(Received 19 July 2023, Accepted 17 November 2023, Scheduled 26 November 2023)

Antenna Notch Structure Optimization Using Deep Neural Networks

Wenjin Liu, Chen Yang*, Jingchang Nan, Mingming Gao, and Hongliang Niu

School of Electronics and Information Engineering, Liaoning Technical University, Huludao 125105, China

ABSTRACT: To address the stressful and time-consuming problem with the current notched antenna modelling optimization tools, an improved deep multilayer perceptron (DMLP) neural network framework is designed. The method introduces an attention mechanism (Attn) layer to improve the interpretability of the model, uses the leaky ReLU activation function to prevent the gradient from vanishing, and optimizes the structure of the DMLP model using an improved particle swarm algorithm (PSO) to improve the model prediction accuracy. Then, the notch structure geometric parameters of the designed double-notch ultra-wideband (UWB) antenna serve as input to predict the return loss S_{11} of the antenna. The experimental results show that the method reduces the root mean square error of prediction for S_{11} by 73.01% compared to the traditional MLP and 64.14% compared to the unimproved DMLP, which provides a solution for modelling notched UWB antennas and helps to optimize the design of this type of antenna.

1. INTRODUCTION

In addition to reducing electromagnetic interference from narrowband communication systems, ultra-wideband (UWB) antennas with notching properties can increase the antenna's bandwidth to a certain degree [1]. In the design of a notch UWB antenna, traditional antenna simulation tools, such as high-frequency structure simulation (HFSS), can accurately calculate various performances of the antenna. However, the HFSS frequently needs to repeatedly invoke the electromagnetic full-wave numerical simulation and analysis module during the optimization process, which increases the design time and difficulty. Therefore, more efficient and feasible methods are needed to optimize the antenna.

Deep neural network (DNN) is a highly nonlinear model with practical applications in optimizing microwave devices as a powerful data-driven tool with high optimization design efficiency [2, 3]. DNN uses sample data to train the neural network, and the trained network can immediately react to the mapping connection between the geometric parameters of the catch structure and the electromagnetic characteristics. The trained network can also quickly provide the matching anticipated output results when new data are input, thereby reducing the number of calls to the electromagnetic full-wave numerical simulation in antenna design and shortening the design time [4]. A novel multilayered neural network model is proposed in [5], which deepens the network depth, applies batch normalization techniques and leaky ReLU activation functions to reduce the network model training time, and applies to wide-range microwave parametric modelling. The results show that the deep network is more accurate when modelling microwave devices than the traditional shallow neural network. Reference [6] proposes using a particle swarm algorithm (PSO) to identify a deep trust network's model structure, followed by a deep model coupled with an extreme learning machine to get the best fractal antenna design. Reference [7] proposes using an adaptive chaotic particle swarm optimization algorithm to train the feed-forward neural network. Reference [8] proposes using genetic algorithms in neural networks to calculate the frequency of a single-shorting-post antenna. These experiments demonstrate that incorporating intelligent optimization algorithms can improve the neural network model's prediction accuracy. Reference [9] proposes a deep neural network behavioral model for power amplifiers based on migration learning, which first acts as a filter to extract the features of the power amplifier and then as an adaptive layer to fit the output of the real power amplifier. A back propagation (BP) neural network model with three hidden layers and the Monte Carlo sampling approach is suggested in [10] for quick prediction of the coupling cross-section of conductor transmission lines.

Most of the aforementioned neural network modelling methods are built based on intuition and experience, which leads to many parameter redundancies in the model's training process and causes computer arithmetic power consumption. Attention mechanism (Attn) is a better method to solve this problem because it can observe the correlation between the decision and the input and the connection weights, achieving the simplification and refinement of the model [11, 12]. Thus, this paper proposes an Attn-based improved deep multilayer perceptron (DMLP) neural network model, which adopts the leaky ReLU activation function to reduce the gradient disappearance problem and uses the Adam optimizer to speed up the algorithm update [13] and the Dropout [14] to prevent the overfitting phenomenon due to the deepening of network layers. Besides, the improved particle swarm algorithm with compression factor (PSOCF) determines the number of hidden layer nodes and the dropout discard rate of DMLP. The advantage of PSOCF

* Corresponding author: Chen Yang (yangchen6526@163.com).

over standard PSO is that it increases the compression factor and expands the influence of inertia weighting to make it faster and more stable in velocity updates. This method is applied to the optimal design of a double-notch UWB antenna model to achieve fast and accurate prediction of S_{11} by antenna notch geometric parameters. The Attn-based PSOCF-DMLP network model put out in this research exhibits fewer errors than the conventional MLP, DMLP, PSO-DMLP, and PSOCF-DMLP networks, according to experimental findings, which attests to the viability of this network model.

2. IMPROVED DMLP NETWORK ARCHITECTURE

2.1. DMLP

DMLP, or "deep neural networks with multiple neuron layers," inserts three or more fully connected hidden layers between the input and output layers. The DMLP's input layer receives data; the middle layer does calculations; and the output layer produces results. A deep network's capacity to fit data is better than a shallow-layer network's. However, arbitrary increases in the number of layers can lead to overfitting and reduced predictive power [15]. Common anti-overfitting means are layer normalization (LN) regularization [16], batch normalization [17], and Dropout. By eliminating specific neurons with probability p, Dropout decreases the intricate co-adaptive interactions between neurons, preventing overfitting and enhancing network generality.

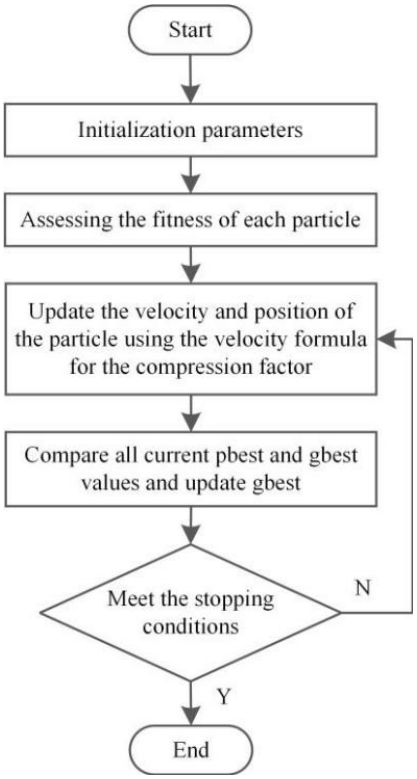


FIGURE 1. PSOCF algorithm flow chart.

2.2. Improved Particle Swarm Algorithm

PSO is a swarm intelligence algorithm that simulates the predatory behaviour of birds. PSO compares the flight space of the bird to the search space for solving the problem and the bird in the search space to the potential solution of each optimization problem, i.e., the particle [18, 19]. Each particle has two properties: velocity and position. Velocity represents the direction and distance the particle will move in the next iteration, while position is the solution to the problem. The expressions for particle velocity and position are as follows:

$$\begin{aligned} V_{id}(t+1) &= wV_{id}(t) + c_1\, rand_1(pbest_{id} - X_{id}(t)) \\ &\quad + c_2\, rand_2(gbest_{id} - X_{id}(t)) \end{aligned} \tag{1}$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \tag{2}$$

where w is the inertia factor; c_1 is the individual learning factor; c_2 is the social learning factor; $rand_1$ and $rand_2$ are [0, 1] uniformly distributed random numbers; $V_{id}(t)$ and $X_{id}(t)$ represent the particle i's velocity and position in generation t; $V_{id}(t+1)$ and $X_{id}(t+1)$ represent the particle i's velocity and position in generation $t+1$; $pbest_{id}$ and $gbest_{id}$ define the particle i's own historical experience and group experience gained via learning. c_1 and c_2 are non-negative constants that control the weights of the local and global optimums, respectively. In the case of setting too large a value of c_1, the particles will stay in the local range for too long as well as the algorithm convergence will be slow, and if the c_2 value is set too large, the particle will quickly converge to the local optimal and easily fall into the optimal local solution. To make an effective balance between the two, the PSO algorithm needs to be improved. PSOCF adds the compression factor φ to the front of the overall velocity equation after removing the inertia weight parameter from the PSO velocity attribute. Such an improvement not only enhances the local space search capacity of the algorithm but also successfully manages the particle flight speed, assures algorithm convergence, and removes the velocity boundary limit [20]. Figure 1 shows a flow diagram of the PSOCF algorithm used to find the optimal value.

The speed update expression of PSOCF is as follows.

$$\begin{aligned} V_{id}(t+1) &= \varphi\Big\{V_{id}(t) + c_1\, rand_1(pbest_{id} - X_{id}(t)) \\ &\quad + c_2\, rand_2(gbest_{id} - X_{id}(t))\Big\} \end{aligned} \tag{3}$$

$$\varphi = \frac{2}{\left|2 - C - \sqrt{C^2 - 4C}\right|} \tag{4}$$

$$C = c_1 + c_2 \tag{5}$$

2.3. Attention Mechanism Layer

Attn is an information selection mechanism widely used in deep neural networks today. It is divided into two stages, first calculating the attention distribution, calculating the probabilities corresponding to different feature vectors according to the weight assignment principle, and then calculating a weighted average of the input information based on the attention distribution and continuously updating and iterating a better matrix

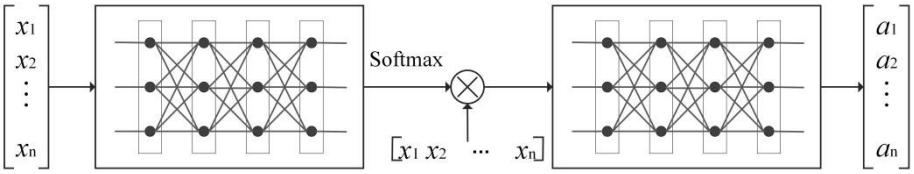


FIGURE 2. Attn layer network architecture.

of weight parameters [21]. The additive model of Attn calculates the weight coefficients in the following way.

$$s_n = v^T \text{Tanh}(W x_n + b) \tag{6}$$

$$\alpha_n = \frac{\exp(s_n)}{\sum_{i=1}^{n} \exp(s_n)} \tag{7}$$

$$attn_n = \sum_{n=1}^{N} \alpha_n x_n \tag{8}$$

In the formula, x_n represents the nth value of the input vector x; s_n represents the matching degree between x_n and the query vector; v and W are the weights; b is the bias; α_n represents the attention probability distribution value determined by the nth input vector; and $attn_n$ is the output after a weighted average of the nth input vector. Figure 2 depicts the structure of the Attn layer.

The input to the Attn layer is an array of features [x_1, x_2, ..., x_n], and the input data is passed through four fully connected layers to obtain the attention distribution of the feature, i.e., the softmax function as the activation of the last layer function to calculate the importance of each characteristic for the task. This attention distribution is then multiplied with the input to obtain the representation of the input data at the next stage, and then the higher-level representation of the input data, the $attn_n$ output, is obtained through four fully connected layers, denoted by [a_1, a_2, ..., a_n].

2.4. Attn-PSOCF-DMLP Neural Network Architecture

The Attn-PSOCF-DMLP neural network model proposed in this paper consists of an input layer, an attention layer, a fully connected layer, and an output layer. The dimensional parameters of the notch structure of the notched UWB antenna are taken as input samples, and the antenna return loss S_{11} is taken as an output sample. Compared with the shallow-layer neural network, this paper deepens the number of hidden layers of the fully connected layer. It has six fully connected layers, including five hidden layers and one output layer. In addition, Dropout operations are embedded among all connected layers to prevent overfitting that occurs during training. The learning rate is automatically updated using the Adam optimizer instead of the traditional stochastic gradient descent algorithm to speed up the model weight update, increase the convergence speed of the algorithm, and jump out of local minima. A variant of the ReLU activation function, leaky ReLU, is applied to solve the problem that neurons do not learn when they enter negative intervals [22].

$$Leaky\,ReLU(x) = \begin{cases} x, & x > 0 \\ \lambda x, & x \le 0 \end{cases}, \quad \lambda \in (0,1) \tag{9}$$

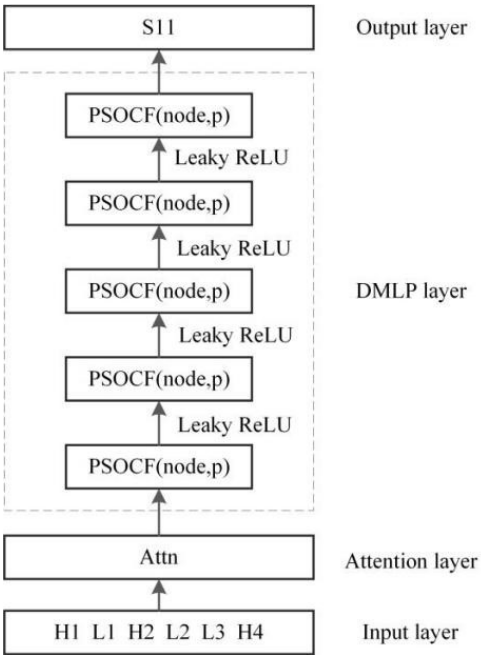
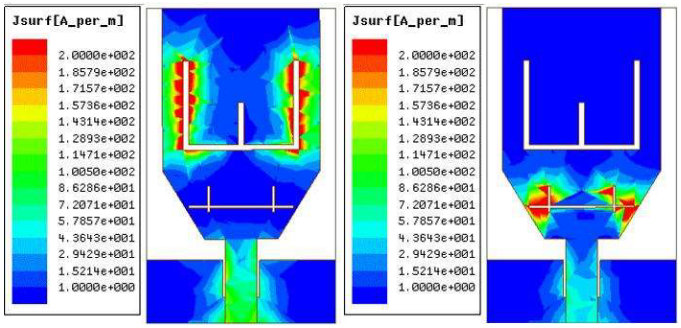


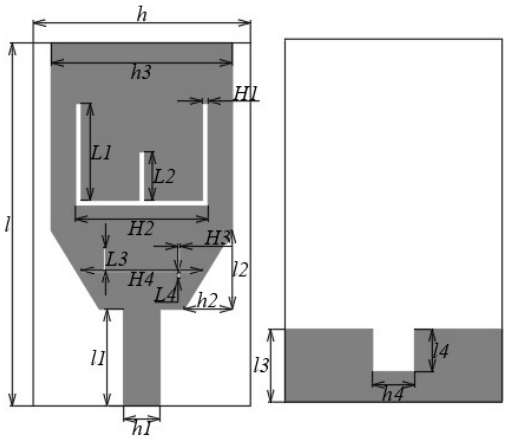
FIGURE 3. Attn-PSOCF-DMLP neural network architecture.

Due to deep neural networks with low explanatory power and many parameter redundancies in the model training process, this paper adds the Attn layer to the deep neural network model to improve the interpretability of DMLP while simplifying the model and improving the model prediction accuracy. Attn is a plug-and-play module. To not change the structure of DMLP, the Attn layer is placed in the first layer of the network in this paper. The input of the Attn layer is the six notches geometric parameters with the most significant influence on antenna performance. The input information is weighted by Attn and input to the network model. The selection of the number of hidden nodes is significant for neural networks. The number of hidden nodes is not only related to the excellent and fast performance of the model but also a direct cause of overfitting during training. Therefore, this paper uses the PSOCF algorithm to optimize the number of hidden nodes and the dropout discard rate p for a DMLP to determine the network architecture. Figure 3 shows the PSOCF-DMLP network architecture diagram based on Attn.



(a) (b)

FIGURE 4. The current distribution in antenna surface. (a) 3.5 GHz and (b) 7.5 GHz.



(a) (b)

FIGURE 5. Dual-notch UWB antenna structure. (a) Front view and (b) back view.

Antenna parameters	l	h	$l1$	$h1$	$l2$
Value (mm)	30	18	8	3	6.5
Antenna parameters	$h2$	$l3$	$h3$	$l4$	$h4$
Value (mm)	4	6	15	3.5	3.5
Notch parameters	$H1$	$L1$	$H2$	$L2$	$H3$
Value (mm)	0.5	8	11	4	0.2
Notch parameters	$L3$	$H4$	$L4$	—	—
Value (mm)	2.5	10	0.2	—	—

TABLE 1. Geometric parameters of the dual-notch UWB antenna.

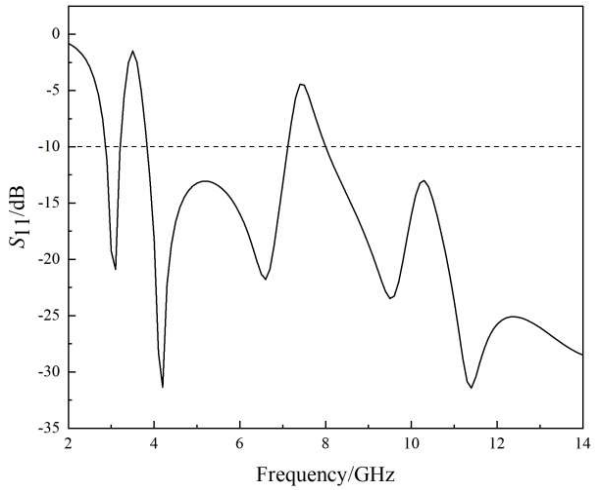


FIGURE 6. Double band-notched UWB antenna S_{11} simulation curve diagram.

3. MODELLING AND OPTIMIZATION OF DUAL-NOTCH UWB ANTENNA

3.1. Analysis and Design of Antenna Structure

This paper selects UWB antenna miniaturization double notch [23] as the research object, the antenna design for Teflon dielectric substrate, with loss tangent 0.02, dielectric constant 4.4, and thickness 1.6 mm. With an inverted E-shaped notch on the upper part of the antenna front and a barbell-shaped notch structure on the lower part, the bandwidth of the antenna reaches $2.8 \sim 14$ GHz, generating $3.2 \sim 3.8$ GHz and $7.1 \sim 8.0$ GHz trap band. Figure 4 shows the antenna current distribution simulation diagram at different notch frequencies. It can be seen that the energy is concentrated at the notch's position so that the antenna's energy cannot be radiated outward, thus the frequency band of the antenna produces a notch. Figure 5 shows the structure of a miniaturized dual-notch UWB antenna.

Figure 6 shows the S_{11} simulation curve derived using HFSS software, and Table 1 shows the antenna geometric parameters. As can be seen from Figure 6, the antenna meets the operating requirements of $S_{11} < -10$ dB in the 2.0–14.0 GHz band and the bandwidth requirements of UWB antennas. Moreover, after the two slots are etched, two notch waves of 3.2–3.8 GHz and 7.1–8.0 GHz are realized, effectively avoiding the interference of the wireless metro area network band and X-band.

3.2. Optimization Steps of Antenna Modelling Based on Improved DMLP

This paper uses Python 1.11 to build network architecture on an i5-7200U CPU. The modelling optimization steps of the miniaturized dual-notch UWB antenna based on AttnPSOCF-DMLP neural network are as follows:

Step 1: Obtain the data set. After extracting simulation samples of the notch structural parameters and modelling a dual-notch, the double-notch UWB antenna was modelled using HFSS software. Following modeling, the notch geometric parameters $H1$, $L1$, $H2$, $L2$, $L3$, and $H4$, which had the greatest influence on antenna performance, were simulated and sam-

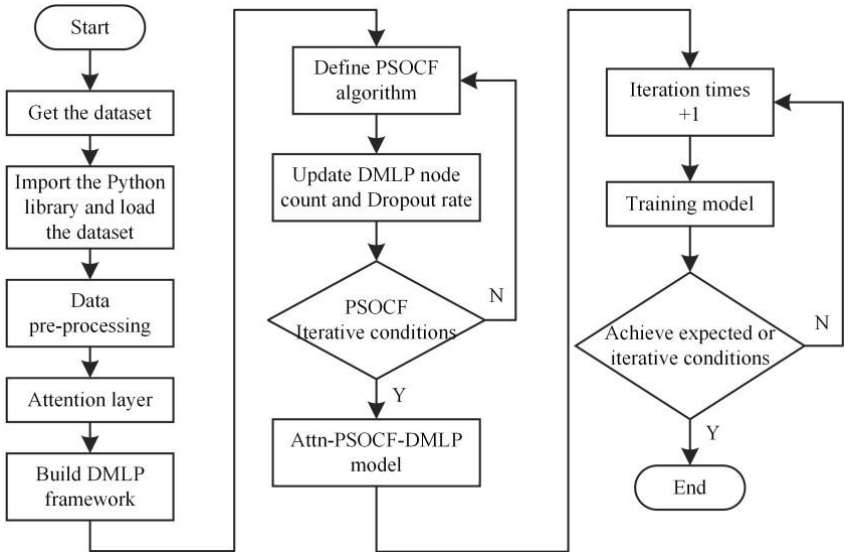


FIGURE 7. Attn-PSOCF-DMLP neural network flow chart.

Parameters	Range/mm	Parameters	Range/mm
$H1$	0.4–0.7	$L2$	3.8–4.1
$L1$	7.9–8.1	$L3$	2.4–2.6
$H2$	10.8–11.0	$H4$	9.9–10.1

TABLE 2. Parameter optimization dimension of notch structure.

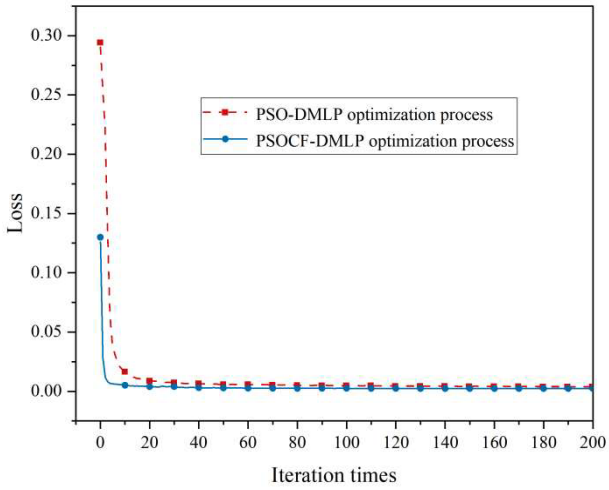


FIGURE 8. PSOCF-DMLP model and PSO-DMLP model optimization process.

pled. Table 2 depicts sample sampling. The antenna frequency is set to $2 \sim 14.0$ GHz, and the step geometry of the samples is set to 0.1 mm. After 1254528 samples are generated with HFSS, transpose every 121 rows of S_{11} to make it into 1 row. The geometric parameter becomes one row reserved for every 121 rows, meaning that six geometric parameters corresponding to $121 S_{11}$ values become a new sample, meaning that the six geometric parameters map a function curve about S_{11}. Finally, 1296 new samples are obtained to form the dataset. The dataset was divided into 80% for training and 20% for testing the model.

Step 2: Data preprocessing. Here min-max is chosen to normalize the data and map the data to the interval [0, 1]. After normalization, the optimization process becomes narrower, and converging to the optimal solution is more accessible.

Min-max data normalization:

$$x_{\text{new}} = \frac{x - \min\ x}{\max\ x - \min\ x} \tag{10}$$

where x_{new} is the normalized input data; x is the actual input data; min$x$ and $maxx$ are the minimum and maximum values of the input data.

Step 3: Build the network model and define the Attn and PSOCF algorithms. PSOCF finds the most suitable solution by calculating the fitness of each particle corresponding to the solution and then iterating over the particles according to the features they need to optimize. A particle is an array of [number of hidden layer nodes: *node*, dropout discard rate: p], with test loss as the fitness. Then the network construction of DMLP is completed after embedding the Attn layer into the input layer of DMLP.

Step 4: Initialization of weights. After establishing the model, by using the uniform distribution of Kaiming to initialize the weights, the scale of the output value of the deep neural network can be maintained within a specific range, which can alleviate gradient disappearance and gradient explosion in backpropagation [24].

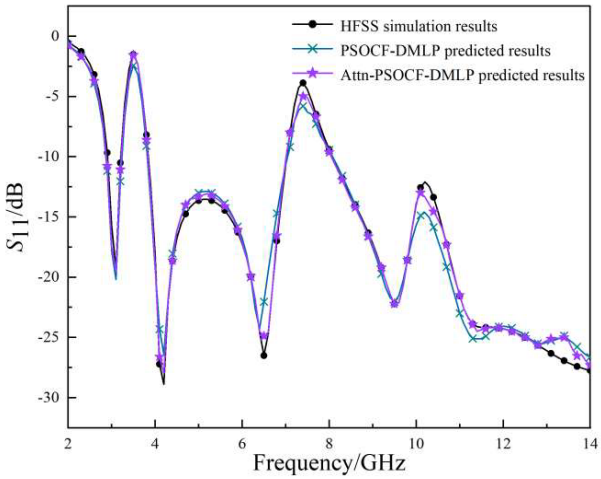


FIGURE 9. Prediction fitting results for network models with and without Attn applied.

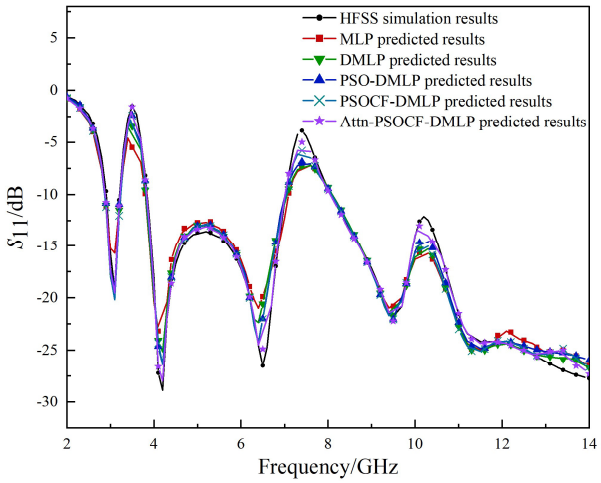


FIGURE 10. Prediction curves of five network models and fitting of HFSS simulation curves.

Step 5: PSOCF algorithm optimization DMLP. In the training process, particle optimization is carried out through PSOCF, and the optimal particle position corresponding to the final DMLP (node, p) is output, then the optimized DMLP model is obtained [6]. Set the particle interval; each node takes the value of [130, 1000]; p takes the value of [0.15, 0.8]; the average accuracy of cross-validation of test samples is selected as the fitness function; the number of PSOCF iterations is 20; both learning factors are 2.05; the number of particles is 5; the batch geometric is 128; the number of neural network iterations is 2000; the learning rate is 0.001; use the MSE loss function and Adam optimizer for model optimization.

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - f(x_i))^2 \tag{11}$$

where y_i $(i = 1, 2, ..., m)$ and $f(x_i)$ $(i = 1, 2, ..., m)$ denote the actual and predicted values of the ith sample, respectively, and m is the number of samples.

Step 6: Train the model and judge whether the output results meet expectations. Verify whether the error meets the expectation through the test set, and if it does, end the training. Save the training model and the data and results generated during the model's training. If the test result does not meet expectations, it returns to step 5 while the number of iterations increases by one, and if the maximum number of iterations is reached, the training ends. In addition, the desired goal can be reached by appropriately adjusting the hyperparameters in step 5.

Data normalization will restore the data recovery formula to achieve the fit of the predicted data and accurate data visualization. Figure 7 depicts this antenna modelling optimization procedure.

$$x = x_{\text{new}}(\max\ x - \min\ x) + \min\ x \tag{12}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

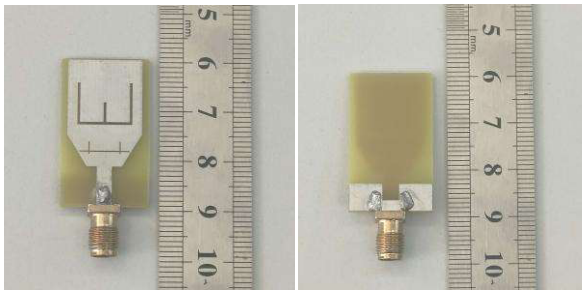
After network learning, the optimal parameter setting of the PSOCF-DMLP neural network model based on Attn is 6-763-296-627-314-395-121, and p is 0.167. In order to verify the superiority of applying the improved particle swarm algorithm and the attention mechanism in deep multilayer perceptual machines, the traditional MLP model, DMLP model, PSO-DMLP model, PSOCF-DMLP model, and Attn-PSOCF-DMLP model were constructed in Pytorch version 1.11 (CPU), respectively. The same dataset was used to perform the five methods of training, and the maximum number of iterations for training is 2000 for all of them. The maximum training times were all 2000 times. Figure 8 shows the loss comparison between the standard particle swarm optimization algorithm and the improved particle swarm optimization algorithm in the process of DMLP.

It can be seen from Figure 8 that the loss of PSOCF-DMLP stabilizes after the 10th iteration, while the loss of PSO-DMLP stabilizes after the 70th iteration, which verifies that the PSOCF algorithm converges faster than the PSO algorithm.

A sample was selected randomly from the test set for the experiments. Figure 9 compares the prediction fit of the DMLP network model with and without applying the attention mechanism to the HFSS simulation values. Figure 9 shows a visual application of the attention mechanism to improve the prediction precision of the model, which is more fitting with the HFSS simulation results.

In addition, Figure 10 compares the output fitting results of all network models and more intuitively proves that deepening the number of network layers, optimizing the network structure with an improved particle swarm optimization algorithm, and adding an attention layer can improve the model's prediction accuracy.

Table 3 shows the performance comparison results of the five network models mentioned above. The evaluation metrics for the experiments are mean absolute error MAE, root mean square error RMSE, mean absolute percentage error MAPE, and coefficient of determination $R2$. Table 3 shows that the Attn-PSOCF-DMLP model has the five models' lowest MAE, RMSE, and MAPE. The root mean square prediction error for S_{11} is also reduced, which is 73.01% better than the conventional perceptron and 64.14% less than the unimproved



(a) (b)

FIGURE 11. Physical diagram of the antenna (a) front view and (b) back view.

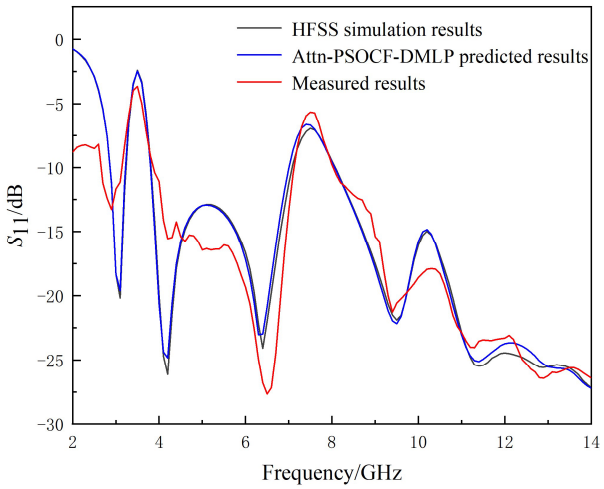


FIGURE 12. S_{11} measurement and simulation and Attn-PSOCF-DMLP prediction.

TABLE 3. Performance comparison of five network models.

Neural network model	MAE	RMSE	MAPE	R2	CPU/s
MLP	4.28957	5.13348	66.52552	0.57254	0.03500
DMLP	2.62315	3.86433	21.61201	0.77296	0.05100
PSO-DMLP	1.09547	1.57689	9.37252	0.95967	0.17000
PSOCF-DMLP	1.07625	1.54214	9.86193	0.96885	0.04900
Attn-PSOCF-DMLP	0.94870	1.38580	8.63200	0.96954	0.10800

depth multilayer perceptron. The coefficient of determination is 0.96954, which is 69.34% better than the conventional perceptron and 25.43% better than the unimproved DMLP, which indicates that the network model has an excellent parallel data processing function and can be more effective in researching complex non-linear interactions between antenna parameters. In addition, Table 3 lists the CPU runtime of different models in predicting S_{11}, and it can be seen that the process is speedy.

The double-notch UWE antenna is physically tested using a vector network analyzer, and the physical antenna is shown in Figure 11. Figure 12 shows the results of the measured S_{11} compared with the HFSS simulated S_{11} and the neural network-predicted S_{11}, from which it can be seen that the antenna produces better trapping characteristics at 3.1–3.9 GHz and 7.0–8.2 GHz. The error in this experimental result may be due to the accuracy of the processing and the loss of the dielectric substrate and the measured environment.

5. CONCLUSION

The improved DMLP network model proposed in this paper uses the leaky ReLU activation function, dropout operation, and Adam optimizer. In the comparison experiments with the traditional perceptron and the unimproved multilayer perceptron model, it is demonstrated that in the optimization of the method for modelling double-notch UWB antennas, the operations of deepening the number of network layers, optimizing the structure of the deep multilayer perceptron model by using the particle swarm algorithm with compression factor, and adding the attention layer all help to improve the prediction accuracy of the model. Among them, improving the particle swarm algorithm accelerates the convergence speed of the algorithm, and the introduction of attention improves the model's interpretability, prediction accuracy, and stability. In conclusion, the model has better fitting accuracy for the S_{11} parameters of the target antenna, which confirms the method's feasibility and provides a fast and effective tool for designing notched UWB antennas.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61971210).

REFERENCES

[1] Jing, H., G. He, J. Sun, and S. Y. Wang, "Design of a reconfigurable band notch antenna for UWB applications," *Progress In Electromagnetics Research C*, Vol. 127, 101–112, 2022.

[2] Rayala, R. K. and S. Raghavan, "Artificial neural network based SIW bandpass filter design using complementary split ring resonators," *Progress In Electromagnetics Research C*, Vol. 115, 277–289, 2021.

[3] Khan, M. M., S. Hossain, P. Mozumdar, S. Akter, and R. H. Ashique, "A review on machine learning and deep learning for various antenna design applications," *Heliyon*, Vol. 8, No. 4, Apr. 22, 2022.

[4] Jin, J., C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, "Deep neural network technique for high-dimensional microwave mod-

eling and applications to parameter extraction of microwave filters," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 67, No. 10, 4140–4155, Oct. 2019.

[5] Cheng, G., M. Wang, W. Zhang, and X. Liu, "Advanced deep neural network technique for microwave parametric modeling," in *2021 IEEE Mtt-S International Wireless Symposium (IWS 2021)*, May 23-26 1–3, 2021.

[6] Nan, J., H. Xie, M. Gao, Y. Song, and W. Yang, "Design of UWB antenna based on improved deep belief network and extreme learning machine surrogate models," *IEEE Access*, Vol. 9, 126 541–126 549, 2021.

[7] Wang, S.-H., M. A. Khan, Z. Zhu, and Y.-D. Zhang, "WACPN: A neural network for pneumonia diagnosis," *Computer Systems Science and Engineering*, Vol. 45, No. 1, 21–34, 2023.

[8] Neog, D., S. Pattnaik, D. Panda, S. Devi, B. Khuntia, and M. Dutta, "Design of a wideband microstrip antenna and the use of artificial neural networks in parameter calculation," *IEEE Antennas and Propagation Magazine*, Vol. 47, No. 3, 60–65, Jun. 2005.

[9] Zhang, S., X. Hu, Z. Liu, L. Sun, K. Han, W. Wang, and F. M. Ghannouchi, "Deep neural network behavioral modeling based on transfer learning for broadband wireless power amplifier," *IEEE Microwave and Wireless Components Letters*, Vol. 31, No. 7, 917–920, Jul. 2021.

[10] Zhou, S. H. and X. Zhao, "Prediction of conductor transmission line coupling cross section based on bp neural network," *Modern Computer*, Vol. 29, 85–90, 2023.

[11] Liu, Z., X. Hu, T. Liu, X. Li, W. Wang, and F. M. Ghannouchi, "Attention-based deep neural network behavioral model for wideband wireless power amplifiers," *IEEE Microwave and Wireless Components Letters*, Vol. 30, No. 1, 82–85, Jan. 2020.

[12] Liu, Z. Y., S. S. Ma, J. Liang, C. M. Zhu, and Y. Lei, "Channel estimation algorithm for millimeter-wave large-scale MIMO systems with attention mechanism CNN," *Systems Engineering and Electronics Technology*, Vol. 44, 307–312, 2022.

[13] Kingma, D. and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.

[14] Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol. 15, 1929–1958, Jun. 2014.

[15] Bejani, M. M. and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artificial Intelligence Review*, Vol. 54, No. 8, 6391–6438, Dec. 2021.

[16] Ying, X., "An overview of overfitting and its solutions," in *2018 International Conference on Computer Information Science and Application Technology*, Vol. 1168, Dec. 7-9, 2019.

[17] Ioffe, S. and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, Vol. 37, 448–456, Jul. 7-9, 2015.

[18] Kennedy, J. and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 — International Conference on Neural Networks*, Vol. 4, 1942–1948, 1995.

[19] Cai, L. G., Y. Q. Hou, Y. S. Zhao, and J. H. Wang, "Application research and improvement of particle swarm optimization algorithm," in *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 238–241, 2020.

[20] Tang, J., S. Q. Zheng, Z. C. Wang, Z. H. Hong, and Y. B. Zhang, "An improved particle swarm optimization algorithm for learning factor and compression factor," *Yunnan Hydropower*, Vol. 38, 77–79, 2022.

[21] Niu, Z., G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, Vol. 452, 48–62, Sep. 10, 2021.

[22] Nayef, B. H., S. N. H. S. Abdullah, R. Sulaiman, and Z. A. A. Alyasseri, "Optimized leaky relu for handwritten arabic character recognition using convolution neural networks," *Multimedia Tools and Applications*, Vol. 81, No. 2, 2065–2094, Jan. 2022.

[23] Gao, M. M., H. L. Niu, and J. C. Nan, "Miniaturized dual band-notched ultra-wideband antenna," in *2021 National Antenna Annual Conference*, 990–993, 2021.

[24] Lyu, Z., A. ElSaid, J. Karns, M. Mkaouer, and T. Desell, "An experimental study of weight initialization and lamarckian inheritance on neuroevolution," in *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*, Vol. 12694, 584–600, Springer International Publishing, Cham, Apr. 7-9, 2021.