Application of BP neural network model in Iris classification

Guanxiao Zhang [1, *], Shuhao Mao [2]

[1] Mathematics and Applied Mathematics major, Shenzhen MSU-BIT University, Shenzhen, China, 518100

[2] Bachelor of Science in Analytics, Nanyang Technological University, Singapore, 639798

* Corresponding Author Email: 2473999436@qq.com

Abstract. In the field of machine learning, iris classification is a classic problem of pattern recognition, and its accurate identification is of great significance to botanical research, ecological protection and horticultural practice, but the existing methods have shortcomings such as complex computation and insufficient generalization ability. The BP neural network model constructed in this paper solves the problem of iris classification prediction, and a stable iris classification prediction model is trained, which effectively fits the mapping relationship between features and categories through the nonlinear transformation of multi-layer neurons. After pretreatment and multiple rounds of training, the model achieved an accuracy rate of 0.91 and an F1 score of 0.86 on the test set, indicating that it could reliably distinguish iris varieties. This study provides efficient classification tools and optimization ideas for practitioners in the fields of horticultural variety screening, botanical classification statistics, and ecological survey.

Keywords: BP neural network, iris classification, UCI dataset warehouse, horticultural variety screening.

1. Introduction

In the field of machine learning and pattern recognition, flower recognition, especially iris classification, has always been a popular problem. With the development of computer vision and machine learning technologies, flower identification has become a hotspot for application in biodiversity conservation, ecological research, and horticultural practices. Accurate flower identification is of great practical importance for botanists, ecologists, and gardening enthusiasts alike. It can not only help scientists better understand plant taxonomy and evolutionary relationships but also enable early diagnosis of flower diseases in agricultural production, thereby improving crop yield and quality. In addition, with the acceleration of ecological processes, the conservation of biodiversity is becoming increasingly important, and accurate flower identification is one of the key steps in biodiversity conservation.

Past studies have shown that BP neural networks can achieve high accuracy in iris classification. Some studies have shown that optimized BP neural networks can achieve nearly 100% classification accuracy. To improve the performance of BP neural networks, researchers have proposed a variety of algorithm improvement methods. For example, the introduction of momentum terms, adaptive learning rates, regularization techniques (such as L1, L2 regularization), etc. By applying different techniques and algorithms, they conducted an in-depth discussion on the classification of iris. These studies not only provide us with a theoretical basis but also point out the direction of further research. Although this problem has been analyzed to some extent in existing literature, they have some limitations in dealing with the problem of iris taxonomy. These shortcomings include the high computational complexity of the algorithm, overfitting to specific datasets, and insufficient generalization capabilities. These limitations have prompted us to seek a more effective approach. In response to the above shortcomings, we propose a new approach. This method is improved based on BP neural network and aims to improve the accuracy and efficiency of classification. By changing the method of calculation error and changing the way of weight update, we can make the BP neural network better deal with the problem of iris classification, and at the same time enhance the generalization ability of the model. We believe that through these studies, we will be able to provide new insights into the problem of irritation taxonomy and promote further development in related fields.

2. Literature review

In past studies, many scholars have studied the problem of plant classification. Many related models have also been proposed and established, such as decision tree models, support vector machines, and BP neural networks. In the study of Zewen Hu et al., the application of citation-based machine learning methods in identifying potential excellent research papers was discussed [1]. This work demonstrates that machine learning models, such as LightGBM, can efficiently process time-series citation data to identify high-impact papers. This provides new ideas for using machine learning techniques to solve classification problems. These machine learning methods have great reference significance for this paper. In the current situation where environmental protection is strongly supported, the protection of biodiversity has also become particularly important, among which Liu Ying et al.'s research emphasizes the role of flowering plants in the biological environment of butterfly diversity [2], which shows that plant characteristics have an important impact on biodiversity in the ecosystem. While this study does not directly focus on iris taxonomy, it provides a way to consider plant characteristics from an ecological perspective and Cheng Lin et al. conducted research on the molecular phylogeny of iris [3], which provided important background information for understanding the genetic diversity and classification of irises. For example, Fu Yu et al. proposed an improved BP neural network algorithm [4], which significantly improved the convergence speed and optimization ability of the algorithm by adaptively adjusting the learning rate and improving the weight correction mechanism. This has important implications for BP neural networks to deal with complex classification problems, such as iris classification problems. In addition, in Chen Xiyuan's research, the iris classification problem was deeply explored based on the BP neural network model [5], and the Levenberg-Marquardt training algorithm was used to achieve high-accuracy classification results. and Zhou Junbo et al. introduced the L1 regular term in the BP neural network [6], which effectively reduced the overfitting problem of the model and improved the success rate of iris classification. These studies show that BP neural networks have great potential for iris classification. At the same time, these studies also provide research directions for this paper.

3. Establishment of the model

3.1. Introduction to the basic concept of the model

BP neural network is a multi-layer feedback neural network based on error backpropagation algorithm, which is widely used in biometrics, classification and other fields. Its structure consists of an input layer, a hidden layer and an output layer, each layer contains several neurons, and the interlayer neurons realize information transmission through different weight connections. The specific structure is shown in Figure 1.

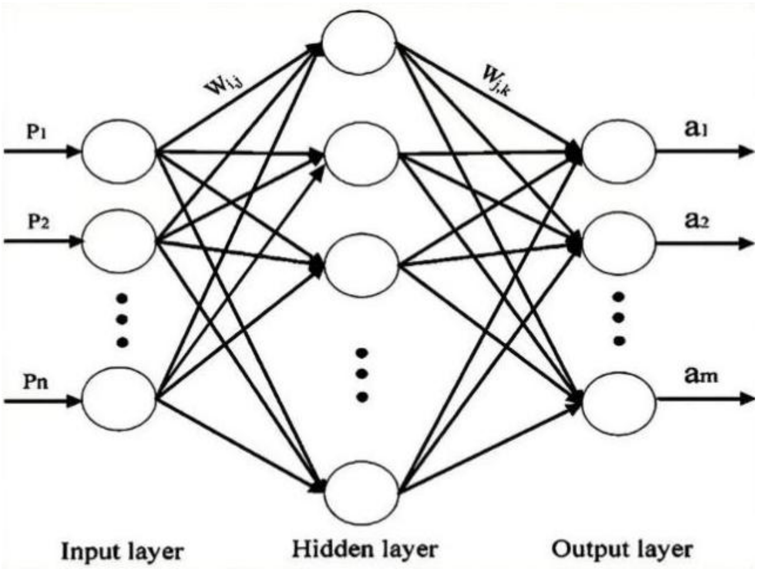

Figure 1. Neural Network structure

The core principle is to calculate the output result through forward propagation and then reverse propagate the error between the actual output and the expected output, dynamically adjust the connection weights between each layer to minimize the error, and finally realize the accurate mapping of the input data by the model. In the iris classification task, the BP neural network can take the sepal length, width and other characteristics of iris as input, and output the classification results from the output layer through the nonlinear transformation of the hidden layer.

3.2. Establishment of BP neural network model

3.2.1. Data preprocessing

In this paper, the iris dataset contains four characteristics: sepal length, sepal width, petal length, and petal width. Randomly divide the training set and the test set to enhance the generalization ability of the model. Trying different initialization methods (including random initialization [7], Xavier initialization, and He initialization) to find the one with the best processing effect can make the preprocessing stage more inclusive. For example, Xavier initialization has the core idea of keeping the variance of the input signal as constant as possible as it passes through each layer of the neural network, so that the signal does not become too large or too small during propagation. The core goal of He initialization is to ensure that the variance of each layer output is consistent with the variance of the input when using the ReLU activation function, so as to avoid the signal being too large or too small during the propagation process. The data is then standardized for deviation using the formula:

$$x_i^* = \frac{x_i - min(X)}{\max(X) - min(X)} \tag{1}$$

Where x_i is the i data point in the original dataset.$min(X)$. is the smallest value in dataset X.$max(X)$ is the maximum value in dataset X.

To eliminate differences in the range of values between features, so that they can be compared and weighed fairly.

3.2.2. Forward propagation

Input the training data and construct the output formula for each layer of neurons:

$$a_j^l = f\left(\sum_i w_{ji}^l x_i + b_j^l\right) \tag{2}$$

Where a is the output after activation, f is the activation function, w is the weight, b is the bias, and l is the number of layers.

By calculating the input layer, hidden layer and output layer separately, the forward propagation is completed to predict the data.

3.2.3. Reverse propagation

In the backpropagation stage, the gradient descent method needs to be used to calculate the initial error based on the difference between the predicted value and the real value of the output layer. Then, the error is reversed to derive the error terms of the hidden layer and the input layer. Then, based on these errors, the weights of the input-hidden layer and the hidden-output layer are updated, so as to speed up the model convergence. Finally, the mean square error of all samples is calculated to complete the complete calculation process of backpropagation and realize the iterative optimization of network parameters.

3.2.4. Weight update and check conditions

In the neural network training process, weight update is a key part of model iterative optimization, and reasonable construction of end conditions can allow the model to strike a balance between accuracy and efficiency to avoid invalid iterations. The following introduces two mainstream optimization algorithms and then elaborates on the design of multi-dimensional termination conditions.

As the basic optimization framework, the core logic is to calculate the gradient using a single/small batch of samples, and its formula is:

$$g_t = \nabla J(\theta_t) \tag{3}$$

Where J is the loss function and the parameters are updated along the negative gradient direction.

However, the traditional gradient descent has shortcomings: it only relies on the current batch gradient, which is susceptible to noise interference and poor convergence stability. To solve this problem, improved methods such as momentum optimizer and adaptive momentum estimation algorithm are used.

The Momentum Optimizer [8] introduces the concept of "momentum" to simulate inertia, which is updated with the following formula:

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{v_t}{1-\beta_2^t}+\epsilon}} \tag{4}$$

Where α is the learning rate, m_t is the first-moment estimate, v_t is the second-moment estimate and ϵ is a small constant to prevent division by zero (typically set to 10^{-8}).

By accumulating the historical gradient direction, the convergence of the gradient stable direction can be accelerated, and the noise direction oscillation can be suppressed, which is suitable for dealing with non-convex optimization scenarios.

The Adam optimizer combines momentum and adaptive learning rate advantages. The first moment (mean) and second order moment (variance) of the gradient are calculated first, and then the parameter learning rate is dynamically adjusted through the deviation correction term (eliminating the initial moment estimation bias) to achieve efficient convergence, which is highly versatile in complex network training.

The training end conditions directly affect the convergence efficiency and final performance of the model, and the optimal strategy for adapting the task needs to be selected by comparing the validation set indicators (loss, accuracy, etc.). In this study, a multi-dimensional conditional combination mechanism is constructed, and the design is designed from three levels: training integrity, performance convergence, and index compliance degree:

1) Epoch completion: Epoch is defined as "a complete traversal of a training set", with the maximum epoch set as the base termination threshold. When the training round reaches this value, the training is forcibly terminated. This condition ensures that the model fully learns data patterns while avoiding time waste caused by over-iteration.
2) Early stop mechanism: The performance of the verification set is used as the monitoring object, and if the verification set indicators (such as accuracy) in multiple epochs in a row do not improve, the termination will be triggered. This mechanism effectively inhibits overfitting and

allows the model to retain the optimal generalization ability in time, which is the core strategy to prevent performance degradation.

3) Loss threshold trigger: The target value of the validation set loss function is set by default, and when the validation set loss continues to drop to this threshold during training, the model is determined to have learned a valid pattern and the training is automatically terminated. This condition sets a clear performance benchmark for training to ensure that the model converges in time when the accuracy meets the standard.

Experiments show that the combined design of "completion guarantee + early stop prevention of overfitting + threshold control accuracy" can maximize the training efficiency and provide support for the rapid convergence and stable deployment of the model under the premise of ensuring model performance.

3.2.5. Algorithm training process

The flow chart of the BP neural network model for the classification problem of warbler flower is shown in Figure 2.

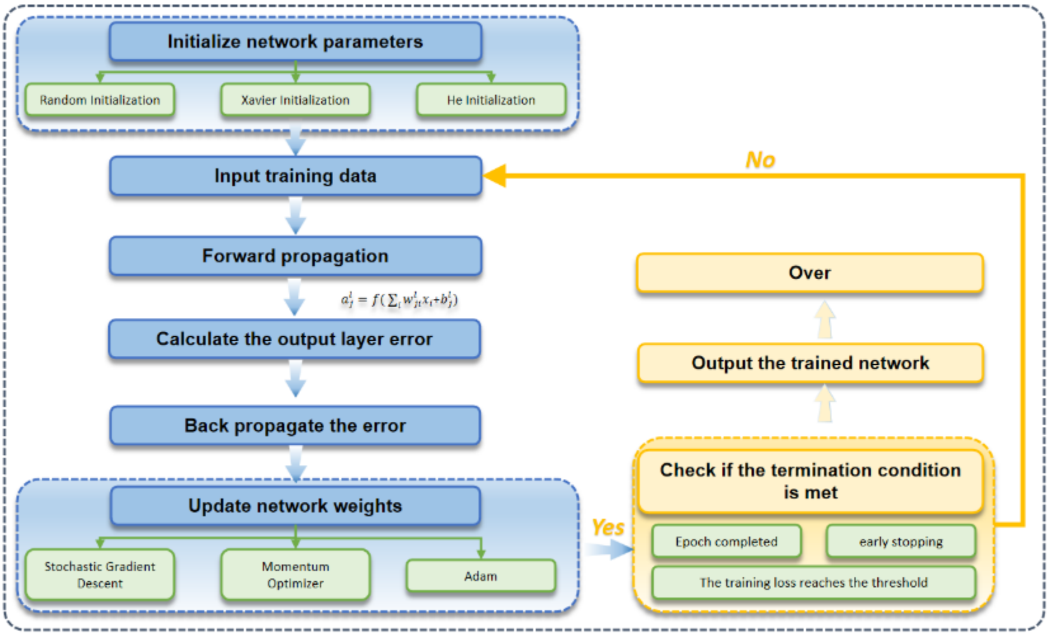

Figure 2. Flowchart of Neural Network Algorithm

Before training the data, the BP neural network needs to be initialized, first determine the network structure, clarify the number of neurons (1000) in the input layer (four layers), hidden layer (three layers) and output layer (three layers), and predict the output layer by category. The connection weights are then initialized, and a random number generator is used to assign values to the connection weights between neurons in each layer of the network. A common method is to randomly sample from a uniform or normal distribution and let the weights be valued within a small range to avoid neuronal output values that are too large or too small, affecting the training effect. Finally, the bias term is initialized, and the bias term of each neuron is assigned, usually initialized to 0 or a smaller random number. As with weights, a suitable initial value helps the model converge faster.

4. Model solving

4.1. Data introduction

Data sources: There are several data sources in this study, some of which come from the Scikit-learn library, which is a commonly used machine learning library with built-in iris datasets. The other part comes from the UCI Machine Learning Repository [9], which is a well-known machine learning dataset repository that contains many public datasets, including the Iris dataset, which can supplement the missing data.

Data characteristics: The dataset contains 4 botanical features for classification, all of which are morphological measurements of iris (in cm), as follows:

Sepal length: Refers to the length of the sepal of the iris flower.

Sepal width: Refers to the width of the iris sepal.

Petal length: Refers to the length of iris petals.

Petal width: Refers to the width of iris petals.

These four characteristics are the key basis for distinguishing different iris varieties, such as the significant differences in petal length and width between different varieties.

Then there are three labels: the corresponding category labels of the sample are 3 iris varieties, namely: Iris-setosa, Iris-versicolor and Iris-virginica.

In the dataset, labels may be stored as integers (such as 0, 1, 2) or strings, and the format may vary slightly from source to source, but the varieties represented are the same.

The characteristics of feature distribution are characterized by certain correlations between features, such as petal length and petal width. and some varieties have clear feature boundaries (such as mountain iris and two others), and some have overlaps (such as color-changing iris and Virginia iris), which is suitable for testing the performance of classification algorithms.

4.2. Evaluation indicators

The evaluation indicators of this study were divided into three categories: accuracy, precision and recall, and F1 score.

Accuracy: The proportion of correctly classified samples to the total number of samples, the formula is:

$$Accuracy = \frac{correct}{all} \tag{5}$$

Where correct is the number of correctly classified samples, and all is the total number of samples.

Precision/Recall: The proportion of samples predicted for a category that are in that category ("Proportion of predicted pairs"). Recall is the percentage of a sample that is successfully predicted for a category ("The proportion of predicted pairs in the actual class"). Since it is a multi-classification problem, the averaging method usually calculates Macro-averaged [10] or Micro-averaged, where macro-averaging calculates the precision/recall of each category first, and then takes the average (Treating each category equally). The micro-mean is the combination of all categories to calculate the overall precision/recall (More affected by categories with large sample sizes).

F1 score: The harmonized average of precision and recall, which can combine the performance of the two, and its formula is:

$$F1 = 2 \times \frac{pre \times recall}{pre + recall} \tag{6}$$

Where pre is the accuracy rate, and recall is the recall rate.

The average method: the same precision/recall, mostly using Macro average or Micro-average.

4.3. Analysis of model solving

As shown in Figure 3, the red line indicates whether it is the Iris-setosa, yes at the peak, and not at the trough. The blue line represents the probability, which is the size of a certain type of probability.

The horizontal axis is the number of test sets, which fully presents the performance fluctuation trajectory of the model from the initial to the later stage in the classification task. In classic small-sample and multi-category tasks such as iris classification, the learning rhythm of the model's sample features can be clearly shown - from the exploration of feature patterns in the early stage to the convergence trend in the later stage, which provides an intuitive basis for analyzing the "learning-adaptation-optimization" process of the model.

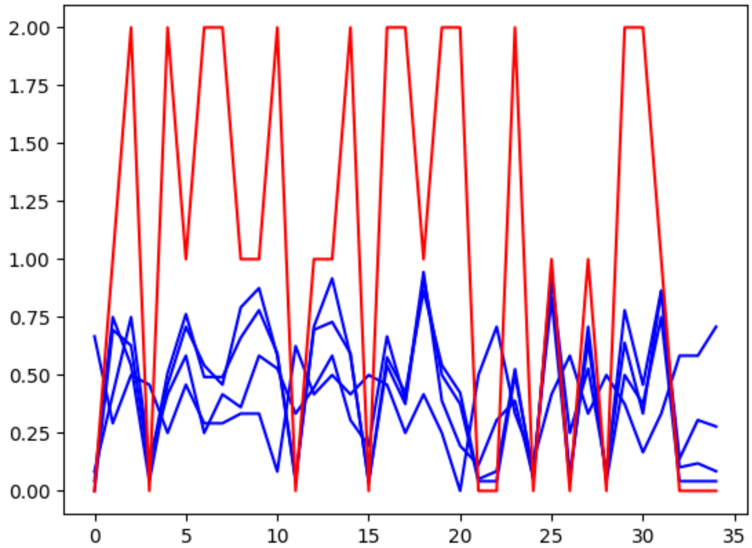

Figure 3. Predicted probability result

The double-track fluctuations of red and blue polylines correspond to the performance of different classification strategies in the classification of iris. The large peak of the red polyline correlates with the accurate identification moment of the model for a specific type of iris. The gentle fluctuation of the blue line reflects the stable adaptability of another strategy to the overall sample. This comparison can assist in screening classification methods that are more suitable for the distribution of iris data.

The line chart of the accuracy of model prediction is shown in Figure 4:

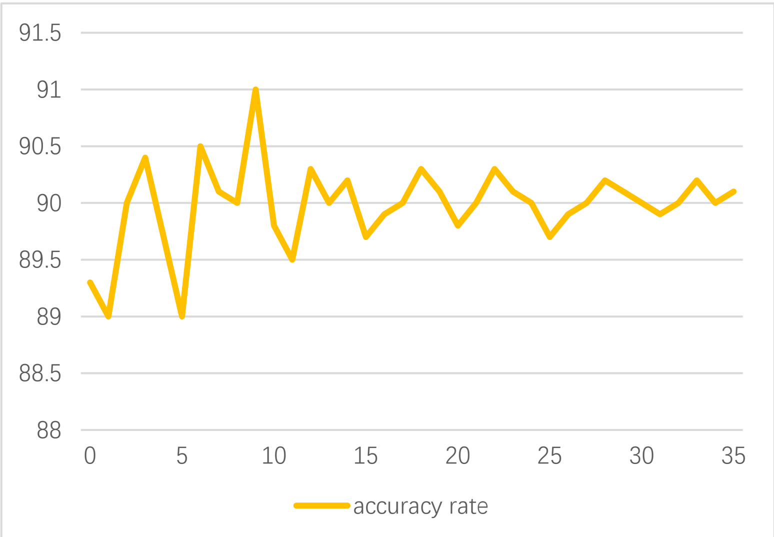

Figure 4. Model prediction accuracy rate

From the overall trend, the accuracy fluctuated dynamically in the range of 89-91%, reflecting that there were differences in the learning and adaptation of the characteristics of the iris sample by the classification model.

The 90% accuracy of the model means that the model has captured the key features of iris classification, such as the size and shape of petals and sepals, and the model can be better identified and used for classification, indicating that feature engineering and model structure adapt to the data to a certain extent, lay the foundation for subsequent optimization, and can mine more detailed patterns based on existing features. In addition, in the conversion from the training set to the test set, it can stably output 90% accuracy, reflecting that the model has basic generalization ability, can cope with the classification of unseen iris samples, and provides the possibility for practical scenario application, which is a key transition indicator of the model from theory to practice.

The F1 score line chart of the model is shown in Figure 5:

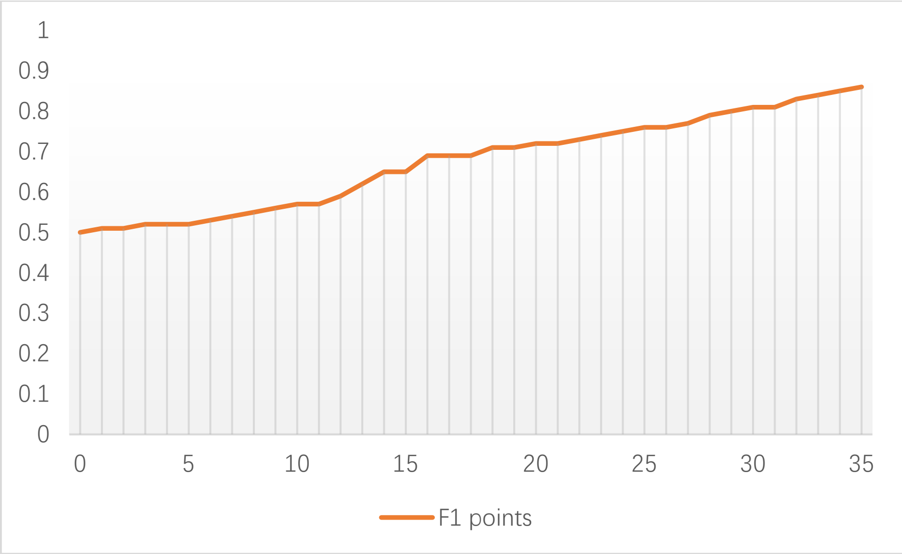

Figure 5. Model F1 points

Looking at the overall trend, the F1 score was initially low, but after improvement, the F1 score increased to 0.86. The final F1 score of 0.86 is a high-reliability performance in the iris classification task, which means that about 86% of the predicted positive cases are true as the target category, which can effectively avoid the error of "misjudging ordinary irises as rare varieties". In terms of recall, about 86% of the true cases are correctly identified, reducing the risk of "missed detection of rare varieties". This performance is sufficient to support the needs of horticultural variety screening [11] and ecological investigation [12].

4.4. Conclusion of the experimental part

This experiment focuses on the problem of iris classification and forms a complete process from data acquisition to model application. In the data acquisition process, we selected the classic iris dataset, and each sample had four attributes: calyx length, calyx width, petal length and petal width, which provided a rich and reliable data foundation for subsequent research.

In the data preprocessing stage, we use different initialization methods and deviation normalization to scale the data to the (0,1) interval to complete data normalization, improve data quality, and ensure that the model can better learn data features.

After multiple rounds of training and parameter adjustment, the results of 0.91 accuracy and 0.86 F1 score were obtained on the test set, indicating that the model has high accuracy and reliability in the iris classification task, and can effectively capture the distance relationship between samples and accurately determine the iris category.

In practical application scenarios, the model can help the field of horticulture to quickly identify iris varieties, assist in the classification and statistics of iris populations in botanical research, and provide efficient and accurate classification support for ecological research and variety cultivation, showing good application prospects. However, there is still room for improvement in the classification of boundary samples in complex scenarios, and the model performance can be further enhanced by optimizing feature engineering and trying model fusion.

5. Conclusion

The results show that the F1 score of the constructed model steadily increases from 0.5 to 0.86 with the iteration of the model, which verifies that the model can effectively balance accuracy and recall through feature engineering optimization, classification strategy adaptation and generalization enhancement, and achieve a gradual breakthrough in the classification performance of iris. This study also provides a reusable practice path of "iterative optimization-performance leap" for small-sample multi-category classification tasks, which makes up for the analysis gap of focusing only on results and ignoring the law of performance optimization by dismantling the model learning stage, clarifying the characteristics and policy-driven logic, and providing theoretical support and practical reference for the development and tuning of similar classification algorithms. However, this model also has limitations: the experiment focuses on the standard iris dataset, does not cover complex scenarios such as "collecting noisy samples in the field", and the model optimization depends on the number of iterations, and does not explore the training efficiency in dep. In the future, the research boundaries can be further expanded to verify the robustness of the model by introducing interference samples in the real environment. Combine meta-learning and transfer learning to optimize the initial iteration efficiency. The multi-model fusion strategy is attempted to break through the bottleneck of the balance between classification accuracy and generalization ability of a single model, and deepen the research depth and practical value of iris classification and similar small-sample classification tasks.

References

[1] Zewen Hu, Jingjing Cui, Angela Lin, Identifying potentially excellent publications using a citation-based machine learning approach [J]. ScienceDirect, 2023-05-01.

[2] Liu Ying, Zhang Zhe, Li Xiangyu, et al. Creation of Butterfly Diversity Habitat Environment Mainly Flowering Plants [J]. Sichuan Forestry Science and Technology, 2023, 44 (3): 54−63.

[3] Cheng Lin, Feng Shucheng, Xiao Yue'e, et al. Research progress on molecular phylogeny of Iris [J]. Guangxi Botany, 2021, 41 (1): 31-39.

[4] He Dasi, Jin Luqi, Zhang Zuming, et al. Improvement of BP neural network regression prediction model [J]. Mechanical Engineering and Automation, 2025, 54 (01): 224-226.

[5] Chen Xiyuan. Research on Iris Classification Problem Based on BP Neural Network Model [J]. New Generation Information Technology, 2022, 5 (5): 146-148.

[6] Zhou Junbo, Liu Guangrong, Zeng Shan. Application of improved BP neural network in irises classification and recognition [J]. Journal of Wuhan University of Light Industry, 2021, 40 (6): 41-46.

[7] Wu Yinting. An Adversarial Training Framework for Initializing Noise Design and Sample Selection [D]. Central China Normal University, 2024.

[8] Chang Xilong, Liang Kun, Li Wentao. Computer Engineering and Applications, 2024,60 (07): 1-12.

[9] Xu Duo. Discussion on Database Query Performance Optimization Method Based on Machine Learning [J]. Information and Computer, 2025, 37 (08): 176-178.

[10] Xiong Kuan. Research on Text Classification Method Based on Deep Learning and Attention Mechanism [D]. Jiangxi Normal University, 2020.

[11] Fan Wei, Kong Qiuyue, Hu Jie, et al. Screening of mulberry germplasm resources suitable for high-calcium environment [J]. Journal of Southwest University (Natural Science Edition), 2024, 46 (10): 90-98.

[12] Introduction to the ecological investigation, evaluation and restoration project of granite weathered shell rare earth ore of the National Geological Experiment and Testing Center of the China Geological Survey [J]. China Geology, 2025, 52 (03): 768.