iJournals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 8 Issue 12 December 2020

STUDYING OF THE FERMENTATION PROCESS IN BLACK TEA PRODUCTION TECHNOLOGY APPLIED BY IMAGE PROCESSING METHOD

Le Thi Thuy Ngan, Ta Minh Long, Tran Thi Thanh Huyen*

Thai Nguyen University of Technology, Thai Nguyen city, Viet Nam

* ***Email:*** *thanhhuyenspkt@tnut.edu.vn*

ABSTRACT

In order to maintain the quality, flavor and color of tea, fermentation is an essential and decisive process for the production of delicious black tea. To solve the problem of fermentation in black tea production technology, applying image processing to implement this process is one of the optimal options. Based on the optimization of the nonlinear regression model using the RF algorithm (Random Forests), we will analyze the image of tea leaves, thereby controlling the temperature, gas flow, conveyor speed to ensure that quality indicators at times of meeting the standards. The control system was simulated by MATLAB SIMULINK and MINITAB software.

Keywords: Black tea; Fermentation; Image processing; nonlinear regression model; Random Forest.

1. INTRODUCTION

Black tea is one of the most popular functional beverages in the world. After the plucked tea leaves are treated by series of processes called withering (removal of moisture by air flow), pre-conditioning and CTC (essentially maceration and cutting of leaves), the leaves are subjected to the process of fermentation by exposing them to air by laying the cut tea leaves on floor, trough or moving conveyor under controlled temperature, humidity and air-flow conditions. During this process, the leaves change color from green to coppery brown and the grassy smell gets transformed into a floral smell. It is critical that the leaves be allowed to ferment only up to the desired limit and both under and over fermentation result in the deteriorated quality of black tea. Out of the two detectable parameters (colour and smell), the smell is significant since a strong, particular fragrance emanates from the leaves once leaves are optimally fermented [1-5]. Tea color and luster is formed when the above pigments are dissolved in water; This change in color can be observed and distinguished by the human visual system, but it isn't easy to define a specific scale.

2. BLACK TEA FERMENTATION MACHINE MODEL

Black tea can be fermented by an interrupted method and is fermented continuously on a conveyor belt, in this section I only present the continuous fermentation method. After the cells are broken and shaped, Tea leaves are solved evenly on the conveyor belt with a thickness of 15 to 20cm. The conveyor belt moves continuously with the right factor; fermentation time from 2 hours to 4 hours [9-12] is enough to complete black tea quality. The right temperature and humidity factors. The fermentation equipment's structure principle is as follows: The fermentation conveyor is composed of many stainless steel blisters and assembled—movements by chain and track system. The continuous fermentation equipment has many advantages, which are complete mechanization of the fermentation process, high yield, and stable tea quality.

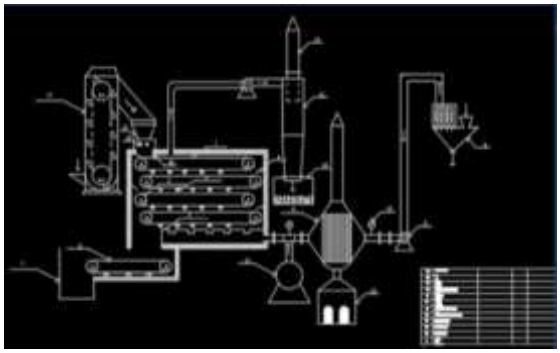

Figure 1: Practical continuous black tea fermentation machine

To measure the tea pigmentation of tea leaves, the author relied on the models of authors Gaozhen Liang and Jiangtao [13] to develop research. Below is the model:

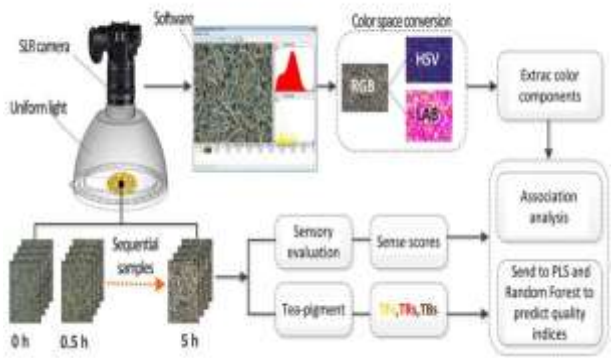

Figure 2: Model for measuring the colors of tea samples

The fermentation cycle is 300 minutes; 20 samples will be taken every 30 minutes; These 20 samples will be taken from different random locations of the fermented tea block; In total, 220 samples will be collected. The content of TF and TR was measured according to Tea leaf pigment measurement - High-performance liquid chromatography (GB / T30483-2013). Samples were lyophilized and finely ground, and a high-performance liquid chromatography (PDGU-20A3, Shimadzu Corporation, Japan) was used for the measurement. Finally, each tea sample's sensory quality was assessed using a code evaluation method based on an official assessment of tea leaves [14-15].

3. CONTROL ALGORITHMS

RF is used quite commonly because of its advantages compared to other algorithms: it can handle data with a large number of attributes, be able to estimate the importance of attributes, often high accuracy in classification (or regression), fast learning process [14].

The RF accuracy depends on the predictive quality of decision trees and the degree of correlation between decision trees. Given a training data set (sample set) containing N data samples, p attribute Xj (j = 1,2, ..., p) and Y∈ {1, 2, .., C} with C ≥ 2 is the dependent variable. RF uses the Gini index to measure sample set mix. During the construction of decision trees, the RF develops child nodes from a parent node based on the Gini index evaluation of a subspace mtry of randomly selected properties from the original property space. The property selected to separate node t is the property that minimizes the sample sets' confusion after division. The formula for calculating the Gini index for a node t is as follows:

$$Gini(t) = \sum_{c=1}^{C} \Phi_c(t)[1- \Phi_c(t)]$$

where Φc (t) is the frequency of class c∈ C in node t. Let s be a value in attribute Xj separating node t into two children: left node tL and right node tR depending on Xj ≤ s or Xj> s; tL = {Xj ϵ t, Xj ≤ s} and tR = {Xj ϵ t, Xj> s}. Then, the total measure of the Gini index of the two nodes tL and tR after using the property Xj to separate node t at s is:

$$\Delta Gini(s,t) = p(t_L)Gini(t_L)+ p(t_R)Gini(t_R)$$

To obtain a good division point, at each RF node, it will find all the mtry variables' possible values to find the point s with the smallest iniGini (s, t) as the node separation point t. The property that contains the t-node separator property is called the t-node split property. Calling ISk (Xj), ISXj is the measure of the attribute Xj's importance in a decision tree Tk (k = 1… K) and in a random forest. The formula for calculating ISk (Xj) and ISXj is as follows:

$$IS_k(X_j) = \sum_{t\in k} \Delta Gini(X_j,t)$$

$$IS_{Xj} = \frac{1}{K}\sum_{k=1}^{K} IS_k(X_j)$$

The min-max normalization is used to convert the attribute importance measure to the segment [0, 1], following the formula:

$$VI_{Xj} = \frac{IS_{Xj} - \min_{j=1}^{M} (IS_{Xj})}{\max_{j=1}^{M} (IS_{Xj}) - \min_{j=1}^{M} (IS_{Xj})}$$

4. SIMULATION AND DISCUSSION

- *Pre-process data and sample division*

20 samples (R: representative) for 11 points in the fermentation:

Min	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
1	118	106	96	89	86	83	81	78	77	83	79
2	121	103	96	85	84	86	86	84	83	84	81
3	119	101	99	78	85	90	89	80	82	86	87
4	119	99	97	86	92	88	78	86	80	85	80
5	122	103	98	94	86	84	86	82	83	75	80
6	116	101	97	89	81	88	79	81	75	79	78
7	125	100	96	89	84	82	85	80	80	79	84
8	118	102	101	83	84	83	83	79	80	83	79
9	123	102	99	88	88	87	82	80	82	84	74
10	123	100	101	82	86	89	82	75	77	81	85
11	116	99	96	88	82	89	90	84	81	79	77
12	115	99	94	89	86	91	80	82	85	84	79
13	123	101	95	87	86	90	85	78	76	80	83
14	123	98	94	91	82	92	83	80	85	82	84
15	117	102	96	80	87	89	84	77	82	80	79
16	117	96	97	87	85	87	84	72	77	82	82
17	119	104	98	87	85	87	83	80	80	85	78
18	115	101	98	93	85	83	85	80	81	78	82
19	119	101	98	86	89	88	81	82	79	85	84
20	117	100	101	91	88	91	85	76	82	84	83

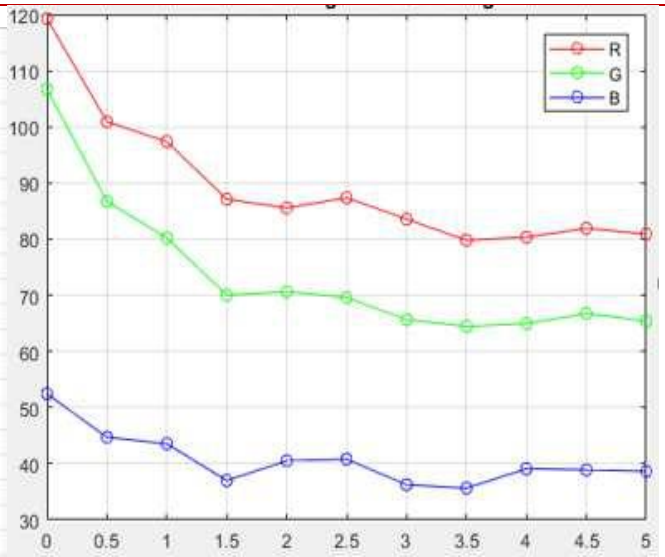

Figure 3: Datasheet and graph showing the change in average RGB value over time

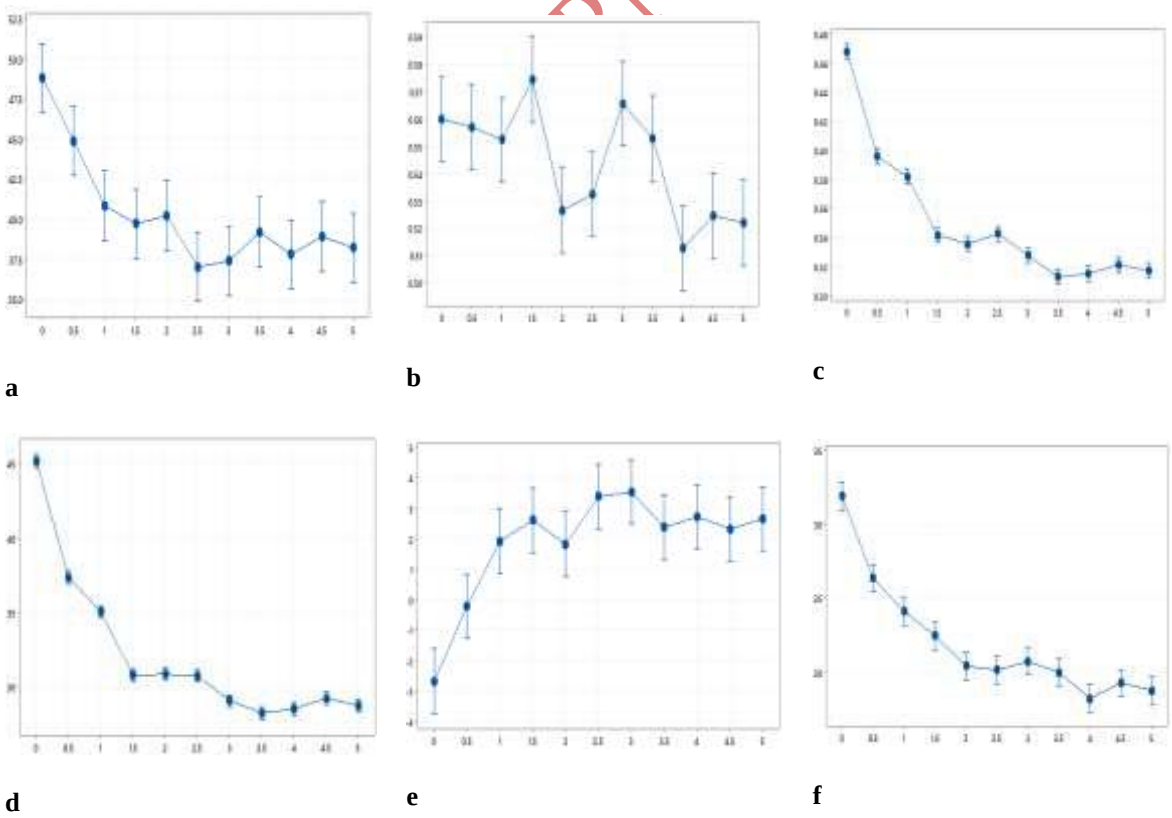

Figure 4: Graph showing the change in value a, H; b, S; c, V; d, L; e, a; f, b averaged over time

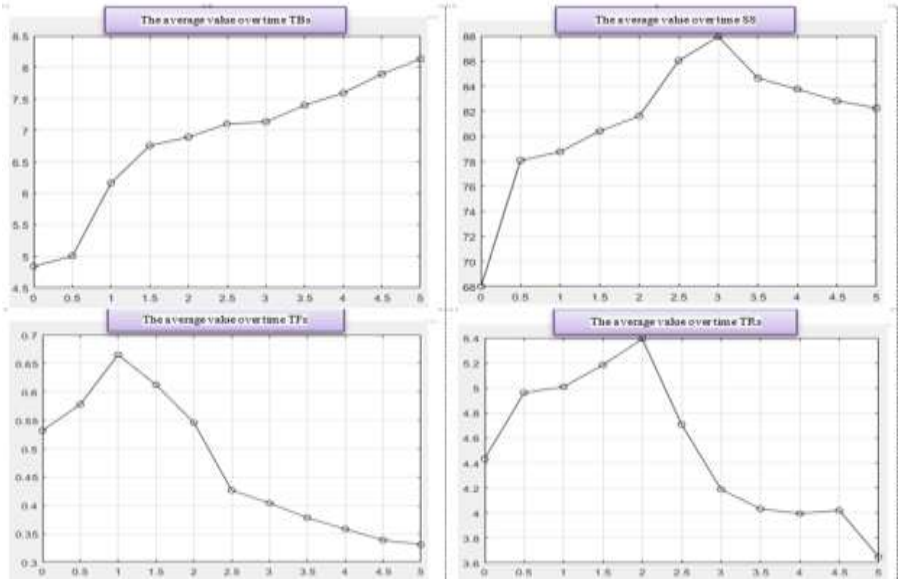

Figure 5: Graph showing the change in the value of TFs quality index; TRs; TBs; SS (sensory score) over time

As shown in Figure 5, all quality indicators observe the rule of change "increase - decrease", and the sensory quality reaches the highest score at 3 o'clock. TFs increased rapidly on fermentation and peaked at 1 hour. After that, it dropped a lot and slowed down after 2.5 hours. TRs increased with fermentation and peaked at 2 o'clock, after which it decreased dramatically. TB increased continuously during the entire fermentation.

Analyze the difference in quality index and color characteristics

Single-factor ANOVA (Analysis of variance) was conducted based on the quality indexes (TFs, TRs, TB and Sensory Score) and visual characterization values during each fermentation stage. The results are shown in Table 1

Table 1. The results of the quality indexes (TFs, TRs, TB ,and Sensory Score)

Variables	Mean Square		F	Sig.
	Between groups	In groups		
R	2895.08	8.81	328.54	<0.001
G	3368.32	8.6	391.55	
B	462.461	8.367	55.27	
H	252.07	23.82	10.58	
S	0.00846	0.001242	6.81	
V	0.044523	0.000136	328.54	
L	537.280	0.867	619.64	
a	64.509	5.727	11.26	
b	327.754	4.256	77.01	
TFs	0.286482	0.000653	438.56	
TRs	6.61264	0.04115	160.69	
TBs	23.3092	0.0914	254.95	
Sensory Score	562.163	4.482	125.42	

Correlation between color characteristics and quality index

Table 2. Analysis of correlation of color features and quality index

	R	G	B	H	S	V	L	a	b_1
G	0.939								
B	0.797	0.786							
H	0.409	0.684	0.281						
S	0.288	0.204	-0.345	0.174					
V	1.000	0.939	0.797	0.409	0.288				
L	0.967	0.996	0.804	0.617	0.220	0.967			
a	-0.397	-0.686	-0.345	-0.984	-0.051	-0.397	-0.615		
b_1	0.859	0.904	0.464	0.701	0.591	0.859	0.899	-0.650	
TFs	0.564	0.523	0.398	0.284	0.257	0.564	0.546	-0.226	0.524
TRs	0.314	0.282	0.213	0.156	0.167	0.314	0.300	-0.106	0.293
TBs	-0.856	-0.853	-0.676	-0.498	-0.262	-0.856	-0.866	0.477	-0.796
SS	-0.807	-0.822	-0.726	-0.460	-0.097	-0.807	-0.827	0.476	-0.700

2

Correlations between quality indices (sensory score and pigment composition) and color-specific variables of the test sample were analyzed as shown in Table 2. The results showed that all quality indexes were significantly correlated with color characteristics ($p < 0.01$), especially with parameters a *, b * and L * in the CIE Lab color model.

Divide training and testing (train/test)

Three algorithms are Random, Kennard-Stone, Spxy with Euclidean distances selected. We see that Spxy works more effectively with the rate of 0.75 train, 0.25 test through experiments with the data set. Output distribution and PCA diagram of the data set are shown in Table 3 and Figure 6

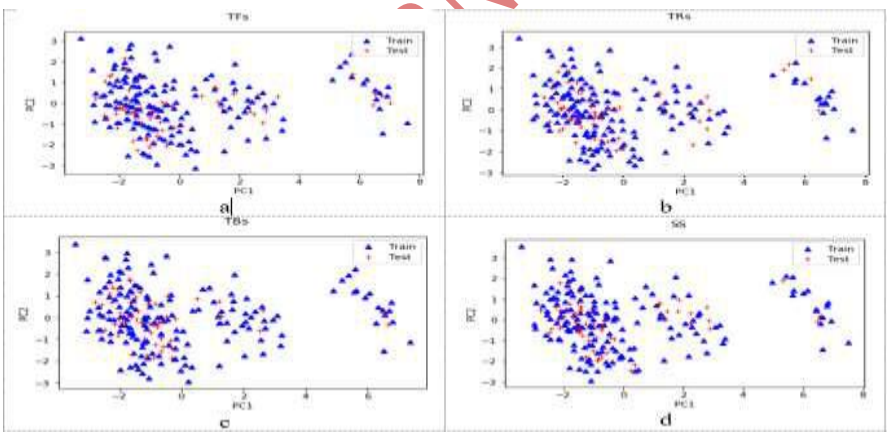

Figure 6: PCA score chart with the first two main components

a, TFs; b, TRs; c,TBs; d, SS

The selected train data covers all test data so that the train data can be obtained is representative. The calibrator range of the quality indices is greater than the predictor range, which can guarantee the predictive model's robustness.

Main component analysis

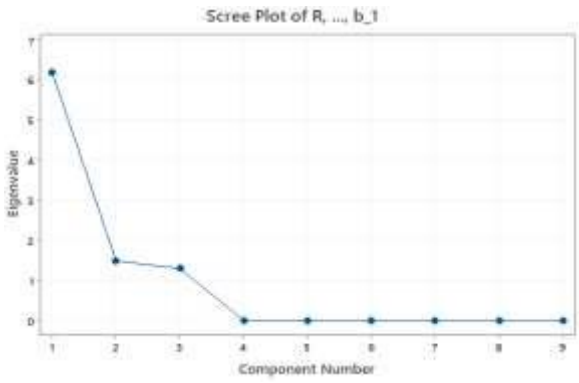

Figure 7: Graph showing the relationship between eigenvalue and the number of major components

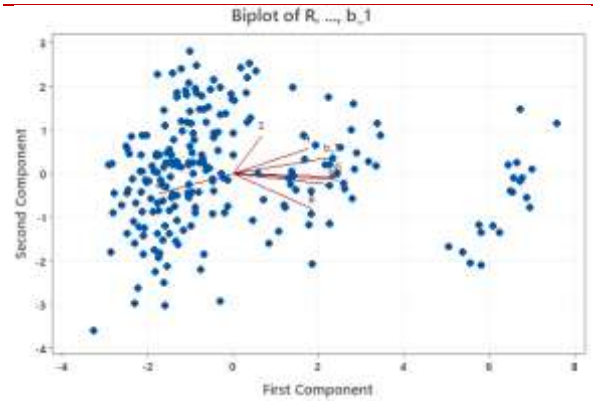

Figure 8: Observations and separate vectors on the first and second main components

After having own vectors, we project normalized data points onto these vectors and obtain a new input data table

Establish a nonlinear RF model for each quality indicator

PC (number of major components) and N (number of decision trees) directly influence the accuracy of the RF model. Therefore, further optimization is required on N and PC (within a certain range). 50N (20–1000, with a step size of 20) and 9PC (1–9, with step size 1) selected respectively to optimize the parameters based on the RMSE of each quality index model.

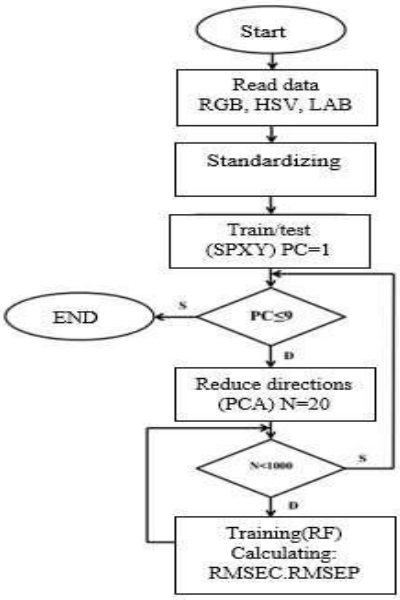

Figure 9: Algorithm flowchart of optimal PCs and N

```
z = rmsec
rmsec = np.array(rmsec).reshape(9,49)
rmsep = np.array(rmsep).reshape(9,49)
min_rmsec = np.min(rmsec)
for h in range(0, 9, 1):
    for k in range(0, 49, 1):
        if rmsec[h][k] == min_rmsec:
            print("RMSEC:",min_rmsec)
            print("RMSEP:",rmsep[h][k])
            print("PC:",h+1)
            print("N:",(k+1)*20)
            break
        if rmsec[h][k] == min_rmsec:break
    if rmsec[h][k] == min_rmsec:break
```

```
RMSEC: 0.028562306002023342
RMSEP: 0.05262137069644376
PC: 7
N: 880
```

Figure 10: The program optimizes the number of major components and the number of trees in RF (representative TFs)

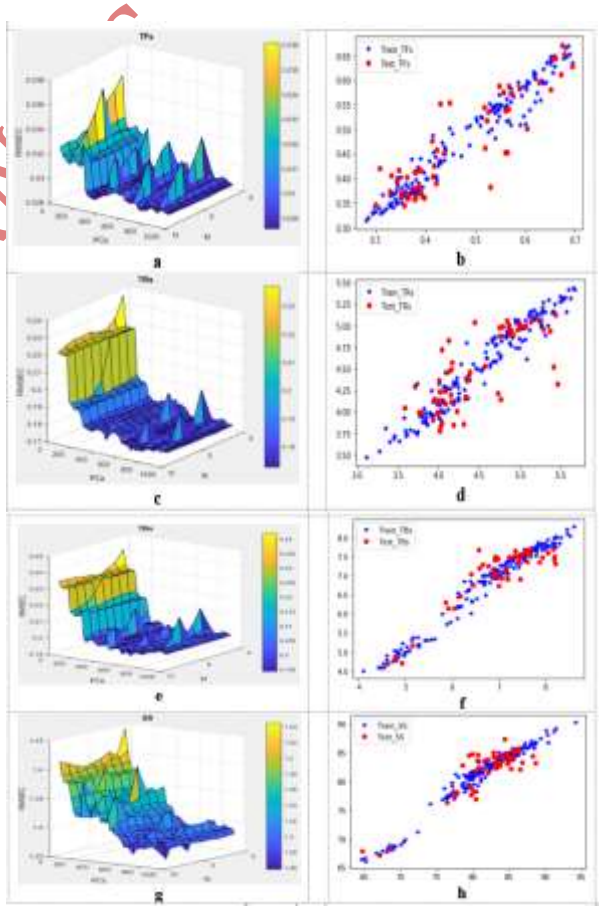

Figure 11: RMSEC value of each quality index for RF model from different PC and N a, representing TF; c, representing TR; e, represents TB and g, represents SS (sensory point), reference value versus predicted value of the RF model: b, represents TF; d, representing TR; f, represents TB and h, represents SS.

Table 3. Summary of parameters after optimizing the prediction model

Parameter	PC	N	RMSEC	Rc	biasc	RMSEP	Rp	biasp	SEP	CV	RPD
TFs	7	880	0.0286	0.9762	0.0004	0.0526	0.8906	0.0007	0.0526	0.2103	1.8685
TRs	7	60	0.1727	0.9687	0.0105	0.3452	0.7356	0.0051	0.3451	0.0953	1.2321
TBs	6	380	0.1943	0.9864	0.0036	0.3789	0.9008	0.0417	0.3766	0.1083	2.0182
SS	8	60	1.2579	0.9773	0.0249	2.165	0.8855	-0.076	2.1637	0.0522	1.972

Through the simulation results we see that:

- In the TFs prediction model, when PC = 7 and N = 880, the model's RMSEC reaches the minimum level (0.0286), Rp, RMSEP, Bias, SEP, CV and RPD of the predictor are 0.8906, 0.0526. , 0.0007, 0.0526, 0.2103 and 1.8685 respectively, and the relationship between the predicted value and the measured value is shown as Figure 4.8a, b.
- In the TRs prediction model, when PC = 7 and N = 60, the model's RMSEC reaches the minimum level (0.1727), Rp, RMSEP, Bias, SEP, CV and RPD of the predictor are 0.7356, 0.3452. , 0.0051, 0.3451, 0.0953 and 1.2321 respectively, and the relationship between the predicted value and the measured value is shown as Figure 4.8c, d.
- In the prediction model TBs, when PC = 6 and N = 380, the RMSEC of the model reaches the minimum level (0.1943), Rp, RMSEP, Bias, SEP, CV and RPD of the predictor are 0.9008, 0.3789 , 0.0417, 0.3766, 0.1083 and 2.0182 respectively, and the relationship between predicted value and measured value is shown as Figure 4.8e, f.
- In the SS prediction model, when PC = 8 and N = 60, the model's RMSEC reaches the minimum level (1,2579), Rp, RMSEP, Bias, SEP, CV and RPD of the predictor are 0.8855, 2.165, - 0.076, 2.1637, 0.0522 and 1.972 respectively, and the relationship between the predicted value and measured value is shown as Figure 4.8g, h.

5. CONCLUSION

The author presented the theoretical basis of black tea fermentation, color characteristics, transition between related color spaces. Apply and execute program writing for data processing algorithms. Optimizing the nonlinear regression model using the Random Forest algorithm, predicting the quality of black tea fermentation with quality indicators, namely meeting over 90% of the quality indicators.

6. ACKNOWLEDGMENTS

This research was funded by the Thai Nguyen University of Technology, Thai Nguyen, Viet Nam.

7. REFERENCES

[1] Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition*, *46*(12), 3483-3489.

[2] Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

[3] Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130-1141.

[4] Yaqub, M., Javaid, M. K., Cooper, C., & Noble, J. A. (2013). Investigation of the role of feature selection and weighted voting in random forests

for 3-D volumetric segmentation. *IEEE transactions on medical imaging*, *33*(2), 258-271.

[5] Deng, H. (2019). Interpreting tree ensembles with intrees. *International Journal of Data Science and Analytics*, *7*(4), 277-287.

[6] Bhattacharyya, N., Seth, S., Tudu, B., Tamuly, P., Jana, A., Ghosh, D., ... & Bhuyan, M. (2007). Monitoring of black tea fermentation process using electronic nose. *Journal of Food Engineering*, *80*(4), 1146-1156.

[7] Yang, Z., Baldermann, S., & Watanabe, N. (2013). Recent studies of the volatile compounds in tea. *Food Research International*, *53*(2), 585-599.

[8] Peris, M., & Escuder-Gilabert, L. (2013). On-line monitoring of food fermentation processes using electronic noses and electronic tongues: a review. *Analytica chimica acta*, *804*, 29-36.

[9] Tüfekci, M., & Güner, S. (1997). The determination of optimum fermentation time in Turkish black tea manufacture. *Food chemistry*, *60*(1), 53-56.

[10] Turkmen, N., & Velioglu, Y. S. (2007). Determination of alkaloids and phenolic compounds in black tea processed by two different methods in different plucking seasons. *Journal of the Science of Food and Agriculture*, *87*(7), 1408-1416.

[11] Samanta, T., Cheeni, V., Das, S., Roy, A. B., Ghosh, B. C., & Mitra, A. (2015). Assessing biochemical changes during standardization of fermentation time and temperature for manufacturing quality black tea. *Journal of Food Science and Technology*, *52*(4), 2387-2393.

[12] Wright, L. P. (2006). *Biochemical analysis for identification of quality in black tea (Camellia sinensis)* (Doctoral dissertation, University of Pretoria).

[13] Muthumani, T., & Kumar, R. S. (2007). Influence of fermentation time on the development of compounds responsible for quality in black tea. *Food Chemistry*, *101*(1), 98-102.

[14] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

[15] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.