processes

MDPI

Article

Non-Destructive Analysis for Machine-Picked Tea Leaf Composition Using Near-Infrared Spectroscopy Combined Chemometric Methods

Qinghai Jiang, Bin Chen, Jia Chen and Zhiyu Song *

Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China; jiangqinghai101@126.com (Q.J.)
* Correspondence: songzhiyu@caas.cn

Abstract: This paper aimed to predict the mechanical composition of machine-picked fresh tea leaves (MPFTLs) using near-infrared spectroscopy (NIRS) rapidly and non-destructively. Samples of MPFTL with different mechanical composition ratios were collected and subjected to NIRS analysis. Subsequently, various preprocessing methods were employed to eliminate extraneous noise information. Next, characteristic spectral information was extracted using the backward interval partial least squares (biPLS) method, which was subsequently subjected to principal component analysis (PCA). Finally, a predictive model was constructed by applying the back propagation artificial neural network (BP-ANN) method, which was tested by external samples to assess its predictive efficacy, and the results were expressed as root mean square error and determination coefficient of prediction (R_p^2). The optimal spectral pretreatment method was the following: (standard normal variate (SNV) + second derivative (SD)). Four characteristic spectral subintervals of ([2, 3, 7, 10]) were screened out, and the cumulative contribution rate of 95.20%, attributable to the first three principal components, was determined. When the tanh transfer function was applied to construct the BP-ANN-NIRS model, the results demonstrated optimal performance, exhibiting a root mean square error and a determination coefficient of prediction (R_p^2) of 0.976 and 0.027, respectively. The absolute values of prediction deviation for all prediction set samples were found to be less than 0.04. The results of the best BP-ANN model for external samples were found to be in close agreement with those of the prediction set model. NIRS technology has successfully achieved the forecasting of the mechanical composition of machine-picked fresh tea leaves rapidly and accurately, providing a fair and convenient new method for purchasing fresh tea raw materials by machines, according to their quality, and promoting the sustainable high-quality and healthy development of the tea industry.

Keywords: machine-picked fresh tea leaves; mechanical composition; near-infrared spectroscopy; principal component analysis; artificial neural network

Citation: Jiang, Q.; Chen, B.; Chen, J.; Song, Z. Non-Destructive Analysis for Machine-Picked Tea Leaf Composition Using Near-Infrared Spectroscopy Combined Chemometric Methods. *Processes* **2024**, *12*, 2397. https://doi.org/10.3390/pr12112397

Academic Editor: Cunshan Zhou

Received: 18 August 2024
Revised: 15 October 2024
Accepted: 25 October 2024
Published: 31 October 2024

1. Introduction

China is the country with the longest history of tea production and consumption, and the discovery and utilization of tea can be traced back by 4000 years [1,2]. The tea industry is a distinctive advantage for China, supplying a vital income source for tea farmers in mountainous regions. It also has a pivotal role in maintaining the ecological balance of green water and green mountains. The tea industry's primary objective is to sustain the economy of tea-producing areas, meet consumer demand, and provide stable and expanding employment opportunities. The national tea industry development plan explicitly emphasizes the importance of standardizing tea leaf picking and processing techniques, as well as enhancing digitalization levels. This is viewed as a significant scientific and technological force that has the potential to drive transformation and advancement within the tea industry [3].

Presently, tea processing has gradually been mechanized and automated. But machine-picked fresh tea leaves have emerged as a key area of interest and a significant challenge within the context of tea industry research [4]. In recent years, the tea industry has faced significant challenges due to the increasing shortage of labor and rising labor costs. The difficulty of tea picking and the rising costs of picking have posed considerable obstacles to the industry's healthy development. The implementation of mechanical techniques for picking tea leaves has the potential to result in significant cost and labor savings in the tea leaf-picking process. This approach effectively addresses the challenge of insufficient tea pickers. Furthermore, the technological advancements and corresponding equipment utilized in this process have reached a level of maturity that has contributed to the emergence of mechanized tea leaf picking as a novel trend [5]. Currently, scholars have conducted extensive research on machine-picked fresh tea leaves [6]. Xiao Xing [7] employed a vibrating fresh leaf screening machine to investigate the impact of machine-picked tea fresh leaves on grading. The optimal leaf feeding rate was determined to be 5 kg per time, with an hourly yield reaching 300 kg. Yang Juan [8] has employed a first-machine selection of fresh leaves prior to sorting, which enhanced the quality of tea processing. Ye Fei [9] conducted a study in which machine-picked and hand-picked fresh tea leaves were applied to make Enshi Yulu tea. The findings indicated that the utilization of processing machine-picked fresh leaves in the production of Enshi Yulu tea not only resulted in a reduction in the costs associated with manual picking, but also ensured the quality of the tea produced and enhanced the efficiency with which the fresh leaves could be processed.

Liu Mingli [10] highlighted that the implementation of machine picking mechanisms throughout the entirety of the tea production process has emerged as a prevailing trend. The aforementioned research works have established a robust scientific and technological foundation for machine-picked fresh tea leaves.

Due to factors such as tea garden management level, tea tree varieties, topography, and equipment, machine-picked fresh tea leaves often have problems including uneven length and tenderness, low uniformity, a high stem content, and so on. This will have an impact on the cost of purchasing fresh tea leaves and the selection of subsequent processing procedures [11]. When tea farmers sell machine-picked fresh tea leaves to tea processing plants, the purchasing personnel typically rely on sensory perceptions, such as odor, vision, and tactile experience, as well as personal experience, to determine the mechanical composition ratio of the fresh tea raw materials, subsequently giving the corresponding purchase price. However, the sensitivity of human sensory organs is susceptible to external influences, including the individual's work background, current physiological status, and surrounding environmental factors such as temperature, humidity, and weather conditions. This subjectivity can lead to discrepancies between the purchasing personnel and tea farmers, which could have adverse effects on the growth of the tea industry.

Generally speaking, to reasonably determine the quality grade of MPFTLs, it is necessary to first determine their mechanical composition [12]. It can be observed that there is a positive correlation between the mechanical composition values and the integrity of the fresh leaves. Furthermore, it can be seen that the grade of machine-picked fresh leaves is directly proportional to the purchase price. Nevertheless, this approach is inherently time-consuming, labor-intensive, and costly in terms of both human resources and materials. Consequently, machine-picked fresh leaves are unable to be withered and subsequently processed in a timely manner. It is therefore imperative to develop a method for objectively, fairly, and rapidly predicting the mechanical composition of machine-picked fresh leaf materials.

Near-infrared spectroscopy (NIRS) is a technique that employs electromagnetic waves with a wavelength in the range of 780–2526 nm, which primarily reflect the X–H chemical bond. It offers the benefits of rapid and non-destructive analysis and has become a prevalent tool in various fields, including agriculture [13–15], the petrochemical industry, the textile industry, and the pharmaceutical industry [16,17]. NIRS has been employed extensively to predict the concentrations of polyphenols, caffeine [18], and other compounds in tea [19], evaluate the quality of fresh tea leaves [20], and differentiate between tea varieties [21]. Nevertheless, there is currently a paucity of works in the literature examining the utilization of near-infrared spectroscopy technology for the prediction of the mechanical composition of machine-picked fresh tea leaves. Consequently, further research is required in this area.

In this study, the fresh leaves of 43 Longjing varieties in the tea garden of Jiangsu Xinpin Tea Industry Co., Ltd. (Changzhou, China) were taken as the research object, and the fresh leaves were machine-picked by a single tea picker, and then the mechanical composition indexes of the raw materials were calculated. On this basis, fresh leaves with different mechanical compositions underwent NIRS. After preprocessing the spectra, partial least squares (PLS), backward interval partial least squares (biPLS), principal component analysis (PCA), and artificial neural network methods were applied to construct a prediction model of the mechanical composition of fresh leaf raw materials. The practical applicability of the model was evaluated through the use of external samples. This study facilitates the rapid assessment of the mechanical composition of fresh leaf raw materials at the acquisition site. Furthermore, the objective is to present a novel approach that offers a fair and convenient means of procuring machine-picked fresh leaf raw materials, with the aim of overcoming the influence of subjective human factors. This will lay a solid foundation for the next step in the development of machine-picked fresh leaf mechanical composition using a portable near-infrared spectrometer.

2. Materials and Methods

2.1. Fresh Tea Leaf Samples and Classification

The fresh leaves were machine-harvested by employing a single Maiyue tea picker (ME-LJ02-330) (Maiyue Co., Ltd., Shenzhen, China) from 9 June 2024 to 10 June 2024, and then the fresh leaves were mixed well, from which about 100 g of representative fresh leaf samples was weighed, and the samples were categorized and weighed according to the criteria of (one bud and one leaf + one bud and two leaves) and (one bud and three leaves+ one bud and four leaves) to compute the ratios between the two. The samples were classified into two distinct categories, namely a calibration set and a prediction set, based on their mechanical compositions. The ratio of the two sets was 3:1, with 60 samples in the calibration set and 20 samples in the prediction set. The samples of the calibration set were applied to construct an NIRS prediction model, while the samples of the prediction set were applied to check the model's predictive efficacy. Moreover, an additional 20 external samples were utilized to assess the efficacy of the predictive model in practice.

2.2. Methods

2.2.1. Near-Infrared Spectra Acquisition

The NIR spectra of fresh leaf samples were scanned using an Antaris II Fourier transform near-infrared spectrometer (Thermofisher Scientific, Waltham, MA, USA), with a spectral range of 4000 cm^{-1} to 10,000 cm^{-1} and a 8 cm^{-1} resolution, coupled with an InGaAs detector. Before scanning the spectra, it was essential to ensure that the instrument was activated and allowed to stabilize for a period of approximately one hour. During the scanning process, fresh leaf samples were loaded into the sample cups that accompanied the instrument and the spectra were scanned using diffused reflectance. To ensure that the full NIR spectral information could be captured for each fresh leaf sample, during the scanning process, a complete rotation of 360° was performed on the sample cup. The loading of each sample was repeated three times, and three spectra were scanned for each sample; then, the sample spectra were averaged, the results of which were used as the final spectrum for that sample (Figure 1).

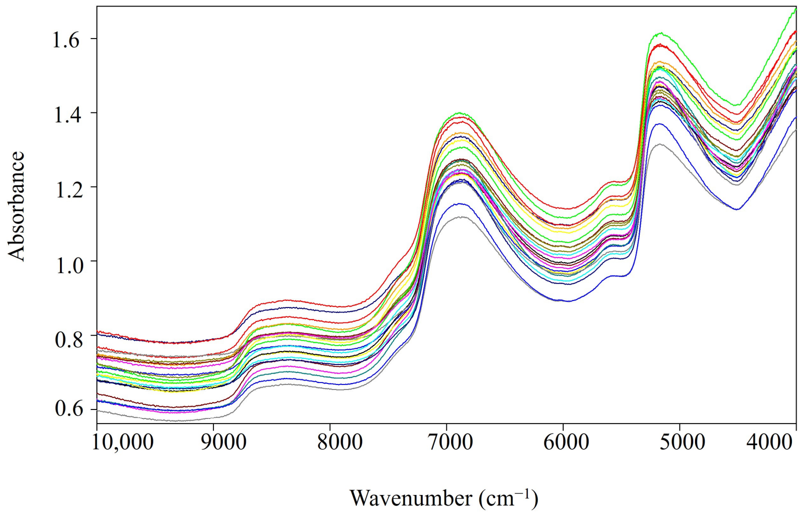

Figure 1. NIRS of machine-picked fresh tea leaves.

2.2.2. Spectral Data Analysis

Each spectrum was transformed into 1557 paired data points, which were stored in an Excel spreadsheet with a data point interval of 3.86 cm^{-1}. The data were then subjected to analysis through the application of TQ Analyst 9.4.45 software and Matlab 2012a software, respectively. In order to achieve the effective removal of a substantial quantity of noise information and the improvement of the signal-to-noise ratio of the spectra, the effects of standard normal variate (SNV), multiple-scatter correction (MSC), first-order derivative (FD), second-order derivative (SD), and their combination of spectral preprocessing methods were subjected to comparison, with a view to selecting the optimal spectral preprocessing method.

2.2.3. Modeling Methods

The biPLS method [22] was an effective wavelength-filtering method. The spectral data were segmented into ten to twenty-five spectral subintervals, after which the partial least squares (PLS) method was employed to develop the models. The most accurate spectral data were obtained when the root mean square error of cross-validation (RMSECV) of the model was at its lowest point.

The RMSECV calculation was as follows:

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i' - y_i\right)^2}{n}} \tag{1}$$

Notably, n is the number of samples in the calibration set; y_i is the true value of sample i; and $y_i{}'$ is the predicted value of sample i in the calibration set.

The obtained optimal spectral subintervals were subjected to PCA, and the NIRS model for the mechanical composition of machine-picked fresh leaves was built by applying the BP-ANN algorithm with the principal components as the input values and the mechanical compositions of machine-picked fresh leaves as the output values, and the findings were quantified using four statistical parameters: the determination coefficient of cross-validation ($R_c{}^2$), the determination coefficient of the prediction set ($R_p{}^2$), the RMSECV, and the root mean square error of prediction (RMSEP). A model with a larger R^2 and a smaller RMSEP was considered to have better predictive capabilities than others.

The RMSEP calculation was as follows:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{n}} \tag{2}$$

Notably, n is the number of samples in the prediction set; y_i is the true value of sample i; and $y_i{}'$ is the predicted value of sample i in the prediction set.

The R^2 calculation was as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i' - y_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{3}$$

Notably, y_i and $y_i{}'$ are the true and predicted values of sample i, respectively, and $\overline{y}$ is the average of the true values for all samples.

3. Results and Discussion

3.1. Analysis of Mechanical Compositions

The mechanical composition of fresh tea leaves refers to the percentage of different tenderness levels of buds and leaves in a single batch of freshly picked leaves from the tree. The mechanical composition of fresh tea leaves serves as the raw material grading standard, providing a foundation for the development of a reasonable tea production process. The analysis method involves a weighing of buds and leaves, with the weight percentage of 100 g of fresh leaves in different standards of buds and leaves being calculated. As can be seen from Table 1, the range of mechanical composition of all samples was from 0.10 to 0.88, like the range of mechanical composition in the calibration set samples, also from 0.10 to 0.88, and the range of mechanical composition in the prediction set samples was from 0.12 to 0.70. The mechanical composition range of the prediction set samples was aligned with that of the calibration set samples, indicating that the classification of the modeling samples was justified, providing the foundation for the development of a robust mechanical composition prediction model for fresh tea leaves.

Table 1. The mechanical composition results of machine-picked fresh tea leaves.

Index	All Samples	Calibration Set	Prediction Set
Max	0.88	0.88	0.70
Min	0.10	0.10	0.12
Average ± SD	0.52 ± 0.26	0.48 ± 0.31	0.41 ± 0.18

Note: Max: maximum; Min: minimum; and SD: standard deviation.

3.2. Comparison of Spectral Pretreatment Methods

Figure 1 shows that there were more spectral absorption peaks in the range of 4000 cm^{-1}–7000 cm^{-1}, which mainly reflected the NIRS information of water-OH in the machine-picked fresh tea leaves, while the spectral information reflecting the mechanical composition of fresh tea leaves was relatively weak and likely masked by the spectral information of water-OH and other spectral data, which, inevitably, might have resulted in the poor prediction of the NIRS mechanical composition model [23]. In this study, nine spectral pretreatment methods were employed to remove extraneous information, and a prediction model was constructed using the PLS method. When the RMSECV was at its minimum value and Rc^2 was at its maximum, the preprocessing method was at its most effective. The results of the nine preprocessing methods are shown in Table 2.

As illustrated in Table 2, the NIRS model obtained without any spectral preprocessing (i.e., the "none" method) yielded the poorest results, with $R_c{}^2$ and RMSECV values of 0.526 and 0.465, respectively. Among the models constructed using a single preprocessing method, the NIRS model built with the MSC preprocessing method exhibited the most favorable outcomes, with $R_c{}^2$ and RMSECV values of 0.685 and 0.440, respectively, but the results were still relatively poor. Among the models of combined preprocessing methods, the best NIRS model results were obtained by the (SNV + SD) combined preprocessing method, with $R_c{}^2$ and RMSECV of 0.732 and 0.434, respectively. Compared with the model results without a spectral preprocessing method, the $R_c{}^2$ was improved by 39.2%, and the RMSECV was decreased by 6.7%. Therefore, the preprocessing of raw spectra prior to modeling can effectively remove some of the noise information, consistent with previous research [24]. Although the results of the NIRS model of mechanical composition were improved after preprocessing, these results remained unsatisfactory and could not satisfy the requirement of accurately predicting different mechanical compositions of MPFTLs. It was thus imperative to conduct a more rigorous screening of the feature spectra data in order to enhance the predictive efficacy of the model.

Table 2. The performance of nine pretreatment methods.

Pretreatment Methods	Rc^2	RMSECV
None	0.526	0.465
SNV	0.551	0.460
FD	0.623	0.451
SD	0.652	0.445
MSC	0.685	0.440
SNV + FD	0.713	0.439
SNV + SD	0.732	0.434
MSC + FD	0.659	0.442
MSC + SD	0.648	0.456

3.3. Feature Spectral Interval Screening

In this study, the biPLS method was employed for the purpose of screening feature spectral intervals. Once all spectral data had been divided into 16 distinct spectral subintervals, the minimum RMSECV was 0.316, obtaining characteristic spectral intervals closely related to the mechanical composition of fresh tea leaves. The results are presented in Table 3.

Table 3. Characteristic spectral regions selected by the biPLS method.

No.	Number of Spectral Subintervals	R_c^2	RMSECV
16	13	0.733	0.433
15	14	0.742	0.420
14	11	0.764	0.408
13	12	0.793	0.360
12	16	0.801	0.358
11	5	0.805	0.355
10	1	0.808	0.352
9	4	0.811	0.350
8	6	0.817	0.348
7	15	0.826	0.347
6	8	0.832	0.342
5	9	0.835	0.331
4	10	0.840	0.316
3	7	0.778	0.391
2	3	0.759	0.415
1	2	0.731	0.439

Note: R_c^2 is the determination coefficient of cross-validation, while RMSECV is the root mean square error of cross-validation.

As shown in Table 3, once the biPLS models were established, it became evident that, as the number of modeled spectral subintervals gradually increased, the results of the model showed a trend of first gradually improving and then gradually deteriorating. When only one subinterval ([2]) was applied to build the model, the results were the worst; at this time, the R_c^2 and RMSECV were only 0.731 and 0.439, respectively. When four spectral subintervals ([2, 3, 7, 10]) were used for modeling, the best model results were obtained at this time, with R_c^2 and RMSECV of 0.840 and 0.316, respectively. The spectral data corresponding to the four spectral subintervals were from 4377.6 cm^{-1} to 4751.7 cm^{-1}, 4755.6 cm^{-1}–5129.7 cm^{-1}, 6262.7 cm^{-1}–6633.9 cm^{-1}, and 7386 cm^{-1}–7756.3 cm^{-1}, respectively, and the proportion of the characteristic spectral data to the total spectral data was found to be 25.00%. In light of the aforementioned findings, it can be posited that, when establishing a model, the spectral data input into the model were not necessarily less or more, but rather there existed relatively reasonable amounts of spectral data. Due to the presence of certain noise information in the spectral data, the results of the model could have been affected. Therefore, it was necessary to extract spectral intervals that were closely related to the modeling indicators as much as possible from numerous instances of spectral data [25]. Additionally, the biPLS method [26] can greatly reduce the spectral data information of the input model and improve the prediction accuracy of the model. The optimal biPLS model, as indicated in Table 3, exhibited a 14.75% higher R_c^2 value and a 27.19% lower RMSECV value in comparison to the optimal PLS model. Consequently, the application of the biPLS method for the selection of characteristic spectral intervals not only reduced the quantity of spectral data necessary for modeling but also enhanced the predictive efficacy of the NIRS model. This benefit can mitigate the complexity and reinforce the robustness of the model, providing a robust foundation for the subsequent step to build a BP-ANN mechanical composition model for machine-picked fresh tea leaves.

3.4. Principal Component Analysis (PCA)

The characteristic spectral intervals obtained in Section 3.3 were subjected to PCA, and the results are presented in Table 4.

Table 4. Cumulative contribution rate of the five initial principal components.

PCs	PC1	PC(1–2)	PC(1–3)	PC(1–4)	PC(1–5)
Cumulative contribution rate/%	82.45	91.65	95.20	96.14	96.90

Note: PCs are the principal components.

Table 4 shows that the contribution of the five initial principal components decreased rapidly after performing PCA on the characteristic spectral intervals. Among them, PC1 had the highest contribution rate at 82.45%, followed by PC2 at 9.20%, PC3 at 1.95%, PC4 at 0.94%, and PC5 at 0.80%, and the cumulative contribution rate of the three initial principal components was 95.20%. In accordance with the PCA principle [27], the data pertaining to the initial three principal components are sufficient to represent the entirety of the information contained within the characteristic spectral intervals, which can be used in the next step to build the BP-ANN prediction model.

3.5. Establishment of BP-ANN Model

An NIRS mechanical composition prediction model for fresh tea leaves was developed by applying the BP-ANN method with the input values of the first three PCs and the mechanical compositions of fresh tea leaves as the output values. In building the model, due to the different transfer functions applied between each transfer layer, the anticipated outcomes of the model also exhibited discrepancies. A learning rate of 0.1 was set, and the results of three transfer functions were compared. They were the linear [0, 1] function, the logistic function, and the tanh function. The prediction results of the model are presented in Table 5.

Table 5. Results of BP-ANN models using three distinct transfer functions.

Transfer Functions	Calibration Set		Prediction Set	
	R_c^2	RMSECV	R_p^2	RMSEP
linear [0, 1]	0.882	0.246	0.853	0.284
logistic	0.931	0.104	0.908	0.129
tanh	0.987	0.022	0.976	0.027

Table 5 shows that the prediction results of the BP-ANN models using three transfer functions were different. The model built using linear [0, 1] functions yielded the worst results, with R_p^2 and RMSEP of 0.853 and 0.284, respectively. This may have been due to the high moisture content in fresh tea leaves, which can mask some information about the mechanical composition of fresh tea leaves. Additionally, the spectra contained a large amount of spectral information from buds, the first leaf, the second leaf, the third leaf, the fourth leaf, and tender stems, resulting in extremely rich spectral information. There was a significant overlap and superposition of information, and the nonlinear relationship between the mechanical composition of the sample and the spectral information was evident. Therefore, the prediction results of the BP-ANN-NIRS model for the mechanical composition of fresh tea leaves built using the linear [0, 1] transfer function were the worst. However, due to the good nonlinear characteristics of logistic function and tanh function, the prediction results of the models built using these two transfer functions were superior to those of model with the linear [0, 1] transfer function. The BP-ANN model, constructed with the tanh transfer function, exhibited the highest degree of accuracy in predictive performance (R_p^2 = 0.976, RMSEP = 0.027), and the model was also the most robust, meaning that it could accurately predict the prediction set samples. The results of the prediction set model are presented in Figure 2.

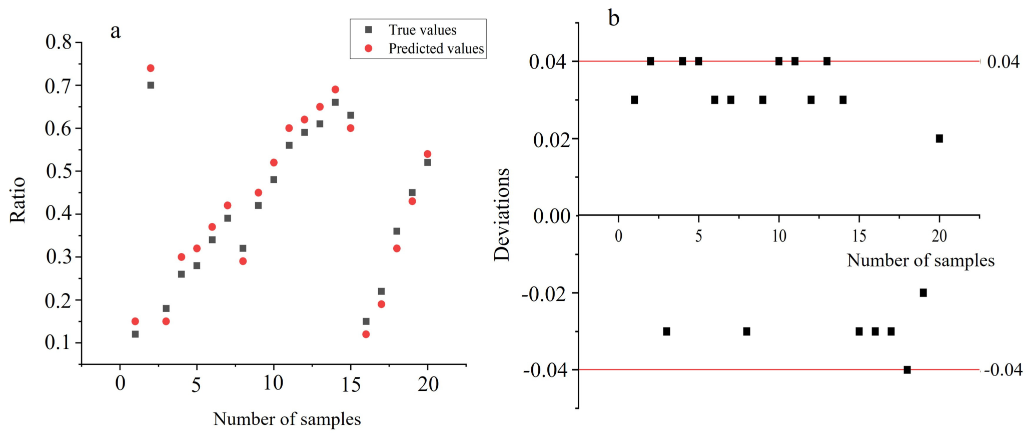

Figure 2. The results of the prediction set samples using the best BP-ANN model: (**a**) the true values vs. the predicted values of the prediction set samples by the best BP-ANN model; and (**b**) the deviations in samples in the prediction set.

Figure 2 shows that, upon testing the robustness of the calibration set model using 20 samples from the prediction set, it was observed that the true and predicted values were almost identical, with a maximum absolute prediction deviation of 0.04. This indicates that the BP-ANN model exhibited a high prediction accuracy and did not exhibit any overfitting phenomena. Furthermore, the application of the tanh function to the BP-ANN model enabled the accurate prediction of the mechanical composition of fresh tea leaves.

3.6. Verification of the Actual Application Effect of the Model

To test the actual prediction performance of the BP-ANN model, the mechanical composition of 20 external machine-picked fresh tea leaf samples was predicted. The results are shown in Table 6 and Figure 3.

As shown in Table 6 and Figure 3, the results demonstrate that the optimal BP-ANN model is an effective tool for accurately predicting the mechanical composition of 20 external fresh tea leaf samples (R^2 = 0.975, RMSEP = 0.027). The results of the predictive analysis are in close alignment with the outcomes of the prediction set model, suggesting that the BP-ANN model, when configured with a tanh transfer function, is capable of accurately predicting the mechanical composition of machine-picked fresh tea leaves.

Table 6. The mechanical composition of 20 external samples by the best BP-ANN model.

Numbers	True Values	Predicted Values	Numbers	True Values	Predicted Values
1	0.13	0.16	11	0.53	0.55
2	0.67	0.68	12	0.57	0.59
3	0.20	0.18	13	0.62	0.60
4	0.27	0.30	14	0.64	0.62
5	0.30	0.33	15	0.60	0.64
6	0.32	0.29	16	0.18	0.21
7	0.35	0.31	17	0.24	0.26
8	0.33	0.30	18	0.35	0.38
9	0.40	0.42	19	0.47	0.50
10	0.46	0.44	20	0.55	0.58

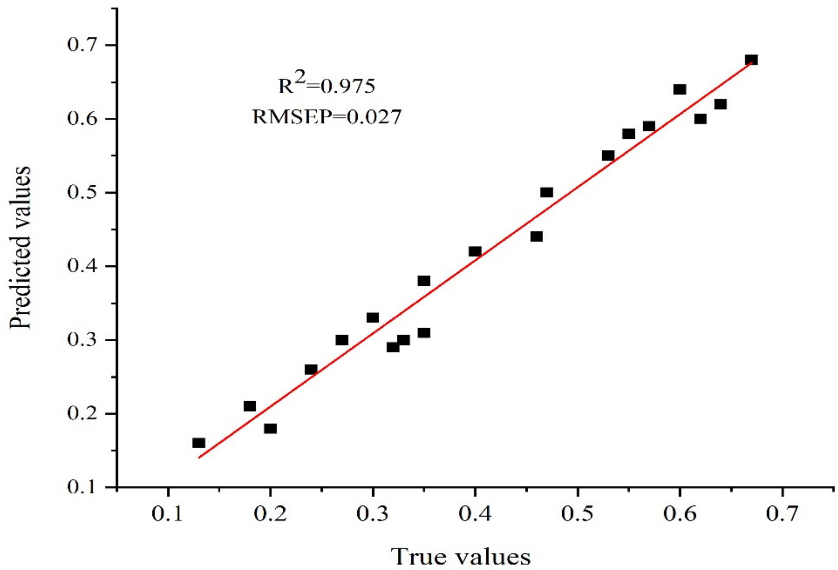

Figure 3. Prediction results of 20 external samples.

4. Conclusions

Machine harvesting fresh tea leaves is the current trend in the tea industry. As the summer and autumn tea leaves harvested by machines are mainly used for processing common tea and fresh leaves mainly come in groups of one bud and three leaves or one bud and four leaves, it is necessary to predict the mechanical composition of fresh leaves in a timely, fast, and accurate manner when purchasing raw materials based on quality, which would also be conducive to the sustained high-quality development of the tea industry. In this paper, the NIRS technique, the biPLS method, PCA, and the BP-ANN method were combined to build a robust prediction model of the mechanical composition of machine-picked fresh tea leaves (R_p^2 = 0.976, RMSEP = 0.027), which could quickly and accurately predict the mechanical composition values of fresh tea leaves in a few seconds without destroying the samples, presenting a novel approach to quality assurance based on the acquisition of mechanically harvested fresh tea leaves. At the same time, in future applications, by utilizing the selected characteristic spectral ranges, a substantial quantity of superfluous spectral data could be excluded. The targeted development of near-infrared spectrometers composed of detection machines for machine-picked fresh tea leaves could also be achieved, without having to use full-wavelength NIRS detectors, which could reduce the research and development costs of the instruments and facilitate their early deployment. Furthermore, to improve the precision and scope of the model, a diverse set of machine-harvested fresh leaf samples from various tea tree varieties and geographical regions, collected over multiple years, were obtained in this study. This expanded the existing model database, enhancing the model's adaptability and robustness.

Author Contributions: Conceptualization, B.C.; software, J.C.; writing—original draft preparation, Q.J.; and funding and writing—review and editing, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by an institute-level basic scientific research business expenses project of the Chinese Academy of Agricultural Sciences (S202404), the earmarked fund for CARS-19, and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-NIAM).

Data Availability Statement: The data are available from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ren, G.X.; Fan, Q.Y.; He, X.J.; Li, W.C.; Tang, X.L. Applicability of multifunctional preprocessing device for simultaneous estimation of spreading of green tea, withering of black tea and shaking of Oolong tea. *J. Sci. Food Agric.* **2020**, *100*, 560–569. [CrossRef] [PubMed]
2. Mei, Y.; Liang, X. Report on the Production and Sales Situation of Chinese Tea in 2023. Available online: https://www.sohu.com/a/772130441_307138 (accessed on 15 May 2024).
3. China Tea Circulation Association. Suggestions for the 14th Five Year Development Plan of China's Tea Industry (2021–2025). Available online: https://m.yunnan.cn/system/2021/04/01/031370270.shtml (accessed on 19 December 2023).
4. Luo, Y.P.; Tang, M.; Cai, W.Z.; Wen, D.H.; Wen, Z.J. Study on the optimum machine-plucking period for high quality tea. *J. Tea Sci.* **2008**, *28*, 9–13.
5. Tang, X.L.; Li, W.C.; Fan, Q.Y. Research on the classification technology and the grading equipment of machine-plucking fresh leaves. *Chin. Tea Process.* **2015**, *22*, 5–8.
6. Chen, Y.Z. The development and future direction of national excellent tea in China. *Chin. Teas* **1993**, *5*, 9–10.
7. Xiao, X.; Yin, L.Q.; Sun, Y.M.; Liang, M.Z.; Li, Z.; Chen, H.Y.; Liu, D.H. Preliminary study on vibrating screen for machine-plucking fresh tea leaves grading effect. *Shanxi Agric. Sci.* **2016**, *44*, 1867–1869.
8. Yang, J.; Li, Z.L.; Yuan, L.Y.; Liu, S.Q.; Wu, X.H.; Sheng, Z.L. Preliminary study on grading technology of machine-plucking fresh tea leaves. *Chin. Tea Process.* **2015**, *22*, 41–45.
9. Ye, F.; Gao, S.W.; Gong, Z.M.; Cui, Q.M.; Dai, J.H. Comparative study on the processing of Enshi Yulu by fresh tea leaves of machine-picked and hand-picked. *Zhejiang Agric. Sci.* **2016**, *56*, 1131–1133, 1136.
10. Liu, M.L.; Peng, Y.; Xu, Y.Y.; Liu, N.; Zeng, X.; Li, M.Y. Research status of mechanized picking mechanism of premium tea in China. *Food Ind.* **2023**, *44*, 287–291.
11. Yang, J.; Xu, Z.; Yuan, L.Y. Research progress on grading technology of machine harvested tea fresh leaves. *Fujian Tea* **2014**, *2*, 2–4.
12. Xing, Y.; Tang, S.H.; Mei, J.F.; Xu, D.L. Comparative study on machine harvested fresh tea leaves of different tea tree varieties. *Chin. Teas* **2020**, *4*, 36–39.
13. Khan, A.; Munir, M.T.; Yu, W.; Young, B.R. Near-infrared spectroscopy and data analysis for predicting milk powder quality attributes. *Int. J. Dairy Technol.* **2021**, *74*, 235–245. [CrossRef]
14. Silva, L.K.R.; Jesus, J.C.; Onelli, R.R.V.; Conceição, D.G.; Santos, L.S.; Ferrão, S.P.B. Discriminating Coalho cheese by origin through near and middle infrared spectroscopy and analytical measures. Discrimination of Coalho cheese origin. *Int. J. Dairy Technol.* **2021**, *74*, 393–403. [CrossRef]
15. Fagnani, R.; Damião, B.M.C.; Trentin, R.P.S.; Vanot, R.L. Predicting adulteration of grated Parmigiano Reggiano cheese with Ricotta using electrophoresis, multivariate nonlinear regression and computational intelligence methods. *Int. J. Dairy Technol.* **2022**, *75*, 239–245. [CrossRef]
16. Guillemain, A.; Dégardin, K.; Roggo, Y. Performance of NIR handheld spectrometers for the detection of counterfeit tablets. *Talanta* **2017**, *165*, 632–640. [CrossRef] [PubMed]
17. Malegori, C.; Marques, E.J.N.; Freitas, S.T.; Pimentel, M.F.; Pasquini, C.; Casiraghi, E. Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta* **2017**, *165*, 112–116. [CrossRef]
18. Yan, S.H. Evaluation of the composition and sensory properties of tea using near infrared spectroscopy and principal component analysis. *J. Near Infrared Spectrosc.* **2005**, *13*, 313–325. [CrossRef]
19. Wang, S.P.; Liu, P.P.; Feng, L.; Teng, J.; Ye, F.; Gui, A.H.; Wang, X.P.; Zheng, L.; Gao, S.W.; Zheng, P.C. Rapid determination of tea polyphenols content in Qingzhuan tea based on near infrared spectroscopy in conjunction with three different PLS algorithms. *Food Sci. Technol.* **2022**, *42*, e94322. [CrossRef]
20. Wang, S.P.; Zhang, Z.Z.; Ning, J.M.; Ren, G.X.; Yan, S.H.; Wan, X.C. Back propagation-artificial neural network model for prediction of the quality of tea shoots through selection of relevant near infrared spectral data via synergy interval partial least squares. *Anal. Lett.* **2013**, *46*, 184–195. [CrossRef]
21. Wu, X.H.; He, F.; Wu, B.; Zeng, S.P.; He, C.Y. Accurate classification of Chunmee tea grade using NIR spectroscopy and fuzzy maximum uncertainty linear discriminant analysis. *Foods* **2023**, *12*, 541. [CrossRef]

22. Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. Interval partial least squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 413–419. [CrossRef]
23. Chu, X.L. *Chemometric Methods in Modern Spectral Analysis*, 2nd ed.; Chemical Industry Press: Beijing, China, 2021.
24. Li, Y.J.; Altaner, C.M. Effects of variable selection and processing of NIR and ATR-IR spectra on the prediction of extractive content in *Eucalyptus bosistoana* Heartwood. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **2019**, *213*, 111–117. [CrossRef]
25. Lu, W.Z. *Modern Near Infrared Spectroscopy Analytical Technology*, 2nd ed.; China Petrochemical Press: Beijing, China, 2010.
26. Feng, H.; Hu, Y.J.; Ge, Y.; Dai, R.Y.; Tian, Z.; Cui, E.H.; Wu, H.; Zhang, Y.W. BiPLS-RF: A hybrid wavelength selection strategy for laser induced fluorescence spectroscopy of power transformer oil. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2024**, *321*, 124693.
27. Wolfgang, H.; Leopold, S.; Chen, S.Y., Translators; *Applied Multivariate Statistical Analysis*, 2nd ed.; Peking University Press: Beijing, China, 2014.