figure 2.

In Figure 2, the original image is first input, then it is represented in grayscale, resized, and an attention map is generated by improving the local binary patterns (LBP) weight generation method. Finally, it is fused into a neural network to generate an enhancement layer. The study designs a neural network architecture based on the results obtained from the above steps, defining the output and number of convolutional layers, pooling layer output, and fully connected layer output. The proposed neural network input representation is shown in the following equation:

$$\begin{cases} \text{Input\_}A = \text{Conv3\_3} + P\text{\_Net} \\ \text{Pool3} = \text{MaxPool}(\text{Input\_}A). \end{cases} \tag{4}$$

In Eq. (4), P_Net represents the enhancement layer of the improved LBP, Conv3_3 represents the third group in the convolutional layer and the output of the third layer, MaxPool represents the max pooling layer, and Pool3 represents the pooling layer of the third layer. The activation function also affects the performance of neural networks. To alleviate the phenomenon of neuron vanishing in the ReLU function, a new activation function has been proposed, as denoted in the following equation:

$$f(x) = \begin{cases} x, x \geq 0 \\ \dfrac{1 - e^{-2x}}{1 + e^{-2x}}, \quad x < 0. \end{cases} \tag{5}$$

This function has two parts, namely the negative half-axis of the tanh function and the positive half-axis of the ReLU function. Based on the above, a CNN architecture for mechanical product images is designed.

Mechanical product datasets and intelligent service discovery models were collected and trained, with acquisition strategies mainly including image occlusion, mechanical product images from different angles, and images with different light intensities [22,23]. Prior to this, the design process of mechanical products has been decomposed, as shown in Figure 3.

Then, data preprocessing is carried out to improve the generalization ability of the neural network and reduce overfitting. Research is conducted on data augmentation of the dataset, including red green blue (RBG) enhancement, noise processing, random pruning, rotation, *etc.* After preprocessing, the quality of the dataset for mechanical products is evaluated, mainly through experimental training and research on the constructed dataset. By comparing the test results, the quality evaluation situation is obtained. Finally, the high-precision dataset is applied to the intelligent service discovery model and neural network model. The intelligent service discovery and optimization framework constructed by the research is shown in Figure 4.

In Figure 4, first, the customer requests mechanical products and services. Then, an intelligent service discovery model is used to search and query databases, followed by service composition. In the server optimization module, a mathematical model is constructed, and a server

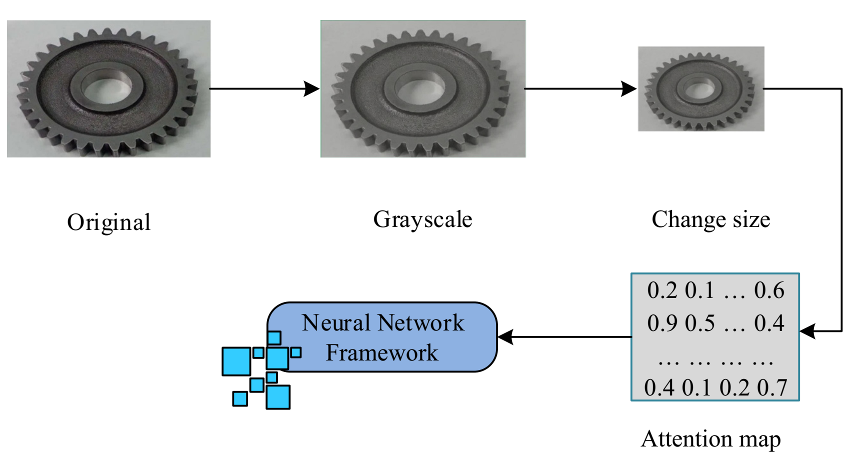

Figure 2: Steps for generating image attention maps.

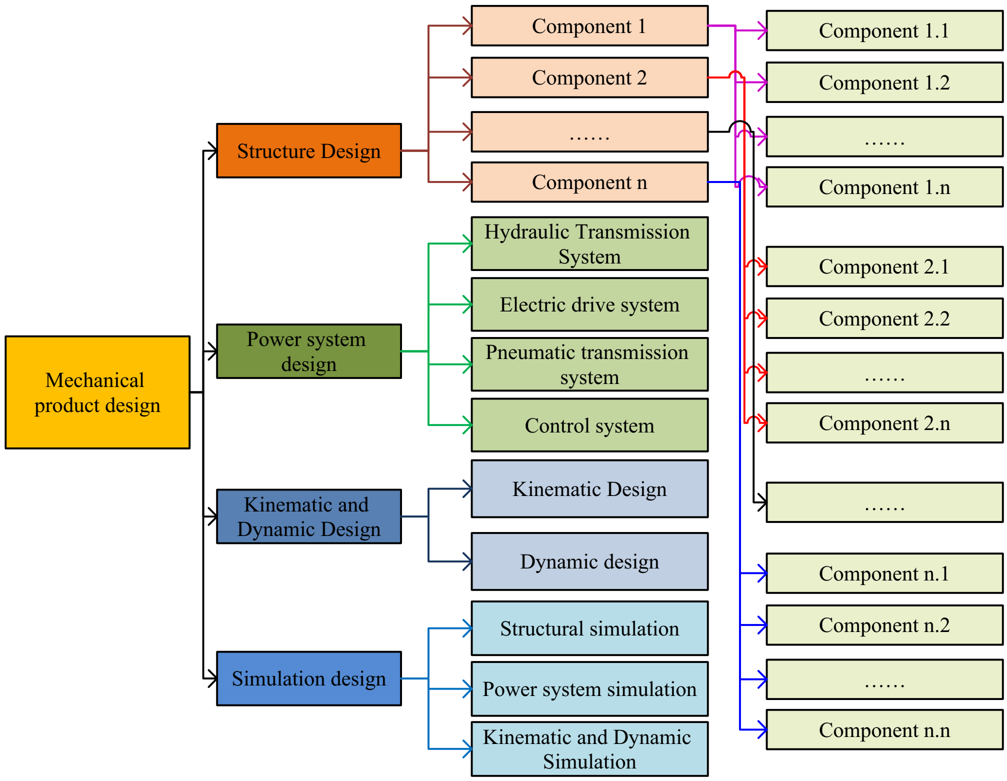

Figure 3: Decomposition of the mechanical product design process.

optimization algorithm is introduced. Finally, feedback is given to the demand side to determine whether they are satisfied. If satisfied, the process ends. If not, the optimization indicators are changed again, and the server optimization algorithm is used for selection. The feedback is then given to the demand side until they are satisfied.

Compared with other methods, CNN typically requires input data to have a fixed size and resolution. In practical

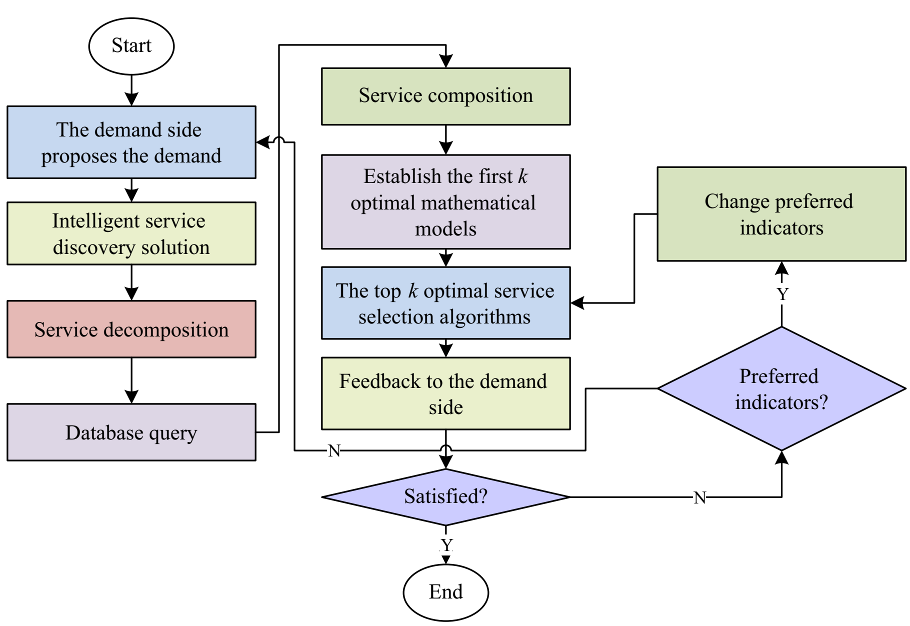

Figure 4: A smart system for finding and improving services.

applications, the size and resolution of mechanical product images may vary, requiring pre-processing of the input data. Model training may require a large amount of data and high-performance computers. However, CNN has strong learning and generalization abilities in research methods and can extract fine features such as edges, textures, shapes, *etc.*, thereby improving recognition accuracy. Combined with optimal service selection algorithms based on directed graphs and Dijkstra's algorithm, it forms a more complete intelligent service discovery solution, providing users with higher quality and more efficient intelligent services.

2.2 Optimal server optimization algorithm in cloud manufacturing mode

In the cloud intelligent service discovery and optimization architecture, mathematical models and server optimization algorithms are applied. This study introduces and designs the model and algorithm. First, mathematical modeling is conducted on k optimal service combination schemes, and optimization indicators are set. The optimal schemes select production services, design services, and product services, and an objective function is constructed [24]. The overall capability of the preferred solution is represented by

$$\begin{cases} S^D = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1}(\alpha_{ij}S_{ij}^D) \\ S^M = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1}(\alpha_{ij}S_{ij}^M) \\ S^{\text{PS}} = \dfrac{1}{2}\left(\dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1}(\alpha_{ij}S_{ij}^D) + \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1}(\alpha_{ij}S_{ij}^M)\right) \\ \alpha_i = \begin{Bmatrix} 1, & \text{if the } i\text{-th servicer is selected} \\ 0, & \text{else} \end{Bmatrix}. \end{cases} \tag{6}$$

In Eq. (6), S^D represents the overall average capability of each sub-service of the design service, S^M represents the overall average capability of each sub-service of the production service, S^{PS} represents the overall average capability of each sub-service of the product service, i represents the i-th sub-task, $i = \{0, 1, 2, \dots, (n-1)\}$, j represents the j-th service provider, $j = \{1, 2, \dots, m\}$.

The corresponding manufacturing and overall capabilities of the service providers selected for the manufacturing capability and comprehensive capability in the objective function of the product portfolio scheme are shown in the following equation:

$$\begin{cases} Q^P = \alpha_i Q_j^P \\ S^P = \dfrac{w_1}{T_j^P} + \dfrac{w_2}{C_j^P} + w_3 Q_j^P + w_4(N_j^P \times R_j^P). \end{cases} \tag{7}$$

In Eq. (7), Q^P represents manufacturing capability and S^P represents overall capability. A mathematical model is constructed as shown in the following equation:

$$\begin{cases} T = \{T^D, T^M, T^{\text{PS}}, T^P\},\ \min|T_N^k|,\ \max|T_N^k| \\ C = \{C^D, C^M, C^{\text{PS}}, C^P\},\ \min|C_N^k|,\ \max|C_N^k| \\ Q = \{Q^D, Q^M, Q^{\text{PS}}, Q^P\},\ \min|Q_N^k|,\ \max|Q_N^k| \\ S = \{S^D, S^M, S^{\text{PS}}, S^P\},\ \min|S_N^k|,\ \max|S_N^k|. \end{cases} \tag{8}$$

In Eq. (8), T, C, Q, and S, respectively, represent the time, cost, manufacturing capability, and overall capability of the design service combination scheme. $\min|T_N^k|$, $\min|C_N^k|$, $\min|Q_N^k|$, and $\min|S_N^k|$ represent the mathematical models of the first k minimum time, cost, manufacturing capability, and overall capability. $\max|T_N^k|$, $\max|C_N^k|$, $\max|Q_N^k|$, and $\max|S_N^k|$ represent the mathematical models of the first k maximum time, cost, manufacturing capability, and overall capability. The analysis of mathematical models and the study of optimal selection of maximum or minimum indicator service combination schemes are represented by directed graphs, as shown in Figure 5.

In Figure 5, R represents the task with 0 resources, M_{ij} represents the j-th service provider of the i-th task, and the resource cost of subtasks, and when there is no resource cost between subtasks, the resource cost calculation is shown in the following equation:

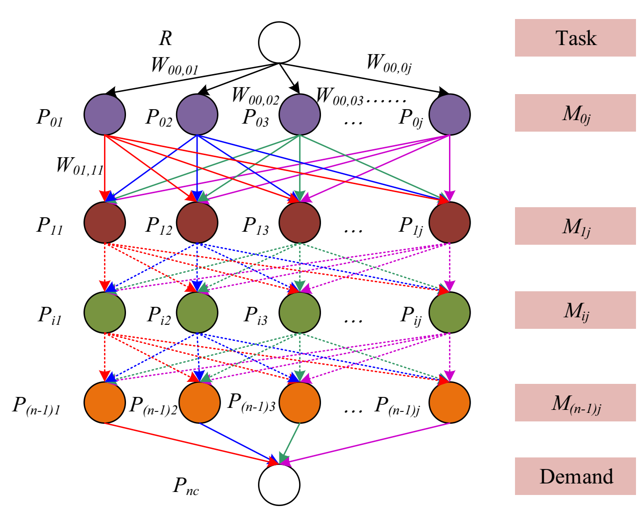

Figure 5: Directed graph for optimal selection of service combination schemes with maximum or minimum indicators.

$$\begin{cases} P_{ij} = \{T_{ij}^{D}, C_{ij}^{D}, Q_{ij}^{D}, S_{ij}^{D}, T_{ij}^{M}, C_{ij}^{M}, Q_{ij}^{M}, S_{ij}^{M}, T_{j}^{P}, C_{j}^{P}, Q_{j}^{P}, S_{j}^{P}\} \\ W_{ij,(i+1),j} = \{T_{ij,(i+1),j}^{M}, C_{ij,(i+1),j}^{M}, C_{jn}^{P}, 0\}. \end{cases} \tag{9}$$

In Eq. (9), R represents the demand side resource cost, and P_{ij} represents the i-th task j-th service side resource cost.

$$\begin{cases} W = \dfrac{1}{K}\left(\sum_{i=0}^{n-1}(\alpha_{ij}W_{ij,(n+1)j} + \beta_{ij}P_{ij}) + \gamma_{(n-1)j}W_{(n-1)j,n1}\right) \\ \beta_i = \begin{cases} 1, & \text{if the } i\text{-th servicer is selected} \\ 0, & \text{else} \end{cases}. \end{cases} \tag{10}$$

In Eq. (10), $W_{ij,(i+1)j}$ represents the resource cost in the combination process of arrow paths, and K represents the number of objective functions. The research is based on directed graphs; therefore, the research is transformed into finding the longest and shortest paths in the directed graph, which have resource costs between tasks in the service composition scheme, to solve the mathematical model [25]. For the first $n-1$ stage subtasks, the resource cost of releasing nodes is shown in the following equation:

$$\begin{cases} W'_{ij,(i+1)j} = W_{ij,(i+1)j} + P_{ij} \\ \text{S.T.}\begin{cases} i = \{1, \ 2, \ \ldots, \ n-1\} \\ j = \{1, \ 2, \ \ldots, \ m\}. \end{cases} \end{cases} \tag{11}$$

The n-th subtask is shown in the following equation:

$$W'_{(n-1)j,n1} = W_{(n-1)j,n1} + P_{n1} = W_{(n-1)j,n1}. \tag{12}$$

Through the operations of Eqs. (11) and (12), all node resource costs are reset to zero, and all paths have resource costs, as shown in Eq. (13).

$$\begin{aligned} W &= \sum_{i=0}^{n-1}(\alpha_{ij}W'_{ij,(n+1)j}) + \gamma_{(n-1)j,n1}W'_{(n-1)j,n1} \\ &= \sum_{i=0}^{n}(\alpha_{ij}W'_{ij,(n+1)j}). \end{aligned} \tag{13}$$

In order to obtain the shortest path, an improved algorithm for the first k shortest paths is proposed, which is based on Dijkstra's algorithm and includes sub-shortest paths and path extensions. The resource cost of the second shortest path is greater than that of the shortest path, so when the path cost is 0 in other stages, as shown in the following equation:

$$\min(S_i - S_0) = \min(x_{ij} - X_{ij}) + 0. \tag{14}$$

The cost of secondary short-path resources is shown in the following equation:

$$S_1 = S_0 + \min(x_{ij} - X_{ij}). \tag{15}$$

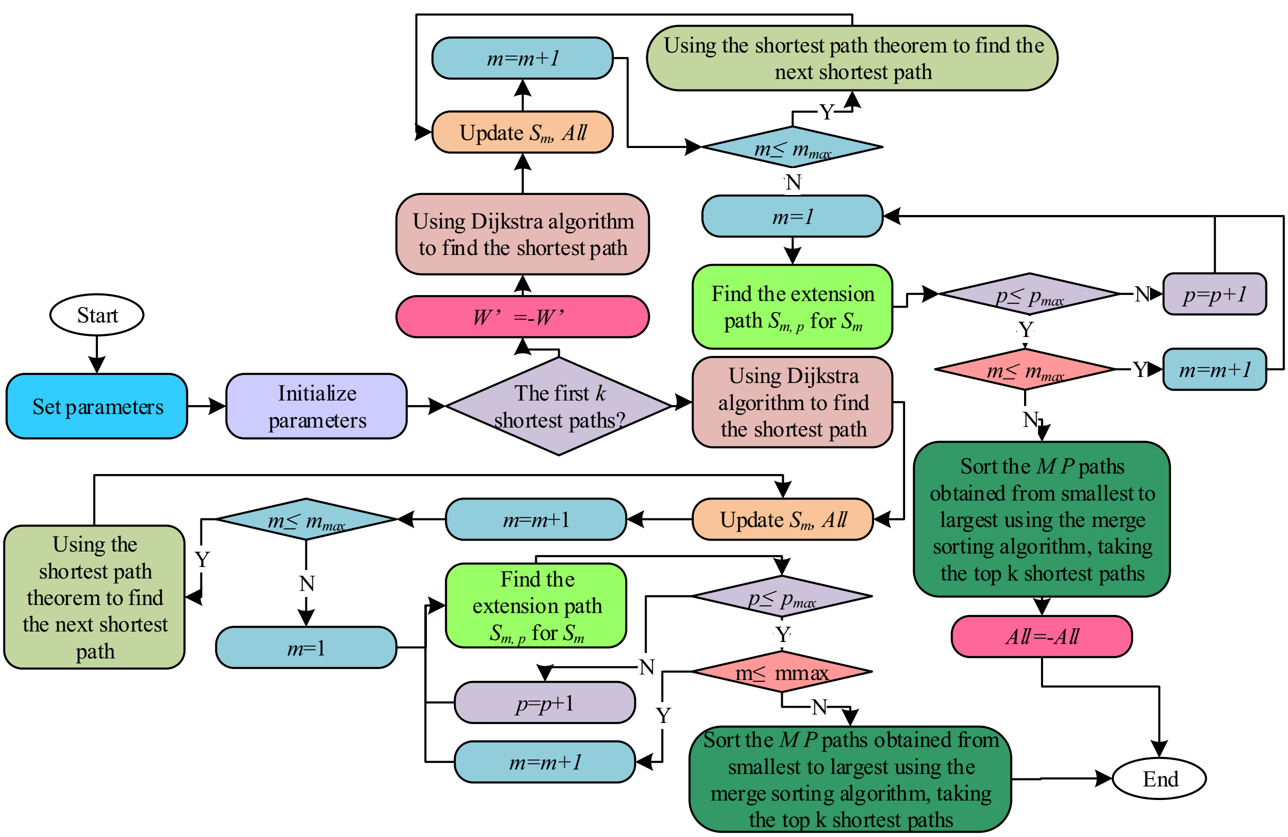

Figure 6: The shortest or longest path algorithm process.

In Eqs. (14) and (15), X_{ij} represents the set of shortest path resource costs, X_{ij} represents the resource cost of the j path in the i stage, x represents the set of second shortest path resource costs, x_{ij} represents the resource cost of the j path in the i stage, S_0 represents the shortest path, and S_1 represents the second shortest path. Based on the above, the algorithm flow for the shortest or longest path is shown in Figure 6.

In Figure 6, all represents the sum of path resource costs. $m_{\max}$ represents the total number of shortest paths, $p_{\max}$ represents the number of extended paths, W' represents the total stage path resource costs, and $S_{m,p}$ represents the extended paths obtained through path extension methods.

3 Results

To verify the proposed mechanical product design method based on CNN and server optimization algorithm, an experiment was conducted to validate it, analyze the corresponding design parameters and experimental data results, verify the advantages and feasibility of the method, and provide reference for product demanders to optimize products.

3.1 Intelligent service discovery and optimization system verification settings

The experiment focused on the mechanical product services of a certain enterprise under cloud manufacturing mode. Table 1 shows the parameter settings of the neural network architecture before and after adopting LBP improvement. The experimental platform had 8 GB of memory, the system was OSXE | Capitan system, and the experimental software was Python 2.7, with a 2.9 GHz Intel i5 processor. In order to verify the performance of the neural network architecture, experiments were conducted on the improved and unmodified networks based on the MNIST dataset. The MNIST dataset is a fundamental dataset for deep learning, sourced from the National Institute of Standards and Technology in the United States. MNIST has a total of 60,000 training sample sets and 10,000 validation sample sets, and its images are handwritten digits with 28 × 28 pixels stored in byte form. Before conducting the experiment, it is necessary to first set the parameters of the neural network. To verify the performance of the research network in extracting image features, design parameters for the neural network. Choose ReLU as the activation function, stochastic gradient descent as the optimization algorithm, and set the maximum training steps for the neural network framework to 2,000.

3.2 Analysis of intelligent service discovery and optimization algorithm results

Figure 7 shows the training accuracy and loss curves of models trained on different networks, compared to traditional CNNs. In Figure 7(a), the correct rate of model training for the research network performs well. The correct rate increases rapidly in the early stage of training and stabilizes at around 0.9 at step sizes greater than 200. This indicates that the research network is able to learn the features of the data more effectively during the training process, thus achieving accurate classification or prediction of the input data. In contrast, the model training

Table 1: Parameter settings for neural network architecture

Input	Number of parameters	Feature map size	Convolutional kernel	Step
Image	30,726	28 × 28	—	—
Convolutional layer 1	65,548	28 × 28	[2 × 2, 96], 2	1
Pooling layer 1	16,374	14 × 14	[2 × 2], 1	2
Convolutional layer 2	32,756	14 × 14	[2 × 2, 128], 2	1
Pooling layer 2	8,188	7 × 7	[2 × 2], 1	2
Convolutional layer 3	16,376	7 × 7	[2 × 2, 256], 3	1
Pooling layer 3	4,084	4 × 4	[2 × 2], 1	2
Convolutional layer 4	8,166	4 × 4	[2 × 2, 384], 3	1
Pooling layer 4	2,048	2 × 2	[2 × 2], 1	2
Convolutional layer 5	2,048	2 × 2	[2 × 2, 512], 3	1
Pooling layer 5	512	1 × 1	[2 × 2], 1	2
Fc 1	4,096	1 × 1	Average pool	—
Fc 2	4,096	11	Average pool	—
Fc 3	10	—	—	—

correctness of traditional CNNs improves slowly and fluctuates greatly during the training process. It is only when the step size reaches about 800 that the correct rate stabilizes at about 0.83, which is lower than the correct rate of the research network. This is due to the limitations of the traditional CNN in feature extraction or model structure, resulting in its poor performance in handling complex data. In Figure 7(b), the model training loss of the research network decreases rapidly at the beginning of training and approaches 0 at a step size of around 20. This indicates that the research network is able to converge to the optimal solution quickly during training, thus achieving accurate fitting to the input data. The model training loss of the traditional CNN decreases slowly and fluctuates greatly during training. The loss approaches 0 only after the step size reaches 800, but is always greater than the model training loss of the study network. This further confirms that traditional CNNs are deficient in training efficiency and model fitting ability. It can be seen that models studying network training have more advantages in accuracy and loss.

Figure 8 showcases the comparison of different activation function curves, as well as the training accuracy and loss of different networks. The “ThLU” function, which combines tanh function and ReLU function, was studied as the activation function. Other compared activation functions include commonly used ReLU function, Leaky rectified linear unit (LReLU), and exponential linear unit (ELU). Figure 8(a) shows the variation curves of each activation function, which showed significant changes at the node with input 0. ThLU and ELU had similar changes, but ThLU had a smaller output, and when the input was greater than 0, the output was close to other functions. The changes in ReLU and LReLU were similar, ranging from −0.2 to −0.1 when the input was less than 0. When the input was greater than 0, the output changed proportionally. In Figure 8(b), the training accuracy and loss of the neural network showed that the training accuracy and testing accuracy of the studied network were 98.89 and 96.88%, respectively, with training loss and testing loss of 0.04619 and 0.08921, respectively. In contrast, the training accuracy and testing accuracy of the compared network were 93.95 and 85.31%, respectively, with training loss and testing loss of 0.3556 and 0.4872. The running time of the two neural networks was 105 and 73 min, respectively. Overall, the accuracy of the studied network was high, with relatively small losses.

Figure 9 shows the training accuracy and loss curves of each activation function on the dataset. In Figure 9(a), the training accuracy of the proposed ThLU function was consistently higher than that of other functions. This was mainly reflected in the rapid improvement of the training accuracy of the ThLU function, which was within the range of 80–100%, while other functions had a slower upward trend, slowly increasing within the range of 40–100%. In Figure 9(b), the training loss of the proposed ThLU function was always smaller than other functions, rapidly decreasing in the range of 0–1.2 and around 0.1, while the training loss of other functions decreased more slowly in the range of 0–1.8. The training loss of other activation functions decreases slowly in the range of 0–1.8 and fluctuates greatly during the training process. The loss values of these functions are significantly higher than those of the ThLU function, indicating their shortcomings in model optimization and feature extraction.

Figure 10 shows the training accuracy, training loss, validation accuracy, and validation loss curves of a mechanical product dataset using a neural network model. The training accuracy and loss fluctuated around 0.97 and 0.068, respectively. The training accuracy, loss curve, validation accuracy, and loss curve all tended to stabilize at a

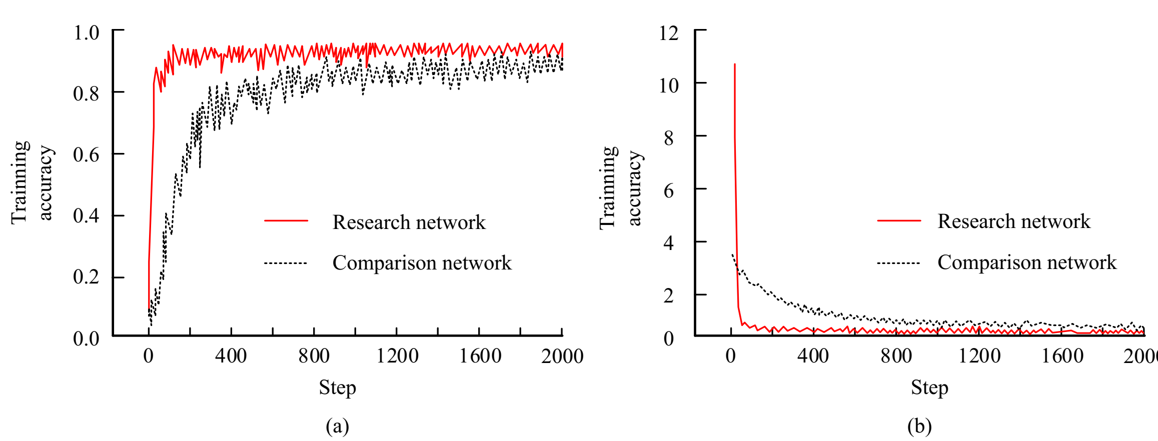

Figure 7: The accuracy and loss curve of model training based on different networks. (a) Accuracy curves of model training based on different networks. (b) Training loss curves of models trained on different networks.

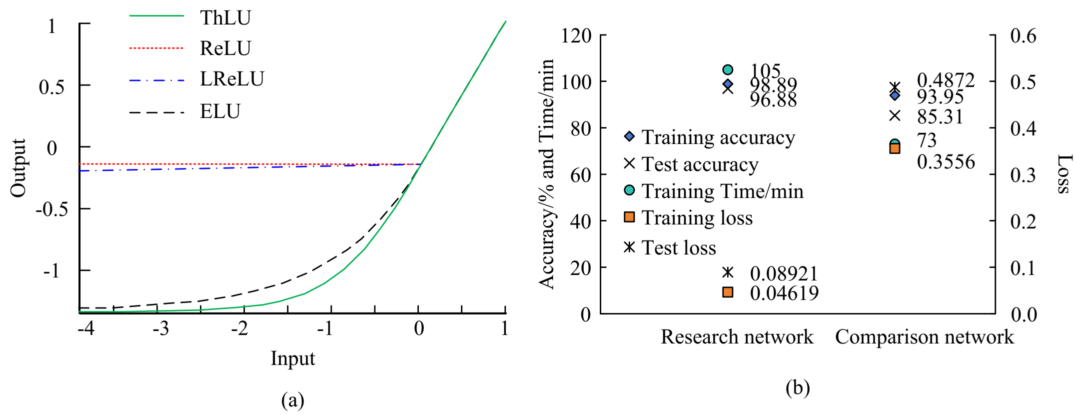

Figure 8: Comparison of different activation function curves and training accuracy and loss of different networks. (a) Different activation function curves. (b) Comparison of training accuracy and loss for different networks.

step size of around 50, while the testing accuracy and loss were around 0.99 and 0.1, respectively. The overall accuracy was close to 1, and the loss was close to 0. It can be seen that the research model effectively handled data training and testing.

To verify the accuracy and efficiency of the shortest path for mechanical product services under cloud manufacturing mode, Figure 11 shows the length and running time of the top 10 paths for different algorithms. The comparative algorithms used in this experiment were deletion algorithm (DA), traditional Dijkstra algorithm, and Yang M algorithm based on multi-attribute decision-making and BP neural network for server optimization in existing research. In Figure 11(a), the path length increased with the increase of the path, from 312 on the first shortest path to 340 on the tenth path, where the path length of the research algorithm was similar to most paths. In Figure 11(b), the running time of the research algorithm was the smallest, at 0.016 s, while the running time of DA was the longest, at 1.23 s. Overall, the research proposed that the algorithm had better performance.

4 Discussion and conclusion

In the cloud manufacturing mode, cloud computing, through its efficient processing capabilities, can process a large amount of data generated during the production and design process of mechanical products in real time, extract valuable information, and provide support for production and design decisions. To effectively feedback the mechanical product requirements of the demand side, a mechanical product design and manufacturing system based on CNN and server optimization was proposed. It combined the service discovery algorithm and optimization

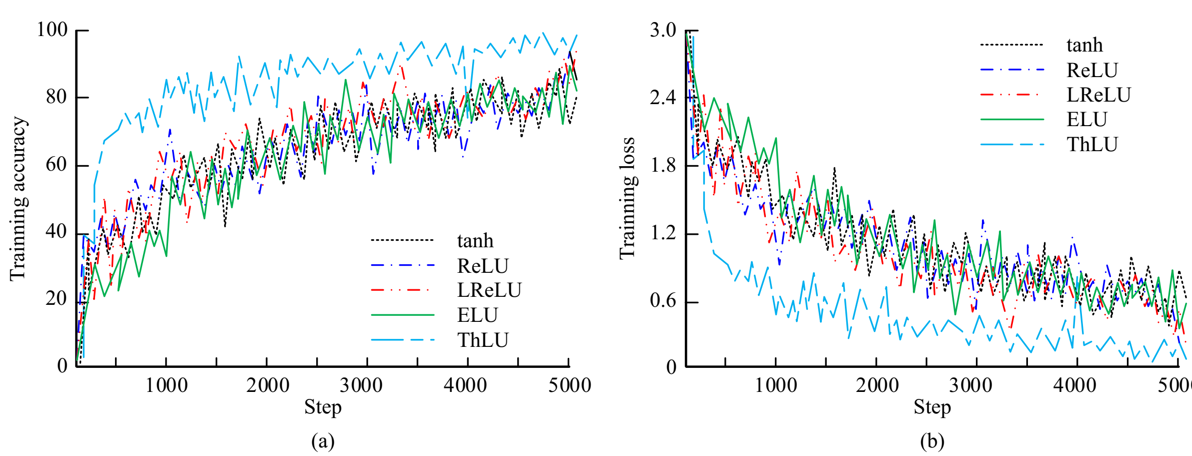

Figure 9: The training accuracy and loss rate curves of each activation function on the dataset. (a) Training accuracy curves of each activation function on the dataset. (b) Training loss curves for each activation function on the dataset.

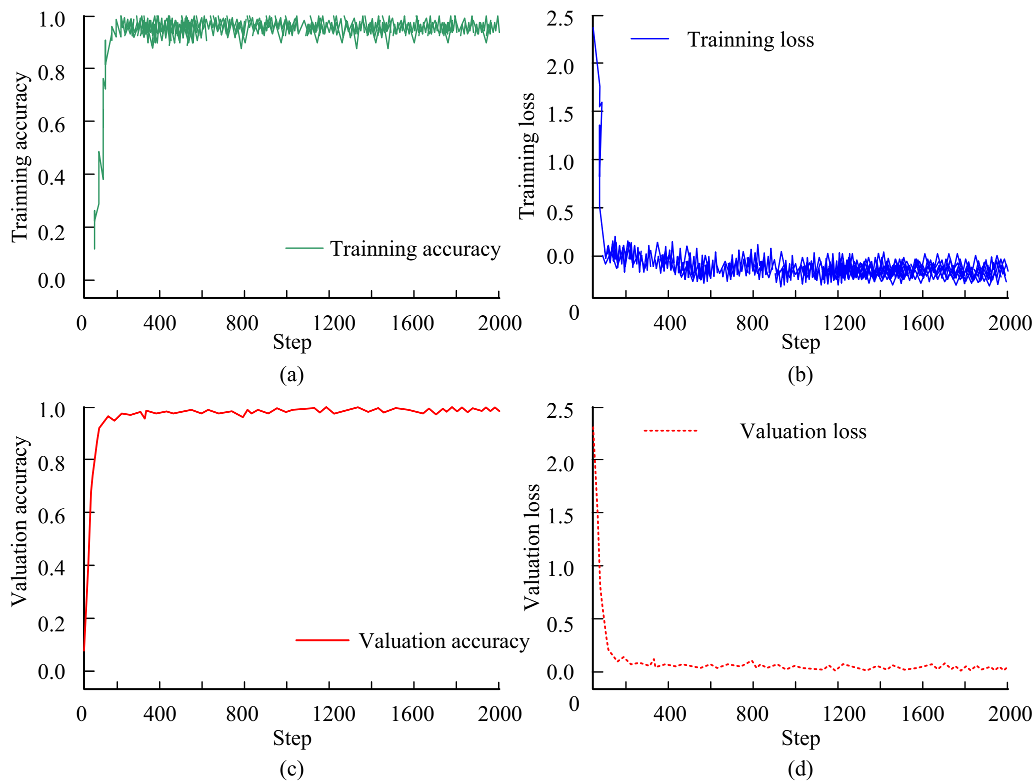

Figure 10: Training and validation curves for neural network models for mechanical product datasets. (a) Training accuracy curve. (b) Training loss curve. (c) Verification accuracy curve. (d) Verification loss curve.

architecture, and adopted directed graph and Dijkstra algorithm for optimal path selection in cloud manufacturing. The research results indicated that the research network had a higher model training accuracy, maintaining around 0.9 when the step size was greater than 200, compared to the network model training accuracy, which remained around 0.83 when the step size was greater than 800. The model training loss of the research network approached 0 at a step size of around 20, while the model training loss of the comparative network approached 0 after a step size of 800, and the loss was always greater than the model training loss of the research network. It can be seen that

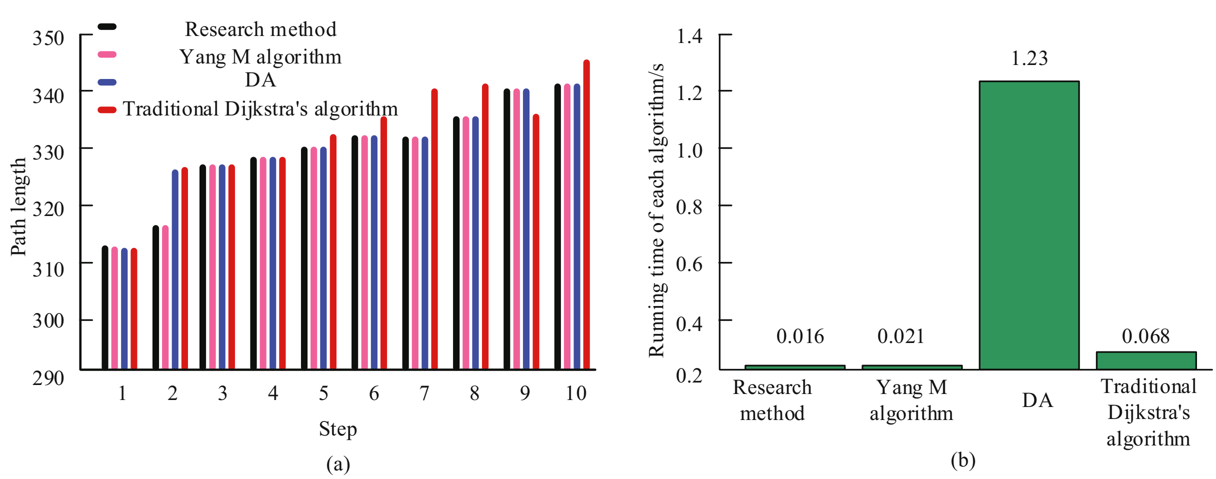

Figure 11: The top 10 path lengths and runtime for different algorithms. (a) The top 10 path lengths for different algorithms. (b) The top 10 shortest path running times for each algorithm.

models studying network training had more advantages in accuracy and loss. The path length of the research algorithm was similar to most paths, but the running time was the smallest, at 0.016 s, while DA had the longest running time, at 1.23 s. Overall, the research proposes that the algorithm has better performance, providing a reference for the optimization and manufacturing of mechanical product design services. However, further research is needed in the selection of optimal indicators and training of datasets, so further exploration of indicators and model training for mechanical products is needed in the future.

Funding information: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions: Lei Tian: study design, data collection, statistical analysis, visualization, writing, and revision of the original draft. Ance Zhao: statistical analysis, revised the manuscript, led and supervised this study. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: All data generated or analyzed during this study are included in this published article.

References

[1] Qi C, Yan L, Yang S, Yuan S. An optimization procedure for concave preform design in rectangular tube hydroforming. Int J Adv Manuf Technol. 2022;120(7–8):5511–24.
[2] Giarmas E, Tzetzis D. Optimization of die design for extrusion of 6xxx series aluminum alloys through finite element analysis: a critical review. Int J Adv Manuf Technol. 2022;119(9–10):5529–51.
[3] Hossain MS, Chakrabortty RK, Sawah SE, Ryan MJ. A multi-objective bi-level leader-follower joint optimization for concurrent design of product family and assembly system. Comput Ind Eng. 2023;177(4):1–18.
[4] Murugiah P, Muthuramalingam A, Anandamurugan S. A design of predictive manufacturing system in IoT-assisted Industry 4.0 using heuristic-derived deep learning. Int J Commun Syst. 2023;36(3):e5432.1–24.
[5] Zhou TL. Multi-objective optimization-based TOPSIS method for sustainable product design under epistemic uncertainty. Appl Soft Comput. 2021;98(1):1–12.
[6] Brito FMD, Junior GDC, Frazzon EM, Basto JPTV, Alcala SGS. Design approach for additive manufacturing in spare part supply chains. IEEE Trans Ind Inf. 2021;17(2):757–65.
[7] Li SY, Wei H, Yuan S, Zhu J, Zhang W. Collaborative optimization design of process parameter and structural topology for laser additive manufacturing. Chin J Aeronaut. 2023;36(1):456–67.
[8] Zhang Z, Xu D, Ostrosi E, Chen H. Optimization of the product-service system configuration based on a multilayer network. Sustainability. 2020;12(2):746–7.
[9] Yang M, Zhu H, Guo K. Research on manufacturing service combination optimization based on neural network and multi-attribute decision making. Neural Comput Appl. 2020;32(3):1691–700.
[10] Lee H, Ryu K. Product and design feature-based similar process retrieval and modeling for mold manufacturing. Int J Adv Manuf Technol. 2021;115(1):1–12.
[11] Ortner P, Tay JZ, Wortmann T. Computational optimization for circular economy product design. J Clean Prod. 2022;362(15):1–12.
[12] Ma X, Foong LK, Morasaei A, Ghabussi A, Lyu Z. Swarm-based hybridizations of neural network for predicting the concrete strength. Smart Struct Syst. 2020;26(2):241–51.
[13] Gao J, Koopialipoor M, Armaghani DJ, Ghabussi A, Baharom S, Morasaei A, et al. Evaluating the bond strength of FRP in concrete samples using machine learning methods. Smart Struct Syst. 2020;26(4):403–18.
[14] Dou B, Zhu Z, Merkurjev E, Ke L, Chen L, Jiang J, et al. Machine learning methods for small data challenges in molecular science. Chem Rev. 2023;123(13):8736–80.
[15] Tajziehchi K, Ghabussi A, Alizadeh H. Control and optimization against earthquake by using genetic algorithm. J Appl Eng Sci. 2018;8(21):73–8.
[16] Zhang X, Ming X, Bao Y, Liao Y, Miao R. Networking-enabled product service system (N-PSS) in collaborative manufacturing platform for mass personalization model. Comput Ind Eng. 2022;163(2):1–16.
[17] Marei M, Zaatari SE, Li W. Transfer learning enabled convolutional neural networks for estimating health state of cutting tools. Robot Comput Integr Manuf. 2021;71(4):1–11.
[18] Sheeba A, Rajalakshmi R, Maheswari BU, Ebenezer AS. Secure smart city application using webservice model and mayfly optimization-based lightweight CNN. Trans Emerg Telecommun Technol. 2024;35(1):e4896.1–22.
[19] Bose A, Garg R. Optimized CNN using manta-ray foraging optimization for brain tumour detection. Procedia Comput Sci. 2024;235(45):2187–95.
[20] Hong Z, Feng Y, Li Z, Hu B, Zhang Z, Tan J. Performance balance oriented product structure optimization involving heterogeneous uncertainties in intelligent manufacturing with an industrial network. Inf Sci. 2022;598(18):126.
[21] Bhosle K, Musande V. Evaluation of deep learning CNN model for recognition of Devanagari digit. Artif Intell Appl. 2023;1(2):114–18.
[22] Ghaderi A, Hassani H, Khodaygan S. A Bayesian-reliability based multi-objective optimization for tolerance design of mechanical assemblies. Reliab Eng Syst Saf. 2021;213(9):1–12.
[23] Mohammed KK, Hassanien AE, Afify HM. Classification of ear imagery database using Bayesian optimization based on CNN-LSTM architecture. J Digit Imaging. 2022;35(4):947–61.
[24] Jamshed A, Mallick B, Bharti RK. Grey wolf optimization (GWO) with the convolutional neural network (CNN)-based pattern recognition system. Imaging Sci J. 2022;70(4):238–52.
[25] Xu J, Song Z, Chen X, Yang Q. Design and optimization of high-density cryogenic supercritical hydrogen storage systems integrating with dual mixed refrigerant cycles. Energy. 2024;290(1):1–20.