WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Predicting Color Schemes Based on Genetic Algorithm and Deep Feedforward Neural Network Model

Yongchao Wang *, Xingcheng Wu, Jianni Hou, Mengda Hu

College of automation, Chongqing University of Posts and Telecommunications, Chongqing, China, 400065

* Corresponding Author Email: wangyongchao1212@petalmail.com

Abstract. To enhance the accuracy and efficiency of color matching for opaque products, a relevant model is employed to predict their color schemes. By extracting the spectral reflectance of dyed panels as input and dyeing ratios as output, it is possible to reduce color matching errors and improve production efficiency, thereby advancing the dyeing technology of opaque products and achieving color matching freedom. This study initially applies a logarithmic transformation to the K/S (absorption coefficient/scattering coefficient) ratio and concentration, converting the resultant equations into linear forms. The least squares method is then employed to solve the linear equations describing the relationship between K/S and concentration. A comparison is made between linear regression fitting and nonlinear regression fitting methods, revealing that the former yields better fitting results and higher correlation coefficients in the range of 400nm-700nm for K/S versus concentration. Thus, linear fitting is chosen. In addressing the issue of predicting color differences with minimal deviation based on target samples, this paper unifies variables using the Beer-Lambert color matching equation, Kubelka-Munk theory, and the Simpson discrete variable integration formula to obtain a training set. Leveraging the CIELAB color space model, a deep feedforward neural network is trained on the L, a, and b values for red, yellow, and blue. Finally, a predictive model combining genetic algorithm and deep feedforward neural network (GA-DFNN) is utilized to forecast color matching schemes for the original samples, resulting in 5 sets of color matching schemes with color differences of no more than 1 unit.

Keywords: Predicting Color Schemes Using Computational Color Theory Algorithms, Least Squares Method, Genetic Algorithm, Deep Feedforward Neural Network.

1. Introduction

Color measurement [1] refers to the measurement of color stimulus values. The human eye itself serves as a colorimeter, capable of discerning minute differences between colors. However, visual inspection has many drawbacks, primarily manifested during the color measurement process, where subjective human judgment often influences the outcome. Additionally, factors such as observation conditions and color adaptation further impact results, leading to variations between individuals and rendering it unsuitable for quantitative analysis and processing of colors [2] With economic development, there is an increasing demand for opaque products in daily life. The coloration of opaque products plays a crucial role in their aesthetic appeal and market competitiveness. However, traditional manual color matching has limitations, characterized by subjectivity and low efficiency. Therefore, researching how to achieve color matching for opaque products through computational methods [3] is of significant importance.

Huang Jing et al. [4] proposed an improved LSTM-based method for predicting wood staining formulations. Utilizing a dataset of stained wood, where the surface spectral reflectance of wood serves as the model input and the corresponding wood staining formulation as the output, an LSTM model was employed to predict staining schemes based on spectral reflectance. Deng Xiaogang et al. [5] introduced a study on neural network-based offset ink color matching. They established a neural network model using the PyTorch framework and achieved a highly accurate ink color matching software through model training and optimization. The model trained using an improved particle swarm optimization BP neural network had a mean squared error of 0.0033 and an average color

difference of 4.7896. The model trained using a partitioned network had a minimum color difference of 0.0731, a maximum color difference of 7.5883, and an average color difference of 2.4957. The predicted color differences were mainly distributed between 1 and 3, with 92.31% of color differences being less than 5.

The rapid development of computer technology and its peripherals has greatly enhanced the industrialization level in our country. The use of computers for color reproduction has become increasingly mature, enabling the rapid representation of the tristimulus values of object colors and color differences in numerical form. This not only avoids errors in measurement results due to personal subjective factors during color measurement but also facilitates quantitative analysis and processing of colors. Therefore, this paper conducts research on the relationship between the wavelength and concentration of the primary colors red, yellow, and blue and the K/S value and color difference based on a predictive model combining genetic algorithms and deep feedforward neural networks.

2. Solving the Relationship between K/S and Concentration

2.1. Preliminary Analysis

This question involves a data fitting problem, aiming to determine the relationship between K/S and concentration. Such a relationship can be identified through statistical methods like regression analysis or curve fitting. Initially, the data in Attachment 2 needs to be cleaned and organized, pairing K/S values with their corresponding concentrations. Subsequently, an analysis of the data for each dye is conducted. Based on the distribution of data for different dyes, an appropriate fitting model is selected, which could be a linear regression model or a nonlinear polynomial regression model.

2.2. Confidence Level of Fitting Coefficients

(1) Statistical Significance: When conducting linear fitting in this study, the estimated coefficients may be influenced by observational errors, leading to inaccurate estimates. This study aims to determine whether these fitting coefficients are parameters truly indicative of a significant linear relationship with the target variable through statistical significance testing. Generally, a 95% confidence level is commonly used for statistical significance. Fitting coefficients at a 95% confidence level imply a 95% level of confidence or reliability, indicating that these coefficients exhibit a significant linear relationship with the target variable [6].

(2) Hypothesis Testing: Hypothesis testing is a method used to make statistical inferences about fitting coefficients. In hypothesis testing, a null hypothesis (H0) is set, assuming that the fitting coefficient is zero or unrelated. Subsequently, the probability of this null hypothesis is evaluated through calculating statistics such as standard errors, t-values, or p-values. Generally, if the fitting coefficient significantly deviates from zero, i.e., the p-value is less than 0.05 (at a 95% confidence level), this study rejects the null hypothesis, considering the coefficient to be significant.

(3) Reliability and Generalization Ability: The confidence level of fitting coefficients is related to the reliability and generalization ability of the model. In most cases, a higher confidence level of fitting coefficients, indicating higher significance, corresponds to better predictive ability and stability of the model. This implies that when applying the model for future predictions or generalizations, the obtained fitting coefficients are more likely to accurately predict the target variable.

Due to the limited accuracy of linear functions, to achieve a confidence level of over 95%, this study not only adopts linear fitting but also utilizes higher-order nonlinear fitting.

2.3. Solving Linear Equations Using the Least Squares Method

Fitting coefficients are typically required to achieve a confidence level of over 95% to ensure the reliability and statistical significance of the fitting results. This study can employ methods such as the

least squares method to determine the parameters of the model, minimizing the error between the predicted values and the actual values. Upon obtaining the model parameters, this study establishes the relationship between K/S and concentration.

The fitting coefficients obtained from linear regression analysis are as follows:

$$a = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{1}$$

$$b = \bar{y} - a\bar{x} \tag{2}$$

Least Squares Method:

Utilizing Euclidean space, we demonstrate the least squares method and provide the algebraic conditions satisfied by the least squares method. Let's denote the fitting equation as:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix} \tag{3}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} \cdot \cdot \; Y = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_1 \\ \sum_{j=1}^{n} a_{2j}x_2 \\ \sum_{j=1}^{n} a_{mj}x_m \end{bmatrix} = AX \tag{4}$$

Using the notion of distance, the least squares method [7] entails identifying the vector Y that minimizes the distance between Y and B. It is understood that vector Y represents the vectors formed by the respective columns of A, denoted as $a_1, a_2, \cdots, a_n$ generating the subspace $\mathrm{L}(a_1, a_2, \cdots, a_n)$.

$$Y = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{bmatrix} \tag{5}$$

Therefore, the least squares problem can be formulated as follows: Find a vector Y in the subspace $\mathrm{L}(a_1, a_2, \cdots, a_n)$ such that its distance to B is shorter than its distance to any other vector in the subspace $\mathrm{L}(a_1, a_2, \cdots, a_n)$.

Applying the conclusion discussed earlier, let's denote the vector Y as:

$$Y = AX = x_1a_1 + x_2a_2 + \cdots + x_na_n \tag{6}$$

If it is the vector that is required,

$$C = B - Y = B - AX \tag{7}$$

It must be orthogonal to the subspace $\mathrm{L}(a_1, a_2, \cdots, a_n)$. To achieve this, it is both necessary and essential to, in accordance with the rules of matrix multiplication, express the aforementioned series of equations in terms of matrix multiplication, namely:

$$(C, a_1) = (C, a_2) = \cdots = (C, a_s) = 0 \tag{8}$$

$$\dot{a_1}C = 0, \dot{a_2}C = 0, \cdots, \dot{a_n}C = 0 \quad (9)$$

And $\dot{a_1}, \dot{a_2}, \ldots, \dot{a_n}$ arranged as rows precisely form matrix $\dot{A}$. Combining the aforementioned series of equations yields $\dot{\mathrm{A}}(\mathrm{B} - \mathrm{AX}) = 0$. The coefficient matrix is denoted as $\dot{AA}$, with the constant term being $\dot{AB}$.

The expressions relating K/S to concentration for the three types of coloring agents at different wavelengths are given in table 1. The fitting relationship between K/S and concentration for the red coloring agent at various wavelengths is provided. The same approach applies to the remaining two colors.

Table 1. Fitting relationship between K/S and concentration for the red coloring agent at different wavelengths.

Coloring Agent Color	Coloring Agent Color	Coloring Agent Color	Coloring Agent Color
Red	400	y = 0.7519x + 0.5226	0.9835
	420	y = 0.6826x + 0.2244	0.9947
	440	y = 0.6472x + 0.1294	0.9949
	460	y = 0.6309x + 0.1437	0.9974
	480	y = 1.0847x + 0.1820	0.9992
	500	y = 1.6852x + 0.2305	0.9998
	520	y = 2.4895x + 0.2822	0.9992
	540	y = 2.6880x + 0.2629	0.9995
	560	y = 2.5908x + 0.2977	0.9997
	580	y = 1.0760x + 0.0375	0.9998
	600	y = 0.2338x + 0.0101	0.9986
	620	y = 0.0499x + 0.0081	0.9987
	640	y = 0.0098x + 0.0063	0.9985
	660	y = 0.0050x + 0.0050	0.9995
	680	y = 0.0018x + 0.0061	0.9691
	700	y = 0.0007x + 0.0064	0.9621

3. Establishment of Color Prediction Model based on Genetic Algorithm and Deep Feedforward Neural Network

3.1. Data Preprocessing

Based on the attached data provided on the official website of Big Data and Mathematical Model of China Society for Future Studies, this paper preprocessed the data first, so as to unify the variables of the data used in the topic, and facilitate the prediction of the sample data to obtain the specific concentration ratio.

Based on the provided supplementary data, this paper preprocesses the data to standardize the variables used in the topic, facilitating the prediction of specific concentration ratios for the sample data.

Utilizing existing textile optical models, the Belc computer color matching equation and the K-M theory [4], R(λ) can be computed.

$$\frac{K}{S} = \frac{(1-R)^2}{2R} \quad 10)$$

$$2R\left(1 + \frac{K}{S}\right) - R * R = 1 \quad (11)$$

From the calculated R values corresponding to the red, yellow, and blue colors, the spectral tristimulus values are computed using the Simpson discrete variable integration formula.

$$X = k\int_{400}^{700} S(\lambda)\bar{x}(\lambda)R(\lambda)d(\lambda) \tag{12}$$

$$Y = k\int_{400}^{700} S(\lambda)\bar{y}(\lambda)R(\lambda)d(\lambda) \tag{13}$$

$$Z = k\int_{400}^{700} S(\lambda)\bar{z}(\lambda)R(\lambda)d(\lambda) \tag{14}$$

After computing the spectral tristimulus values of the observer, this paper employs the CIELAB uniform color space model, as illustrated in figure 1.

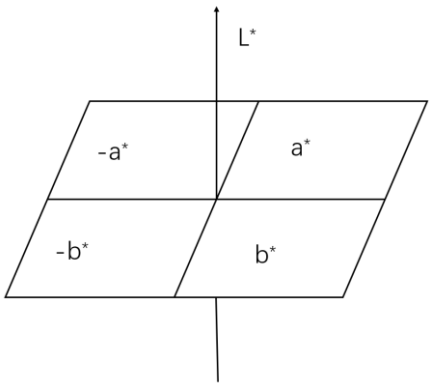

Figure 1. Illustrates the schematic diagram of the CIELAB color space system.

The color parameters L, a, and b are obtained from the formulas.

$$L = 116(\frac{Y}{Y_0})^{\frac{1}{3}} - 16 \tag{15}$$

$$a = 500\left[(\frac{X}{X_0})^{\frac{1}{3}} - (\frac{Y}{Y_0})^{\frac{1}{3}}\right] \tag{16}$$

$$b = 200\left[(\frac{Y}{Y_0})^{\frac{1}{3}} - (\frac{Z}{Z_0})^{\frac{1}{3}}\right] \tag{17}$$

The new sample data after preprocessing, as obtained in this study, are presented in table 2.

Table 2. Sample data after pretreatment.

sample	X	Y	Z	L	a	b
1	57.507563	49.08212313	57.9257143	75.50238965	28.81011992	-5.046603987
2	70.35944545	68.61737881	80.46701718	86.31422773	11.64008369	-5.257073825
3	64.2897443	60.09783437	73.60554093	81.89133773	17.29468405	-7.565159244
4	49.54502552	52.80676033	69.29419823	77.76076974	-1.435243323	-11.17430587
5	55.82478493	58.98341203	74.64978741	81.28247694	-0.273019912	-9.44493766
6	60.07492177	47.83633961	58.12015969	74.72158643	38.38219712	-6.574786097
7	60.00130007	51.50059533	41.73801193	76.98125542	28.46675365	14.35286984
8	68.45197753	67.04337777	70.50519083	85.52584625	10.91025345	1.21276276
9	54.77808522	57.31471298	71.71860355	80.35628109	1.083649909	-8.69184404
10	45.43646525	48.02823889	51.60665417	74.84273686	-0.311794626	-0.034342006

After processing the given data using the above formulas, the data were split into training and testing sets. The training set was utilized for model training and parameter adjustment, while the testing set was employed for evaluating the model's performance. In this context, the preprocessed data serve as the training set, utilized to train the model to achieve the best results, i.e., minimizing color differences. The testing set comprises the original samples, from which the optimal formulations are derived.

3.2. The establishment and solution of the prediction model combining genetic algorithm and deep feedforward neural network (GA-DFNN) are described in this study

By selecting the load prediction results of 403 and 411 lines. We can see that the actual values of the lines basically match the predicted values, but there are also some errors, especially in the peak period of electricity consumption, as shown in table 1.

Training of the neural network on the preprocessed data [8] was conducted to ensure that the model's predictions exhibit a color difference of less than 1 compared to the original samples.

Step 1: Model Establishment and Optimization: A training dataset was constructed using the given concentration data and their corresponding Lab* color data. The concentration data were utilized as inputs (Xtrain), while the normalized color difference data (normalizedcolordiff) were employed as the target outputs (Ytrain). Leveraging the preprocessed data, a feedforward neural network model was constructed using the feedforwardnet function. This function utilizes a vector parameter (layers), where each element denotes the number of nodes in the corresponding layer. In this model, two hidden layers were incorporated, each comprising 10 nodes. The train function was utilized to train the neural network based on the training dataset. This function takes the neural network model (net) and the input-output data (Xtrain, Ytrain) as parameters, as depicted in figure 2.

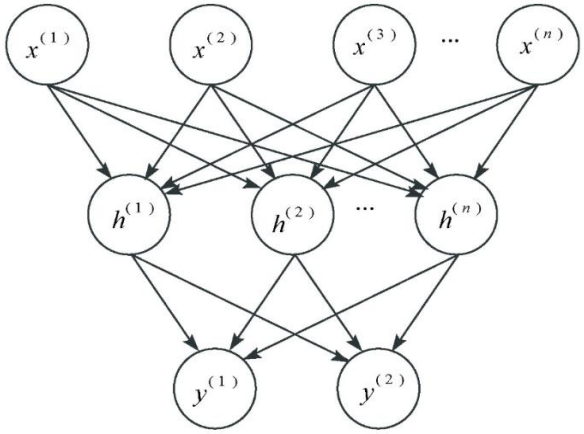

Figure 2. Schematic diagram of feedforward neural network.

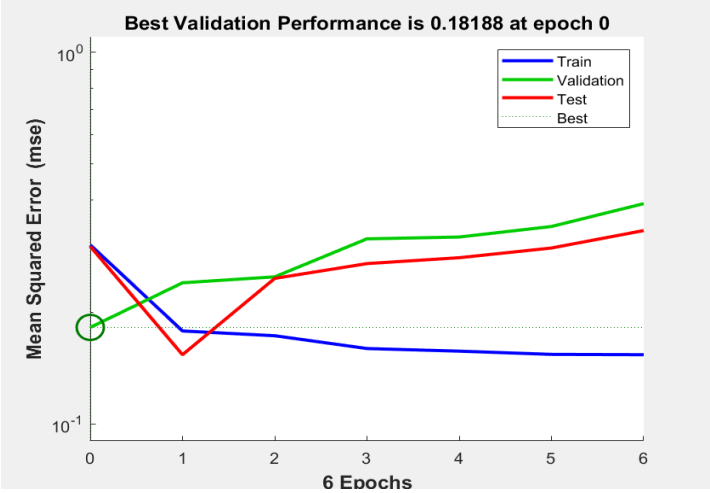

Figure 3. Optimum performance training diagram.

Step 2: Constraint Specification: The color difference between the target and original samples is constrained to be less than or equal to 1.

Step 3 Model Solution: The objective of the code is to optimize the concentrations of three colors (red, yellow, blue) based on concentration data. By training a neural network model, the optimal trained network model is obtained as illustrated in figure 3. The aim is to minimize the color difference for specific target samples represented in the Lab* color space. The optimize Concentrations function is invoked to compute the optimal solution using the pre-trained neural network model [9]. This function likely implements an optimization algorithm by adjusting the concentrations (X, Y, Z) to minimize color disparities. During iterative loops, the code computes multiple optimal solutions and outputs the corresponding optimal concentrations along with the minimum color differences [10]. The performance metrics of the prediction model are depicted in figures 4, 5, and 6.

Calculation of CIELAB Color Difference:

$$\text{Brightness difference:}\Delta L = L_{sp} - L_{std} \tag{18}$$

$$\text{Saturation difference:}\Delta C_s = C_{sp} - C_{std} \tag{19}$$

$$\Delta C_s = \left[a_{sp} + b_{sp}\right]^{\frac{1}{2}} - \left[a_{std} + b_{std}\right]^{\frac{1}{2}} \tag{20}$$

When the difference ΔCs between the tested sample and the standard sample is negative, it indicates that the color of the standard sample is more vivid than the sample. When ΔCs is positive, it indicates that the color of the sample is more vivid than the standard sample.

$$\text{Chrominance difference:}\Delta C_c = (\Delta a^2 + \Delta b^2)^{1/2} \tag{21}$$

$$\text{Colour contrast:}\Delta H = \left(\Delta {C_c}^2 - \Delta {C_s}^2\right)^{1/2} \tag{22}$$

$$\text{Total Chromatic Aberration} \begin{cases} \Delta E = (\Delta L^2 + \Delta a^2 + \Delta b^2)^{1/2} \\ \Delta E = \left(\Delta L^2 + \Delta {C_c}^2\right)^{1/2} \\ \Delta E = \left(\Delta L^2 + \Delta {C_s}^2 + \Delta H^2\right)^{1/2} \end{cases} \tag{23}$$

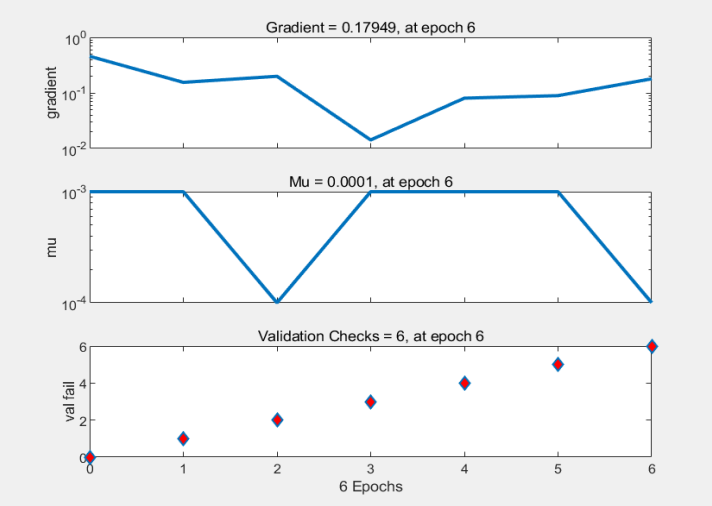

Figure 4. Training index.

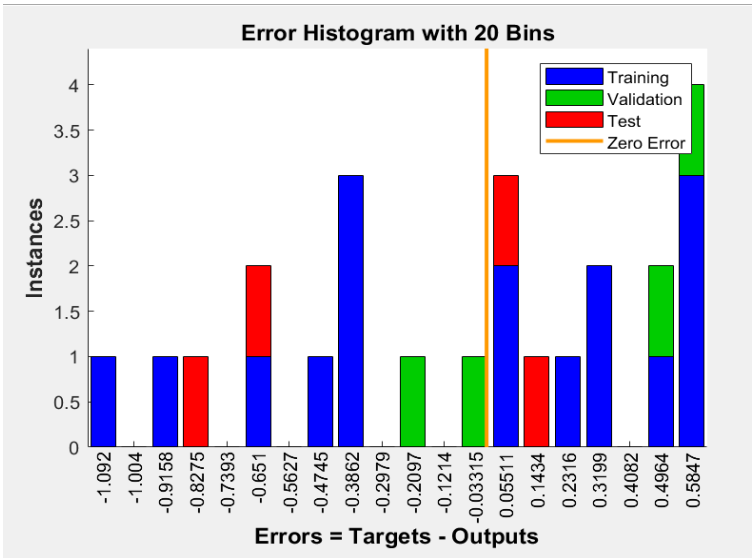

Figure 5. Error analysis diagram of training test.

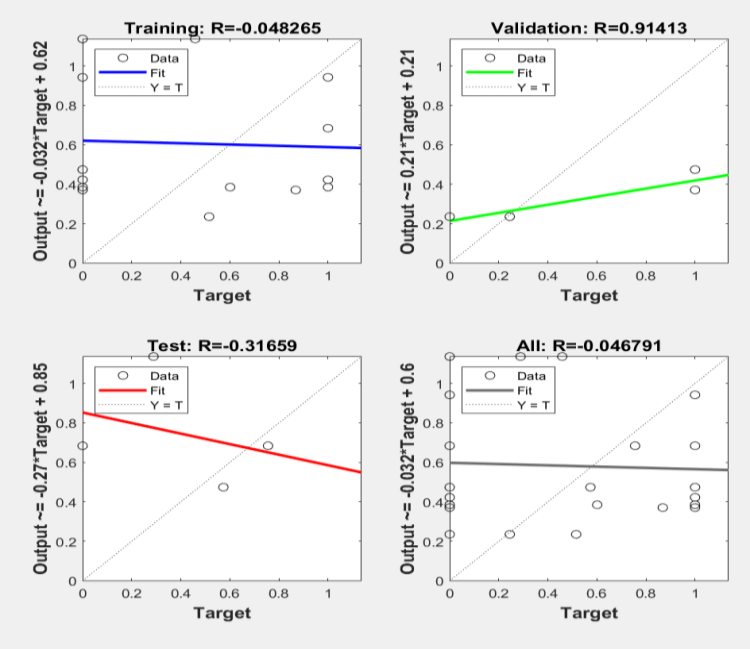

Figure 6. Predicted values for different scenarios.

Using the trained model, the predicted color schemes for five sets of samples are presented in table 3 to table 7.

Table 3. Formulation of sample 1.

Sample	Red	Yellow	Blue	Chromatic Aberration
1	0.5471	0.7506	0.9069	0.2396
2	0.2485	0.5471	0.5337	0.2456
3	0.5471	0.2513	0.7953	0.2212
4	0.7072	0.5471	0.0909	0.2231
5	0.5471	0.7011	0.7458	0.3213
6	0.5471	0.9882	0.2658	0.1222
7	0.7369	0.5471	0.7536	0.1432
8	0.5471	0.854	0.1954	0.4231
9	0.6857	0.5471	0.108	0.3233
10	0.1981	0.0036	0.5471	0.3112

Table 4. Formulation of sample 2.

Sample	Red	Yellow	Blue	Chromatic Aberration
1	0.158	0.9349	0.1767	0.5559
2	0.966	0.5019	0.1583	0.4399
3	0.158	0.3281	0.1527	0.4531
4	0.158	0.1591	0.9799	0.5654
5	0.165	0.6838	0.1583	0.6651
6	0.158	0.5469	0.4536	0.4231
7	0.222	0.1583	0.1321	0.5831
8	0.932	0.9622	0.1583	0.5509
9	0.472	0.4342	0.1583	0.2351
10	0.158	0.2639	0.0453	0.4423

Table 5. Formulation of sample 3.

Sample	Red	Yellow	Blue	Chromatic Aberration
1	0.6586	0.6279	0.9189	0.6754
2	0.8452	0.226	0.6279	0.4569
3	0.6224	0.3427	0.6279	0.7765
4	0.5963	0.1002	0.6279	0.7688
5	0.538	0.6279	0.3038	0.7234
6	0.9007	0.6279	0.5232	0.6547
7	0.6279	0.6545	0.0324	0.7665
8	0.618	0.6279	0.5909	0.6754
9	0.2871	0.3769	0.6279	0.6749
10	0.6342	0.6279	0.9802	0.8544

Table 6. Formulation of sample 4.

Sample	Red	Yellow	Blue	Chromatic Aberration
1	0.012	0.3931	0.3888	0.1318
2	0.408	0.0541	0.3888	0.1098
3	0.389	0.7095	0.151	0.2109
4	0.464	0.3328	0.3888	0.2213
5	0.389	0.7621	0.3174	0.3201
6	0.513	0.5049	0.3888	0.0889
7	0.389	0.4177	0.3889	0.1009
8	0.149	0.1941	0.3888	0.1024
9	0.434	0.3888	0.4318	0.2107
10	0.502	0.3909	0.3888	0.2213

Table 7. Formulation of sample 5.

Sample	Red	Yellow	Blue	Chromatic Aberration
1	1	0.9077	0.6321	0.4763
2	0.0765	1	0	0.6897
3	1	0.6952	0.7941	0.4535
4	0.5454	1	0.5821	0.5465
5	0.1171	0.2409	1	0.4533
6	1	0.2901	0.7125	0.5443
7	0.2193	1	0.1147	0.4544
8	1	0	0.1959	0.5427
9	0.8022	0.2076	1	0.5465
10	1	0.4868	0.3119	0.5464

4. Conclusions

This study aims to enhance the accuracy and efficiency of color matching for opaque products. Initially, logarithmic transformations were applied to K/S values and concentrations, and the resulting equations were linearized. The linear relationship between K/S and concentration was then determined using the least squares method. A comparison between linear and nonlinear regression fitting methods revealed that linear fitting performed exceptionally well within the wavelength range of 400nm to 700nm, exhibiting the highest fitting coefficient. To address the challenge of predicting color schemes with minimal color differences based on target samples, this study unified various theoretical models including the Berlekamp color matching equation, Kubelka-Munk theory, and the Simpson discrete variable integration formula, constructing a training set. Building upon the CIELAB color space model, a deep feedforward neural network was employed to train the L, a, and b values for red, yellow, and blue. Finally, a predictive model combining genetic algorithms and deep feedforward neural networks (GA-DFNN) successfully predicted color matching schemes for 10 sets of samples with color differences not exceeding 1 unit.

In practical computer color matching processes, various factors may introduce interference. The limitations of linear regression lie in its assumption of linear relationships, making it inadequate for handling nonlinear relationships and noise. Therefore, careful analysis of the data characteristics and fitting requirements of practical problems is necessary, along with consideration of whether more complex function forms or fitting methods could yield better results. Future endeavors will focus on improving the accuracy and efficiency of opaque product color matching by exploring advanced data analysis techniques and models, optimizing the relationship between K/S and concentration, and considering additional influencing factors such as environmental and material characteristics to enhance the practicality of color matching schemes. Furthermore, there will be efforts to refine models and methods, expand the scale of the dataset, and enhance predictive and generalization capabilities. Collaborative efforts with industry and academia will be pursued to jointly develop new technologies, address challenges, and provide reliable support for industrial development.

References

[1] How much do you know about the four methods of measuring the color of printed matter? [J]. Screen Printing Industry, 2022(03):54-55.

[2] Gong Xuping, Huang Min, Shen Yuetong, et al. Study on color measurement methods of optical variable ink [J]. Acta Optica, 2023, 43(22): 188-195.

[3] Feng Nuoya. Design and development of computer intelligent color matching system based on machine learning [D]. Zhongyuan Institute of Technology, 2024.

[4] Huang Jingyi. Research on wood dyeing formula prediction method based on improved LSTM [D]. Northeast Forestry University, 2024.

[5] Deng Xiaogang. Research on ink color matching based on neural networks [D]. Xi'an University of Technology, 2022.

[6] Li Tieding, Lei Yuze, Cheng Shuyi, et al. Research on jump calculation method based on linear fitting and interpolation compensation [J]. Sichuan Electric Power Technology, 2020, 43(01): 6-11+15.

[7] Song Haiyu, Li Wei, Wei Wei, et al. Design and implementation of research-based experimental study on least squares method [J]. Experimental Science and Technology, 2023, 21(03): 7-13.

[8] Ren Gang, Li Xin, Liu Xiaojie, et al. Training algorithm for genetic algorithm deep feedforward neural network based on Spark big data computing model [J]. Journal of Henan Institute of Technology, 2023, 31(05): 14-22.

[9] Liu Xinpeng, Qin Yuhua, Zhang Fengmei, et al. Study on near-infrared model transfer method based on sparrow search algorithm combined with deep feedforward neural network[J]. Journal of Analysis and Testing, 2022, 41(11).

[10] Liu Ting. Analysis of computer tristimulus value measurement and color matching method and its application in leather [J]. China Leather, 2023, 52(12): 34-37.