

Prediction of Milk Powder Raman Spectra and Ammonium Sulfate Adulteration Concentration Based on Neural Network Model

Sihan Zeng *

Shunde No.1 Middle School Southwest School, Guangdong Province, China

* Corresponding author: zengsihan@hhu.edu.cn

Abstract. This article aims to explore the changes in Raman spectra of milk powder when different concentrations of ammonium sulfate are adulterated. By conducting neural network tests on milk powder samples with different ammonium sulfate concentrations, we studied the impact of ammonium sulfate adulteration on milk powder. The research results show that moderate concentrations of ammonium sulfate adulteration can improve the solubility and flowability of milk powder, but high concentrations of adulteration may reduce the protein content of milk powder and have negative effects on its sensory quality and safety, causing significant adverse effects on human health. Therefore, the adulteration concentration of ammonium sulfate should be strictly controlled in the production and processing of milk powder to ensure its quality and safety.

Keywords: milk powder, ammonium sulfate, Raman spectroscopy technique, BP neural network model.

1. Introduction

Milk powder is a common dairy product with a long shelf life and easy storage. However, in the production and processing of milk powder, some additives may be added to improve its physical properties and flowability. Ammonium sulfate, as one of the common additives, is used to improve the solubility and flowability of milk powder. However, the adulteration concentration of ammonium sulfate may affect the quality and safety of milk powder. Traditional methods for detecting ammonium sulfate in food, such as chromatography, often result in a lot of reagent loss and time consumption, making the detection work inconvenient. With the popularization of spectroscopy techniques, the issue of ammonium sulfate adulteration has been greatly improved. However, different spectroscopy techniques have different requirements for detection, and the spectroscopic detection of ammonium sulfate has not been unified for a long time. By integrating some detection techniques for ammonium sulfate in food and conducting neural network tests, this study provides a reference for the popularization of food safety.

Raman spectroscopy is based on the scattering spectrum produced when the incident light interacts with the molecules of the test substance, which changes the frequency. It is usually used to characterize the structural characteristics of molecules, also known as "molecular fingerprint spectra".

Handheld Raman spectrometers are a type of spectral analysis instrument based on the Raman scattering effect. The principle is simple: when light is irradiated onto the sample, the molecules of the sample cause scattering of the incident light. Most of the scattered light does not change in frequency, only changes direction, which is called Rayleigh scattering. A small amount of scattered light has both frequency and direction changes, which is called Raman scattering. The frequency difference between the scattered light and the incident light is called Raman shift. Raman shift can confirm the composition and molecular structure of substances, and it is the basis for qualitative analysis. The intensity of Raman peaks is the basis for quantitative analysis of substance concentration and total amount. It is easy to operate and commonly used for on-site qualitative and quantitative detection of harmful substances in various foods and drugs.

Optical data processing is an important step in spectral detection techniques. The complex matrix in food can produce spectral interference, such as matrix fluorescence interference in Raman spectra and scattering interference in near-infrared spectra. Suitable preprocessing methods are needed to

perform baseline correction, error elimination, and feature extraction on the spectra. Finally, the optimal modeling method is selected to establish classification or regression models and neural models according to different experimental requirements, so as to effectively utilize the data information collected by spectral detection techniques.

2. Method

2.1. Introduction to Neural Network Models

A neural network model is an artificial intelligence model inspired by biological neural systems, used to simulate and solve various tasks and problems. It is a network structure composed of a large number of artificial neurons (or nodes) and their connections. Neural network models typically include input layers, hidden layers, and output layers. The input layer receives external input data, the hidden layer is used to process and transform the input data, and the output layer produces the model's prediction results or decisions. Each neuron receives inputs from the previous layer of neurons and processes these inputs through activation functions to generate outputs. The connections between neurons correspond to weights, which determine the degree to which input signals affect the output of neurons. The weights are adjusted during the training process to optimize the performance of the model. The neural network model is trained through a backpropagation algorithm, which adjusts the weights and parameters of the network by comparing it with real labels (or expected outputs) to minimize prediction errors. This training process involves calculating gradients and using optimization algorithms such as gradient descent to update network parameters. Neural network models have strong fitting and generalization abilities, which can handle nonlinear relationships and learn and model complex data patterns. It has achieved significant results in many fields such as image recognition, natural language processing, speech recognition, recommendation systems, and has become one of the core technologies in the field of artificial intelligence today.

2.2. Establishment of Neural Network Model for Ammonium Sulfate in Food

The establishment of a neural network model is also an important step in the detection of ammonium sulfate. We trained a BP neural network model and generated data ranging from 100 to 3000, and the model was subjected to data correction and noise elimination. Then, relevant data on ammonium sulfate adulteration concentrations were obtained from the literature, and the number of layers in the neural network model was unified to three. The data under this layer were relatively stable, and the number of nodes and hidden nodes in the same layer was changed sequentially, with node numbers of 5, 10, 15, and 21, respectively.

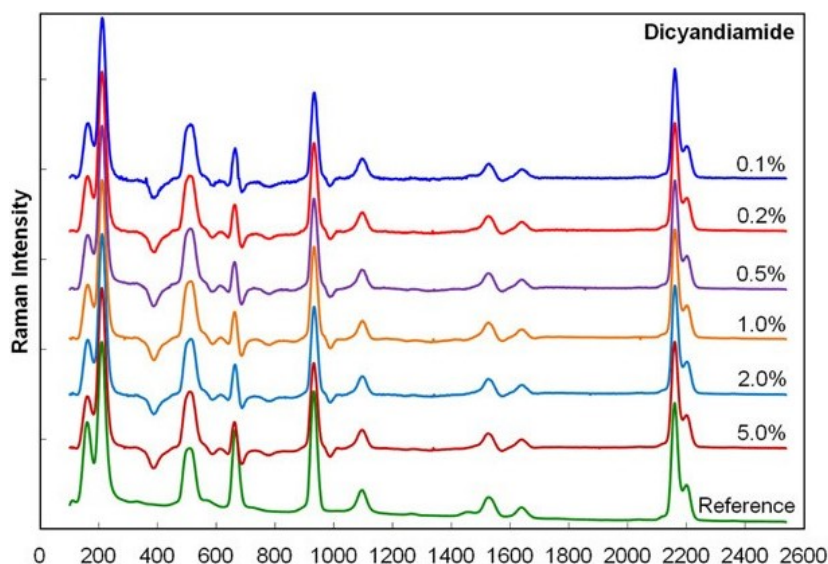


Figure 1. Raman intensity relationship diagram under different ammonium sulfate concentrations

2.3. Evaluation Metrics

In the BP neural network model, we will standardize the concentration of ammonium sulfate and observe the accuracy of the neural network model under different conditions, using accuracy as an evaluation metric. In addition, we will input different data into the neural network model and demonstrate the corresponding ammonium sulfate concentration.

$$\text{Concentration (\%)} = (\text{mass of ammonium sulfate (g)}/\text{mass of milk powder (g)}) \times 100 \quad (1)$$

In this formula, the mass of ammonium sulfate refers to the mass of ammonium sulfate contained in the milk powder, and the mass of the milk powder refers to the total mass of the sample to be tested.

It should be noted that this formula assumes that ammonium sulfate is completely dissolved in milk powder. If ammonium sulfate only partially dissolves, it is necessary to consider the effect of its solubility to correct the calculation results.

2.4. Software Used for Establishing the Neural Network Model

We used Matlab, which is a commercial mathematical software developed by MathWorks in the United States. It is used in various fields such as data analysis, wireless communication, deep learning, image processing and computer vision, signal processing, quantitative finance and risk management, robotics, and control systems. Matlab stands for matrix & laboratory, combining the words “matrix” and “laboratory,” and it mainly serves as a high-tech computing environment for scientific calculation, visualization, and interactive program design. It integrates powerful functions for numerical analysis, matrix computation, scientific data visualization, and modeling and simulation of nonlinear dynamic systems in a user-friendly windowed environment. It provides a comprehensive solution for scientific research, engineering design, and various scientific fields that require effective numerical computation, greatly overcoming the editing mode of traditional non-interactive programming languages such as C and Fortran.

3. Result

3.1. Extraction and Application of Image Data

By investigating the images of milk powder and the corresponding ammonium sulfate adulteration concentration, we used “Getdata” to digitize the images and obtained the relationship between Raman shift and Raman intensity. The following experimental data were obtained (as shown in Figure 2).

Table 1. Key number of experimental data used

21x146 double						
1	2	3	4	5	6	7
549.8370	549.8790	557.7600	565.6220	576.1070	581.3520	573.5220
466.0720	466.1050	476.5990	484.4710	494.9600	497.5920	481.9040
371.8370	371.8690	384.9860	400.7060	411.1960	405.9790	400.7660
288.0670	288.1000	301.2120	314.3280	319.5820	309.1200	298.6680
199.0580	201.6940	201.7130	212.2020	220.0690	225.3190	230.5680
115.2890	112.7030	123.1970	131.0690	136.3140	141.5590	141.5770
36.7550	44.6176	39.4005	36.8057	42.0598	52.5401	60.4073
549.8370	549.8790	557.7600	565.6220	576.1070	581.3520	573.5220

Key data is data that has significant impact and significance in data analysis and decision-making processes. For key data, the following are some aspects of requirements and analysis: Correlation:

Key data should be closely related to specific issues or goals. They should provide a deep understanding of the problem and have meaningful relevance for decision-making or solutions. Reliability: Key data should be accurate, reliable, and verifiable. The source, collection methods, and processing of data should be strictly controlled and validated to ensure the quality and credibility of the data. Integrity: Key data should cover the entire range of the problem area. The lack of important data may lead to misleading analysis results and inaccurate decision-making. Therefore, the process of collecting and organizing data should ensure the integrity of the data. Timeliness: Key data should be the latest and most timely. Over time, data may change, so data analysis should be based on the latest data to obtain more accurate conclusions and decisions. Understandability: Key data should be easy to understand and interpret. The results and conclusions of data analysis should be clear and concise, so that decision-makers and relevant stakeholders can accurately understand and apply them. When analyzing key data, the following methods can be used: Descriptive statistical analysis: By calculating statistical indicators such as mean, median, and standard deviation of key data, the overall description and summary of the data are carried out. Data visualization: Use charts, graphs, and visualization tools to present key data in a form that is easy to understand and analyze, helping to discover potential patterns and trends in the data. Hypothesis testing and statistical inference: Using hypothesis testing and statistical inference methods, statistical significance testing is conducted on key data to verify the relationships and differences between the data.

Prediction and modeling: By using statistical models, machine learning algorithms, etc., predictions and modeling are made based on the historical patterns and trends of key data for future decision-making and planning. The requirements and analysis methods for key data will vary depending on specific business and issues. The key is to ensure the quality and reliability of data, and transform it into meaningful information and insights to support effective decision-making and action.

3.2. Test the neural network model with the same number of layers and different nodes:

Objectively speaking, neural network models located at multiple levels may have an impact due to certain irresistible forces. Therefore, we have unified the number of layers for the neural network model to three, and the neural network model at this level is relatively stable and not prone to overfitting. Afterwards, we tested and trained different nodes with different numbers of 5, 10, 15, and 21. A rule has been discovered: in cases where the number of nodes is not one, the higher the number of nodes, the higher the accuracy (i.e. R to the 2nd power). Of course, this is not absolute either. When the number of nodes exceeds a certain value, the accuracy will plummet. Therefore, after multiple tests, we found that when the number of nodes is equal to 21, the accuracy can reach over 99%, while the accuracy of the subsequent number of nodes is less than 50%, and the error value is also increasing; In addition, the smaller the number of nodes, the greater the measurement error for individual values, making it difficult to accurately achieve noise elimination.

3.3. Noise Cancellation and Extreme Points

Noise cancellation is an important task in signal processing, aimed at extracting effective information from the original signal with noise. The analysis and planning of extreme points are often used in noise cancellation to identify and process outliers or noise in signals, in order to improve signal quality. Extreme points can be divided into two types: peak and valley. The peak represents the maximum value of the signal, while the valley represents the minimum value of the signal. The analysis and planning of extreme points can be carried out using the following methods: Smooth filtering: By applying low-pass filters or moving average filters, high-frequency noise can be removed from the signal, smoothing the signal, and reducing the amplitude changes of extreme points. Threshold denoising method: Based on the statistical characteristics of the signal, set appropriate thresholds, treat extreme points below or above the threshold as outliers or noise, and eliminate or replace them. Based on peak and valley analysis: By detecting the peak and valley values of the signal, the amplitude and distribution characteristics of noise can be estimated. Then, appropriate filtering algorithms or denoising methods can be selected based on this information. Statistical analysis: Using

statistical methods such as mean, variance, autocorrelation function, etc. to analyze and model signals. By analyzing the statistical characteristics of the signal, the distribution type of noise can be determined and denoised accordingly. Wavelet transform: Wavelet transform is a multi-scale analysis method that can decompose signals into sub signals of different frequencies. By removing or suppressing high-frequency wavelet coefficients, the impact of noise can be reduced and the original features of the signal can be restored. The above methods can be combined to select appropriate strategies based on specific noise characteristics and signal analysis needs. The goal of noise cancellation is to remove noise as much as possible, preserve useful information of the original signal, and improve signal quality and accuracy of subsequent signal processing tasks.

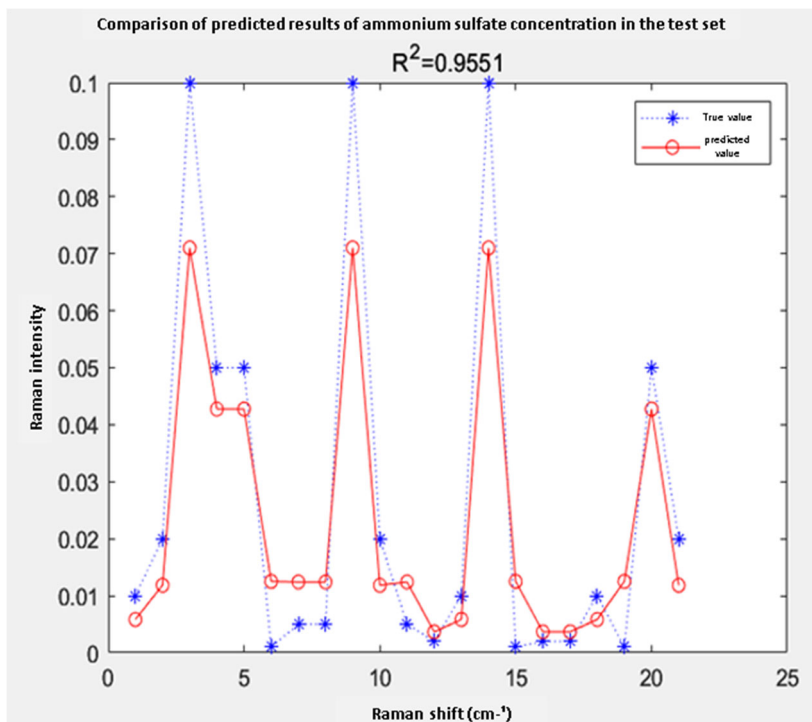


Figure 2. When the number of layers is 3 and the number of nodes is 5

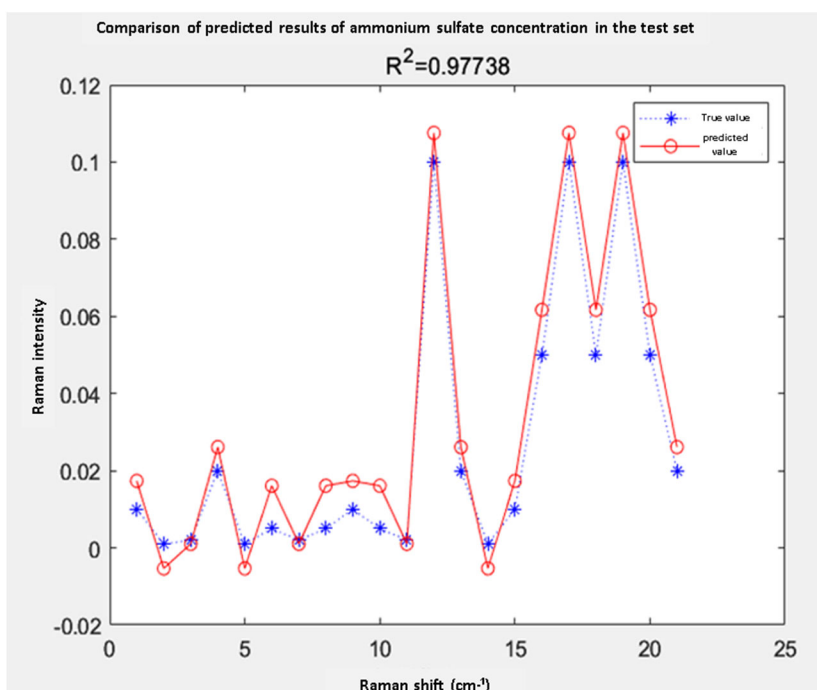


Figure 3. When the number of layers is 3 and the number of nodes is 10 (Accuracy 97.7%)

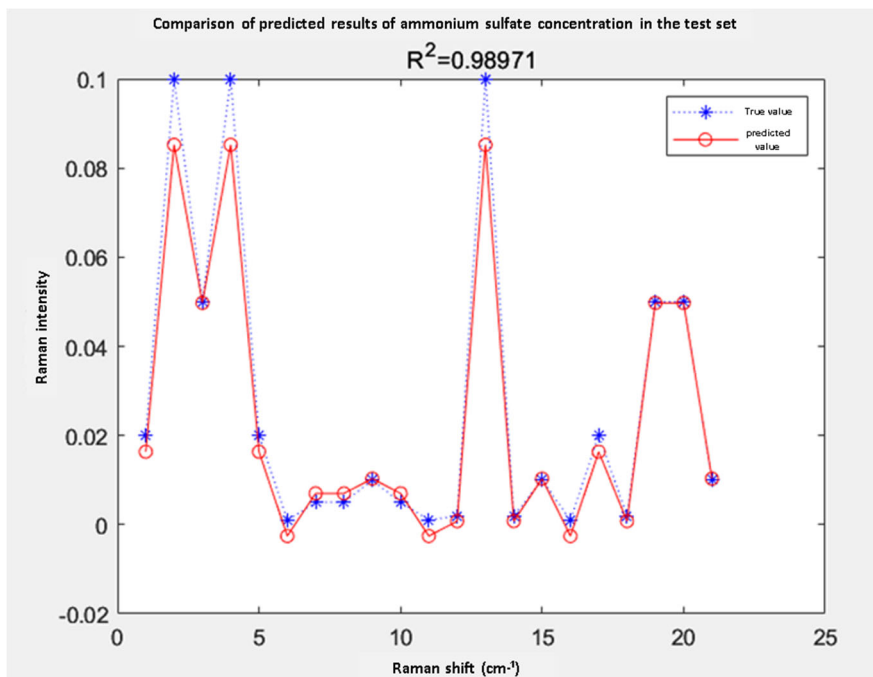


Figure 4. When the number of layers is 3 and the number of nodes is 15(Accuracy 98.9%)

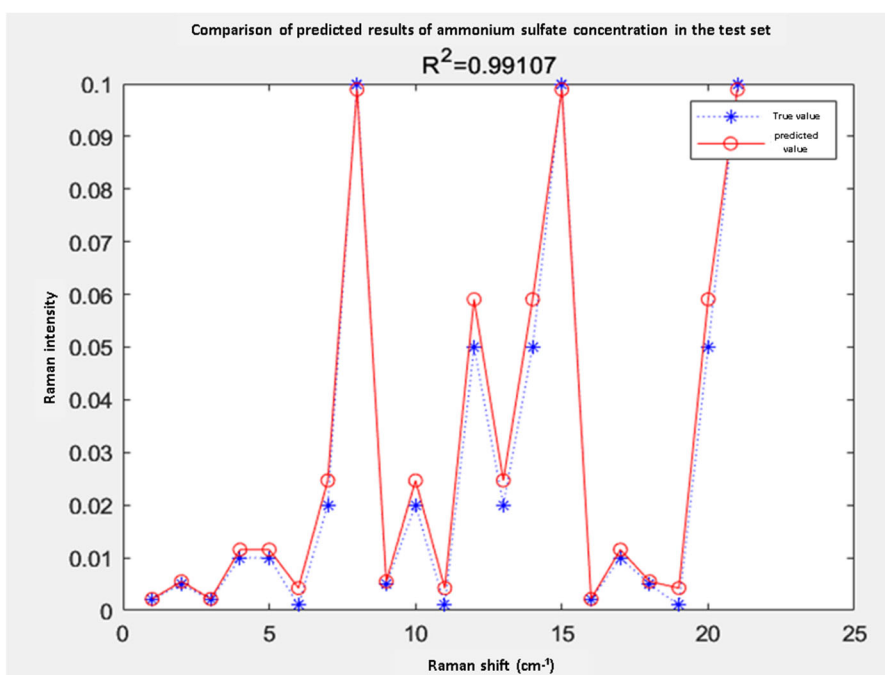


Figure 5. When the number of layers is 3 and the number of nodes is 21 (Accuracy 99.1%)

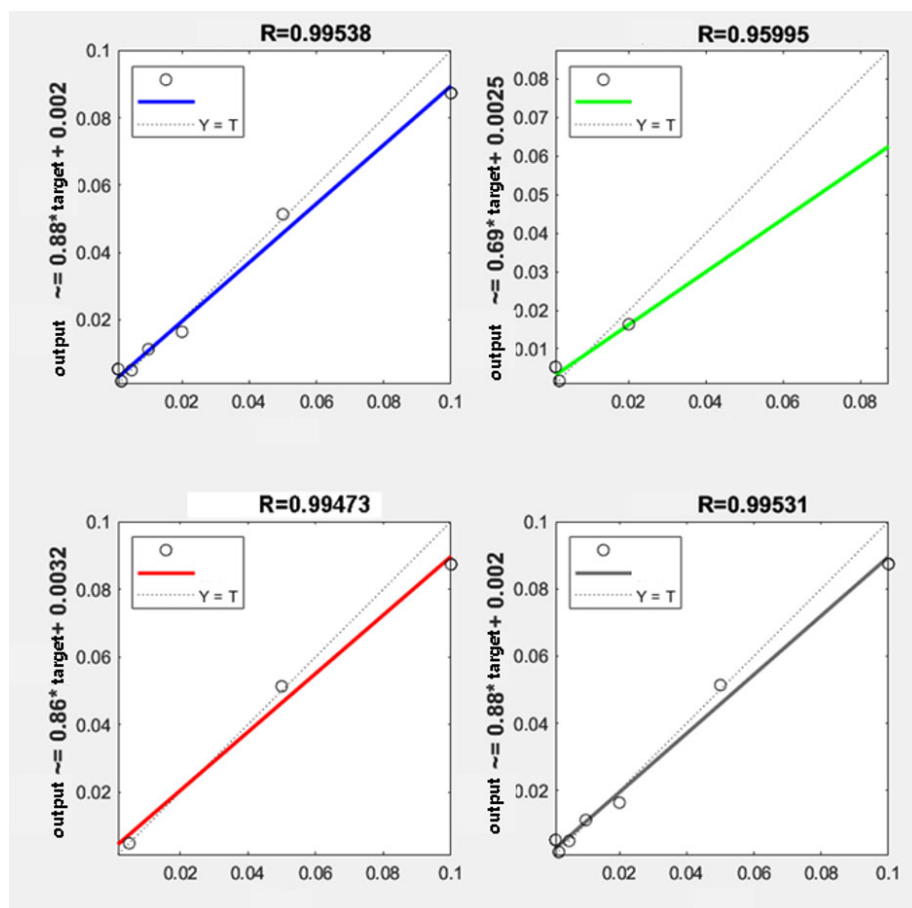


Figure 6. Display diagram of various data when the number of nodes is 21

Through a detailed analysis of the above charts, we can clearly observe that in Figure 1 and Figure 3, the neural network model has poor analysis and fitting ability for higher extreme points; In Figure 2, on the contrary, it has poor analysis and fitting ability for lower extreme points; Finally, in Figure 4, the analysis ability defects of the two extreme points were repaired one by one under the influence of a large number of nodes; In the four small images in Figure 5, the training, validation, testing, and overall accuracy were analyzed in sequence, and fitting and actual graph lines were drawn based on this. In summary, the conclusion was drawn that the neural network model has strong analytical ability for ammonium sulfate in milk powder when the number of nodes is 21. And this has led to the trend of analyzing other harmful impurities. Make significant contributions to component testing.

4. Discussion

Advantages: Nonlinear: able to perform well on complex nonlinear problems; High fault tolerance: Strong tolerance for noise and errors in input data; Scalability: The model's expressive ability can be improved by increasing the number of layers and nodes; Adaptive: It can automatically adjust parameters based on training data, avoiding the trouble of manually designing features.

Disadvantages: The internal complexity of the model makes it difficult to explain the operational mechanism, resulting in a lack of interpretability;

Easy overfitting: For small-scale data, overfitting is prone to occur; **Requires a large amount of computing resources and time:** Training based on large-scale data requires a significant amount of computing resources and time.

5. Conclusion

Ammonium sulfate is a commonly used chemical mainly used in industrial and agricultural fields, including fertilizers, gunpowder, dyes, etc.

The production process of milk powder is strictly regulated and quality controlled, including raw material procurement, processing, and packaging. Regulatory agencies usually conduct sampling tests on milk powder to ensure product quality and safety. Therefore, we can confidently purchase milk powder products that meet the standards. An appropriate concentration of ammonium sulfate doping can improve the solubility, fluidity, and sensory quality of milk powder. However, high concentrations of doping can reduce the protein content of milk powder, potentially affecting its taste and safety. Therefore, the doping concentration of ammonium sulfate should be strictly controlled during the production and processing of milk powder to ensure the quality and safety of the milk powder.

However, as consumers, we should also remain vigilant, purchase milk powder products from legitimate channels and brands, and pay attention to the production date, shelf life, and packaging integrity of the products. If there are any quality issues or doubts, they should be reported and complained to relevant departments or manufacturers in a timely manner.

References

- [1] Alkanesulfonates; Research Conducted by P. Feng and Co-Researchers Has Updated Our Knowledge about Alkanesulfonates [Quantification of Whey Protein Content in Milk-Based Infant Formula Powders by Sodium Dodecyl Sulfate-Capillary Gel Electrophoresis (SDS-CGE):] [J]. *Chemicals & Chemistry*.
- [2] Chen Xiuming, Liu Qing, Liang Haoxin, et al. A rapid identification method for illegal additives in infant formula milk powder based on near-infrared spectroscopy [J]. *Food Industry Technology*, 2020, 41 (01): 224 - 228+283. DOI: 10.13386/j.issn1002-0306. 2020. 01. 036.
- [3] Li Zhuo, Zhu Jie, Zhou You. Research progress on detection techniques for harmful components in dairy products [J]. *Modern Food*, 2018 (21): 114 - 116. DOI: 10.16736/j.cnki.cn41-1434/ts.2018. 21. 036.
- [4] Hu Jieming. Spectra-Quad achieves online detection of melamine content [J]. *Food Safety Journal*, 2008 (02): 56 - 57.
- [5] Li Zheng. Food should undergo full component testing [J]. *Brand and Standardization*, 2010 (17): 40.
- [6] Xu Zongliang, Xia Yuanjun, Gong Jun. Study on the rapid detection of the composition of milk powder and whey powder using near-infrared spectroscopy [J]. *China Dairy Industry*, 2002 (02): 40 - 42.
- [7] Liu Yongfeng, Extensive and Accurate Quantitative Detection Method for Milk Components in Sheep Milk Powder Shaanxi Province, Shaanxi Normal University, 2019 - 10 - 01.
- [8] Veterinary Medicine; Researchers from Benha University Discuss Findings in Veterinary Medicine [The Camel Rhinarium: A Study Revealing the Presence of the Nasal Plane in Dromedary Camel (*Camelus dromedarius*), with Special Reference to Its Epidermal Structure] [J]. *Veterinary Week*.
- [9] Wu Yujie, Research and application of rapid detection methods for multiple nutritional components in imported and exported milk powder Inspection and Quarantine Technology Center of Guangxi Entry-Exit Inspection and Quarantine Bureau, Guangxi Zhuang Autonomous Region, 2016 - 12 - 06.
- [10] Zhu Lixian, Lin Qinbao, Zhong Huaining, et al. Detection of Hydroxyapatite in Milk Powder [J]. *Food and Machinery*, 2019, 35 (03): 67 - 71. DOI: 10.13652/j.issn.1003-5788. 2019. 03. 012.