**Supplementary Material**

**Evaluation of methods for detection of** **hazardous substances in food based on machine learning**

Long-yu Zhu, a Lijuan Yan, b Fang Zhao, c Xuewen Guo,\*a Dunming Xu,\*b Jingzhang Lv, c Lin Ding, b Na Niu, c Jun-qin Qiao, a Shumian Ma, c Xindi Huang, c Huiling Liu c and Hong-zhen Lian\*a

a *State Key Laboratory of Analytical Chemistry for Life Science, School of Chemistry & Chemical Engineering and Center of Materials Analysis, Nanjing University, Nanjing 210023, China*

b *Technology Center, Xiamen Customs, Xiamen 361026, China*

c *Shenzhen Academy of Inspection and Quarantine, Food Inspection and Quarantine Technology Center of Shenzhen Customs District, Shenzhen 518045, China*

Catalogue

[Fig. S1. The learning curves of BPNN, CART, SVR and RFR model on training and validation datasets of chemical hazard detection methods 3](#_Toc152004485)

[Fig. S2. The learning curves of BPNN, CART, SVR and RFR model on training and validation datasets of biological hazard detection methods 4](#_Toc152004486)

[Table S1. Scoring table for detection methods of chemical hazards 5](#_Toc152004487)

[Table S2. Goodness of fit on the test set with different structure of hidden layer 6](#_Toc152004488)

[Table S3. Different hyperparameters for the BPNN, CART, SVR, RFR evaluation model of chemical and biological hazard detection methods 7](#_Toc152004489)

[Table S4. Mean MSE of evaluation models for chemical and biological hazard detection methods during ten repetitive random sub-sampling processes 8](#_Toc152004490)



Fig. S1. The learning curves of BPNN, CART, SVR and RFR model on training and validation datasets of chemical hazard detection methods.



Fig. S2. The convergence curves of BPNN, CART, SVR and RFR model on training and validation datasets of biological hazard detection methods.

Table S1. Scoring table for detection methods of chemical hazards

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hazard | Detection method | I | II | III | IV | V | VI | VII | VIII | Tag value |
| SO2 | GB 5009.34-2016 |  |  |  |  |  |  |  |  |  |
| SO2 | GB/T 5009.49-2008 |  |  |  |  |  |  |  |  |  |
| SO2 | GB/T 5009.49-2008 |  |  |  |  |  |  |  |  |  |
| … | …… |  |
| Inorganic As | GB 5009.11-2014 |  |  |  |  |  |  |  |  |  |

Table S2. Goodness of fit on the test set with different structure of hidden layer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Combination | Single hidden layer nodes | Goodness  of fit | Double hidden layer I node | Double hidden layer II node | Goodness  of fit |
| 1 | 5 | 0.793 | 5 | 5 | 0.240 |
| 2 | 6 | 0.761 | 6 | 6 | 0.897 |
| 3 | 7 | 0.514 | 4 | 5 | 0.337 |
| 4 | 12 | 0.868 | 5 | 10 | 0.897 |
| 5 | 12 | 0.868 | 6 | 7 | 0.817 |

Table S3. Different hyperparameters for the BPNN, CART, SVR, RFR evaluation model of chemical and biological hazard detection methods

|  |  |  |  |
| --- | --- | --- | --- |
| model | Hyperparameter | Chemical detection methods | Biological detection methods |
| BPNN | Number of nodes in hidden layer | 9 | 10 |
| Activation function | relu | relu |
| Weight optimizer | adam | adam |
| Learning rate | 0.1 | 0.08 |
| CART | max\_depth | 14 | 7 |
| min\_samples\_leaf | 3 | 1 |
| min\_samples\_split | 2 | 8 |
| SVR | C | 3 | 200 |
| epsilon | 0.01 | 0.0009 |
| gamma | 1 | 0.002 |
| RFR | n\_estimators | 600 | 500 |
| max\_depth | 14 | 6 |
| min\_samples\_leaf | 1 | 1 |
| min\_samples\_split | 2 | 10 |

Table S4. Mean MSE of evaluation models for chemical and biological hazard detection methods during ten repetitive random sub-sampling processes

|  |  |  |
| --- | --- | --- |
| Model | Chemical detection methods | Biological detection methods |
| MSE of BPNN model | 0.00101 | 0.000212 |
| MSE of CART model | 0.00127 | 0.00388 |
| MSE of SVR model | 0.000405 | 0.00000166 |
| MSE of RFR model | 0.000599 | 0.00268 |