AMMCS 2022

CS432

Risk classification study of carbofuran in vegetables based on K-means++ algorithm

Tongqiang Jiang [a,b], Qi Yang [a,b], Tianqi Liu [a,b], Wei Dong [a,b], Yongchun Jiao [a,b], Qingchuan Zhang [a,b]

[a] School of E-Business and Logistics, Beijing Technology and Business University, Beijing, China
[b] National Engineering Research Centre for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing, China

Introduction

Carbofuran is a common pesticide used for the control of pests in crops. Carbofuran is harmful to humans because it can be absorbed by the body, and the distribution and metabolism in the body are rapid. Carbofuran poisoning includes symptoms such as headache, nausea, and convulsions.

Food risk classification is an emerging form of risk assessment that integrates factors such as contaminant concentration levels, consumers' dietary exposures, and the degree of public health hazards, and then ranks the risks in a hierarchy that allows for more scientific and rapid identification of risk levels and facilitates better resource allocation by decision makers.

In the existing studies, food risk classification methods all suffer from the problem of quantifying risk values and determining risk levels too subjectively. The clustering algorithm is data-driven and classifies samples according to their similarity, which can effectively reduce the influence of subjective factors.The K-means++ algorithm is an optimized and improved method that can well solve the problem that the clustering results are too dependent on the initial cluster centroid selection . Therefore, in this study, the K-means++ clustering algorithm was used to conduct a risk classification study of carbofuran in vegetables.

Data

Data on pesticide residue contamination levels in vegetables in this study were obtained from national food safety sampling data in 2020, with a total of 12,940 samples. The data on vegetable consumption of residents in each region were taken from the corresponding provinces and cities in the Fifth China Total Diet Study. In this study, vegetables were classified into six categories, namely leafy vegetables, root and potato vegetables, bean vegetables, melon vegetables, bulb vegetables and eggplant vegetables. Toxicological data were obtained from the Joint Meeting on Pesticide Residues.

Food Risk Assessment Model

Chronic risk assessment indicators

$$HQ_{i,j} = \frac{EDI^{50}_{i,j}}{ADI} \quad (1)$$

$$EDI^{50}_{i,j} = \frac{FC_{i,j} \times X^{50}_{i,j}}{W} \quad (2)$$

Acute risk assessment indicators

$$HI_{i,j} = \frac{EDI^{95}_{i,j}}{ARfD} \quad (3)$$

$$EDI^{95}_{i,j} = \frac{FC_{i,j} \times X^{95}_{i,j}}{W} \quad (4)$$

Nemerow integrated pollution index

$$NIPI_{(i,j)} = \sqrt{\frac{P^2_{\max(i,j)} + P^2_{ave(i,j)}}{2}} \quad (5)$$

$$P_{i,j} = \frac{X_{i,j}}{S_j} \quad (6)$$

The above three indicators were constructed to quantify the risk factors of contaminants by integrating the likelihood of exceedance, exposure and hazard of food contaminants, and to establish a food safety risk assessment model.The K-means++ algorithm is applied in this study, and the better clustering method is selected by comparing the silhouette coefficients. The data-driven approach to risk classification in this study can eliminate the subjectivity of risk ranking.

Results

The number of dietary intake risk classification of carbofuran in vegetables was determined by silhouette coefficients, and the silhouette coefficients of different categories are shown in Figure 1.Therefore, in this study, the risk level of dietary intake of carbofuran in vegetables was classified into 3 levels: low, medium and high.

Figure 1. Silhouette coefficients for clustering number of classes 2-6

The K-means++ algorithm was used to determine the risk level of dietary intake of carbofuran in vegetables, and the risk grading results of various vegetable and province combinations were obtained, as shown in Figure 2.The combinations of medium and high risk levels were ranked from high to low, as shown in Table 1.

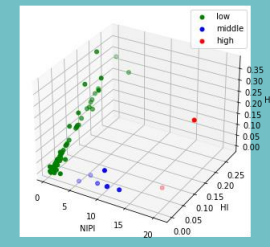

Figure 2. K-means++ clustering results

Table 1. Combination of province-vegetable categories with medium and high risk levels

Province	Vegetable Category	Risk Level
Hebei	Bulb vegetables	High
Shaanxi	Leafy Vegetables	High
Guangdong	Legumes	Medium
Guangxi	Legumes	Medium
Henan	Bulb vegetables	Medium
Zhejiang	Legumes	Medium
Fujian	Legumes	Medium
Fujian	Bulb vegetables	Medium
Liaoning	Legumes	Medium

Conclusion

In this study, we established a risk classification model for carbofuran based on the exposure assessment method, and achieved the risk classification of 20 provinces and different vegetable categories across China by K-means++ clustering algorithm. Compared with the traditional dietary exposure assessment method, the risk grading model enables us to consider the combined effects of multiple indicators in a comprehensive and objective manner. The study conducted a comparative study through the silhouette coefficients and finally selected the K-means++ algorithm to determine the number of risk levels and achieve the risk classification of the combination of provinces and vegetable categories. The results show that the model in this study can be effective for the risk classification of carbofuran, but more validation tests are needed to see if this model is applicable to other hazard categories.