Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/50/20241477

Neural network-based Raman spectral recognition model for milk powder and wheat flour

Peichen Li[1,4,5], Mingxuan Wu[2], Weibing Wu[3]

[1]Wuhan Foreign Language School, Wuhan, Hubei, China
[2]Chongqing Nankai High School, Chongqing, China
[3]Kunming No.10 High School, Kunming, Yunnan, China

[4]Aaron2690639041@ldy.edu.rs
[5]Corresponding author

Abstract. The problem of counterfeit milk powder has occurred in recent years, and one of the standard techniques is to adulterate milk powder with wheat flour to reduce the cost. Since the morphology and colour of wheat flour are similar to that of milk powder, it is not easy to distinguish it from the naked eye, so a rapid method for detecting wheat flour adulteration in milk powder is urgently needed. This study used Raman spectral analysis as a novel and efficient screening technique. With the advantages of high sensitivity, non-destructive detection, low cost, no contamination, and online analysis, this method can play an essential role in rapid food detection and production control. By processing and analyzing the spectral data, a neural network model for identifying milk powder and wheat flour based on Raman spectroscopy has been developed in this study, and the model's performance has been evaluated. The results show that the model can quickly and accurately identify the proportion of wheat flour in milk powder to determine whether milk powder is adulterated or not. The results of this study are of great significance in ensuring the safety of infant milk powder. In the future, the method can be further optimized, for example, to improve detection speed and accuracy and to provide a more reliable quality control method for the milk powder industry.

Keywords: Reinforcement learning, Raman spectroscopy, neural network, rapid method for detecting.

1. Introduction

As an essential source of nutritional intake for infants and young children in addition to breast milk, milk powder supplements infants and young children with all kinds of substances they need, and its safety importance is self-evident. However, nowadays, fake and shoddy products mix into the market from time to time [1]. Many unscrupulous traders adulterate milk powder with wheat flour [2]. Given that the market price of wheat flour is usually lower than the market price of milk powder, and the shape and colour of the two are the same, it is difficult to see the difference through the naked eye so that unscrupulous traders can make illegal profits to a great extent. Therefore, finding a rapid method to determine the adulteration of wheat flour in milk powder is significant. In the testing methods promulgated by the state, the determination of starch in food is by enzymatic hydrolysis, and the determination of sucrose includes high-performance liquid chromatography and acid hydrolysis [3].

Although the traditional laboratory monitoring methods have high detection sensitivity, they could be more convenient for sample processing, time-consuming, labor-intensive, and costly, making it difficult to screen milk powder in large quantities rapidly. Therefore, developing a new and efficient screening technology is urgently needed. As a high-sensitivity rapid analysis method developed in recent years, Raman spectroscopy has been widely used in agriculture, pharmaceuticals, biochemicals, petroleum products, and other fields. This analytical method has the superiority of non-destructive detection, low cost, no pollution, and on-line analysis, a rapid analytical method that is better than the traditional laboratory food testing method and more suitable for production control [4]. It is worth mentioning that this technique organically combines the functions of small-scale micro-area scanning and large-scale area screening, which can efficiently collect information on milk powder samples for research and modelling [5].

In this paper, for the collected samples of multiple groups of milk powder and wheat flour, the Raman spectrometer was used to scan them in the range of 100-2500 cm-1, and the spectral curves of each sample contained a total of 240 wave number points spaced at ten wave number intervals. In this study, milk powder and flour are taken as the objects of study, and a neural network model for identifying milk powder and wheat flour based on Raman spectroscopy is developed. The research objective of this paper is to quickly identify the proportion of wheat flour in milk powder, thereby determining whether the milk powder is adulterated.

2. Collection

2.1. Data sources

In this experiment, Raman spectral data were obtained from two papers, "Agricultural Sciences in China, ASA" and "Simultaneous detection of multiple adulterants in dry milk using macro-scale Raman chemical imaging Jianwei Qin, Kuanglin Chao, Moo," in which the authors purchased the samples from local shops, which are generally representative and increase the credibility of the data [6, 7].

2.2. Data extraction tools

This study used getdata software for extraction, Getdata is professional and powerful graphical digitisation software. Adopting advanced automated numerical algorithms, it not only can quickly carry out the transformation of data, but also can help users to quickly convert pictures in many formats into vector graphics, extract pictures and icons from articles, and extract digital information from them, supporting four image formats: BMP, JPEG, and PCX. Because of its powerful features and convenient characteristics so in this data extraction is part of the selection of this software.

The extraction steps are as follows:

(1) Import the image

Next, this image is extracted. Use the first tool to import from the toolbar in the upper left corner of Figure 1.

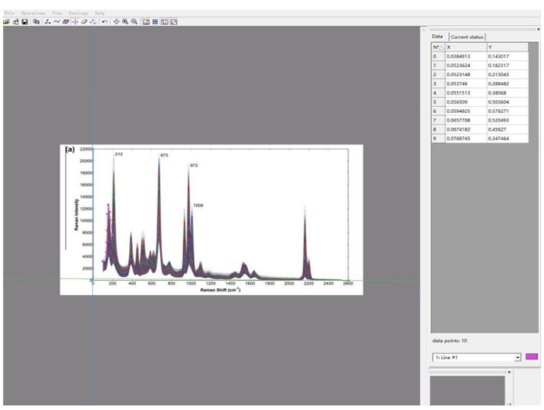

Figure 1. Getdata software interface.

(2) Fetch Points

Next, this study must label the x and y axes to pick up the points. Figure 3 from the upper left corner of the toolbar to find the fourth tool to mark out the x and y axes, the sixth and eighth tools to take the data point

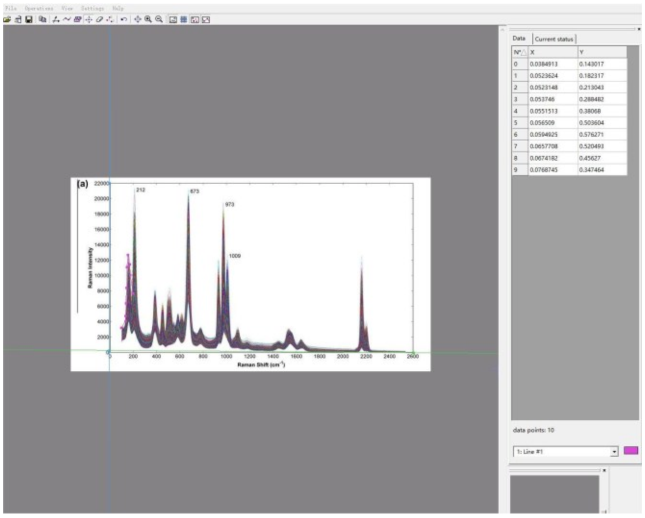

Figure 2. Picture of the data with x and y axes labelled and points taken.

(3) Exporting data

Next, this study requires exporting the data.

Open File from the top of the toolbar in Figure 2 and find Export data to get Figure 4.

(4) Data checking

Finally, this study needs to bring the data back to MATLAB for checking to avoid any error in taking the points.

3. Data preprocessing and neural network building and training

3.1. Data preprocessing

3.1.1. Adding perturbation. The 15 sets of unprocessed data obtained through Getdata in this study were expanded with random perturbations. Since the data is not extensive, the random perturbation number range here is set in the range of -0.2~0.2. The data were successfully expanded from 15 to 75 by adding the perturbation.

3.1.2. Generating training and test sets. In this study, the author set the training set to 65 random data and the test set to 10 data to reflect the correctness of the model while ensuring a sufficient database.

3.1.3. Normalisation of data. After randomly generating the training and test sets, the data is mapped to the interval [-1,1] using normalisation to make the neural network converge faster and shorten the training time

3.2. Neural network establishment and training

3.2.1. Neural network creation and training. Use the new function to create a feedforward neural network model net, where P_train and T_train are the input and target output of the training set. Then the training parameters: the number of iterations for training, the error target, and the learning rate are set by modifying the parameters in the net. trainParam structure [8].

DOI: 10.54254/2755-2721/50/20241477

For accuracy, the number of iterations chosen in this experiment is 1000, the error target is 0.001, and the learning rate is 0.01.

3.2.2. Simulation. After the neural network model is constructed and programmed, the next key step is to train the model using the training function. The training function uses the backpropagation algorithm to optimise the weights and biases of the model to fit the training data better.

In the training function, the author inputs the training dataset into the model and compares it with the predicted results to calculate the error. A gradient descent algorithm then updates the model parameters to minimise the error. This iterative process will be repeated several times until the prediction results of the model reach the expected accuracy.

Once the training is complete, the trained model is assigned to the variable net. this allows the test set to be simulated using the net. By inputting the test set into the net, the simulation result t_sim is then obtained.

Finally, in order to improve the accuracy of the results, this experiment rounds the prediction result t_sim. This converts the continuous prediction results into discrete values to make them more realistic.

4. Results and discussion

4.1. Model performance.

This section introduces the performance of this BP neural network model, including its performance on the training set, validation set, and test set.

4.1.1. Training and Validation Losses. This study used parameter settings with a learning rate of 0.01 and a training period of 100 0 for training.

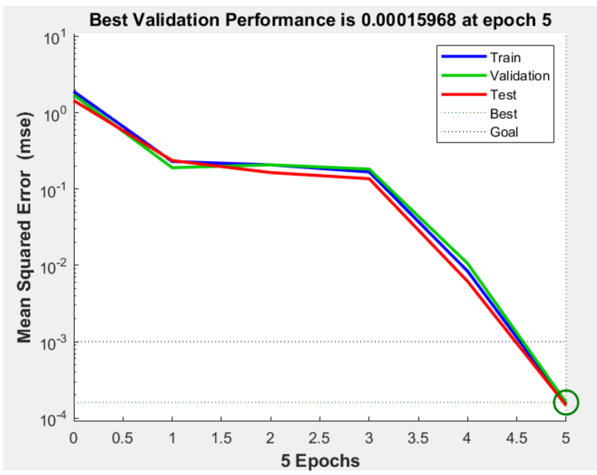

Figure 3. Training and validation loss curves.

Figure 3 shows that as the training period increases, the training loss and verification loss of the training set, verification set, and test set gradually decrease. The best mean square error is reduced to about 0.0002, lower than the required 0.001. This shows that this neural network model shows good fitting ability on both the training and validation sets, which also means that the model has high prediction accuracy.

4.1.2. Accuracy. Next, this study uses samples in the test set to verify the model classification accuracy.

DOI: 10.54254/2755-2721/50/20241477

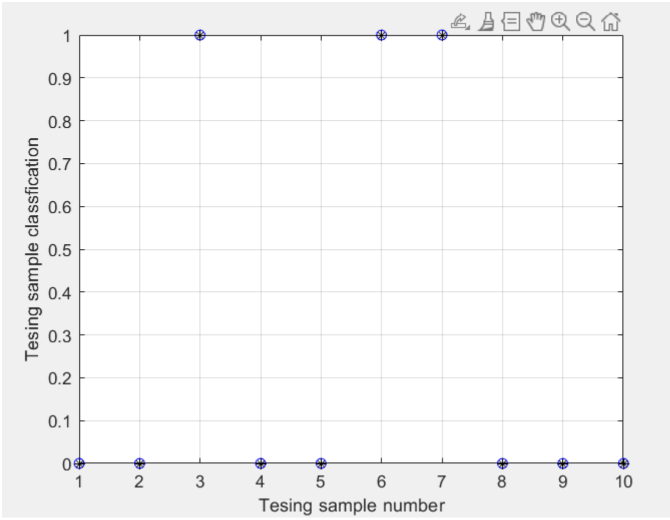

Figure 4. Classification diagram.

As shown in Figure 4, circles and asterisks represent calculated and actual values, respectively. All ten samples overlap, meaning the model's classification accuracy on the test set is 100 %, achieving extremely high accuracy. This means that the model can identify milk powder and wheat flour effectively.

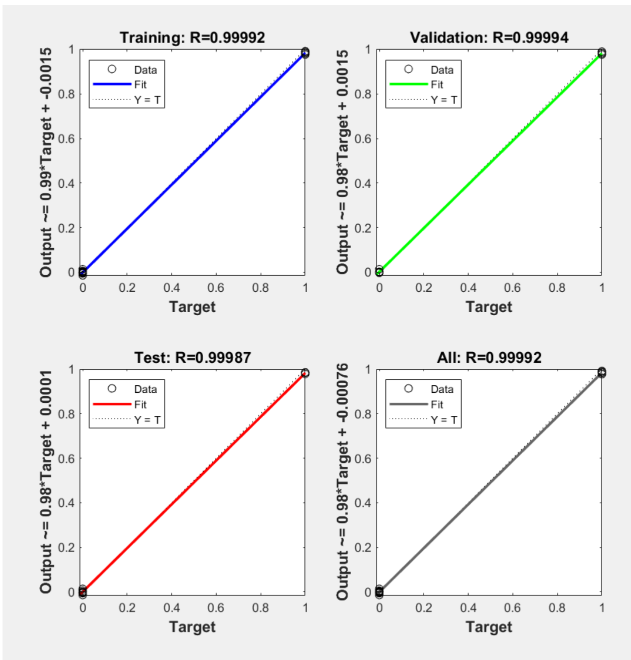

Figure 5. Regression.

It is not difficult to see from Figure 5 that this model's training set, test set, and verification set have extremely high similarity and fitting degree to the target, showing that its prediction accuracy is very high.

4.1.3. Gradient. Specifically, the gradient is the rate of change of the loss function relative to the model parameters. If in the graph of the gradient, the value of the gradient gradually decreases as the number of iterations increases, then this means that the model parameters will be adjusted in the direction of reducing the loss function in each iteration, so the value of the loss function will slow shrieking. As can be seen from the figure, the gradient graph of the model shows a downward trend, indicating that the model continues to improve during the training process and gradually approaches the optimal state.

4.2. Comparison with other methods

4.2.1. Other methods. Methods including mass spectrometry, high-performance liquid chromatography (HPLC), and gas chromatography (GC) electrochemical techniques are usually used to detect various components in milk powder, such as fat, protein, lactose, Additives, etc [9].

1) Biological method: This method is usually used to detect microbial contamination. By culturing and analyzing the microorganisms in the sample, it is possible to determine whether bacteria, mold, or other microbial contamination is present in the milk powder.

2) Physical property testing: The physical properties of milk powder, such as particle size, density, and fluidity, can be measured through physical testing methods.

3) Sensory testing: Professional tasters can taste the milk powder to detect any abnormal tastes or odors.

4.2.2. Comparison. Compared with traditional detection methods, the neural network method based on Raman spectroscopy to identify milk powder and wheat flour has the following advantages:

1) Non-destructive: Raman spectroscopy is a non-destructive analysis method, which means that the sample does not need to be chemically treated or destroyed so that the integrity of the sample can be maintained [10]. Traditional methods may require breaking down or changing the properties of the sample.

2) Rapidity: Raman spectroscopy detection can be completed quickly, while traditional methods may require more time for sample preparation, testing, and analysis.

3) Multivariate analysis: Raman spectroscopy can provide multivariate analysis and detect multiple chemical components simultaneously, so a more comprehensive understanding of the composition and properties of the sample can be obtained.

4) No chemical labels required: Unlike some traditional methods, Raman spectroscopy does not require chemical labels or stains to identify or analyze compounds. This reduces the complexity and cost of analysis.

5) Wide applicability: Raman spectroscopy suits various sample types, including solids, liquids, and gases [11]. It is also suitable for various application areas, from food safety to pharmaceuticals and materials science.

4.3. Model limitations and improvement directions

Although the neural network model in this experiment showed good performance in recognition, there are still some limitations and room for improvement.

Data volume issue: Current data sets are relatively small. To further improve the performance and generalization ability of the model, more Raman spectroscopy data can be collected and the diversity of the data set can be ensured.

Parameter optimization: The experimental model uses a set of initial parameters for training, but there may be better parameter combinations. The model's performance is optimized by systematically adjusting parameters such as learning rate, batch size, and number of hidden layer neurons.

5. Conclusion

In summary, this study constructed a BP neural network model based on Raman spectrum data and successfully identified milk powder and wheat flour. The model performs well on the test set with high accuracy and confidence. Raman spectroscopy data from different sources, including milk powder and wheat flour samples, were collected. Each sample contains measurements at multiple wavenumber points within the spectral range. These measurements are used as input features to train the model. To ensure the reliability of the model, this study also performed data preprocessing. This includes adding random perturbations to increase the diversity of the data, generating training and test sets for model evaluation, and normalizing the data to speed up the training process.

Compared with traditional detection methods, Raman spectroscopy technology provides a fast, accurate, and non-destructive method for food quality detection. However, there is still room for improvement, including expanding the size of the data set and further optimizing model parameters.

This study provides a valuable reference for the field of food safety and provides directions for future research. As a non-destructive analytical method, Raman spectroscopy technology can detect food samples without damaging them. In the future, this will continue to be an essential advantage for food quality testing, especially for high-value foods or those that need to preserve their original quality.

In addition, Raman spectroscopy technology can analyze multiple components simultaneously. In the future, it may become more accurate and able to detect more components, thereby providing more comprehensive food quality information. Moreover, with the development of machine learning and artificial intelligence, future Raman spectroscopy technology will better customize the processing of large amounts of data and conduct customized analysis and pattern recognition. This will help identify food quality issues, improving production processes more accurately. Raman spectroscopy technology has broad application prospects in food quality testing. With the continuous advancement and innovation of technology, it will continue to provide more reliable, efficient, and safe testing methods for the food industry, helping to ensure Food quality and safety.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Zhang S P, Lu J 2013 Research on the overall market development of China's dairy industry *China Dairy Industry* **41** 11 33-37

[2] Shi R T 2016 To explore the government's public opinion guidance strategy in public food safety incidents -- Taking the "fake milk powder" incident in Shanghai in 2016 as an example *Journalism Research Guide* **12** 334

[3] Xiong J F, Zhou G M, Xu L, Zhang L 2012 Determination of glucose, sucrose and lactose in milk powder by ion chromatography *Food Science* **33** 8 176-179

[4] Huang Y X, Tsai Y H, He W X 2023 Application of surface-enhanced Raman spectroscopy in food safety detection *Modern Food* **29** 13 93-98

[5] Chen D, Huang Z X, 2017 A new method for detection of adulterated milk powder by Raman spectral imaging *Han S Nanotechnology and Precision Engineering* **15** 01 26-30

[6] Zhao J 2016 Research on non-destructive real-time detection of harmful additives in flour based on Raman spectroscopy imaging *China Agricultural University*

[7] Qin J, Chao K, Kim S 2013 Simultaneous detection of multiple adulterants in dry milk using macro-scale Raman chemical imaging *M Food Chemistry* **138** 2-3

[8] Wang Y L 2021 Research on digital Recognition Technology based on BP neural Network *Shanxi Electronic Technology* **05** 77-80

[9] Wang Y, Bao S L, Sun L R, Do B X, Xin J Y, Zhang N 2023 Research progress in the application of electrochemical technology in food detection *Food Science* 1-22

[10] Liu Y 2022 Study on non-destructive testing of egg freshness by Raman spectroscopy *Jiangnan University*

[11] Wen Y Q, Li Z Y, Ning Y, Wang H Y 2023 Advances in the application of Raman spectroscopy in sensitive detection of odorant substances *Analytical Laboratory* 1-13