

Neural Network Model for Surface Raman Spectroscopy and Pesticide Concentration of Spinach

Yi Wang *

The Affiliated High School to Hangzhou Normal University, Hangzhou, China

* Corresponding Author Email: huxiaoshuang@cncitc.com.cn

Abstract. At present, there are over 1000 chemical pesticides. The widespread use of these pesticides not only causes serious environmental pollution, but also poses a threat to human health. Currently, there are over 1000 types of chemical pesticides. The widespread use of these pesticides has caused serious environmental pollution. Meanwhile, due to some workers not effectively removing pesticides from food during the picking and processing stages of the farm plants, so the long-term consumption of such food will also pose a threat to human health. An analytical technique called Raman spectroscopy is based on the Raman scattering effect, which was discovered by Indian scientist C.V. Raman. It is used in molecular structure research to get data on molecule vibration and rotation by analyzing scattering spectra at various frequencies from incident light. As a newborn way, the Raman spectroscopy analysis method has higher sensitivity. In this thesis, I put it and the BP neural network together and set up the model, to predict pesticide residues in spinach.

Keywords: BP neural network, pesticide residue, raman spectroscopy.

1. Introduction

There are over 1000 types of pesticides used worldwide, aimed at ensuring that food is not damaged or destroyed by pests. Each pesticide has different characteristics and toxicological effects. When people are exposed to large amounts of pesticides, the result may be acute poisoning or long-term health effects, which may include adverse effects on cancer and reproduction. The World Health Organization has two goals related to pesticides: 1. to ban the most toxic pesticides to humans and the pesticides that remain in the environment for the longest time. 2. to set maximum limits for pesticide residues in food and water to protect public health. [1]

The BP neural network consists of an input layer, a hidden layer, and an output layer. Input layer: The input point of information. Hidden layer: The middle part of information processing, which can change the number of layers of this hidden layer. Output layer: The output of information, the result. Compare the output results with the expected results. If two results match each other, it indicates that the hidden layer does not need to be modified. If the two results differ significantly, it indicates that the parameters we set are not appropriate and need to be modified. Because the signal of the BP neural network propagates forward and the error propagates backward, we need to calculate the two differences from back to front in order to change the parameters until the obtained input results match the expected values. At present, deep learning has been widely studied and applied. The introduction of deep learning has greatly improved the classification accuracy on image datasets. [2]

Artificial neural network is a new type of information processing system or computing system developed based on modern neuroscience research achievements and basic functional characteristics of the human brain, attempting to imitate the functions or structures of biological neural systems. [3] This system has its unique function in handling various fuzzy, random, large quantity, low precision information, etc. [4]

In China, there are still certain deficiencies in pesticide residue detection technology. Typical techniques include spectroscopy, chromatography, and enzyme inhibition. First, is the spectroscopy. The main application object of this technology is organic phosphorus pesticides, which undergo oxidation and other reactions with chromogenic agents in specific environments to produce specific wavelengths of color. This method has relatively low requirements for the treatment of the detection object, and this detection technology also has certain limitations, that is, the object is single and only

suitable for qualitative analysis. 2. Chromatography. This technology mainly includes gas chromatography, high-performance liquid chromatography, etc. Among them, gas chromatography is the most common method in the detection process. Its basic principle is to quantitatively analyze the performance of purified and concentrated Ops in the gas chromatography column. Compared with other technologies, this technology has a relatively high accuracy, and the tandem mass spectrometry detection derived from this method has higher accuracy and comprehensiveness in practice. 3. Enzyme inhibition method. The basic principle is to conduct pesticide residue analysis based on the inhibitory effect of pesticides containing organophosphorus and carbamate on acetylcholinesterase, and its application scope is relatively narrow. So, all in all, the Traditional laboratory testing techniques are not appropriate for quick testing on vegetable sales locations because to issues with large sample quantities, lengthy testing cycles, and high prices. Numerous industries, including agriculture, pharmaceuticals, biochemistry, and petroleum products, have adopted the Raman spectroscopy analysis technique, which has only recently been developed as a very sensitive and quick analytical technique. It is better suited for production control demands because to its benefits, which include non-destructive testing, low cost, pollution-free, and online analysis. That is which way used in this thesis.

2. Method

2.1. Collect information

2.1.1. Basic introduction

There are 5 kinds of ways to collect the information: 1. Observation method; 2. Survey questionnaire method; 3. Experimental method; 4. Interview method; 5. Literature method; Each method has its unique advantages and limitations, suitable for different scenarios and needs.

Observation can provide intuitive, real-time, and dynamic data, especially suitable for behavioral research. The survey questionnaire method is a common method of data collection, which distributes a list of questions to the respondents and then collects answers. This method can collect a large amount of data and is easy to quantitatively analyze. The experimental method collects data by controlling and manipulating variables in order to study the causal relationship between variables. Experimental methods can provide strong causal evidence. The interview method involves asking questions and obtaining answers orally or face-to-face. Interviews can obtain in-depth and detailed data, but may be influenced by the subjectivity of both the interviewer and the interviewee. Forms are mainly used to collect data in online applications. A form for collecting data usually consists of three main parts: title, description, and questions used to collect data. The types of questions used to collect data are diverse, mainly including fill in the blank questions, multiple-choice questions, attachment questions, etc. [5] The literature method collects data by analyzing existing documents and records, such as books, articles, reports, web pages, social media posts, etc. This method can obtain a wide range of information and is cost-effective.

2.1.2. The effect of the collect informations

All in all, collecting information can letting us to make appropriate decisions. They are based on reliable information. So, it is important to ensure or examine whether the information that be collected are accurate.

2.2. Data vetting

The appropriate decisions are based on the reliable information. The legitimacy and correctness of the results can only be guaranteed by getting good data. The amount of data we encounter in actual problems is frequently tremendous. We must filter the data to make the solution simpler while ensuring that the method we employ can be implemented with support from raw data. In order to get more accurate and reliable results, we also need to make sure that the data we choose is representative.

Data filtering's ultimate purpose in the computer business is to get ready for data mining. Data extraction, data cleaning, and data downloading are the three elements of data filtering.

In the research and analysis of a certain case, there must be a lot of information, which can be called variables or indicators. The more information there is, the more overlapping or multicollinearity of information may occur [6]. In order to analyze, study, and solve problems more comprehensively and systematically, we hope to use less information to approximate the original information and provide a reasonable explanation in a context that is in line with practical significance.[7]

2.3. Data analysis

2.3.1. Basic introduction

Data analysis, in its simplest form, is the process of examining the gathered data for a certain issue and identifying economic value. In order to develop the usefulness and worth of the data, data analysis is the process of using the right statistical analysis techniques to examine a sizable amount of obtained data. Getting rid of useless information and drawing conclusions is the key goal.

There are three primary functions of data analysis. Current Situation Analysis, it merely informs you of what happened in the past. Cause Analysis, it just implies explaining to you why a specific circumstance has transpired. Last is the Predictive Analysis, it merely informs you of what will occur in the future.

Data management mainly involves listing and processing data information, such as definition and standardization. When explaining relevant information, it can be combined with the demands of business processing to achieve intelligent operation of data quality, data management, and data storage.[8] [9]

2.3.2. The effect of the data analysis

Classification, prediction and predictive that analysis include methods such as classification, prediction, association rules, collaborative filtering, and pattern recognition (clustering).

Data reduction and dimensionality reduction typically improve the performance of data mining algorithms when the number of variables is limited and a large number of sample data can be classified into similar groups.

2.4. My dataset

It contains 7 * 348 pieces of data. The following is a part of my dataset.

Table 1. Dataset

340.6140	351.4731	342.0619	351.8315	342.0619	348.5773	348.5773
291.0241	289.9381	308.0367	296.4536	284.8706	296.4536	312.0183
234.1947	236.7285	232.7468	233.1088	236.0046	244.3299	246.1397
180.2612	180.9851	183.8809	197.9977	183.5189	187.5006	180.2612
125.9656	127.7755	125.6037	121.9840	127.0515	132.1191	128.8614
68.4124	69.8603	76.7377	68.7743	78.1856	58.6392	91.2165
9.7732	12.6690	22.8041	9.7732	11.2211	35.8351	19.5464

This project aims to scan 20 sets of samples with different pesticide contents collected using a Raman spectrometer, with a scanning range of 541-700cm-1. I chose 7 sets in my dataset, shown by table 1. And then figure 1 is the source of my dataset.

The parameters need to be changed repeatedly. To have the parameter that making the predict's results mostly close to the reality. After finishing the matlab's program and training Predictive variables, run the program, get the results.

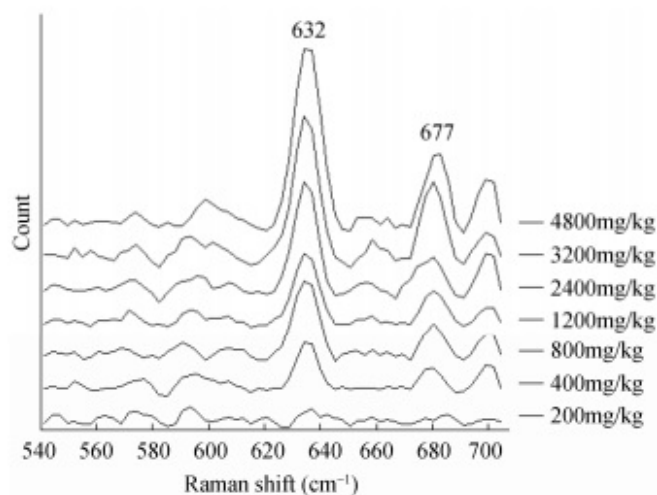


Figure 1. Raman shift

3. Results

This following picture is the result.

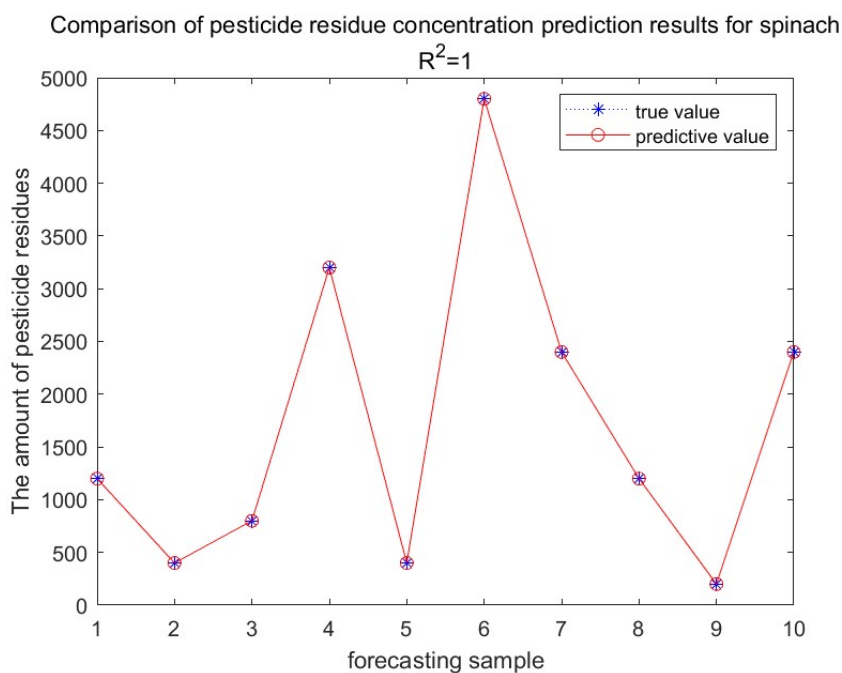


Figure 2. Comparison results

In the figure 2, the blue one is the true value and the red one is the predictive value. The turning points of the predictive value and the true value are very similar. Finally, obtaining ten prediction samples. It can be seen that the predicted values are very similar to the actual values, basically the same. It indicates that the program parameters no longer require to have any significant changes. Because we modify our parameters based on the calculation results (the difference between the predict values and the true values). All in all, this graph that about the Comparison of pesticide residue concentration prediction results for spinach clearly and conveniently show to us the prediction of the spinach's pesticide residue by using Roman spectral values. And its errors are little in the technical areas. [10]

4. Discussion

The reason that the comparisons of pesticide residue concentration prediction results for spinach are small is because I try to using different parameter in the program repeatedly when training Predictive variables

In the process, the information that the dataset just few. It may not enough or too few to training predictive variables. The figure that obtained in the end may not accurate. It has a limited error and it may be different with the facts because of few information. I should expand the range of the compilation and collect more information, or I could do less in the data vetting, to reduce the errors as far as possible.

5. Conclusion

In the end, I obtain the figure that about the comparison of pesticide residue concentration prediction results for spinach and I also establish the model about it. The predictive values are very close to the true values after the training. The model can be used in place that only have the spinach's Roman spectral values and want the pesticide residue of the spinach. This figure is convenient.

This model provides a new way of thinking. Maybe we can also predict the substance's other attribute by using the Roman spectral values, not just the pesticide residue. And maybe this technology can not only used in spinach, but it can be also used in other things. Because the Raman spectroscopy analysis method can used to measure many different kinds of substances. Such as the Solids, powders, liquids, colloids, ointments, gases; inorganic materials, organic materials, biological materials; Pure substance, mixture, solution. And it can't be used to measure the metals and their alloys. We can see this technology can used for many different substances. So, you can measure the pesticide residue of other vegetables and fruits by this technology easily.

In the process, the information that the dataset have is maybe little. The figure that obtained in the end may not accurate. It has a limited error and it may be different with the facts. I should expand the range of the compilation and collect more information, to reduce the errors as far as possible. Also, although the Raman spectroscopy analysis method has a limited high sensitivity, it is still not enough. In practice, some residues may be still masked to some extent. So, improve the sensitivity of Raman spectroscopy analysis method is also a good way to increase the accuracy of the model. Artificial neural networks have advanced quickly in recent years in both theoretical study and real-world applications. However, it is important to recognize that artificial neural networks are merely a basic simulation of the human brain and that there are still a number of issues that need to be solved.

References

- [1] Ying Xu, Discussion on the application of gas chromatography-tandem mass spectrometry method for standardized detection of food pesticide residues, 2023, 7th/, page: 175 - 177.
- [2] Jiajin Yan, Small Sample Image Classification Based on Data Augmentation and Neural, NetworksModern Information Technology 2022, 6, page: 77 - 80.
- [3] Hagan M T, Menhaj M B. Training Feed Forward Networks with Marquarl Algorithm [J]. IEEE Trans. On Neural Networks, 1994, 5 (6): 989 — 993.
- [4] Hecht—Nielsen R. Theory of Back Propagation Neural Network [M]. Proc, of IJCNN, 1989. 593 — 603.
- [5] MENG Yu-liang ZHANG You-hong, Design and Implementation of a Universal Form Data Collection System, Computer application - Information Technology and Informatization, 2021, 125 – 126.
- [6] Quan Liu, Research on ARIMA Based Multivariate Time Series Neural Network Prediction Model [J], Statistics and decision-making.
- [7] Dan Shen, A Method Based on Multivariate Statistical Data Filtering, Journal of Henan Institute of Education (Natural Science Edition) 2019, 28, 6 - 9.
- [8] JunHua Li, Research on the Development of the New Energy Vehicle Industry Based on Big Data Analysis, Technology Information, 2023, 241 - 244.

- [9] Tianfeng Xu, Yankun Peng, Yongyu Li, DHAKAL Sagar, Chen Di. Detection of Chlorpyrifos Residues in Spinach Based on Raman Spectroscopy Technology, Journal of Food Safety and Quality, 2014, 5th, page: 707 - 711.
- [10] Zheng Zhou, Overview of the Development Status of BP Neural Networks, Shanxi Electronic Technology, 2008, page: 90 - 92.