A Multi-Mode Reasoning For Predicting Tobacco Baking Quality

QingE Wu（✉ wqe969699@163.com ）
Zhengzhou University of Light Industry

Tao Zong
Zhengzhou University of Light Industry

Hu Chen
Zhengzhou University of Light Industry

Lintao Zhou
Zhengzhou University of Light Industry

Yingbo Lu
Zhengzhou University of Light Industry

Penglei Li
Zhengzhou University of Light Industry

Yong Li
Zhengzhou University of Light Industry

A Multi-mode Reasoning for Predicting Tobacco Baking Quality

QingE Wu[*], Tao Zong, Hu Chen, Lintao Zhou, Yingbo Lu, Penglei Li, Yong Li

School of Electrical Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China

Corresponding author: wqe969699@163.com

Abstract: In order to reduce the number of defective products caused by the unreasonable baking time during the tobacco production process, this paper proposes a method for establishing a multi-model reasoning tobacco baking quality prediction model. Conduct data mining and analysis on the data of various indicators of the original tobacco, and screen out the data that have an impact on the quality of tobacco baking. In order to reduce the complexity of the model and eliminate the influence between different dimensions, the data are carried out and standardized processing. Next, the normalized data is explored for the multi-input and multi-output mapping relationship. Finally, a mapping matrix is given for the multi-input and multi-output mapping relationship so as to establish a tobacco baking quality prediction model. The test results show that the predicted value of this model is basically the actual value, and the prediction accuracy rate is more than 90%. It has a high prediction accuracy rate. The cured tobacco leaves are basically the same as the actual cured yellow expected value. This model provides a practical guide method for tobacco baking, which has certain practical value in actual tobacco baking.

Keywords: Tobacco baking, Multi-modal reasoning, Data mining, Standardization, Relationship matrix

Introduction

With the development of informatization and digitization, many countries and even the global manufacturing industry are facing a new round of challenges—"data explosion, lack of information", that is, the explosive growth of production data in various industries, but the existing calculation methods capture the information is insufficient. Therefore, the useful information contained in the big data requires more efficient data mining and analysis methods. It forces researchers and various industries to rethink computing solutions. To obtain information generated in various real production and life scenarios, the useful insights of massive data[1] were made to be intelligent. In the context of growing population and increasingly tight resources, it was of great significance to carry out research on tobacco baking quality prediction models aimed at saving resources and improving efficiency. Compared with other industries, the tobacco industry has a higher level of automation. Tobacco baking is an important for tobacco leaf production. With the advancement of data collection technology, the data obtained during the baking process of tobacco has exploded. Efficient utilization,

making production safer, low-cost, and high-efficiency has become an urgent problem to be solved.

An intelligent control system for tobacco status monitoring was designed in the paper[2] based on real-time images of tobacco leaves to monitor the status changes of the tobacco leaves during the baking process. In this paper, neural network method was used to identify and predict the set value of dry and wet bulb temperature and the time to the next set value. However, the neural network method may have local optimization or over-fitting, and the stability of the system needs to be studied. In recent years, with the continuous development of machine learning, integrated learning methods such as boosting[4] had attracted great attention from researchers in industrial product prediction[3]. Shi et al[5] proposed a tobacco quality prediction model based on an ensemble learning algorithm on the basis of data mining. The paper first used the random forest algorithm to reduce the dimensionality of the data features, and then used the XGBoost algorithm to establish a tobacco quality prediction model. From the perspective of experimental results, the model had a good effect on tobacco quality prediction. However, the prediction stability of this model in high-dimensional feature data sets was poor. Some scholars have also explored the relationship between the appearance of tobacco leaves and its chemical composition[6,7]. The appearance quality of tobacco leaves was closely related to its internal chemical composition. The main determinant of the quality of flue-cured tobacco lies in its chemical composition. Therefore, some scholars[8] judged the quality of tobacco leaves by using BP neural network, and indirectly judged the quality of flue-cured tobacco through the neural network to determine the appearance quality of tobacco leaves. However, the appearance quality was only part of the characterization of tobacco leaf quality, and couldn't directly indicate the quality of the tobacco leaf. Lu et al[10] used tobacco leaf color as the input index and flue-cured tobacco sensory quality as the output index to build a neural network to predict and evaluate the sensory quality of flue-cured tobacco. Yang et al[11] explored the application of machine vision in the prediction of flue-cured tobacco quality, using machine vision to distinguish the color difference of tobacco leaves after baking to predict the quality of tobacco leaves. Li Zheng, Lu Xiaochong, Yang Xiaoliang and others all used appearance to indirectly predict the quality of flue-cured tobacco, and did not directly involve the quality indicators of flue-cured tobacco. Fan et al[9] designed a cascade control method for the moisture control of the re-drying of tobacco leaves. This method was robust in response speed and robustness. There was a high improvement in performance and residuals. However, the moisture re-drying process of tobacco belonged to the follow-up project of tobacco leaf baking, and this method had no obvious effect on the prediction of tobacco leaf quality during the baking process. At present, most of the predictions of flue-cured tobacco quality were based on neural network appearance prediction, without considering the direct judgment of tobacco leaf quality from the tobacco baking data. Analyzing the tobacco leaf data was an effective method for improving predictive efficiency during the baking process and directly predicting the tobacco leaf quality.

This paper analyzes the collected baking tobacco data by data mining, and establishes a multi-mode reasoning tobacco baking quality prediction model, which is

used to monitor the state of tobacco leaves in real time during the tobacco baking process. The quality prediction model can be established by discussing the best mapping relationship between input data and output data. Moreover, in order to avoid the influence of different dimensions and data range for establishing the model, the input and output data are standardized before and after tobacco baking. Finally, through model evaluation and some experiments, compares some different prediction methods, the tobacco baking quality prediction model obtained by using multi-mode reasoning can ensure the accuracy, rapidity and anti-interference of the prediction.

Tobacco baking prediction model establishment

The effectiveness of the tobacco quality prediction model is related to a series of major issues such as "quality production prediction, prediction accuracy, precise control of tobacco quality, and intelligent decision-making of tobacco products". The use of data serves for tobacco quality prediction and reduces the labor of production personnel. Strength, greening and environmental protection, reducing resource waste, saving manpower, material and financial resources, and promoting the rapid development of the enterprise economy. In order to ensure the success of the system construction, the initial top-level design of the system development and construction must be done before the system development and construction are formally started. The overall scheme of model establishment is shown in Figure 1.
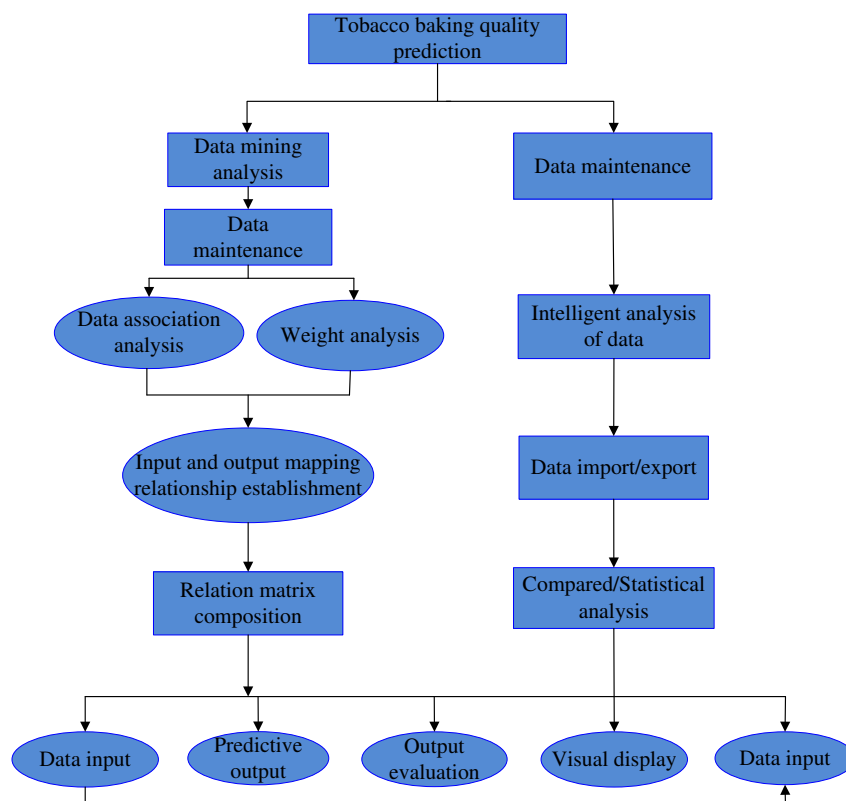
Figure 1. Tobacco baking data model establishment overall plan

In Figure 1, the quantitative and qualitative analysis of tobacco baking data is carried out to find the correlation between the data; the main factor data is mined and extracted; then the mined data is standardized, because of the type, unit, and

characteristics of the data If the relationship between input and output is not unified, the data must be standardized; the relationship between input and output data must be analyzed, and the proportion of the main and secondary factors of the input to the output will be mined; the establishment of multiple inputs and Multi-output mapping relationship matrix, that is, the establishment of input and output models; through this relationship matrix for testing, that is, the user's input is combined with the relationship matrix, and finally the user can reflect some relevant index values of the tobacco input. Bake the predictive index value, and give the change curve graph of the predictive index value.

Standardized processing of baking data

Data analysis

Analyze input and output data, including type, dimension, number of input and output, numerical characteristics, data comprehensiveness, output expectations, etc., according to the characteristics of the data, propose multiple models of data relationships, conduct experiments, and finally determine a reasonable one Algorithm. The specific mode of inference prediction is:

If "input x is A, output y is B";

Then "If the input x is A', what should be the B' corresponding to the output y".

We can define: $B' = A' \circ R_{(A \to B)}$, that is, the conclusion B' can be obtained by combining A' with the inference relationship from A to B. Among them, $R_{(A \to B)}$ represents the mapping relationship obtained by obtaining B from A, and $\circ$ represents the operator combining A' and $R_{(A \to B)}$ to obtain B'.

After obtaining the basic idea of modeling, it is also necessary to screen the input and output indicators of tobacco leaves. Through the research of material providers and Liu et al.[12], we have learned that the input indicators that affect the baking of tobacco leaves are the temperature of the dry and wet bulbs. , Absolute temperature, relative temperature and absolute humidity; The quality indicators of tobacco leaves after baking are red, green, blue and moisture respectively. After the data indicators are selected, the data can be modeled.

Standardized processing

For the actual production process, the collected data is different in dimension and value range. Direct modeling will lead to uneven spatial distribution of sample data, which will affect the prediction results of the model. Therefore, the collected data needs to be preprocessed first. There are various types of preprocessing methods. The conventional one is to normalize or standardize the data. Among them, there are many

methods of data standardization, such as deviation standardization, 0-1 standardization, etc. Here we use the most value comparison method to standardize the data. which is:

$$x'_{t,i'} = \frac{x_{t,i'}}{x_{t,\max}} \tag{1}$$

Among them, $x'_{t,i'}$ represents the weight of the $i'-th$ data of the $t-type$ index, $x_{t,i'}$ represents the $i'-th$ data of the $t-type$ index, and $x_{t,\max}$ represents the maximum value of the $t-type$ index, where $t=1,2,\cdots,T$, T is the number of types of indicators, $i'=1,2,\cdots,I$, I is the number of data of type t indicators.

The standardized data enables different indicators to have the same value range, and also eliminates the influence between different dimensions, and realizes the normalization of the data while retaining the complete information of the data.

Establishment of multi-modal reasoning model

According to the input data and output data during the baking process, a state equation of expected output is established, that is, based on the actual output state value, the analysis and establishment of an expected output state equation are:

$$X(k+1) = f[k, X(k)] + G(k)W(k) \tag{2}$$

here, $X(k)$ is the initial value vector of tobacco baking input at time k; $f[k, X(k)]$ is the nonlinear function of state $X(k)$ given in this article; $W(k)$ is the interference vector during the tobacco baking process, and $G(k)$ is the process interference distribution matrix generated by the input deviation in the tobacco baking process given in this paper. In order to facilitate modeling and processing, it is assumed that there is no control input at the beginning of tobacco baking, and it is assumed that the input deviation in the baking process, that is, the process noise is additive zero-mean white noise, and the interference distribution matrix $G(k)$ is known, that is:

$$E[W(k)] = 0, \quad E[W(k)W'(j)] = Q(k)\delta_{kj}$$

Among them, $E[g]$ is the expected function, $Q(k)$ is the zero-mean positive definite covariance matrix of the process noise, δ_{kj} is a small number of deviations, and j is any moment different from k.

The estimated output data is the existing measurement output data in Excel table,

and the measurement equation is established as follows:

$$Z(k) = h[k, X(k)] + V(k) \tag{3}$$

Among them, $Z(k)$ is the actual tobacco index value vector output at time k, $h[k, X(k)]$ is the non-linear function of the measurement output given in this article, and $V(k)$ is the deviation interference vector measured during the tobacco baking process. For the convenience of standardization processing, $V(k)$ is set as additive white noise with zero mean value, that is:

$$E[V(k)] = 0, \quad E[V(k)V'(j)] = R(k)\delta_{kj}$$

here, $R(k)$ represents the zero-mean positive definite covariance matrix of the measurement noise.

It is assumed that the interference vector and the measured deviation interference vector column in the tobacco baking process are independent of each other, and there is an initial state estimate $\overset{\Lambda}{X}(0|0)$ and a covariance matrix $P(0|0)$.

Then there is an estimate at time k:

$$\overset{\Lambda}{X}(k|k) \approx E[X(k)|Z^k] \tag{4}$$

Among them, $\overset{\Lambda}{X}(k|k)$ is an approximate conditional mean, and the associated covariance is $P(k|k)$. In view of the fact that $\overset{\Lambda}{X}(k|k)$ is not an exact conditional mean, strictly speaking, $P(k|k)$ is the approximate mean square error, not the covariance, and Z^k represents the actual output vector at time k.

The state prediction during the tobacco baking process from time k to time $k+1$ is:

$$\overset{\Lambda}{X}(k+1|k) = f[k, \overset{\Lambda}{X}(k|k)] + \frac{1}{2}\sum_{i=1}^{n} e_i tr[f_{XX}^i(k)P(k|k)] \tag{5}$$

Among them, $\overset{\Lambda}{X}(k+1|k)$ represents the state estimation value at $k+1$ relative to k, e_i is the $i-th$ basis vector in the Cartesian Coordinate System, n is the dimension of the state vector $X(k)$, $tr[g]$ is the trace of the matrix, and $f_{XX}^i(k)$ is the Hessian Matrix of the $i-th$ component of the vector f, and its expression is:

6

$$f_{XX}^i(k) \overset{\Delta}{=} [\nabla_X \nabla_X f^i(k,X)]'_{X=\hat{X}(k)} \tag{6}$$

$\nabla_X \nabla_X$ means that the function performs the second derivative of x, $f^i(k,X)$ means the value of the non-linear function of the state X of the $i-th$ component at time k, and $\hat{X}(k)$ means the estimated value of the state at time k.

The prediction error equation during the tobacco baking process is:

$$\tilde{x}_i(k+1|k) = x_i(k+1) - \hat{x}_i(k+1|k) = f[k,\tilde{x}_i(k|k)] \tag{7}$$

Among them, $\tilde{x}_i(k+1|k)$ represents the prediction error of the $i-th$ component at time $k+1$ relative to the state estimate at time k, $x_i(k+1)$ represents the actual state value of the $i-th$ component at time $k+1$, and $\hat{x}_i(k+1|k)$ represents the estimated value of the state of the $i-th$ component at time $k+1$ relative to time k, and $\tilde{x}_i(k|k)$ represents the prediction error of the $i-th$ component at time k.

The covariance associated with the prediction error is:

$$P(k+1|k) = f_X(k)P(k|k)f_X'(k) \tag{8}$$
$$+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}e_ie_jtr[f_{XX}^i(k)P(k|k)(f_{XX}^j(k|k)P(k))'] + G(k)Q(k)G'(k)$$

Where, $P(k+1|k)$ represents the covariance at time $k+1$ relative to time k, $G'(k)$ is the derivative of the process interference distribution matrix, $f_X(k)$ is the Jacobian Matrix of vector f, which is obtained from the estimation of the current tobacco baking state, and $f_X'(k)$ is its first-order derivative, its specific expression is as follows:

$$f_X(k) \overset{\Delta}{=} [\nabla_X f(k,X)]'_{X=\hat{X}(k)} \tag{9}$$

∇_X means that the function takes the derivative of x.

Similarly, for the second-order filtering, the predicted value of the actual measurement during the tobacco baking process is:

$$\hat{Z}(k+1|k) = h[k+1,\hat{X}(k+1|k)] + \frac{1}{2}\sum_{i=1}^{m}e_itr[h_{XX}^i(k+1)P(k+1|k)] \tag{10}$$

Among them, $\hat{Z}(k+1|k)$ represents the predicted value of the actual measurement of

the output vector at time $k+1$ relative to the output vector at time k, m is the dimension of the measurement vector $Z(k)$, and $h^i{}_{XX}(k+1)$ is the Hessian Matrix of the $i-th$ component of the vector h, and its expression is :

$$h^i_{XX}(k+1)\overset{\Delta}{=}[\nabla_X\nabla_X h^i(k+1,X)]'_{X=\hat{X}(k+1|k)} \tag{11}$$

Same as the vector f, $h^i{}_{XX}(k+1,X)$ represents the value of the non-linear function of the state X of the $i-th$ component at time $k+1$.

The covariance associated with the measured predicted value, i.e. the approximate mean square error, is:

$$S(k+1) = h_X(k+1)P(k+1|k)h'_X(k+1) \tag{12}$$
$$+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}e_i e_j tr[h^i_{XX}(k+1)P(k+1|k)(h^j_{XX}(k+1)P(k+1|k))']$$
$$+R(k+1)$$

Where, $S(k+1)$ represents the covariance matrix of the measured predicted value at time $k+1$, $h'_X(k+1)$ represents the first derivative of the Jacobian matrix of vector h at time $k+1$, and the Jacobian Matrix of vector h is:

$$h_X(k+1)\overset{\Delta}{=}[\nabla_X h(k+1,X)]'_{X=\hat{X}(k+1|k)} \tag{13}$$

And the filter gain is:

$$K(k+1) = P(k+1|k)h'_X(k+1)S^{-1}(k+1) \tag{14}$$

Among them, $S^{-1}(k+1)$ represents the inverse of $S(k+1)$.

The state update equation during the tobacco baking process is:

$$\hat{X}(k+1|k+1) = \hat{X}(k+1|k) + K(k+1)w(k+1) \tag{15}$$

Where, $w(k+1)$ represents the measurement residual, and its expression is as follows:

$$w(k+1)\overset{\Delta}{=}\tilde{Z}(k+1|k) = Z(k+1) - \hat{Z}(k+1|k) = Z(k+1) - h_X(k+1)\hat{X}(k+1|k) \tag{16}$$

The covariance updating equation is, that is, the filtering error covariance updating equation corresponding to equation (13) is expressed as:

$$P(k+1|k+1) = P(k+1|k) - P(k+1|k)h'_X(k+1)S^{-1}(k+1)h_X(k+1)P(k+1|k) \tag{17}$$
$$= P(k+1|k) - K(k+1)S(k+1)K'(k+1)$$

Equations (13) and (15) are the measurement indexes for evaluating the tobacco quality, color and processing degree in the tobacco baking process, and also control some parameter indexes and attribute values of the baking process through the output

of these two equations.

Further, the final actual output vector of tobacco baking is:

$$\hat{y}_i\left(k+1|k\right) = h[k+1, \overset{\Lambda}{X}(k+1|k)] + \frac{1}{2}\sum_{i=1}^{m} e_i tr[h_{XX}^i(k+1)P(k+1|k)] \qquad (18)$$

The error of the final actual output vector of tobacco baking is:

$$\tilde{y}_i\left(k+1|k\right) = y_i\left(k+1\right) - \hat{y}_i\left(k+1|k\right) = h[k+1, \tilde{X}\left(k+1|k\right)] + v_i\left(k+1\right) \qquad (19)$$

Where $\tilde{y}_i\left(k+1|k\right)$ is the residual of the actual output.

Thus, the error covariance of the final actual output of tobacco baking is:

$$S(k+1) = h_X(k+1)P(k+1|k+1)h_X'(k+1)$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} e_i e_j tr[h_{XX}^i(k+1)P(k+1|k+1)\left(h_{XX}^j(k+1)P(k+1|k+1)\right)'] \qquad (20)$$

$$+ R(k+1)$$

At this time, $S(k+1)$ is the covariance matrix of the actual output.

Finally, a complete multi input prediction multi output model is established, that is, the relationship from input vector X to output vector Y is determined as a specific mapping relationship R_T through a function transformation T, and the model can be simply expressed as the following mode:

Set the input to $X = \{x_1, x_2, \cdots x_n\}$ and the output to $Y = \{y_1, y_2, \cdots y_m\}$. Given a mapping relationship R from X to Y, a function transformation from X to Y can be determined, that is:

$$T_R\left(A\right) = A \circ R \qquad (21)$$

Wherein, $T_R\left(A\right)$ represents the synthetic operation result of the input matrix A and the relationship matrix R between input and output, that is, the prediction matrix obtained by the operation of A and mapping relationship R; A represents the input matrix, R represents the mapping relationship matrix from input to output, and $\circ$ represents the synthesis operator.

Evaluation of model establishment

The evaluation method given in this paper is as follows:

(1) Establish factor set $U = \{u_1, u_2, \cdots u_n\}$ and decision set $V = \{v_1, v_2, \cdots v_m\}$.

(2) Establish single factor evaluation matrix

For each factor u_i, first establish a single factor judgment set:

$D_i = \{d_{i1}, d_{i2}, \cdots d_{im}\}$, that is, collection elements d_{ij} , $0 \le d_{ij} \le 1$ represents the judgment of v_j on the factor u_i , so that the single factor judgment matrix $D = (d_{ij})_{n \times m}$ is obtained.

(3) Establish comprehensive evaluation matrix

According to the weight $A = \{a_1, a_2, \cdots a_n\}$ of each factor and the single factor evaluation matrix D , the comprehensive evaluation matrix can be obtained: $B = A \oplus D = \{b_1, b_2, \cdots b_m\}$, which is a fuzzy subset of V . According to different definitions of operation $\oplus$, different models can be obtained.

Let's evaluate this model.

The set of factors in this model are: tobacco leaf color, baking temperature, humidity, and water content, namely $U = \{u_1, u_2, u_3, u_4\}$, where u_1, u_2, u_3 and u_4 are tobacco leaf color, baking temperature, humidity, and water content, respectively;

The decision set is: smoke color and water content, namely $V = \{v_1, v_2\}$, where v_1 and v_2 are smoke color and water content respectively.

First of all, according to the above method of establishing the single factor evaluation matrix, a single factor evaluation D_i is established for each factor u_i , in order to reduce the subjectivity of expert evaluation, the average score of ten different experts is used for each d_{ij} . Thus, the single factor evaluation matrix is:

$$D = \begin{bmatrix} 0.82 & 0.12 \\ 0.91 & 0.89 \\ 0.10 & 0.11 \\ 0.20 & 0.15 \end{bmatrix}$$

Secondly, the weight of each factor in the expected state of tobacco, that is, the expected value of each state in the tobacco baking process is A : $A = \begin{bmatrix} 0.8 & 0.1 & 0.1 & 0.1 \end{bmatrix}$.

Finally, through the synthetic operation of the weight value of each factor and the single factor evaluation matrix, the comprehensive evaluation B is obtained. The algorithm in this paper is: $b_j = \vee \begin{bmatrix} a_i \wedge d_{\bullet j} \end{bmatrix}, (i = 1, 2, 3, 4; j = 1, 2)$, where, $d_{\bullet j}$ represents a column of single factor evaluation matrix, and the comprehensive evaluation matrix of this model is calculated as $B = \begin{bmatrix} 0.8 & 0.12 \end{bmatrix}$.

The tobacco baking quality prediction model established in this paper is theoretically feasible according to the evaluation criteria of high-quality tobacco products for tobacco color and moisture content. Next, we will verify the practical feasibility of this model through experiments.

Application of multi-modal reasoning tobacco baking prediction model in actual tobacco baking

Source of sample data

The experimental samples were tobacco leaves of the same grade provided by a tobacco processing company, including various data indexes of tobacco leaves, including 553 data of upper leaves, middle leaves and lower leaves respectively. Since the establishment of this model is to explore the mapping relationship between input and output, that is, infer the quality index after baking according to the input data of tobacco. Therefore, using the upper, middle and lower leaf data of tobacco has little impact on the experimental results. Therefore, only the upper leaf data of tobacco leaf is used to establish a model to verify the effect of the algorithm.

Construction of multi-mode reasoning tobacco baking prediction model and its application in tobacco baking process

550 samples of upper leaf data were taken to establish the training set and test set with a ratio of 10:1. 500 samples were taken from 550 upper leaf data as the model training set, and the remaining 50 samples were used as the test set to verify the feasibility of the model.

The multi-mode reasoning model is programmed. The system uses MATLAB software platform to build the input and output module interface. Through the built platform interface, it is convenient to input and select input indicators, and can clearly see the prediction results and the trend of each output quality index data. The developed reasoning model interface is shown in Figure 2. The left half of the interface is the input indicator display and operation button area, and the right half is the output indicator display area. After the training, the model can predict a single group of data or batch predict the data. During the prediction of a single group of data, the user can independently input the corresponding value in the corresponding space. After clicking run, the output box on the right will display the corresponding output index.
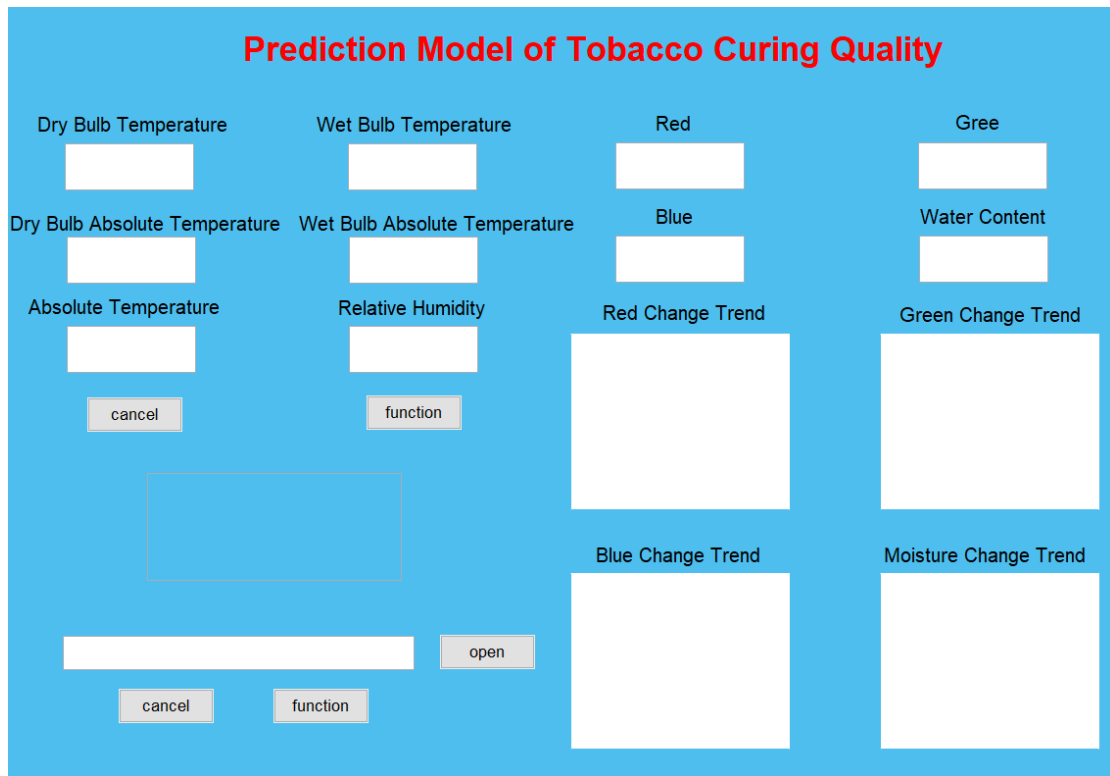
Figure 2. The interface of the developed predictive model software system

As shown in Figure 2, the model can also import the corresponding input parameter values from the excel table to predict the quality of tobacco baking in batches, and the change curve of each quality index after tobacco baking can be displayed in the block diagram on the right side of the interface Show clearly.

Before the model can predict normally, the model needs to be trained and tested first. When training the model, you only need to click the "Open" button, select 500 sets of training data from the pop-up data set, and then click the "Run" button to train the model. After the model is trained, the test set can be used to test the model training results. When performing model testing, select the corresponding test set and click "Run" to test the trained model.

As shown in Figure 3, a single data prediction is performed on the trained model. Enter a group of tobacco leaf corresponding data in the indicator input box, and click Run to display the set of tobacco quality output indicators in the output box.
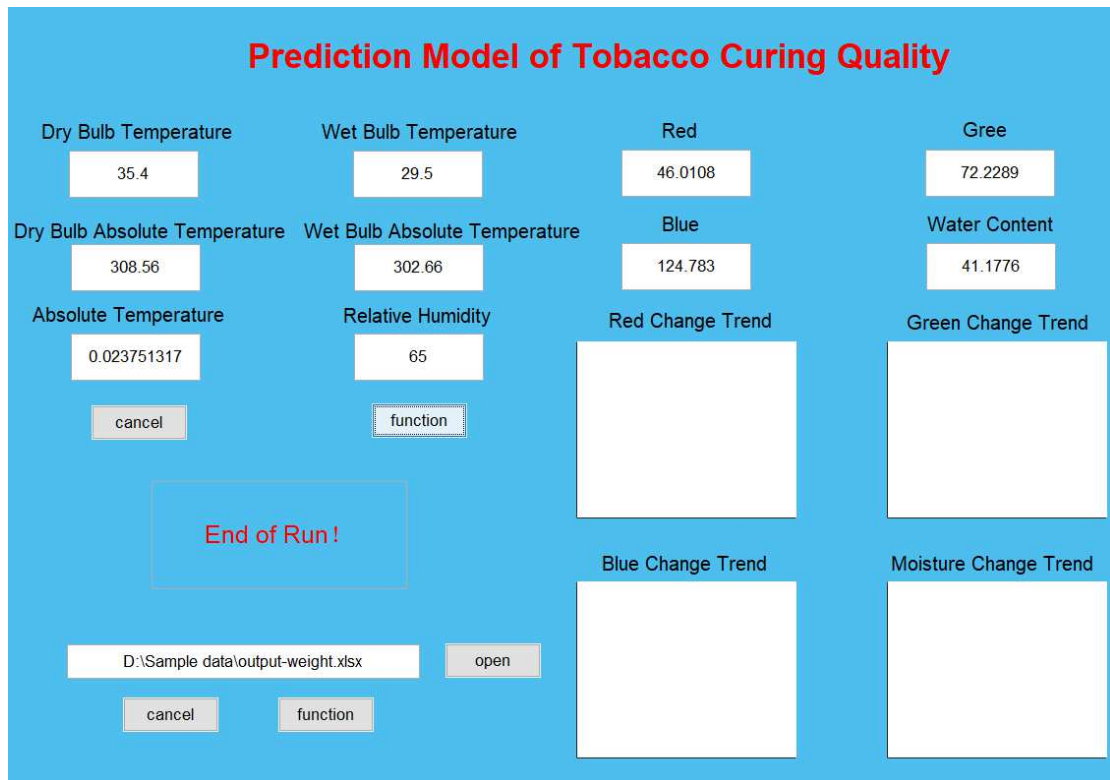
Figure 3. The output of a single set of data predictions

It can be seen from the prediction result that the predicted output value when comparing the single data prediction with the original data is basically consistent with the given index value of the tobacco leaf. However, due to the small amount of data processed, the single group data prediction is inefficient in the actual tobacco production and has no practical significance. Therefore, it is only used as development test. This paper mainly discusses the prediction of batch data.

As shown in Figure 4, it is the result of predicting the test data in a set of test sets after model training. It can be clearly seen from the figure that the last group of data output of various indicators will be displayed in the corresponding box in the figure. In addition, the change trend of all indicators will be displayed in the corresponding trend change diagram, and all predicted values will be exported to Excel table for subsequent use of data.

After analyzing the prediction results, we will verify the accuracy of the model, as shown in Figure 5, which is the comparison diagram of the predicted values of red, green, blue and moisture with the given values of the original quality indicators. It can be clearly seen in the comparison diagram that the predicted output values of each index of the model are basically consistent with the original values of the data, and the accuracy of the model prediction reaches more than 90%, The validity of this model in tobacco baking quality prediction is proved again.
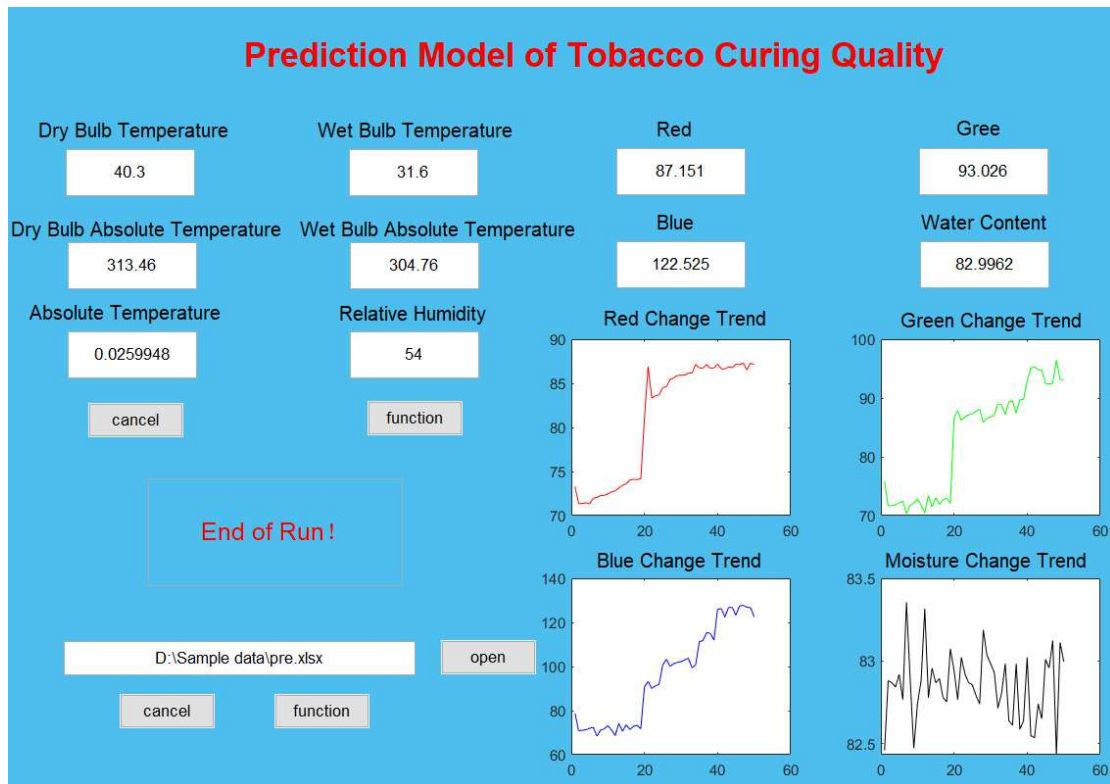
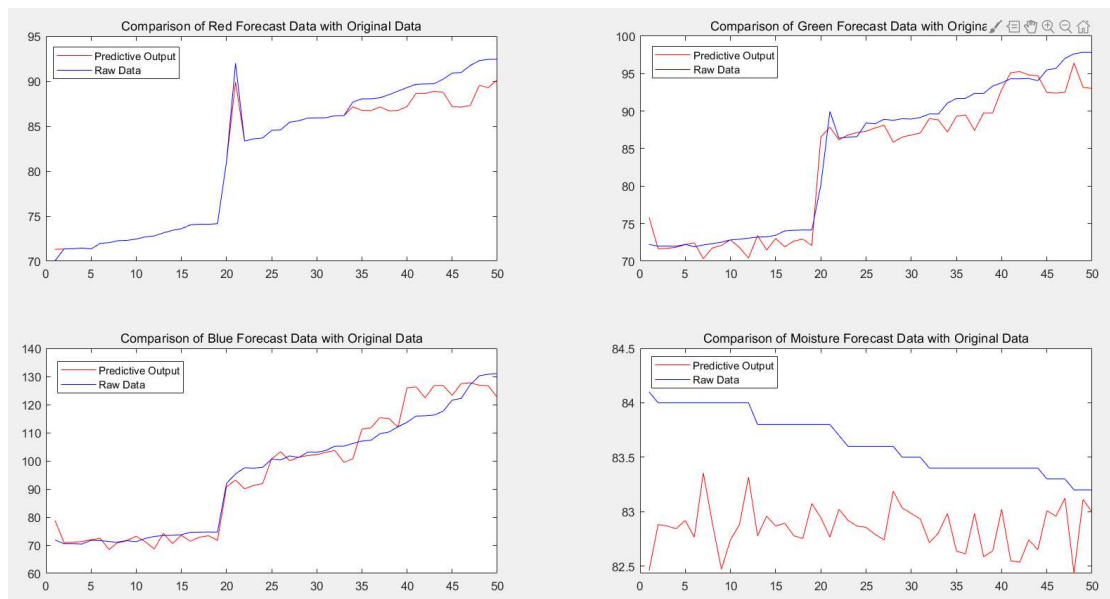Figure 4. Tobacco baking quality batch prediction results

Figure 5. Comparison of the predicted data of each indicator with the original data

This model not only has high prediction accuracy, but also has certain anti risk ability and strong robustness. The multi-mode reasoning method is adopted to make full use of all the data information provided by tobacco data, so as to make the model more scientific and objective. At the same time, when the input format is wrong, the developed system will prompt the user to input the correct data type, as shown in Figure 6. The user directly runs the system without inputting the training value. At this time, the system will report an error and remind the user to select the input value.
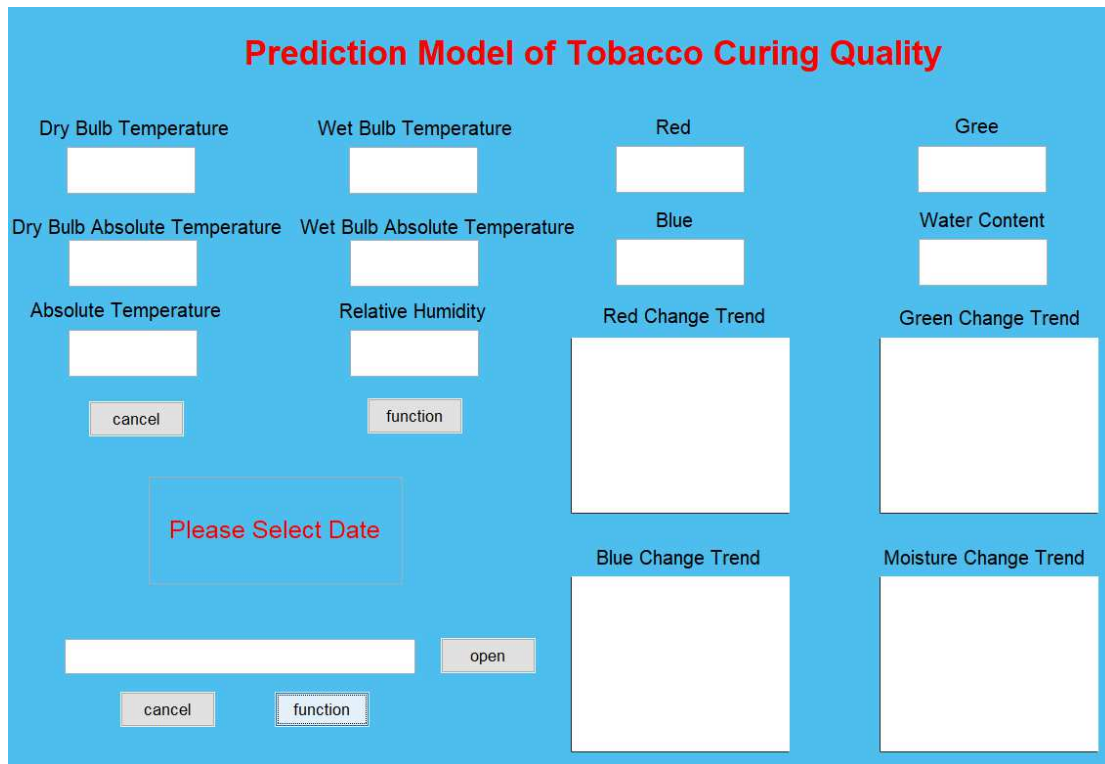
14

Figure 6. The software system will prompt an error without entering a value

Comparison of the constructed predictive modeling method with existing methods

The control of tobacco baking process is a mapping relationship between multiple inputs and multiple outputs, that is, if the input matrix is m, the mapped output matrix is n. For such control, it is difficult to establish a specific function to reflect the relationship between them, but the input parameters affecting the output index are coupled. That in order to achieve better control and accurate prediction, each input data must be analyzed. In this paper, all input parameters and output indexes are taken as a whole to establish a relationship model between them, and a multi-mode reasoning state system method is proposed to establish the state model. At the same time, the model is comprehensively evaluated in the process of establishing the model, and three kinds of errors are corrected by the evaluation method, That is, output index correction method, mapping relationship matrix correction method and curve correction method. Finally, through the iterative learning of samples by machine, a more accurate tobacco baking prediction model system is established and developed. The model is tested in the value range of each input parameter, and the error of each output index is controlled below 10%, as shown in Figure 7.
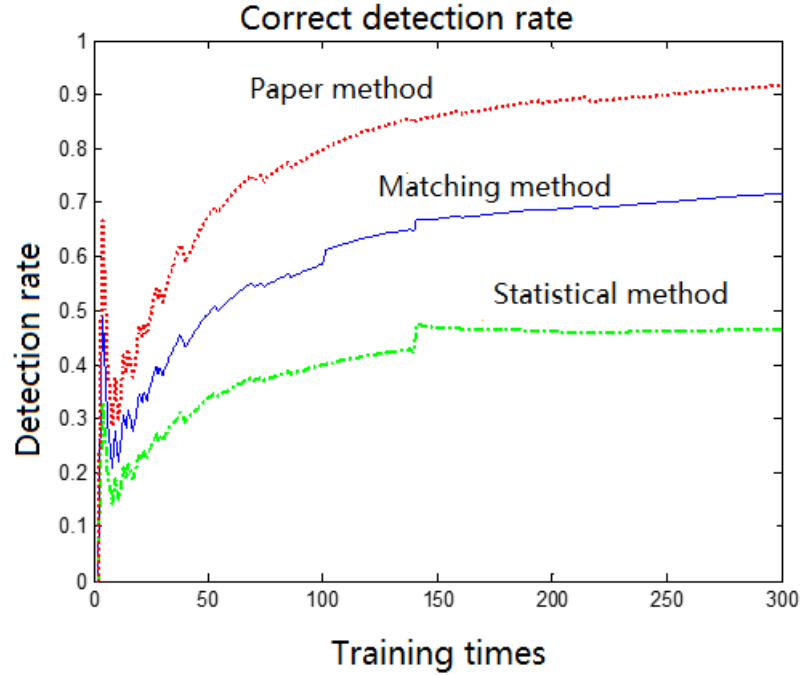
Figure 7. Comparison of correct prediction rate between the proposed method and the existing prediction method

After theoretical and experimental verification, the data modeling method given in this article is used in tobacco baking monitoring. The experimental results established from the model show that the correct prediction rate is as high as 90% and above. This application method is better than the previous application method. The prediction accuracy rate of is higher by about 10%, indicating that the establishment of this model is effective and feasible.

In particular, if the main factor is changed, the output result changes significantly, and if the change is a secondary factor, the output result does not change significantly, because the tobacco baking system is a multi-input multi-output functional relationship, and the input parameters are Coupled, not independent of each other. Under a certain input, some input parameters play a major role, some play a secondary role, and some do not play a role. Therefore, when changing the input parameters, if it is the main factor of the change, the output index changes more obviously.

Conclusion and future work

In this paper, a mapping relationship between multi input and multi output is established by using the modeling method of multi-mode reasoning state. When establishing the model, the data are filtered and irrelevant variables are eliminated; Then, in order to avoid the influence of variable dimension, the maximum value comparison method is used to normalize the data; When modeling, the expectation model is established for the first-order filter and the second-order filter respectively, and finally the complete prediction model is established. Compared with the previous prediction methods, this model removes the redundancy of data, fully extracts the data information, obtains the relationship between data input and output more accurately, and further improves the accuracy of model prediction. The experimental results prove

that compared with other algorithms, the algorithm established in this paper has greatly improved the accuracy of the prediction of tobacco baking quality, and the accuracy of the model's prediction basically reaches more than 90%. This algorithm improves the prediction accuracy of tobacco baking quality, provides technical support for the intellectualization of tobacco production, and provides a guiding method for quality monitoring in tobacco baking. In addition, this model can not only be used in the tobacco baking process, but also provides a good algorithm for the baking of dried fruit, and provides a reliable idea for the automation of life and production.

References

[1] Ianni M, Masciari E, Mazzeo G M, Mezzanzanica M, Zaniolo C. Fast and effective Big Data exploration by clustering. Future generation computer systems. **102**, 84-94(2020).

[2] Wu J, Yang S X, Tian F. A novel intelligent control system for flue-curing barns based on real-time image features. Biosystems Engineering. **123**, 77–90(2014).

[3] Jiang Kun, Ding Xueming. Industrial product quality prediction based on integrated learning algorithm. Software Guide. **18**, 124-127(2019).

[4] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 22nd ACM SIGKDD International Conference. 785-794(2016).

[5] Shi Q, Wang H, Xu X, Lu H, Zhang Z. The Application of Tobacco Product Quality Prediction Using Ensemble Learning Method. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference. IEEE. **1**, 1780-1784(2019).

[6] Li Yongzheng, Yang Hongqi. Research progress on the correlation between appearance quality and internal quality of tobacco leaves. Crop research. **26**, 147-150(2012).

[7] Wang Fang, Shi Gaili, Zhang Qingming, Liu Yang, Ding Gensheng, Ji houwei, Zhang Li. Study on the correlation between appearance quality and internal quality of tobacco leaves. China agronomy bulletin. **30**, 82-88(2014).

[8] Li Zheng, Wang Jianfeng, Cheng Xiaoqiang, Duan Shijiang, Shi Wenqiang, Hu ronghua, Xiao Ronggui, Shen Hongtao. Prediction model of Flue-cured Tobacco Appearance Quality Based on BP neural network. Southwest Agricultural Journal. **32**, 653-658(2019).

[9] Fan Z, Ren Z, Chen A. A Modified Cascade Control Strategy for Tobacco Re-Drying Moisture Control Process with Large Delay-Time. IEEE Access. **8**, 2145-2152(2019).

[10] Lu Xiaochong, Li Hao, Su Jiaen, Pei Xiaodong, Meng Xiangyu, Yang Xiaoliang, Wang Jian'an, He fan, song chaopeng. Construction of sensory quality prediction model of Flue-cured Tobacco Based on color characteristics of flue-cured tobacco. Journal of Henan Agricultural University. **50**, 500-505(2016).

[11] Yang Xiaoliang, He fan, Xie Lilei, Liu Zheng, sun Shuguang, Gong Changrong. Application of machine vision technology in quality prediction of early flue-cured tobacco. Jiangxi Journal of agriculture. **29**, 95-98, 104(2017).

[12] Liu Shijiang. Research on high quality tobacco curing technology. Rural practical

technology. **4**, 92-93(2020).

[13] Ahmad M H, Nache M, Waffenschmidt S, Hitzmann B. A fluorescence spectroscopic approach to predict analytical, rheological and baking parameters of wheat flours using chemometrics. Journal of Food Engineering. **182**, 65-71(2016).

[14] Yuan Xiangyun, Li Da, Ge Wen, Xu Renjie, Yuan Peng, Cui Yuxiang, Li Ligang, Chen Zhen. Application prospect of big data of process control in tobacco manufacturing process. China management informatization. **20**, 55-61(2017).

[15] Chen J, Li Y, He X, Jiao F, Xu M, Hu B, Jin Y, Zou C. Influences of different curing methods on chemical compositions in different types of tobaccos. Industrial Crops and Products. **167**, 1-11(2021).

[16] Cotrim W, Minim V, Felix L B, Minim L A. Short convolutional neural networks applied to the recognition of the browning stages of bread crust. Journal of Food Engineering. **277**, 1-8(2020).

[17] Yang L, Xinwei J I, Folin L I, Lyu B, Huang K, Enping P U, Chen Z, Chen J, Yang X, Duan J. The Quality Difference of Tobacco Leaves Between Different Flue-curing Processes in Honghe Tobacco-growing Area. Agricultural Biotechnology: English version. **9**, 50-53(2020).

[18] Naghavi E A, Dehghannya J, Ghanbarzadeh B. 3D computational simulation for the prediction of coupled momentum, heat and mass transfer during deep-fat frying of potato strips coated with different concentrations of alginate. Journal of Food Engineering. **235**, 64-78(2018).

[19] Li Zhigang, Dian Ruili, Wang Tingxian, Pan Feilong, Yang Xiaoliang, Yang Zhanwei, Qian Chuanlin, Sun Shuguang. Change analysis of color parameters of flue-cured tobacco leaves and main veins during baking. Journal of Hunan University of Arts and Sciences. **31**, 72-78(2019).

[20] Hilal Isleroglu and Selami Beyhan. Prediction of baking quality using machine learning based intelligent models. Heat and Mass Transfer: Wärme- und Stoffübertragung. **56**, 2045-2055(2020).

[21] Iribe-Salazar R, Caro-Corrales J, Y Vázquez-López. Analysis of random variability in Tortilla shells baking. Journal of Food Engineering. **293**, 1-8(2021).

Acknowledgement

Statement:

All formats and methods of this paper are in accordance with the format requirements and regulations of this journal.