*Research Article*

# Application of Intelligent Taste Analysis Based on Random Forest Algorithm in Food Quality Inspection

**Xinghua Zhang,[1] Yongjie Sun,[2] and Yongxin Sun** [ID][3]

[1]*Changchun Sci-Tech University, Changchun 130600, JiLin, China*
[2]*College of Life Science, Changchun Sci-Tech University, Changchun 130600, JiLin, China*
[3]*College of Physics and Electronic Information, Baicheng Normal University, Baicheng 137000, Jilin, China*

Correspondence should be addressed to Yongxin Sun; sunyongxin@bcnu.edu.cn

Food safety is a major concern that has an impact on the national economy and people's lives. The food industry has grown in quality and innovation in tandem with the rapid development of the economy and society. The emergence of new food technologies, as well as changes in dietary habits, has increased public concern about food safety. With the emergence of various counterfeit and substandard products, food quality and safety testing have become even more important. Traditional testing methods rely on sensory analysis and physical and chemical analysis. This approach is subjective and poorly adapted to the general public. It requires a high level of technical operation and is difficult to carry out on a large scale. To address this situation, this paper proposes an intelligent approach to food safety quality testing. The core idea is, first, to use sensors to collect data on the various components of the sample to be tested. Second, the random forest (RF) model used in this paper is trained. Third, the trained model is used to classify and identify the test samples. Based on the classification results, a conclusion is drawn as to whether the food product is a variant or a counterfeit. The advantage of this study is that the training model used is a weighted RF algorithm based on mutual information. The correlation between any two decision trees is calculated using mutual information, and for the more correlated decision trees, only the one with the highest evaluation accuracy is retained to form a new RF, and the evaluation accuracy is converted into voting weights, resulting in an RF model with less redundancy and higher evaluation accuracy. The experimental results show that the method used in this paper can successfully identify spoiled or counterfeit products and has good practicality.

## 1. Introduction

Food safety usually means that the items people consume are nontoxic; that is, they do not cause any harm to the body after consumption, and they provide the body with the nutrients it needs [1, 2]. Food safety is not just something that the producer needs to be aware of since there are multiple components involved in the process of food from production to consumption. Food safety is a matter that is jointly guaranteed in many fields. Eating is one of the most important things for human beings. Therefore, food safety and security are of the utmost importance. With the development of technology, many unscrupulous businessmen use high technology to make some poor-quality food products for sale to earn high profits. The packaging of these foods often looks attractive, and the food looks fresh and tasty. In order to achieve the purpose of attracting customers, unscrupulous businessmen will use some ingredients that are toxic and harmful to the human body. And consumers often focus only on the packaging or indication of the food and lack of food safety awareness. Consumers should therefore increase their awareness and knowledge of food safety, develop a healthy and scientific view of consumption, and actively participate in the regulation of food safety to ensure food safety and health. The protection of food safety is the legal responsibility of all food producers and operators, governments at all levels, and relevant regulatory agencies. Regulation should be continuously

strengthened to promote a steady improvement in the internal management capacity of producers and operators by enhancing constraints from society. At the same time, the whole community must take an interest in and maintain food safety and support the work of the government. At present, many consumers have little awareness of the importance of safety management in food businesses, knowledge of consumer culture is still lacking, and the basis for social development of food safety is still not strong. When we are in food safety and find illegal problems, we should take positive action and take the initiative to report them to the regulatory authorities, rather than just complaining and grumbling, not to mention creating false information out of thin air, which will only interfere with normal regulatory work and bring about consumer panic, with the ultimate victim being the consumer.

Food quality and safety are very important, and the government is also the introduction of relevant policies to protect the quality of food. The current policy is mainly from the following aspects. The first is to strengthen the quality control of food sources and strict food into the market trading system. Food safety must start at the source, from planting, breeding, production and processing, wholesale and retail, and consumer four links to carry out a full range of supervision. The use of computers and other modern technical means is to establish a series of systems such as commodity labeling and quality tracing. The second is to establish an industry self-regulatory mechanism to ensure food safety. The credit self-regulation of food production and operation enterprises is particularly important to ensure food safety. Therefore, it is necessary to accelerate the construction of a food safety credit system and establish industry credit standards for food distribution enterprises. We should strengthen the legal system construction and integrity education for operators and urge enterprises to improve their awareness of integrity, quality, and legal operation. The third is the establishment of food safety standards and inspection and testing systems. Product and health standards are established comprehensively related to food safety, and a food safety standards system is built. At the same time, the food safety testing system is improved to ensure that the hygienic quality of food listed for sale meets the requirements of health standards. Fourth, a traceability system that can be traced back to the source is established. Food safety is a complex process that must consider all stages of the food chain, from farm to fork.

The government has given a series of safeguards from the production of food to its final consumption in people's mouths. The most important part of the whole safeguarding process is the occasional inspection of food quality. Traditionally, food quality checks have relied on manual inspection by quality inspectors, which is both energy and time consuming and inefficient. As a result, in recent years, smarter methods of food quality inspection have emerged one after another. The core idea is to identify the quality of food by recognizing the taste of the food. Reference [3] proposed an improved K-means algorithm for detecting spoiled food. The core idea of the study is to segment the pictures of food products in order to determine the degree of spoilage. Reference [4] uses deep learning algorithms based on computer vision techniques to detect and analyze the quality of complex food products. Reference [5] devised a relationship between food quality and preservation conditions based on Fourier's law of conduction. Reference [6] proposed a quality inspection technique for frozen foods. The core technique used in this study was Raman spectroscopy. Reference [7] introduced blockchain technology for the storage and preservation of food products, thus guaranteeing food quality and safety. Reference [8] summarizes a variety of deep learning techniques applied to the inspection of food quality such as fruits, vegetables, and meat. Reference [9] uses image processing techniques applied to the detection of egg quality. The study combines hyperspectral imaging, multivariate analysis, and image processing to identify the freshness of eggs as well as the breakage of eggs. Experimental results showed that this aspect was able to achieve a detection rate of over 97%. For quality detection of palletized packaged food, reference [10] used a deep neural network based on a principal component analysis network. Experiments were conducted using support vector machines (SVM) [11] as well as K-nearest neighbours (KNN) [12] for experimental comparison, which showed that the study has advantages not only in terms of detection accuracy but also speed. From the above studies, it can be seen that for food quality inspection, more and more scholars are introducing intelligent algorithms such as machine learning algorithms [13–15] and deep learning [16–18] algorithms. The use of intelligent technologies has indeed improved the efficiency and accuracy of food quality inspection and saved human and material resources. In this paper, RF algorithm is introduced for food quality inspection based on taste analysis. Experiments on vinegar demonstrate the effectiveness of the method in this paper.

## 2. Relevant Knowledge

*2.1. Taste Analysis Methods.* It is well known that the taste signals received by the human brain are produced by taste receptor cells in the mouth (mainly the tongue). These taste receptor cells interact with specific chemicals in food. Different types of taste cells are sensitive to specific flavour substances. This property enables humans as well as animals to recognize different tastes. Early artificial taste systems used generic sensors or methods to recognize the physicochemical properties of food. For example, for salty tastes, they can be measured using conductive agents, and for sour tastes, their pH can be measured directly to characterize the acidity. The most primitive calibrations were obtained using a comparison of the human evaluation of salty and sour flavours and the quantitative values detected by the sensors. AsHumans have vague descriptions of taste, e.g. "a little sour", "very sour." The machine, on the other hand, obtains precise measured values. This method therefore only yields results that are not uniform. Later research, towards multisensor multicomponent analytical chemistry methods, which rely on specific chemical sensors, is difficult to implement and difficult to generalize.

On the other hand, there is ongoing research into the underlying perceptual mechanisms of the taste nerve. It has

been found that taste cells are not actually sensitive to only one chemical substance but respond to a variety of substances. At the same time, when different taste nerve cells perceive a stimulus at the same time, they transmit their respective signals to the center, which ultimately decides what type of taste stimulus to receive. A diagram of this process is shown in Figure 1. This mechanistic understanding is the cornerstone of modern taste sensing and analysis, resulting in the approach of cross-response sensor arrays. That is, a small number of types of sensors, through a mechanism similar to that of a neural network, are ultimately able to identify a larger number of types of taste stimuli. The number of stimulus types is an order of magnitude greater than the number of sensor types. Importantly, this approach draws on bionic ideas and breaks through the constraints of strongly selective taste sensors.

The traction in this mechanistic understanding has led to a move away from reliance on specific taste sensors for taste sensing. Tests have been carried out using less selective sets of sensors, which are analyzed by chemical analysis methods as well as by pattern recognition methods. This approach allowed pattern recognition to be used naturally in taste recognition scenarios. This was followed by the development of artificial taste sensing. And nowadays, the conceptual definition of taste sensors and electronic tongues and the mathematical processing methods used have gradually become clearer. With the development of materials, computers, and physical-chemical testing techniques, more sensors are available, and the accuracy and range of tests have advanced dramatically. A schematic representation of modern methods of artificial taste analysis is shown in Figure 2.

*2.2. Application of Taste Analysis Methods.* With the rapid development of electronic tongues and taste sensors, the recognition of types of taste is no longer limited to the recognition of simple basic characteristics such as sweet, sour, bitter, spicy, and salty. Application scenarios now include flavour and quality recognition, such as flavour recognition of alcohol and beverages, determination of the quality of liquid and solid foods, and can even be used to distinguish minor differences between different manufacturers' brands of the same type of food. More work has been carried out on the artificial taste analysis of liquid beverages than on the taste analysis of solid foods. A large body of literature deals with, for example, fruit juices, alcoholic beverages, and liquid flavourings. Reference [19] used multiple sensors to capture the composition of red wine in order to correctly differentiate the taste of red wine. Pattern recognition techniques were also used to perform operations such as preprocessing and classification of the collected data from the red wines. The experimental results show that the method used in this literature can successfully identify red wines as well as nonred wines. Reference [20] used a highly cross-selective sensor to capture the composition of red wines and combine it with principal component analysis and clustering algorithms to model and identify changes in the taste of red wines during storage. Reference [21] used an electronic tongue to capture the concentration of bisulphite
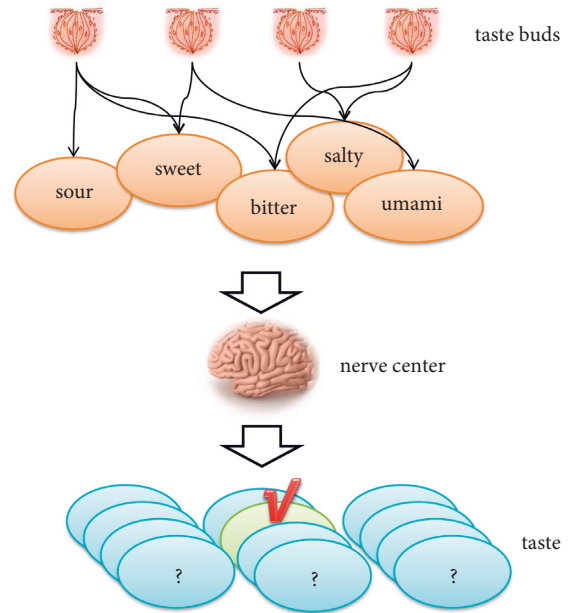


Figure 1: Nerve center determines taste stimuli.

in wine and trained a prediction model using the PLS method. The results showed that the method was effective in predicting the bisulphite concentration. PLS combined with linear discriminant analysis (LDA) was used in reference [22] to determine the taste of yellow wine (a Chinese rice wine). The data were used to analyze the taste characteristics of yellow wine from various regions.

## 3. Improved RF-Based Food Quality Detection Algorithm

*3.1. Traditional RF.* In the 1980s, Breiman et al. developed the classification tree algorithm, which was much less computationally intensive than repeatedly dichotomizing data for classification or regression. Breiman combined the classification trees into an RF in 2001 by randomly generating many classification trees and then aggregating the classification tree results. RF improves prediction accuracy without requiring significantly more computing power. RF is considered one of the best current algorithms because they are insensitive to multicollinearity. RF can deal with noisy and unbalanced data very well, which is one of the important reasons why the algorithm can be widely used.

RF selects $n$ samples from the $N$ training samples, which are typically much smaller than $N$, and repeats the process $C$ times to generate $C$ decision trees. When splitting each node of the decision tree, $d$ attributes are chosen at random from all $D$ attributes, and the best attribute from the $d$ attributes is chosen as the splitting attribute. The CART algorithm is commonly used to select the splitting attributes. The Gini index can determine the splitting attributes in this algorithm. The Gini value of $V$ is calculated by (1). Assuming that the proportion of the $i$-th category in the current sample set $V$ is $s_i$, where $S_i = 1$ and $k$ denotes the total number of categories.
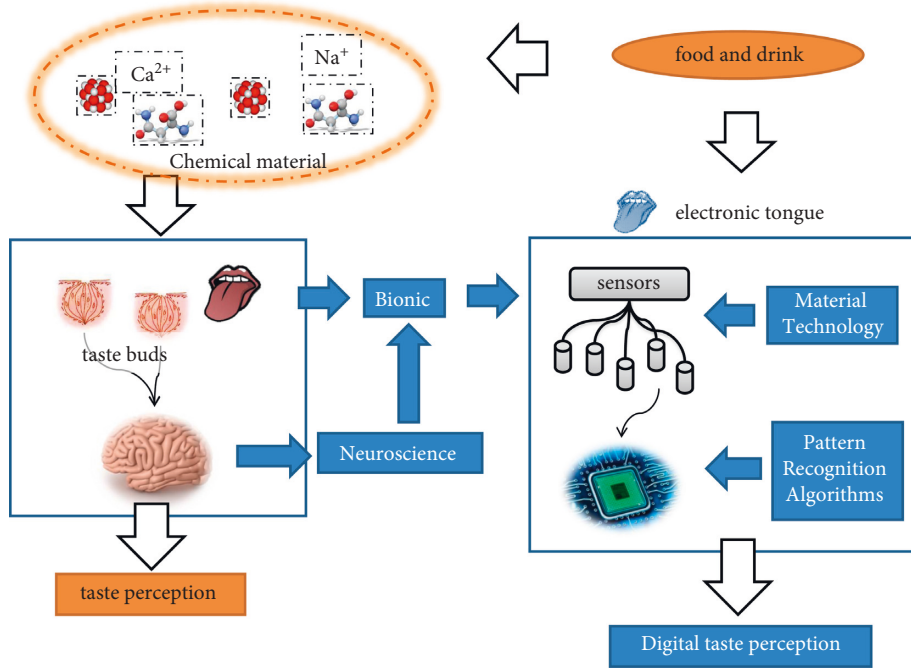
FIGURE 2: Artificial taste analysis.

$$\text{Gini}(V) = 1 - \sum_{i=1}^{|k|} s_i^2. \tag{1}$$

The Gini value is taken to be between 0 and 1. The test attribute $\beta$ is used to dichotomize the random variable $V$ into two categories $V_1$ and $V_2$, and then, the Gini index for attribute a is calculated as follows:

$$\text{Gini}(V, \beta) = \frac{|V_1|}{|V|}\text{Gini}(V_1) + \frac{|V_2|}{|V|}\text{Gini}(V_2). \tag{2}$$

Assuming that the set of candidate attributes is $(B = \beta_1, \beta_2, \ldots, \beta_d)$, then the attribute that minimizes the Gini index after division is selected as the optimal divided attribute.

$$\beta^* = \arg\min(\text{Gini\_index}(V, \beta)). \tag{3}$$

*3.2. Improving RF.* Traditional RF generates decision trees with high similarity, rendering the trained model obsolete. Furthermore, in the voting process, all decision trees are given the same weight by default, ignoring any differences in performance between different decision trees. To address these issues, this paper proposes IRF that incorporates mutual information and a weighting strategy to optimize the RF. The correlation between any two decision trees is calculated using mutual information, and only the one with the highest evaluation accuracy is retained for the more correlated decision trees, forming a new RF and converting the evaluation accuracy into voting weights to obtain a RF model with less redundancy and higher evaluation accuracy.

Mutual information is used to measure the interdependence between two variables. For two sets of given random variables $X_1$ and $X_2$, their mutual information is expressed as

$$MI(X_1, X_2) = L(X_1) \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} p(x_1, x_2)\log\frac{p(x_1, x_2)}{p(x_1)p(x_2)},$$
$$= L(X_1) + L(X_{21}) - L(X_1, X_2), \tag{4}$$

where $p(x_1, x_2)$ is the joint probability distribution of $X_1, X_2$, $p(x_1)$ and $p(x_2)$ are the marginal probability distributions of $X_1, X_2$, respectively, and $L(X_1)$ is the information entropy of $X_1$, which is calculated as

$$L(X_1) = -\sum_{x_1 \in X_1} p(x_1)\log p(x_1), \tag{5}$$

where $p(x_1)$ denotes the probability of event $x$ occurring; $L(X_2)$ is the information entropy of $X_2$, and $L(X_1, X_2)$ is the joint entropy, which is calculated as

$$L(X_1, X_2) = -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} p(x_1, x_2)\log p(x_1, x_2). \tag{6}$$

For a decision tree $l_i$ ($i = 1, 2, \ldots, C$) in a RF, $MI(l_i, l_c)$ ($c \neq i$) represents the mutual information of the decision tree $l_i$ and $l_c$. In this paper, $MI(l_i, l_c)$ ($c \neq i$) is used to calculate the correlation between decision trees $l_i$ and $l_c$. The formula is

$$MI(l_i, l_c) = MI(z_i, z_c), \tag{7}$$

where $z_i$ ($i = 1, 2, \ldots, C$) is the $i$-th decision tree's output state. The greater the value of $MI(l_i, l_c)$, the greater the correlation between the two decision trees, and the greater the degree of information overlap. A group of decision trees with mutual information values greater than a threshold $\varepsilon$ will be

combined by calculating the mutual information of any two decision trees. $MI(l_i, z)$ represents the mutual information between the decision tree $l_i$ and the actual label $z$, that is, the correlation between the decision tree $l_i$'s output evaluation result and the actual evaluation result. The equation is

$$MI(l_i, z) = MI(z_i, z). \qquad (8)$$

The larger the value of $MI(l_i, z)$, the higher the evaluation accuracy of the decision tree $l_i$. Finally, the decision trees with less correlation and higher accuracy are formed into an IRF. The steps of IRF implementation are as follows:

(1) Step 1: get the training set $X = \{(x_1, y_1), (x_1, y_1), \ldots, (x_N, y_N)\}$, validation set $X' = \{(x_1, y_1), (x_1, y_1), \ldots, (x_L, y_L)\}$, $x_i$ is the sample, and $y_i$ is label.

(2) Step 2: draw $n$ samples from $X$ by bootstrap sampling and repeat $T$ times to obtain $T$ training sets.

(3) Step 3: for each internal node of the decision tree, $d$ attributes are randomly selected from the sample $D$ attributes, and the best splitting attribute among the $d$ attributes is chosen as the test attribute of the node.

(4) Step 4: generate $T$ completely random decision trees to form a RF $R = \{l_1, l_2, \ldots, l_t\}$.

(5) Step 5: input the validation set $X'$ into $R$ to get the result.

(6) Step 6: based on the results, the correlation between each decision tree and the rest of the decision trees is calculated in turn. All the decision trees with $MI(l_i, l_c)(i \neq c)$ greater than the threshold $\varepsilon$ are combined into one decision tree group. If the correlation between $h$, and all other decision trees is less than or equal to the threshold $\varepsilon$, then the decision trees are grouped separately.

(7) Step 7: repeat step 6 for the out-group decision trees in order until all decision trees are grouped.

(8) Step 8: obtain the decision tree with the highest accuracy in each group based on the accuracy $MI(l_i, z)$.

(9) Step 9: form a new RF from the obtained decision trees $R' = \{l_1', l_2', \ldots, l_t'\}$.

When evaluating the input sample data, the traditional RF has the same voting weight for each decision tree, which ignores the influence of the evaluation accuracy of different decision trees on the final result and reduces the evaluation accuracy of the RF as a whole. In order to increase the number of decision trees with high evaluation accuracy and reduce the influence of decision trees with low evaluation accuracy on the final evaluation results, this paper proposes a weighted voting method to convert the evaluation accuracy into the voting weights of decision trees. After the traditional RF is streamlined, the evaluation accuracy matrix $A$ of decision trees in the IRF is

$$A = \begin{bmatrix} a_{11} & a_{11} & \cdots & a_{11} \\ a_{21} & a_{22} & \ddots & a_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{G1} & a_{G2} & \cdots & a_{GC} \end{bmatrix}, \qquad (9)$$

where $a_{gc}$ denotes the evaluation accuracy of the $c$-th decision tree for the $g$-th running state, where $c = 1,2, \ldots, C$, $g = 1,2, \ldots, G$, $G$ is the number of classes of evaluation results, and $C$ is the number of improved decision trees which obtained by substituting the validation set $X'$ into IRF and calculating the correctness of the output of each decision tree for each class.

Based on the evaluation accuracy matrix, the weight matrix $Q$ is defined as

$$Q = \begin{bmatrix} q_{11} & q_{11} & \cdots & q_{11} \\ q_{21} & q_{22} & \ddots & q_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ q_{G1} & q_{G2} & \cdots & q_{GC} \end{bmatrix}, \qquad (10)$$

where $Q_{gc}$ is the weight of the $p$ th tree for the $g$-th running state, which is calculated as

$$q_{gc} = a_{gc}. \qquad (11)$$

The specific flow of the IRF is shown in Figure 3.

## 4. Experimental Analysis

*4.1. Experimental Background.* In recent years, there has been an increase in the number of counterfeit and shoddy products. These products often have identical outer packaging, but the quality of the product varies. This paper uses a taste-based food quality detection method. To verify the effectiveness of the method, data from different brands of vinegar are collected and analyzed, and the brands of the respective products are identified. If the results obtained by the RF algorithm for a product captured do not belong to an existing product class, this indicates that there is a problem with the quality of the product. Based on this idea, a Chinese food product vinegar was experimentally analyzed in the experimental part. The evaluation metrics used during the experiments to quantify the effectiveness of the classification models used in this paper are shown in Table 1.

Since the sample size collected during the experiment was not very large, a 5-fold cross-validation method was used to ensure the validity of the experiment, and 60% of the sample overview was selected as the training sample and 40% of the sample as the test sample. The comparison algorithms used were all classical classification algorithms, mainly SVM [23], RF (RF) [24], radial basis function neural network (RBFNN) [25], and Naive Bayes (NB) [26].

*4.2. Vinegar Quality Inspection Experiment.* The taste of different types of table vinegar is different because the raw materials and processes used to make them are different. Edible vinegar is an acidic condiment that is often used when people cook ingredients. Protein, fat, carbohydrates, ash, water, energy, riboflavin, sodium, potassium, calcium, magnesium, iron, manganese, zinc, copper, phosphorus, and selenium are the main components of table vinegar. Leucine, isoleucine, lysine, methionine, cystine, phenylalanine, tyrosine, threonine, tryptophan, arginine, histidine, alanine, valine, glycine, proline, glutamic acid, serine, and aspartic
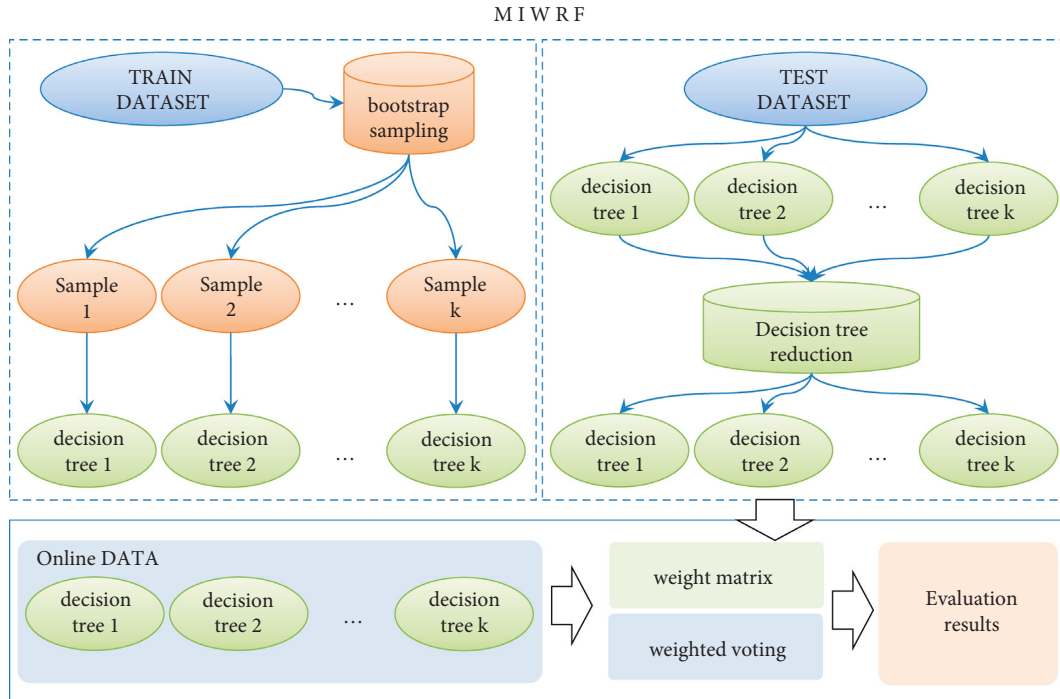
M I W R F



FIGURE 3: IRF flow chart.

TABLE 1: Introduction to evaluation indicators.

| Name | Expressions |
|---|---|
| Accuracy | Accuracy = (Number of correctly classified samples/total number of samples) |
| Precision | Precision = (true sample number/true sample number + false positives number) |
| Recall | Recall = (true sample number/true sample number + false negatives number) |
| $F1$ score | $F1 = 2^*\text{Precision} * \text{Recall}/\text{Precision} + \text{Recall}$ |

TABLE 2: Quantitative indicators for table vinegar.

| Index\Type | Vinegar 1 | Vinegar 2 | Vinegar 3 | Vinegar 4 | Vinegar 5 |
|---|---|---|---|---|---|
| $I1$ (%) | $3.86 \pm 0.57$ | $7.55 \pm 0.73$ | $14.51 \pm 0.28$ | $14.64 \pm 0.63$ | $7.17 \pm 0.81$ |
| $I2$ (%) | $4.82 \pm 0.62$ | $6.46 \pm 1.03$ | $5.45 \pm 0.53$ | $5.35 \pm 0.55$ | $4.65 \pm 1.02$ |
| $I3$ (%) | $0.94 \pm 0.38$ | $1.28 \pm 0.84$ | $1.82 \pm 0.44$ | $1.79 \pm 0.84$ | $1.42 \pm 0.87$ |
| $I4$ (%) | $0.80 \pm 0.92$ | $1.17 \pm 091$ | $2.66 \pm 0.38$ | $2.74 \pm 0.97$ | $1.54 \pm 0.53$ |
| $I5$ | $2.53 \pm 0.86$ | $2.49 \pm 0.40$ | $3.32 \pm 0.64$ | $3.54 \pm 0.63$ | $3.38 \pm 0.62$ |

acid are all found in it. In this paper, a total of five types of edible vinegar were selected: white vinegar, rice vinegar, aged vinegar, balsamic vinegar, and sugar vinegar. The quantitative indexes of each edible vinegar are shown in Table 2. $I1$, $I2$, $I3$, $I4$, and $I5$ represent total soluble sugar, total acid, salt, sugar-acid ratio, and pH, respectively. The sugar-acid ratio was determined by averaging total sugar and total acid.

Principal component analysis was conducted on the five component indicators of the above five types of vinegar, and the principal component score plots are shown in Figures 4 and 5. The variance contribution of the first principal component was 56.28 percent, the variance contribution of the second principal component was 25.93 percent, and the cumulative variance contribution of the first and second principal components was 83.54 percent, explaining the majority of the variance. Therefore, the first 2 principal components can only represent 82.21% of the information contained in the original 5 physical and chemical indicators. In addition, the five types of samples could not be well separated on the first principal component score axis, only white vinegar and sugar vinegar could be distinguished from the other types of vinegar, and there was great dispersion between the different brands of rice vinegar, while aged vinegar and balsamic vinegar were seriously confused with each other, which indicated that aged vinegar and balsamic vinegar had some similarity, while white vinegar and sugar vinegar were very different from the other types and showed specificity, and the samples of rice vinegar were more different from each other. It is difficult to group them together because of the relatively large differences between the samples of rice vinegar. This indicates that the first principal
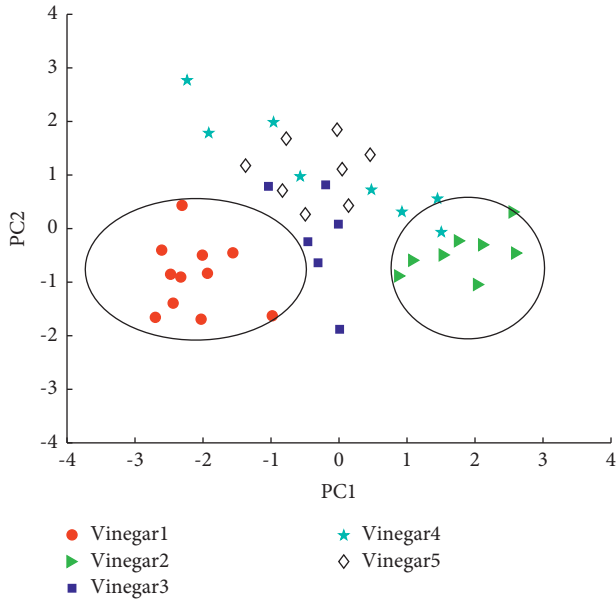
Figure 4: Schematic diagram of principal component analysis of component indicators.
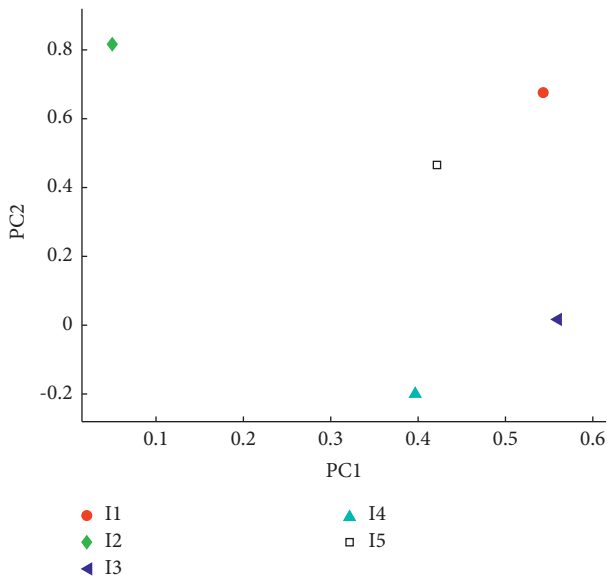


Figure 5: Factor loading diagram.

Table 3: Classification results of different models for vinegar.

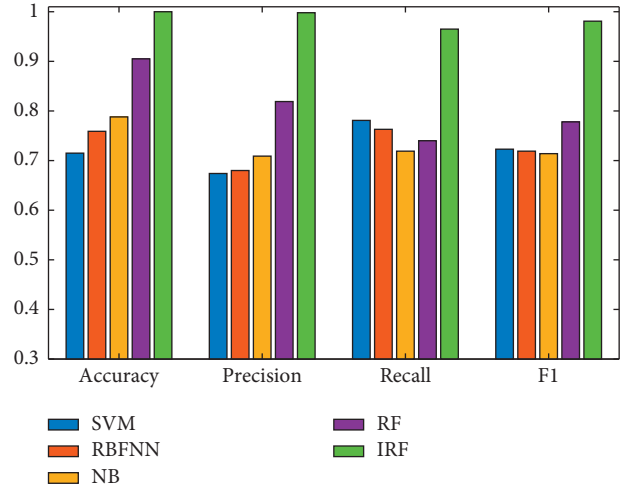| Model\Index | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| SVM | 71.49% | 67.35% | 78.10% | 72.33% |
| RBFNN | 75.87% | 67.99% | 76.25% | 71.88% |
| NB | 78.83% | 70.89% | 71.92% | 71.40% |
| RF | 90.53% | 81.93% | 74.03% | 77.78% |
| IRF | 100% | 99.83% | 96.47% | 98.12% |



Figure 6: Comparison of classification results of different models.

obtained from the physicochemical indicators can be interpreted as acidity (sourness) and that the second principal component is positively correlated with acidity. In general, the results of the principal component analysis obtained from the physicochemical indicators were mainly related to the brix and acidity characteristics of vinegar, which is not unrelated to the fact that vinegar is an acidic condiment. However, these five physicochemical indicators alone are not sufficient to distinguish the different types of vinegar.

Sixty samples were taken from each category of edible vinegar. The total number of samples was 300. Eight indicators were collected for each type of edible vinegar. These 8 indicators were collected by 8 sensors. The size of the samples was $300 * 8$. Since there were 5 types of edible vinegar in total, the number of categories was 5. The results of the experiment are shown in Table 3 and Figure 6.

In terms of recognition accuracy, the IRF method used in this paper improves 0.40, 0.32, 0.27, and 0.10 compared to other SVM, RBFNN, NB, and RF. The improvement ratios fully demonstrate the superiority of this method. In terms of accuracy, the IRF method improves 0.48, 0.47, 0.41, and 0.22 over SVM, RBFNN, NB, and RF, respectively. The high $F1$ scores indicate that the model used in this paper has a good recognition rate for both positive and negative samples. The value range of the four indicators used in this paper is between 0 and 1. The larger the value, the better the performance of the model. The experimental results show that the IRF algorithm has the largest value on the four indicators and is far ahead of other comparison models. The experimental results verify the effectiveness of the IRF algorithm in the detection of edible vinegar quality.

component obtained from the physicochemical indicators can be interpreted as the sugar content (sweetness) and that the first principal component is positively correlated with the sugar content. Furthermore, the distribution from left to right is roughly in the ascending order of the sugar-acid ratio, indicating that the first principal component can also be interpreted as the sugar-acid ratio and that the first principal component is positively related to the sugar-acid ratio. On the second principal component score axis, the five types of samples are also differentiated, and intuitively, the five types are distributed on the second principal component score axis from bottom to top in an ascending order of total acidity, which indicates that the second principal component

## 5. Conclusion

There have been numerous issues with food quality and safety in recent years. Although the government has introduced relevant policies, it has not been able to eliminate all kinds of counterfeit and shoddy products. On the other hand, many food products are prone to deterioration during storage. If people do not detect these food products in time, they can be harmful to their health. This paper proposes the use of an improved RF algorithm to improve the speed and accuracy of food quality detection. The algorithm uses mutual information to calculate the correlation of any two decision trees, and for the more correlated decision trees, only the decision tree with the highest evaluation accuracy is retained, thus forming a new RF, converting the evaluation accuracy into voting weights, and finally obtaining an RF model with less redundancy and higher evaluation accuracy. The experimental results show that the method used in this paper can successfully identify spoiled or counterfeit products. However, although the algorithm in this paper is more intelligent than traditional food detection methods, there are some problems. One is that the model used does not achieve the desired results for high-dimensional data. Second, it is unknown how well the food testing method proposed in this paper can detect the quality of other food products, as the core ingredients of different food products are different. The determination of the core components has a great impact on the final identification effect. For the above two problems, this paper will be followed by an in-depth study.

## Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

## Disclosure

Xinghua Zhang and Yongjie Sun are the same work, and they share the same article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

## References

[1] N. Casado, S. Morante-Zarcero, and I. Sierra, "The concerning food safety issue of pyrrolizidine alkaloids: an overview," *Trends in Food Science & Technology*, vol. 120, pp. 123–139, 2022.

[2] M. O. Caglayan, S. Sahin, and Z Ustundag, "Detection strategies of zearalenone for food safety: a review," *Critical Reviews in Analytical Chemistry*, vol. 52, no. 2, pp. 294–313, 2022.

[3] L. Goel, S. Raman, S. S. Dora et al., "Hybrid computational intelligence algorithms and their applications to detect food quality," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1415–1440, 2020.

[4] Y. Liu, H. Pu, and D. W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends in Food Science & Technology*, vol. 113, pp. 193–204, 2021.

[5] P. S. Mkwae, S. A. Ogundipe, M. Jozela, N. Revaprasadu, and S. S. Nkosi, "The heat rate kinetics on the liquefied hydrocarbon gases sensing and food quality control detecting strategy," *Materials Chemistry and Physics*, vol. 277, Article ID 125550, 2022.

[6] W. Zhang, J. Ma, and D. W. Sun, "Raman spectroscopic techniques for detecting structure and quality of frozen foods: principles and applications," *Critical Reviews in Food Science and Nutrition*, vol. 61, no. 16, pp. 2623–2639, 2021.

[7] M. Creydt and M. Fischer, "Blockchain and more - algorithm driven food traceability," *Food Control*, vol. 105, pp. 45–51, 2019.

[8] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of deep learning in food: a review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 18, no. 6, pp. 1793–1811, 2019.

[9] K. Yao, J. Sun, C. Chen et al., "Non-destructive detection of egg qualities based on hyperspectral imaging," *Journal of Food Engineering*, vol. 325, Article ID 111024, 2022.

[10] M. Benouis, L. D. Medus, M. Saban, G. Łabiak, and A. Rosado-Muñoz, "Food tray sealing fault detection using hyperspectral imaging and PCANet," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 7845–7850, 2020.

[11] O. Bader, D. Haddad, A. Y. Kallel, T. Hassine, N. E. B. Amara, and O. Kanoun, "Identification of communication cables based on scattering parameters and a support vector machine algorithm," *IEEE Sensors Letters*, vol. 5, no. 7, pp. 1–4, 2021.

[12] B. Sowmia and G. Suseendran, "Anomaly detection in business process event using KNN algorithm," *Lecture Notes in Networks and Systems*, vol. 248, pp. 189–199, 2021.

[13] S. Shrivastava, R. Singh, C. Jain, and S. Kaushal, "A research on fake news detection using machine learning algorithm," *Smart Innovation, Systems and Technologies*, vol. 235, pp. 273–287, 2022.

[14] G. Manjula, G. Govinda Rajulu, R. Anand, and J. T. Thirukrishna, "Implementation of smart parking application using IoT and machine learning algorithms," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 75, pp. 247–257, 2022.

[15] R. G. Siqueira, G. V. Veloso, E. I. Fernandes-Filho, M. . R. Francelino, C. E. G. R. Schaefer, and G. R. Corrêa, "Evaluation of machine learning algorithms to classify and map landforms in Antarctica," *Earth Surface Processes and Landforms*, vol. 47, no. 2, pp. 367–382, 2022.

[16] T. H. H. Aldhyani and H. Alkahtani, "Attacks to automatous vehicles: a deep learning algorithm for cybersecurity," *Sensors*, vol. 22, no. 1, p. 360, 2022.

[17] D. Baskakov and D. G. Arseniev, "On the computational complexity of deep learning algorithms," *Smart Innovation, Systems and Technologies*, vol. 220, pp. 343–356, 2021.

[18] M. İnce, "Automatic and intelligent content visualization system based on deep learning and genetic algorithm," *Neural Computing & Applications*, vol. 34, no. 3, pp. 2473–2493, 2022.

[19] A. Legin, A. Rudnitskaya, Y. Vlasov, C. Di Natale, E. Mazzone, and A D'Amico, "Application of electronic tongue for quantitative analysis of mineral water and wine," *Electroanalysis*, vol. 11, pp. 814–820, 1999.

[20] V. Parra, A. A. Arrieta, J. A. Fernandez-Escudero, M. Iniguez, J. A. Saja, and M. L Rodriguez-Mendez, "Monitoring of the ageing of red wines in oak barrels by means of an hybrid electronic tongue," *Analytica Chimica Acta*, vol. 563, no. 1-2, pp. 229–237, 2006.

[21] M. Gutiérrez, A. Llobera, A. Ipatov et al., "Application of an e-tongue to the analysis ofmonovarietal and blends of white wines," *Sensors*, vol. 11, no. 5, pp. 4840–4857, 2011.

[22] H. Yu, J. Zhao, F. Li, H. Tian, and X Ma, "Characterization of Chinese rice wine taste attributes using liquid chromatographic analysis, sensory evaluation, and an electronic tongue," *Journal of Chromatography B*, vol. 997, no. 1, pp. 129–135, 2015.

[23] N. Narkhede, S. Mathur, and A. A. Bhaskar, "Automatic classification of music genre using SVM," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 75, pp. 439–449, 2022.

[24] B. Hamdaoui, A. Elmaghbub, and S. Mejri, "Deep neural network feature designs for RF data-driven wireless device classification," *IEEE Network*, vol. 35, no. 3, pp. 191–197, 2021.

[25] G. M. Rajathi, "Optimized radial basis neural network for classification of breast cancer images," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 17, no. 1, pp. 97–108, 2021.

[26] E. Hato, "Impact of feature selection for data classification using naive Bayes classifier," *Journal of Physics: Conference Series*, vol. 1879, no. 2, Article ID 022088, 2021.