IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2021, accepted August 31, 2021, date of publication October 4, 2021, date of current version October 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3117261

An Industrial Dyeing Recipe Recommendation System for Textile Fabrics Based on Data-Mining and Modular Architecture Design

XIANAN QIN[1], (Member, IEEE), AND XIAOMING JOHN ZHANG[2]

[1]National Engineering Laboratory of Textile Fiber Materials and Processing Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101400, China

Corresponding authors: Xianan Qin (qin@zstu.edu.cn) and Xiaoming John Zhang (zhangxiaoming@bimsa.cn)

ABSTRACT We report a new fabric dyeing recipe recommendation system which is based on mining industrial dyeing manufacturing data and a system design with modular architecture. Unlike traditional dyeing recipe recommendation systems, our method does not rely on labor-intensive calibration works between dye concentrations and the color. Also, the system is generally designed for different dyeing tasks. We describe the framework of our method and discuss strategies that are used for building the system. The system is built in the form of modular architecture which is made up of multiple gradient boosting regression tree models (GBRT). Each GBRT has been trained for predicting dye concentrations of a dye combination set (DCS) for a fabric type. Methods for model training and typical model performance are reported in the paper as well.

INDEX TERMS Data mining, manufacturing industries, textile technology.

I. INTRODUCTION

Dyeing is an essential manufacturing process for the textile industry. In a typical dyeing process for textile fabrics, a solution containing multiple dyes is applied to the fabric. After several processes with careful control of temperature and aqueous environment (eg. pH, salt concentration, etc.), dyes in the dyebath are absorbed by the textile fiber, and permanent color finally stays on the fabrics [1], [2].

The ultimate goal of the dyeing process design is to make fabrics have desired color. To decrease the risk of causing large chromatic inaccuracy on the dyed textile materials, careful dyeing recipe prediction must be performed. In 1966, E. Allen proposed the first computer-aided color matching algorithm for fabric dyeing [3]. This algorithm relies on the famous Kubelka-Munk theory [4], in which the calibration between dye concentrations and the colorimetric information is a necessity [5], [6]. The processes of obtaining the calibration are usually time-consuming and are prone to involve errors and omissions. In these processes, each individual dye must be made into solutions with multiple levels of concentrations. These standard dye solutions will be applied into separated dyeing processes for dozens of fabric samples, each of which will be measured for colorimetric information to get a calibration between the dye concentration and the color. Complicating calculations based on such calibration information must be conducted in order to screen out proper dye combination sets (DCS) with correct dye concentrations. In addition, careful correction to the calibration (for example, the Saunderson correction) must be made to ensure the accuracy in the recipe recommendation [7].

Recently, many efforts have been made towards developing better recipe recommendation methods for the dyeing-printing industry using modern information techniques [8], [9]. These existing works have made important contributions for increasing the accuracy in the recipe prediction. However, for conventional manufacturing tasks, the accuracy is not the only central research niche. In fact, colors or dye concentrations become not sensitive when the inaccuracies are lowered down to some degree. The laborious situation during the manufacturing processes, for example, the acquisition of the standard calibration between the dye concentration and the colorimetric information for the fabric dyeing, is also an important problem that should be taken into consideration. There are also some existing works, eg. ref[10], building mathematical relations between dye concentrations and the color and not relying on the calibration works between dye concentrations and the color. However, these

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif.

TABLE 1. Pros(or highlights) and Cons of typical existing dyeing recipe recommendation methods and the proposed method in this paper.

	Pros (or highlights)	Cons
E. Allen [3]	Iteratively approaching an exact match of dye concentrations to any desired degree of accuracy, using a linear solution that is based on a matrix inversion technique.	Calibrations between dye concentrations and the color must be performed.
Chaouch *et al* [8]	(1) Using of genetic algorithm. (2) Minimizing the CMC color difference.	
Kandi *et al* [9]	(1) Using of genetic algorithm. (2) Decreasing the color difference under second illuminant.	
Zhang *et al* [10]	(1)Calibrations between dye concentrations and the color are not needed. (2) Building mathematical relations between the colorimetric information (RGB) and dye concentrations for unknown types of dyes and fabrics.	(1) Using a single regression model to recommend recipes for different fabrics/dyes. (2) Relying on historical dyeing data.
Our method	(1) Calibrations between dye concentrations and the color are not needed. (2) Building a modular system in which different fabrics vs. different dye combination sets are built with individual regression models.	Relying on historical dyeing data.

existing works mostly use a single regression model to recommend recipes for multiple types of dyes and fabrics. This may lead to practicality problems of the method, because dyes can have very different abilities to get absorbed on different fabrics. Table 1 summarizes the ''Pros & Cons'' of some typical existing dyeing recipe recommendation methods.

In this paper, we report a new recipe recommendation system for fabric dyeing which is based on mining industrial dyeing manufacturing data and a modular architecture system design. Our system does not rely on the calibration works between dyes and the color. In addition, the system is generally design for different types of dyeing tasks. We consider the real application scenarios that happen in most dyeing-printing companies in nowadays. What the dyeing-printing companies really want is a practical system that can help pick out proper dye combination sets for the fabric types that they daily work on and give accurate dye concentrations. Another fact is that most of these companies have records of the historical manufacturing data including the types (or material components) of fabrics, the dye catalogs, the dye concentrations and the color measurement results of the final fabric products. Due to the lack of professional data scientific training, these data are mostly not well organized and not well utilized by the companies. These patchy data contain valuable information in terms of the quantitative relationship between the color of dyed fabrics and the concentrations that are used in the dyeing. This motivates us to use the data mining approach in the development of dyeing recipe recommendation systems.

Our method uses the type of fabrics and the target colorimetric information (such as full K/S spectra and RGB values) [4] of a DCS as inputs, and outputs predicted dye concentrations. The dyeing recipe recommendation system is built under a modular architecture design, in which individual modules are made up of multiple regression models, and each of the models is built by training for a specific type of fabric under a single dye combination set. In the rest parts of this paper, we will first describe the overall framework of our dyeing recipe recommendation system. Since all regression models under this framework are based a same architecture (we use the gradient boosting regression tree model (GBRT) for regression because of its relative insensitivity to hyperparameters and great performance) [13], we will not cover the details for all of them. As an example, we will describe the model building for one of the regression models which is for the rayon fabric under the dye combination set of Colvaceton reactive dye-navy blue CF (CRD-navy blue), Colvaceton reactive dye-bright red 3BSN150% (CRD-red) and Colvaceton reactive dye-yellow 3RS150% (CRD-yellow). The settings of hyperparameters and the numerical evaluation results will be reported for this example model. In fact, the rest of regression models are built based on similar strategies and achieve similar performance.

Our work is a successful case of applying modern data mining and artificial intelligence (AI) techniques to solve practical problems for industry [11], [12]. Using our method, the labor-intensive works of acquiring standard calibrations between dye concentrations and the color of fabrics will be no longer needed. In addition, thanks to the modular architecture design, our recipe recommendation system can be extended to different types of fabrics and dye combination sets, making it have great practicality for real dyeing manufacturing tasks. Our method can be applied to practical dyeing manufacturing tasks to find correct dyeing recipes within one or few shots, which can help the dyeing-printing companies decrease the economic costs and lessen the pollution producing.

II. THE OVERALL FRAMEWORK OF THE DYEING RECIPE RECOMMENDATION SYSTEM

We used a modular architecture design for our dyeing recipe recommendation system (Figure 1). The system was made up of multiple modules, and each module was corresponded to a type of fabric. In each fabric type module, there were multiple dye combination sets (DCS) with which the company had the capability to conduct the dyeing processes. For each pair of fabric type vs. DCS, there was a regression model that was trained for predicting dye concentrations using colorimetric information as inputs. The regression models were trained using gradient boosting regression tree (GBRT). We used GBRT for the regression tasks because of its relative insensitivity to hyperparameters and great performance [13], [14].

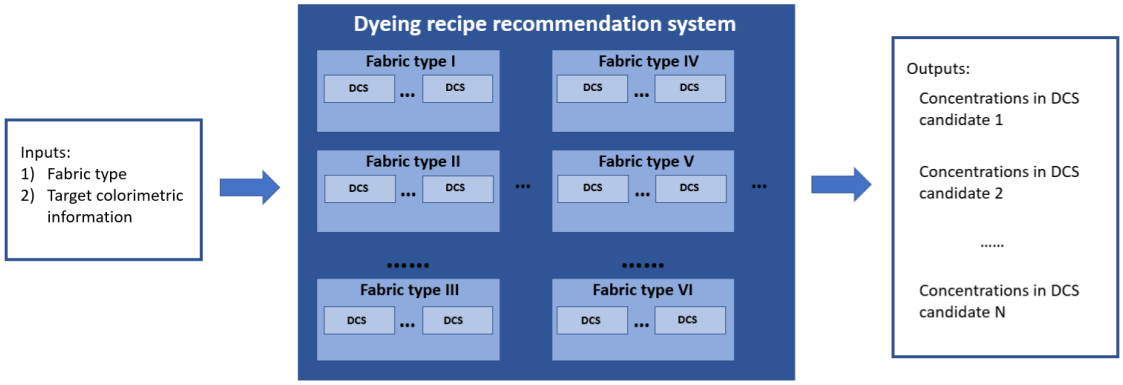


FIGURE 1. The overall framework of the dyeing recipe recommendation system.

III. BUILDING REGRESSION MODELS FOR DYE COMBINATION SETS (DCS)

A. DATA COLLECTION

All datasets used to train regression models were from the dyeing industry. The data was kindly provided by Shaoxing Xingming dyeing & printing Co., Ltd. at Zhejiang Province of China. To build the regression model of rayon fabrics dyed under the DCS of CRD-navy blue, CRD-red and CRD-yellow, we used 810 records of historical dyeing manufacturing data that had been performed data pre-processing to remove data records with outliers or missing values. The colorimetric information of the fabrics was obtained using commercially available spectrophotometers (Datacolor, USA) under D65 light source. The concentrations of dyes were recorded in the unit of o.w.f which stands for ''on weight of fabric''.

B. THE NON-LINEARITY BETWEEN DYE CONCENTRATIONS AND THE COLORIMETRIC INFORMATION

The original colorimetric data given by the spectrophotometers was in the form of K/S spectra [4]. To better understand the color measurement data, we first transformed the spectra into different color spaces, such as RGB space and CIE-Lab space [15]. Then, we plotted the colorimetric information and dye concentrations on 2D graphs to gain insights into the model building tasks. We observed significant non-linearity between the colorimetric information and dye concentrations. In Figure 2 and Figure 3, we show the non-linearity between dye concentrations and the colorimetric information for rayon fabrics. These rayon fabrics were dyed using Colvaceton reactive dye-navy blue CF (CRD-navy blue), Colvaceton reactive dye-bright red 3BSN150% (CRD-red) and Colvaceton reactive dye-yellow 3RS150% (CRD-yellow) as dye combination set (DCS). Though non-linear regression models such as ensemble-based decision trees and neural networks are capable of modeling almost any types of non-linear relations, it may take too long calculation time to find the optima. To increase the linearity, we next took logarithm to dye concentrations as well as the colorimetric information, and then used the logarithmic data in later model training. When CIE-Lab values were used for model training, the colorimetric values were added with an absolute value of 500 to guarantee that the logarithms can be taken. The concentrations were recovered back into normal scale in the final outputs.

C. THE REGRESSION MODELS

In the regression models, three types of colorimetric information were used as inputs in these GBRTs: full spectra (400-700nm, every 10nm for one record), RGB values and CIE-Lab values. The model parameters varied for different DCSs on different fabric types. However, these regression models shared similar strategies and settings in the model training. All models were trained using the ''GradientBoostingRegressor'' package of the Scikit-learn [16]. The loss function used in model training was least-square. Lists of following hyperparameters were traversed when the models were trained: learning rate, the maximum depth for the regression trees (max_depth), the fraction of finding the feature numbers for the best split (max_features), the minimum number of samples that is required to be at a leaf node (min_samples_leaf), the number of estimators (n_estimators), and the fraction of samples that are used for fitting individual base learners (subsample).

Here, we report a typical setting for hyperparameters from one of the GBRT models, which was built for rayon fabrics under CRD-navy blue, CRD-red and CRD-yellow as DCS. Table 2 summarizes the hyperparameters that we used for this model. By traversing lists of model parameters, we found that learning rates at $\sim$0.05, max_depth at 5-7, max_features at 0.1-0.3, min_samples_leaf at 1-3, n_estimators at 500-700 and subsample at $\sim$0.5 could make individual GBRTs have good performance.

Before the model training, datasets were first divided into training sets and testing sets under a ratio of 60%-40%. The models were then trained using the training sets in which 5-fold cross-validation was applied. The models were next tested using the testing sets. Three types of error were calculated to evaluate the model performance: mean absolute error

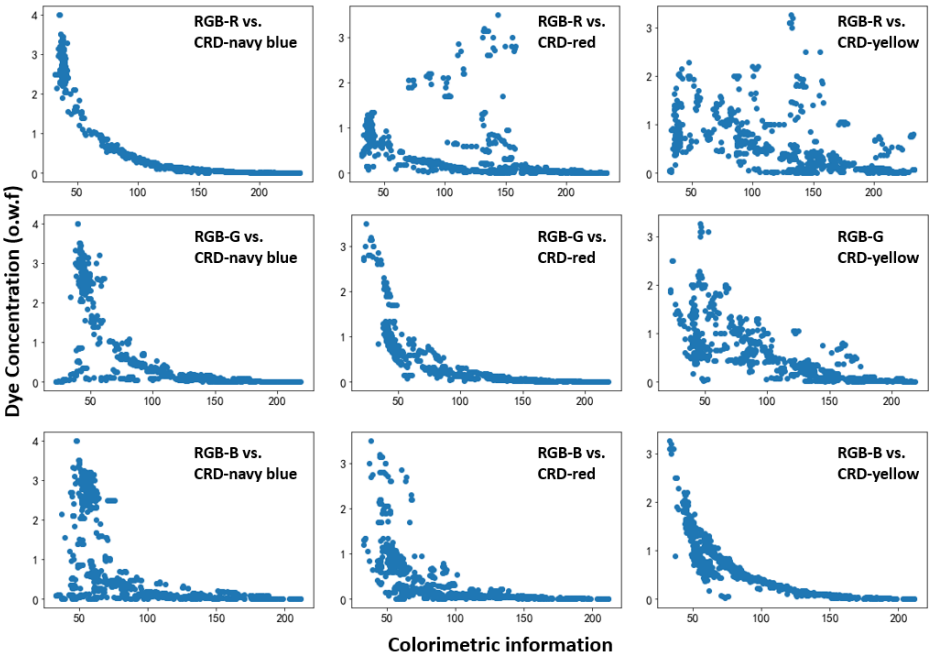


FIGURE 2. Plots of dye concentrations of CRD-navy blue, CRD-red and CRD-yellow vs. RGB values on rayon fabrics.

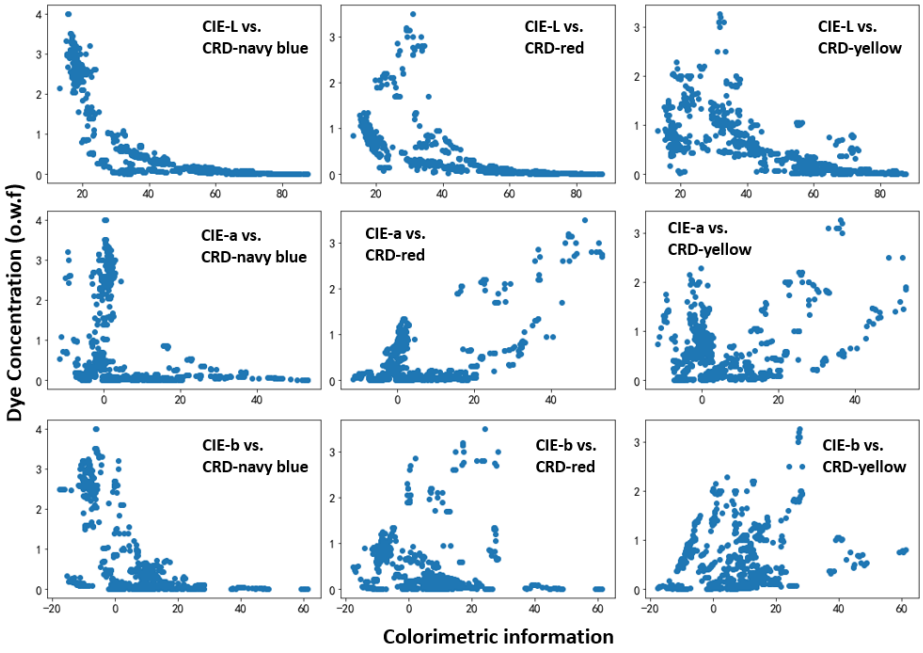


FIGURE 3. Plots of dye concentrations of CRD-navy blue, CRD-red and CRD-yellow vs. CIE-Lab values on rayon fabrics.

(MAE), mean absolute percent error (MAPE) and weighted absolute percent error (WAPE) [17], [18].The definitions of MAE, MAPE and WAPE are shown as below:

$$MAE = \frac{1}{N}\sum_{i}\left|c_{pred,i} - c_{target,i}\right|$$

$$MAPE(\%) = \frac{100}{N}\sum_{i}\left|\frac{c_{pred,i} - c_{target,i}}{c_{target,i}}\right|$$

$$WAPE(\%) = 100\frac{\sum_{i}\left|c_{pred,i} - c_{target,i}\right|}{\sum_{i} c_{target,i}}$$

where N is the total number of recipes in the testing set, "c_{pred}" denotes the "predicted concentrations" and "c_{target}" denotes the "target concentrations".

We could observe ∼10% of MAPE and WAPE for the GBRTs trained in our system. As an example, we report the errors of the GBRT for rayon fabrics, which were dyed

TABLE 2. Hyperparameters used in the gradient boosting regression tree models for the rayon fabric using CRD-Navy blue, CRD-red and CRD-yellow as dyeing combination set.

Colorimetric info	Dye	max_depth	max_features	min_samples_leaf	n_estimators	subsample
Full spectra	CRD-Navy blue	5	0.3	1	700	0.5
	CRD-red	7	0.3	3	700	0.5
	CRD-yellow	5	0.3	1	700	0.3
CIE-Lab	CRD-Navy blue	5	0.1	1	700	0.5
	CRD-red	7	0.1	1	700	0.5
	CRD-yellow	7	0.1	1	700	0.3
RGB	CRD-Navy blue	7	0.1	3	700	0.5
	CRD-red	7	0.1	3	700	0.5
	CRD-yellow	7	0.1	3	700	0.5

TABLE 3. Errors in the dyeing recipe prediction for the rayon fabric using CRD-Navy blue, CRD-red and CRD-yellow as dyeing combination set.

Colorimetric info	Dye	MAE	MAPE	WAPE
Full spectra	CRD-Navy blue	0.054	8.6%	7.6%
	CRD-red	0.027	9.3%	7.0%
	CRD-yellow	0.039	8.5%	7.5%
CIE-Lab	CRD-Navy blue	0.067	12.8%	9.3%
	CRD-red	0.033	11.3%	8.6%
	CRD-yellow	0.053	11.8%	9.6%
RGB	CRD-Navy blue	0.072	11.1%	10.2%
	CRD-red	0.041	14.8%	10.8%
	CRD-yellow	0.051	10.5%	9.1%

under CRD-navy blue, CRD-red and CRD-yellow as DCS (summarized in Table 3). Interestingly, the GBRT models using spectra as inputs seemed to have the best performance comparing to models using other two types of colorimetric information (RGB and CIE-Lab) as inputs. MAPE and WAPE as good as ∼7-8% could be achieved using spectra as inputs. These two types of errors could increase to ∼10% when RGB or CIE-Lab were used as inputs (Table 3). In fact, RGB and CIE-Lab values can be viewed as the dimension reduced versions of spectra data. The superiority of using spectra might be because spectra contain more useful and complete colorimetric information than RGB and CIE-Lab.

IV. DISCUSSION AND CONCLUSION

In this paper, we have reported a new recipe recommendation system for industrial fabric dyeing. Our system is based on mining industrial dyeing and color measurement data, and is developed under a modular architecture system design. In the system, each single module contains multiple gradient boosting regression tree (GBRT) models for a fabric type versus different dye combination sets (DCS). These GBRTs can achieve ∼7-10% as prediction errors for predicting dye concentrations. The proposed method in this paper has two major advantages. First, calibrations for dye concentrations and the colorimetric information are no longer required in our system. Second, thanks to the using of the modular architecture system design, the system can be easily extended for new fabric types and DCSs by added new dyeing manufacturing data. Namely,our method is generally designed for different types of dyeing tasks.

In addition to the using of GBRT, we also tried using neural networks (NN) to build regression models between dye concentrations and the colorimetric information. To compare the model performance of GBRT and NN, we built a NN architecture of which the number of hidden layer was set as one (a typical artificial neural network, ANN), and the number of nodes in the hidden layer was set as 64. Full-spectra colorimetric information was used as inputs. In the NN model, mean absolute error (MAE) was used as the loss function, ''rmsprop'' was used as the optimizer, ''relu'' was used as the activation function, 4-fold cross validation was applied in the model training, and the number of epochs was set as 200 (where the loss could reach the seeming convergence). We could achieve MAPE at ∼20% (27.1% for CRD-Navy blue, 16.1% for CRD-Red and 25.7% for CRD-yellow), WAPE at ∼10% (13.0% for CRD-Navy blue, 8.8% for CRD-Red and 11.2% for CRD-yellow) and MAE at ∼0.06 (0.093 for CRD-Navy blue, 0.034 for CRD-Red and 0.063 for CRD-yellow). It can be seen from the prediction errors that GBRT models can achieve better performances than NN. However, it should be noted that the NN model used in this case was without careful designing of model architecture. Better performance is expected to have if the model architecture of NN can be finely tuned.

For future study, since our current dyeing recipe recommendation system only provides dyeing recipes but no more additional information, more user-friendly functions, such as automatic calculations for economic costs based on the prices of dyes, electricity and waste treatment may be integrated into the system to help the dyeing company better make management decisions. Also, since our system relies on industrial manufacturing data, a convenient dyeing & color measurement data collection system may be added into the system to help update the datasets and the models. Finally, our system could be integrated into enterprise resource planning (ERP) systems [19] that are used by the dyeing companies for better daily management and decision making.

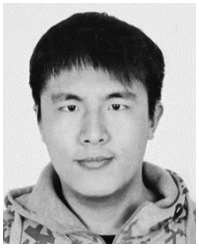
ACKNOWLEDGMENT

The authors would like to thank Dr. Minghua Wu for very helpful discussion regarding the design of the system framework, Yaguang Wang for performing data cleaning for the industrial dataset used in this work, Zhiwen Tu for regression model training, and Yawen Yin for the proof-reading.

REFERENCES

[1] M. A. H. Yasassri, W. D. C. Weerasinghe, and S. V. Udayakumara, ''Dyeing of cotton fabric with a natural dye extracted from floral part of acalypha hispida plant,'' in *Proc. Moratuwa Eng. Res. Conf. (MERCon)*, Moratuwa, Sri Lanka, Jul. 2019, pp. 90–94.

[2] M. I. bin Ab Kadir, W. Y. Wan Ahmad, M. R. Ahmad, M. I. Misnon, and H. A. Jabbar, "Fastness properties and colorimetric characteristics of low temperature dyeing of natural dyes from the barks of ixonanthes icosandra Jack on polyester fabric," in *Proc. IEEE Bus. Eng. Ind. Appl. Colloq. (BEIAC)*, Langkawi, Malaysia, Apr. 2013, pp. 448–452.
[3] E. Allen, "Basic equations used in computer color matching," *J. Opt. Soc. Amer.*, vol. 56, no. 9, pp. 1256–1259, 1966.
[4] P. Kubelka and F. Munk, "An article on optics of paint layers," *Z. Tech. Phys*, vol. 12, pp. 259–274, Aug. 1931.
[5] H. R. Davidson and M. Taylor, "Prediction of the color of fiber blends," *J. Opt. Soc. Amer.*, vol. 55, no. 1, pp. 96–100, 1965.
[6] H. R. Davidson, H. Hemmendinger, and J. L. R. Landry, "A system of instrumental colour control for the textile industry," *J. Soc. Dyers Colourists*, vol. 79, no. 12, pp. 577–590, Oct. 2008.
[7] J. Saunderson, "Calculation of the color of pigmented plastics," *J. Opt. Soc. Amer.*, vol. 32, no. 12, pp. 727–736, 1942.
[8] S. Chaouch, A. Moussa, I. Ben Marzoug, and N. Ladhari, "Application of genetic algorithm to color recipe formulation using reactive and direct dyestuffs mixtures," *Color Res. Appl.*, vol. 45, no. 5, pp. 896–910, Oct. 2020.
[9] S. Gorji Kandi and M. Amani Tehran, "Color recipe prediction by genetic algorithm," *Dyes Pigments*, vol. 74, no. 3, pp. 677–683, Jan. 2007.
[10] B. Zhang, Y. Shi, H. Yang, X. Liu, and A. Zhang, "Research on application of the minimum error average fitting method in computer color matching," in *Proc. IEEE Int. Conf. Intell. Secur. Informat.*, Beijing, China, Jul. 2011, pp. 293–296.
[11] M. Keen, "Successful applications of AI in manufacturing industry," in *Proc. IEE Colloq. Ind. Appl. AI (Artif. Intell.)*, London, U.K., Jan. 1992, pp. 5-1–5-4.
[12] Y. Zhang, W. Huang, and Q. Zhou, "A data-mining approach for input-output table: A case study of mechanical equipment manufacturing industry in China," in *Proc. 2nd Int. Symp. Knowl. Acquisition Modeling*, Wuhan, China, 2009, pp. 49–52.
[13] Z.-H. Zhou and J. Feng, "Deep forest," *Nat. Sci. Rev.*, vol. 6, no. 1, pp. 74–86, 2019.
[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
[15] M. Tkalcic and J. F. Tasic, "Colour spaces: Perceptual, historical and applicational background," in *Proc. IEEE Region Comput. Tool (EUROCON)*, Ljubljana, Slovenia, vol. 1, 2003, pp. 304–308.
[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[17] P. F. Pai and C. H. Liu, "Predicting vehicle sales by sentiment analysis of Twitter data and stock market values," *IEEE Access*, vol. 6, pp. 57655–57662, 2018.
[18] S. Meng, R. Pan, W. Gao, J. Zhou, J. Wang, and W. He, "Woven fabric density measurement by using multi-scale convolutional neural networks," *IEEE Access*, vol. 7, pp. 75810–75821, 2019.
[19] M. A. Rothenberger and M. Srite, "An investigation of customization in ERP system implementations," *IEEE Trans. Eng. Manage.*, vol. 56, no. 4, pp. 663–676, Nov. 2009.



XIANAN QIN (Member, IEEE) received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, in 2015, and the Ph.D. degree in physics from The Hong Kong University of Science and Technology (HKUST), Hong Kong SAR, China, in 2019. Since 2020, he starts working with Zhejiang Sci-Tech University, Hangzhou.

XIAOMING JOHN ZHANG received the B.S. degree in computational fluid mechanics from Zhejiang University, the M.S. degree in computational fluid mechanics from Beijing University, and the Ph.D. degree in computational fluid mechanics from Massachusetts Institute of Technology (MIT). From 2017 to 2020, he was the Director of the Big Data Center for Textile Industry, Zhejiang China Light & Textile Industrial City Group Company. He is currently a Senior Research Scientist with Beijing Institute of Mathematical Sciences and Applications (BIMSA).