DOI: 10.1111/risa.14120

ORIGINAL ARTICLE

Weighted Bayesian network for the classification of unbalanced food safety data: Case study of risk-based monitoring of heavy metals

Xinxin Wang[1,2] | **Yamine Bouzembrak**[2] | **A. G. J. M. Oude Lansink**[1] | **H. J. van der Fels-Klerx**[1,2]

[1]Business Economics, Wageningen, University, Wageningen, The Netherlands

[2]Wageningen Food Safety Research, Wageningen, The Netherlands

Correspondence
H. J. van der Fels-Klerx, Business Economics, Wageningen University, Hollandseweg 1, 6706 KN, Wageningen, The Netherlands.
Email: ine.vanderfels@wur.nl

Funding information
Chinese Scholarship Council and Netherlands Ministry of Agriculture and Food Quality

Abstract

Historical data on food safety monitoring often serve as an information source in designing monitoring plans. However, such data are often unbalanced: a small fraction of the dataset refers to food safety hazards that are present in high concentrations (representing commodity batches with a high risk of being contaminated, the positives) and a high fraction of the dataset refers to food safety hazards that are present in low concentrations (representing commodity batches with a low risk of being contaminated, the negatives). Such unbalanced datasets complicate modeling to predict the probability of contamination of commodity batches. This study proposes a weighted Bayesian network (WBN) classifier to improve the model prediction accuracy for the presence of food and feed safety hazards using unbalanced monitoring data, specifically for the presence of heavy metals in feed. Applying different weight values resulted in different classification accuracies for each involved class; the optimal weight value was defined as the value that yielded the most effective monitoring plan, that is, identifying the highest percentage of contaminated feed batches. Results showed that the Bayesian network classifier resulted in a large difference between the classification accuracy of positive samples (20%) and negative samples (99%). With the WBN approach, the classification accuracy of positive samples and negative samples were both around 80%, and the monitoring effectiveness increased from 31% to 80% for pre-set sample size of 3000. Results of this study can be used to improve the effectiveness of monitoring various food safety hazards in food and feed.

KEYWORDS
Bayesian network model, effectiveness, food safety economics, food safety hazards, sampling

1 | INTRODUCTION

In order to protect animal and human health, the European Union (EU) has set maximum legal limits (MLs) for the presence of certain food safety hazards in feed and food materials and their derivatives. For instance, the MLs for chemical hazards such as mycotoxins, heavy metals (HMs), and dioxins in feed materials are specified in Directive 2002/32/EC (European Commission, 2002). Food safety monitoring is in place to make sure these MLs are met (Focker et al., 2018; van Asselt et al., 2018). A monitoring plan prescribes how monitoring resources are allocated: which food and feed products and which food safety hazards should be sampled and analyzed, and how many samples should be collected (Wang et al., 2022). EU regulation 2017/625 recommends risk-based monitoring, that is, allocating resources toward feed or food batches and contaminants that pose a high risk for animal and/or human health. This risk-based approach entails monitoring those batches that have a high probability of being contaminated. Risk-based monitoring could reduce the probability that contaminated food or feed will enter the food chains and, consequently, avoids expensive recalls or harm to animal or human health.

Feed batches with a high probability of contamination can be identified using predictive models based on historical food safety data (Kuhn & Johnson, 2013). However, historical food safety monitoring data are often unbalanced, that is, the majority of the analytical results refer to low concentrations of the food safety hazards in food/feed products (e.g., below their respective MLs or other thresholds, representing batches with a low risk of being contaminated, the negatives) and the minority of the analytical results refer to analytical results of high concentrations of the food safety hazards (e.g., above respective MLs or other thresholds, representing batches with a high risk of being contaminated, the positives). Using unbalanced historical data to set up future monitoring plans through classification modeling may lead to a low classification accuracy for the minority class (positives), and high classification accuracy for the majority class (negatives). This is particularly problematic when the aim of risk-based monitoring is to focus on batches with a high probability of contamination, in other words, when requiring also a high classification accuracy of the minority class.

Previous studies have applied Bayesian network (BN) modeling for the prediction of food safety contamination in food, with food safety hazards present in high concentrations usually being the minority class (Liu et al., 2018; Liu et al., 2021). They showed that the developed models can more accurately predict the samples with low-level contamination than those with high-level contamination. Several methods can be applied to deal with the unbalanced nature of data to improve the prediction performance of the model for the high-level contaminated samples, such as thresholding (Sheng & Ling, 2006), weighting (Ting, 2002), adjusting prior probabilities (Weiss & Provost, 2001), sampling[1] (Chawla et al., 2002; Sheng & Ling, 2007), or cost-sensitive learning (Johnson & Wichern, 2002). Some studies have explored cost-sensitive BN (CSBN) to deal with unbalanced data (Akila & Srinivasulu Reddy, 2018; Jiang et al., 2014; Nashnush & Vadera, 2017; Xu et al., 2018). However, these studies face the problem of trade-off between the classification accuracy of the minority versus the majority class. That is, an improvement of the classification accuracy of the minority class leads to a reduction of the accuracy of the majority class. Any of the abovementioned methods can obtain a series of classification accuracies (by applying different model parameters), but to date no study has been conducted to find the optimal settings.

The aim of this study was to explore the use of weight-added BN classifier to improve the model classification accuracy in the context of unbalanced food safety monitoring data. WBN models were applied to the case of historical monitoring data related to the presence of HMs in feed, with the monitoring effectiveness—the probability of identifying the contaminated feed batches—as the criterion to select the optimal weight.

2 | MATERIALS AND METHODS

Historical data on sampling and analyses results for food and feed batches can be used as an information source for designing risk-based food safety monitoring plans. Such data contain information on, for example, product category, product name, product country of origin, hazard name (e.g., contaminant or pathogen), and the analyzed concentration of the food safety hazard. Different batches have different likelihoods of being contaminated (i.e., the hazard in the batch is present in a concentration above a predefined threshold), due to differences in origin, the hazard of interest, processing procedures, weather conditions, and other possible reasons. Modeling can identify and analyze patterns in such datasets to classify a batch as contaminated/positive or uncontaminated/negative. Due to the unbalanced nature of food safety monitoring data, a BN classifier was firstly used in this study to make this classification, and then WBN was applied to find a better balance between the positive classification accuracy (A_p) and the negative classification accuracy (A_n) (see Section 2.3). By improving the balance between A_p and A_n, the monitoring effectiveness was improved (see Section 2.4).

2.1 | Data

A food safety monitoring dataset containing the analytical results for the presence of HMs (e.g., arsenic, cadmium, lead, and mercury) in animal feed batches (e.g., feed and feed additives) was retrieved from the official national control program animal feeds (15,672 records) and private industry monitoring (5804 records) in the Netherlands. The results span a 12-year period from January 1, 2005 to December 31, 2016 (21,476 records in total). Most of the monitoring data were collected following Regulation (EC) No. 333/2007 (European Commission, 2007). The monitoring records were retrieved from the Quality of Agricultural Products (KAP) database[2].

Each row (record) in the HM dataset represents the analytical result from one sample from a feed batch analyzed for one HM. Each record covered the following information: product name, product subgroup, hazard name, country of origin, country of analysis, analysis method, determined concentration, analysis lab, and respective legal limit (added manually). The determined concentration of the HM in the feed sample was compared to the respective legal maximum level (ML), as laid down in Commission Regulation (EU) 2019/1869 amending to Directive 2002/32/EC, to represent

[1] This terminology of sampling approach is different from the sampling for food safety monitoring. Here, sampling refers to the approach to improve the balance across minority class and majority class.

[2] https://www.narcis.nl/research/RecordID/OND1304218/Language/en.

TABLE 1 Variable names, descriptions, and labels of the heavy metal (HM) dataset

Nodes	Descriptions	Labels
Year	Registration year	[2005, 2006, …,2016]
Product name (Pn)	Type of feed product	[P_1, P_2, …,P_n]
Product subgroup (Ps)	The feed product subgroup	[Ps_1, Ps_2, …,Ps_n]
Hazard name (Hc)	The particular heavy metal	[H_1, H_2, …,H_4]
Country of origin (Co)	The country of origin of the feed product	[Co_1, Co_2, …,Co_n]
Country of analysis (Ca)	The country where the product was analyzed	[Ca_1, Ca_2, …,Ca_n]
Analysis method (Am)	The analytical method applied	[Am_1, Am_2, …,Am_n]
Analysis lab (Al)	Laboratory where analysis was conducted	[Al_1, Al_2]
Legal limit (Ll)	The maximum legal limit for the hazard—product	[L_1, L_2, …,L_n]
Above or below ML	Above or below the maximum legal limit	[Positive, negative]

the compliance condition (positive/negative). In this case, one record in the HM dataset was assumed to represent one feed product batch. The analytical result was assumed to represent (with 100% certainty) whether the batch was contaminated or not. These records were assumed to represent the actual distribution of HMs in feed. The variables of the HM dataset, together with their corresponding descriptions and labels, are listed in Table 1. The HM dataset was highly unbalanced; it included 138 positive records and 21,338 negative records (an unbalanced ratio of 0.65%).

2.2 | Introduction of Bayesian network and weight-added Bayesian network

2.2.1 | Bayesian network

BN is a powerful classification model that has been widely used in the prediction of food safety hazards in food and feed (C. Liu et al., 2018; N. Liu et al., 2021; Marvin & Bouzembrak, 2020; Marvin et al., 2020; Wang et al., 2022; Xu et al., 2018). It is a probabilistic graphical model that consists of two steps (Pearl, 1987): (i) learning the graphical structure between the nodes/variables and linking them to represent their relationships and (ii) learning the model parameters such as the conditional probabilities to quantify the extent of the relationships between the nodes/variables (Neapolitan, 2004). This study used the tree-augmented naive bayes (TAN) structure learning algorithm (Friedman et al., 1997). Earlier research demonstrated its promise for food safety monitoring (Bouzembrak and van der Fels-Klerx, 2018). For applying TAN, conditional probabilities are calculated using the expectation–maximization algorithm (Dempster et al., 1977). The posterior probability for each class label is calculated as the product of the prior probability and the conditional probability. Given a set of observations X_1, …, X_n, with known class labels, the BN model estimates the posterior probabilities for each class label: $P(C|X_1, \ldots, X_n)$ for any X_i. Then, a posterior probability is calculated with the TAN algorithm as:

$$P(C|X_1, \ldots, X_n) = P(C) \times P(X_{\text{root}}|C) \prod_{i=1}^{n} P\left(X_I|C, X_{\text{parent}}\right), \quad (1)$$

$P(C|X_1, \ldots, X_n)$ represents the posterior probability of one event belonging to class C. The sum of all posterior probabilities of one event equals one. $P(C|X_1, \ldots, X_n)$ is calculated as the product of the prior probability $P(C)$, the conditional probability of the root variable $P(Xroot|C)$, and the conditional probability of all the attributes $P(X_i|C, Xparent)$. The classification result is calculated based on the posterior probability, meaning one sample is classified into the class label with the highest posterior probability. Usually, the classification accuracy for each class is unbalanced when analyzing unbalanced data.

2.2.2 | Weight-added Bayesian network

WBN classifier can balance the classification accuracy for each class. The WBN classifier is designed by applying Bayes Minimum Risk (BMR) theory to a BN classifier. BMR theory assigns different weights to the posterior probabilities obtained from the BN classifier, and thus changes the original classification decision, such as changing C_i into C_j, to result in a balanced classification accuracy (Ghosh et al., 2006). BMR theory consists of the classification weight matrix, loss function, and minimum risk classification.

The classification weight matrix (Elkan, 2001) includes the assigned differential weight values for each classification decision. The weight matrix is represented by a $|C|*|C|$ rank matrix, where $|C|$ is the number of target class labels. The loss occurs when the predicted class is different from the actual class.

The loss function is defined as:

$$L(C_i, x) = \sum_{j=1}^{c} W\left(C_i, C_j\right) P(C_j|X_1, \ldots, X_n), i = 1, \ldots c, j = 1 \ldots .c, \quad (2)$$

where C_i and C_j represent classes i and j, and $W\ (C_i,\ C_j)$ represents the weight of classifying C_j into C_i. $L(C_i, x)$ represents the classification loss when classifying one sample into C_i. $L(C_i, x)$ is calculated as the product of $W(C_i, C_j)$ and the posterior probability $P(C_j|\ X_1, \ldots, X_n)$.

The classification decision determines to which class label a sample should be assigned. The classification result is considered optimal when the classification losses are minimal (Elkan, 2001; Juang et al., 1997). The classification formula is defined as:

$$\text{Class} = argminL\left(C_i, x\right), i = 1, \ldots, c. \tag{3}$$

The classification performance is evaluated using a confusion matrix and the classification accuracy. A confusion matrix compares the actual label with the predicted label of the target class. The classification accuracy is calculated based on the confusion matrix and describes the prediction accuracy for each class.

2.2.3 | Binary classification task

In food safety monitoring, classifying whether one feed/food batch is contaminated/positive or uncontaminated/negative is a binary classification task. For the sake of similarity, formula (2) was reformulated into formula (4) and formula (3) into formula (5) for the specific case of a binary task. Then, the classification weight matrix is expressed as $W\ (C_n,\ C_p)$, $W\ (C_n,\ C_n)$, $W(C_p,\ C_n)$, and $W(C_p,\ C_p)$, and the classification losses as $L(C_p, x)$ and $L(C_n, x)$:

$$\begin{array}{c} L\left(C_p,\ x\right) = W\left(C_p, C_n\right) P(C_n|X_1, \ldots, X_n) \\ +W\left(C_p, C_p\right) P(C_p|X_1, \ldots, X_n) \\ (C_n, x) = W\left(C_n, C_p\right) P(C_p|X_1, \ldots, X_n) \\ +W\left(C_n, C_n\right) P(C_n|X_1, \ldots, X_n) \end{array}, \tag{4}$$

$$\text{Class} = argminL\left(C_i, x\right), i = p, n. \tag{5}$$

According to the BMR decision, if $L(C_p, x) \geq L(C_n, x)$, the sample is classified into class C_n, else C_p.

Applying the BMR decision to the BN model to develop the WBN model could change the original classification decision, such that C_n changes into C_p, or C_p changes into C_n. For example, when $W\ (C_n,\ C_p)$, $W\ (C_n,\ C_n)$, $W(C_p,\ C_n)$, and $W(C_p,\ C_p)$ were set at, respectively, 100, 0, 1, and 0, the posterior probability $P(C_n\ |\ X_1, \ldots, X_n)$ and $P(C_p\ |\ X_1, \ldots, X_n)$ of one sample, calculated by the BN model, was 0.9 and 0.1, respectively. If only the BN result was used, the sample was classified as C_n. When applying WBN, $L(C_p,\ x)$ was calculated as $1 \times 0.9 + 0 \times 0.1 = 0.9$, and $L(C_n,\ x)$ was calculated as $100 \times 0.1 + 0 \times 0.9 = 10$; because $L(C_p,\ x) < L(C_n,\ x)$, the sample was classified as C_p.

2.3 | BN and WBN for food safety monitoring

The BN model was constructed using the HM dataset to identify the compliance condition (positive/negative) of a feed batch, meaning whether it presents a high or low risk of being contaminated with the particular HM. Model input variables were product name, product subgroup, hazard name, country of origin, country of analysis, analysis method, analysis lab, and respective legal limits. The model output variable was the compliance condition (positive/negative). Five-fold cross-validation (a process when all data are randomly split into k folds, in our case $k = 5$, and then the model is trained on the $k - 1$ folds, while one fold is left to test a model) was performed to present the overall performance of the model, and one of the five-fold cross-validation results was used as input for WBN model development. Since this study uses unbalanced data and aims to improve the model performance and the monitoring effectiveness, additional analysis was undertaken, such as value of information analysis and sensitivity analysis. The results of this additional analysis of the Bayesian network were obtained using Hugin 8.9 software (https://www.hugin.com/) and presented in Appendices figure A2 and figure A3. Value of information aims to assess the effect of a specific variable on the output variable node (Cover & Thomas, 2006), which in this case is the HM concentration being "above or below legal limit." Sensitivity analysis aims to assess the effect of changing a specific input variable on the change of an output variable (Castillo et al., 1997). For example, if a reduction in the likelihood of the causal factor attached to one feed product is assumed, sensitivity analysis estimated the reduction in the probability of concentration of HM in feed product above the legal limit.

A WBN was constructed with the parameter setting of weight value $W\ (C_n,\ C_p)$, $W\ (C_n,\ C_n)$, $W(C_p,\ C_n)$, and $W(C_p,\ C_p)$ to balance the classification result from BN. Food safety expert knowledge indicated that the weight of misclassifying contaminated samples to the negative class was much higher than the weight of misclassifying uncontaminated samples to the positive class[3]. Then, $W(C_p,\ C_n) = 1$ and $W(C_n,\ Cp) = 1, 10, \ldots, 1 \times 10^7$ were set for the sake of simplicity. In this way, the ratio of $W(C_p,\ C_n)$ to $W(C_n,\ C_p)$ was changed by adjusting the value of $W(C_n,\ C_p)$. At the same time, we set $W(C_p, C_p) = 0$ and $W(C_n, C_n) = 0$ because no loss occurred in case of a correct classification. Based on the abovementioned weight values and formulas (4) and (5), the classification results were obtained for the actual contaminated feed/food batch (C_p) classified as being contaminated (C_p) (true positive (TP), the actual uncontaminated batch (C_n) classified as contaminated (C_p) (false positive [FP]), the actual uncontaminated batch classified (C_n) as uncontaminated (C_n) (true

[3] Misclassifying contaminated samples to the negative class increases the chance of human consumption of contaminated food; misclassifying uncontaminated samples to the positive class increases the chance that more food/feed products are collected for sampling and analysis, wasting resources in the case of risk-based monitoring.

negative [TN]), and the actual contaminated product (C_p) classified as uncontaminated (C_n) (false negative [FN]).

The classification accuracy was calculated as:

$$A_P = N_{TP} \div (N_{TP} + N_{FN}), \quad (6)$$

$$A_n = N_{TN} \div (N_{TN} + N_{FP}), \quad (7)$$

$$A_t = (N_{TP} + N_{TN}) \div (N_{TN} + N_{FP} + N_{TP} + N_{FN}), \quad (8)$$

where A_p represents the prediction accuracy of the positive samples, A_n represents the prediction accuracy of the negative samples, and At represents the prediction accuracy of all (positive and negative) samples. N_{TP}, N_{TN}, N_{FP}, and N_{FN} represent the number of samples classified as *TP, TN, FP*, and *FN*. Given the different settings of $W(C_n, C_p)$, the balance between A_p and A_n was different.

2.4 | Monitoring effectiveness as a criterion for optimal balance

Given limited resources, a monitoring plan usually has a pre-set number of samples (N_s) that can be collected and analyzed. Risk-based monitoring recommends collecting more samples from batches with a higher likelihood of being contaminated (and fewer samples from batches with a lower likelihood of being contaminated) to increase the probability of detecting batches that are contaminated. The effectiveness of the monitoring plan (E_m) was defined as the ratio between the number of identified contaminated batches relative to the total number of actual contaminated batches, expressed as a percentage. E_m varies both under different N_s and according to the balance between A_p and A_n, and thus can be used as a criterion to determine the optimal weight.

The effectiveness of the monitoring plan (E_m) was expressed as:

$$E_m = \frac{\text{Number of identified contaminated batches}}{\text{Number of actual contaminated batches}}.$$

When there is no pre-set N_s, the number of identified contaminated batches is N_{TP}. The number of actual contaminated batches is $N_{TP}+N_{FN}$. However, N_s is usually pre-set in practice, implying only a small quantity of batches can be sampled and analyzed. When the number of batches classified as contaminated ($N_{TP}+N_{FP}$) $\geq N_s$, N_s samples are collected from the $N_{TP}+N_{FP}$; when ($N_{TP}+N_{FP}$) $< N_s$, N_s samples are collected from $N_{TP}+N_{FP}$ and the remaining ones ($N_s - N_{TP} - N_{FP}$) from $N_{TN}+N_{FN}$. The probability of an actual contaminated batch (N_{TP}) being chosen from the "flagged" contaminated batches ($N_{TP}+N_{FP}$) is $P_p = N_{TP}/(N_{TP}+N_{FP})$; the probability of an actual contaminated batch being chosen from the predicted un-contaminated batch is $P_n = N_{FN}/(N_{TN}+N_{FN})$.

TABLE 2 Examples of Bayesian network (BN) classifier's results

Observational information									Actual class	Posterior probability		Predicted class
Year	Pn	Ps	Hc	Co	Ca	Am	Al	Ll	Above or below ML	Negative	Positive	Above or below ML
2005	Rabbit food	PET	Arsenic	FR	NL	Am_1	WFSR	2 mg/kg	Negative	0.99	1.47×10^{-6}	Negative
2005	Pig feed	PIG	Arsenic	NL	NL	Am_2	PVVR	10 mg/kg	Negative	0.99	2.9×10^{-7}	Negative

Note: "Negative probability" represents the posterior probability of being negative class. "Positive probability" represents the posterior probability of being positive class.

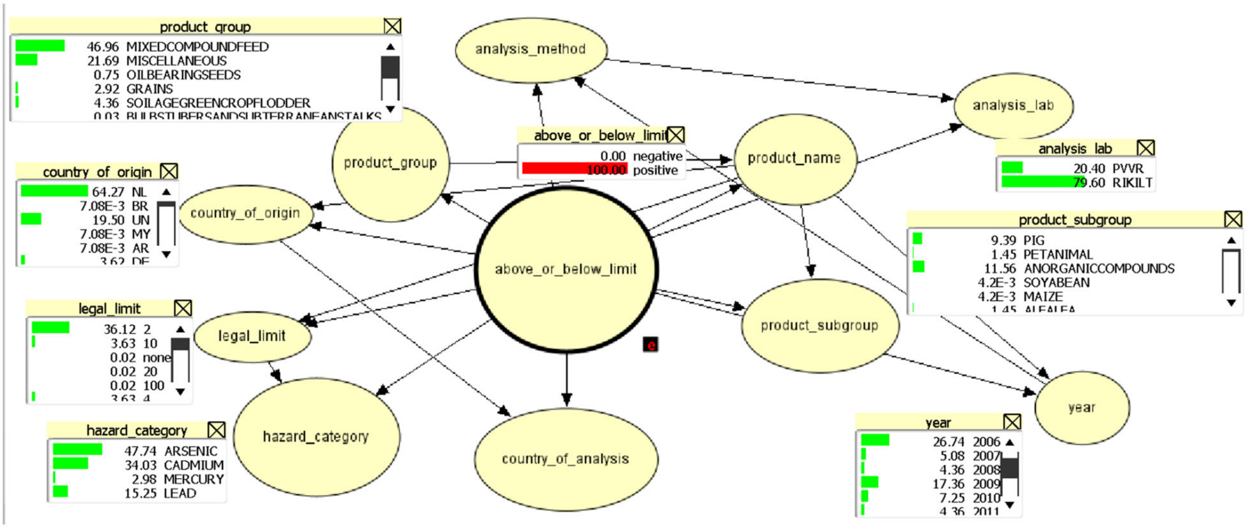

FIGURE 1 The structure of the Bayesian network (BN) model and attributions of several model parameters for the prediction of the presence of heavy metal in feed products (positive).

Then, we get:

$$E_m = \begin{cases} \frac{N_{TP}}{N_{TP}+N_{FN}} \times \frac{N_s}{N_{TP}+N_{FP}}, N_p > N_s \\ (N_{TP}+N_{FP}) \times P_P + \frac{[(N_s-(N_{TP}+N_{FP}))\times P_n]}{N_{TP}+N_{FN}}, N_p < N_s \end{cases}. \quad (9)$$

N_s was set as 50, 100, 5000, 1000, 2000, … 6000, an approximate annual range in historical monitoring records. The classification results (*TP*, *TN*, *FP*, and *FN*) and the balanced classification accuracy between A_p and A_n, as well as E_m were calculated using formulas (4–9). The optimal weight value was determined as the one that resulted in the highest E_m.

3 | RESULTS

3.1 | BN model result

The BN model was composed of (i) a set of 11 variables, namely year, product name, product subgroup, product group, hazard name, country of origin, country of analysis, analysis method, analysis laboratory, the legal limit, and above or below respective limits; (ii) a set of labels for each variable (i.e., the node above or below ML has the following labels: positive and negative); and (iii) a set of directed links between the variables with an assigned conditional probability for each variable. Figure 1 presents the BN model structure and the attributions of several model parameters for the prediction of the presence of HMs in feed products. The table list for each variable reflects the variation (based on posterior probability) of model parameters given the evidence of "above or below legal limit" is "positive." For example, the yearly variation for positive records over the years was presented next to the node "year." Heavy metals present in feed at "above legal limits" and "below legal limits" displayed simply random variation but decreased in general over the years 2005–2016. One reason was that fewer samples of feed were collected for analyzing the presence of HMs. The yearly variation of HMs present in feed at "above legal limits" and "below legal limits" over the 12 years period are presented in Appendix figure A1. Heavy metals present in feed at "above legal limits" generally have low levels. However, when the plants (used to process feed) grow in a strongly contaminated environment, the HM concentration in feed materials of plant origin can increase due to adhering soil particles and/or uptake of the plant (EFSA, 2004a, 2004b, 2005, 2008). A significant increase of HMs present in feed at "above legal limits" was seen in 2006 and 2013.

Table 2 shows the first few records of the classification results derived from the BN classifier, including the observation information of the feed batch, the actual class of the feed batch, the posterior probabilities of each class, and the predicted class of the feed batch. For instance, the first row shows the information for a feed batch of rabbit food, belonging to the pet food group, originating from France, and analyzed in the Netherlands in the laboratory of WFSR for the presence of Arsenic using the "Am_1" analytical method in 2005. The actual class was negative, meaning that the determined concentration of arsenic was below ML (ML for rabbit food is 2 mg/kg). The posterior probabilities (calculated by the BN classifier) indicating the probability that the batch was classified as negative and positive, respectively, were 0.99 and 1.47×10^{-6}. The batch was classified into the negative class because the posterior probability of the feed batch being positive was higher than the posterior probability of it being negative. The overall classification accuracy using five-fold cross-validation was 0.99.

Table 3 presents the classification result using the BN classifier. The classification accuracies A_p (20%) and A_n (99%) were highly unbalanced. In the confusion matrix, N_{TP}, N_{FP},

TABLE 3 Confusion matrix and classification accuracy of the classification results using the Bayesian network (BN) classifier

Classification result		Number of samples and classification rate
Confusion matrix	N_{TP}	28
	N_{FP}	28
	N_{TN}	21,310
	N_{FN}[1]	110
Classification accuracy	Negative	99.87% (21,310/21,338)
	Positive	20.29% (28/138)
	Total	99.36% (21,200/21,338)

[1]The numbers of batches predicted into the four categories of TP, FP, TN, and FN.

TABLE 4 Confusion matrix and classification accuracy of the classification result using the weighted Bayesian network (WBN) classifier

		Weight value							
Classification result		1	1×10^1	1×10^2	1×10^3	1×10^4	1×10^5	1×10^6	1×10^7
	N_{TP}	28	62	75	87	100	110	111	117
	N_{FP}	28	203	535	1001	1530	2298	3293	4473
Confusion matrix	N_{TN}	21,310	21,135	20,803	20,337	19,808	19,040	18,045	16,865
	N_{FN}	110	76	63	51	38	28	27	21
	Negative	0.99	0.99	0.97	0.95	0.93	0.89	0.85	0.79
Classification accuracy	Positive	0.2	0.45	0.54	0.63	0.72	0.8	0.8	0.85
%	Total	0.99	0.99	0.97	0.95	0.93	0.89	0.85	0.79

N_{TN}, and N_{FN} were, respectively, 28, 110, 28, and 21,310, meaning that only 28 positive batches were predicted correctly out of a total of 138 actual positive batches, and 21,310 negative batches were predicted correctly out of a total of 21,420 actual negative batches.

3.2 | WBN model result

Table 4 shows the classification results using the WBN classifier. The WBN classifier provides more balance between A_p and A_n compared to the BN classifier (Table 3). For example, given the weight value ($W(C_n, C_p)$) of 1×10^5, the values of N_{TP}, N_{FP}, N_{TN}, and N_{FN} were 110, 2298, 19,040, and 28, respectively. The classification accuracy of 85% for A_p and of 79% for A_n (using WBN), compared to a classification accuracy of 20% for A_p and of 99% for A_n (using BN). Although the balance of A_p and A_n improves using WBN, it is still hard to identify the optimal weight. Therefore, we used the monitoring effectiveness (E_m) to determine the optimal weights (see Section 4.3).

3.3 | Monitoring effectiveness

Table 5 and Figure 2 present the E_m under different pre-set sample sizes (N_s). E_m changed with different weight values (and its corresponding classification accuracy of A_p and A_n). In this case, the highest E_m was used as the criterion to choose the best balance between A_p and A_n. For instance, when the N_s was pre-set at 3000, E_m was the highest at 80% with a weight value of 1×10^5. Thus, here the corresponding classification accuracy A_p of 80% and A_n of 89% (Table 4) were the optimal values. For the average number of about 2000 samples per year in the historical monitoring dataset, the *Em* varies from 0.28 to 0.73 (Table 5).

4 | DISCUSSION

In our study, we used the WBN classifier to improve the balance between the positive and negative classification accuracy when using unbalanced food safety data and to find the criterion to determine the best balance. We demonstrated this approach for the case of monitoring HMs in animal feeds, where we explored the use of the monitoring effectiveness to determine the optimal balance between the negative and positive classification accuracy. To the best of our knowledge, this is the first study that explored choosing the best balance using monitoring effectiveness as the criterion.

Some studies apply techniques prior to (e.g., sampling) or after (e.g., thresholding) an existing total accuracy-based classifier to deal with unbalanced data, to get a more balanced classification accuracy for each class (Ling, 2006 ; Sheng & Ting, 2002). Our study applied the BMR method after the existing accuracy-based BN classifier which is in line with

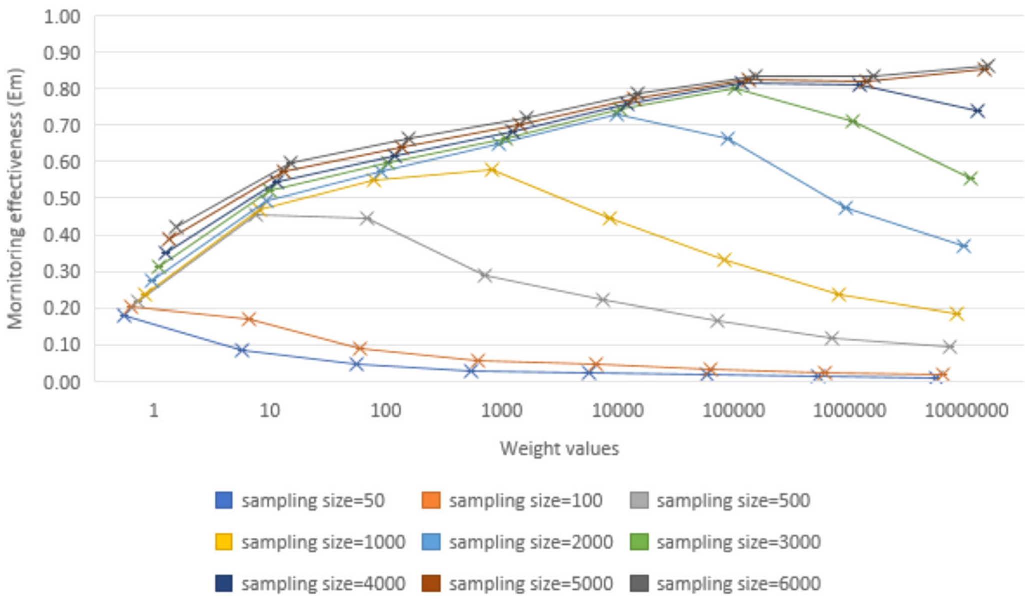

FIGURE 2 Monitoring effectiveness of the weighted Bayesian network (WBN) classifier using different weight values and pre-defined sample sizes.

previous methods. BMR allows for cost-sensitive learning, as it considers the misclassification cost by adding weight to each classification decision. Some studies have shown that the CSBN performs better than normal BN (Jiang et al., 2014; Nashnush & Vadera, 2017; Xu et al., 2018). These studies evaluated the model performance and the impact of weight values in terms of misclassification cost and classification accuracy. Our study results are consistent with the findings of these studies. The applied WBN highly improved the modeling classification accuracy as compared to BN. Xu et al. (2018) compared the CSBN model with BN, and the Synthetic Minority Over-sampling Technique (SMOTE) sampling method. Their results showed that the CSBN model performed the best. Nashnush and Vadera (2017) and Jiang et al. (2014) compared the CSBN classifier to the original BN classifier on more than 30 datasets, showing that the performance of CSBN classifier is indeed superior. These studies showed that it is possible to balance the classification accuracy for each class, but did not investigate how to choose the optimal weight. Compared to the abovementioned studies, our study first confirmed that the WBN could result in a more balanced prediction accuracy and, second, it explored the criterion (the highest monitoring effectiveness) for choosing the optimal weight. Using another method such as thresholding can achieve similar results. In this study, we provide the WBN method, instead of thresholding, because (1) WBN works for binary classification and multi-classification cases as we provided in the method section using formula (1–5). The threshold method works easily for binary classification cases but not for multi-classification cases. (2) There is no clear principle for threshold selection using the thresholding method, but WBN could use expert knowledge to assess the loss of misclassification for each class and thus set the weight for each class.

In our study, we used the structure of the TAN Bayesian network (Friedman et al., 1997) in which the class variable has no parents and each attribute has as parents the class variable and at most one other attribute. Thus, each attribute can have one augmenting edge pointing to it. Studies in food safety domine (Marvin et al., 2016 ;Wang et al., Wang et al., 2022) and other domains used a similar structure of TAN (Jayech & Mahjoub, 2012; Madden, 2009). This TAN structure was learned based on a data-driven approach which captured the directed probabilistic relationship (conditional probability tables) and not the causal relationship. TAN differs from a causal BN, which captures the causal relationship between attributes and specifies what happens under any variable intervention. The causal BN structure could be built using expert judgment to decide the causal relationships between attributes and replace the conditional probability for the variables with a new table. Both approaches of building BN structures have their limitations: (1) expert knowledge elicitation is time-consuming and relies on experts having knowledge of the full domain and (2) a data-driven approach often leads to inefficiencies given small datasets (Flores et al., 2011). We compared the model performance of these two structures (data driven vs. expert based, here the expert knowledge means the causal relationships were built based on the authors' knowledge), and the model performance was almost the same. This may be because the current dataset was large enough to derive the causal relationships between attributes. However, using export knowledge to build the structure is an advantage of BN compared to other machine learning (ML) methods, especially when a "black swan" issue could happen (i.e., extreme and relatively unknown circumstances) and no related data are available in the context of food safety monitoring. For example, the conflict between Russia and Ukraine, which started in early 2022, resulted in

TABLE 5 Monitoring effectiveness of the weighted Bayesian network (WBN) classifier using different weight values and different pre-defined sample sizes

Sample size	**Monitoring effectiveness (E_m) according to different weight values**							
	1	**1×10^1**	**1×10^2**	**1×10^3**	**1×10^4**	**1×10^5**	**1×10^6**	**1×10^7**
50	0.18	0.08	0.04	0.03	0.02	0.02	0.01	0.01
100	0.20	0.17	0.09	0.06	0.04	0.03	0.02	0.02
500	0.22	0.46	0.45	0.29	0.22	0.17	0.12	0.09
1000	0.24	0.47	0.55	0.58	0.44	0.33	0.24	0.18
2000	0.28	0.49	0.57	0.65	0.73	0.66	0.47	0.37
3000	0.31	0.52	0.60	0.67	0.74	0.80	0.71	0.55
4000	0.35	0.55	0.62	0.68	0.76	0.81	0.81	0.74
5000	0.39	0.57	0.64	0.70	0.77	0.82	0.82	0.85
6000	0.42	0.60	0.66	0.72	0.79	0.84	0.83	0.86

disturbances in the agro-food supply chains and severe damage to farming operations in these regions, thus changing the availability of feed commodities, and patterns of trade. Therefore, it is highly recommended to incorporate expert knowledge as new information into a data-driven monitoring plan, especially under extreme and relatively unknown circumstances. What is more, due to very limited "positive records," our study focused on dealing with unbalanced data to improve the model performance and the monitoring effectiveness, and using extra data (either simulated data or collected data) to test the model performance was not taking into account in our study. Further research could explore the predictive accuracy of the method with a simulated power analysis, by hypothesizing a change in the contamination rate for the next year. Using new data when available to sequential update the model (maybe once per year) is recommended. The periodic updating could allow quicker detection of the change in the data and keep the model parameters up to date. Our proposed monitoring plan focuses on collecting samples from batches with a high likelihood of being contaminated, assuming that the sample collection and analyses allow detection of batch contamination (when present) with certainty. Besides the performance of the analytical method, detecting a contamination also depends on the distribution of the hazard in the batch. The likelihood of sampling a contaminated spot in the batch is higher with homogenously distributed food safety hazards such as HMs, but much lower for heterogeneously distributed food safety hazards such as mycotoxins. Bouzembrak and van der Fels-Klerx (2018) investigated various sampling strategies and showed that the detection probability was affected by the numbers of collected samples, the contamination level, and the sampling strategy (simple random sampling, stratified random sampling, and systematic sampling). Therefore, in the design of a complete sampling plan for food safety monitoring, it is important to combine research insights on the identification of high-risk batches and on the sampling strategy, especially for heterogeneously distributed food safety hazards (such as mycotoxins).

In addition to food safety monitoring data, the approach explored in our study can be applied in other classification tasks to deal with unbalanced data in different contexts, for example, using disease diagnosis data or fraud detection data. Moreover, our study provided a methodology of finding the criterion for determining the optimal weight values. This methodology can provide insights for researchers in other research fields to consider the practical criterion for determining an optimal balance.

5 | CONCLUSION

This study explored the use of a WBN classifier for balancing the classification accuracy of each class in the context of unbalanced food safety monitoring data. In our specific case study, based on historical data on HMs in feed, the optimal weight for the modified BN classifier was set as the value that maximized the monitoring effectiveness. Results showed the WBN resulted in a better balance between the positive and negative classification accuracy (both around 80%) compared to the BN classifier (positive classification accuracy of 20% and negative classification accuracy of 99%). The approach proposed in this study can also be used to increase the monitoring effectiveness for other hazards in food, animals, or plants.

AUTHOR CONTRIBUTIONS

All authors participated in the conceptualization of the study. X. Wang has defined the methodology, performed the analysis and data interpretation, developed the model, and drafted the manuscript. H. J. van der Fels-Klerx, Y. Bouzembrak, and A. G. J. M. Oude Lansink contributed to the methodology and the writing of the manuscript. All authors read and approved the submitted manuscript.

ACKNOWLEDGMENTS

This research received financial support from the Chinese Scholarship Council and the Netherlands Ministry of Agriculture and Food Quality (WOT-02-004-012). The authors thank Paulien Adamse for making the dataset available for this study, under project WOT-02-004-012.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data were obtained from the official control program of food safety hazards in animal feed in the Netherlands (via project 'WOT-02-004-012′) and, hence, cannot be shared freely by the authors.

ORCID

Xinxin Wang https://orcid.org/0000-0002-8633-8971

REFERENCES

Akila, S., & Srinivasulu Reddy, U. (2018). Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. *Journal of Computational Science*, *27*, 247–254. https://doi.org/10.1016/j.jocs.2018.06.009

Bouzembrak, Y., & van der Fels-Klerx, H. J. (2018). Effective sampling strategy to detect food and feed contamination: Herbs and spices case. *Food Control*, *83*, 28–37. https://doi.org/10.1016/j.foodcont.2017.04.038

Castillo, E., Gutiérrez, J. M., & Hadi, A. S. (1997). Sensitivity analysis in discrete Bayesian networks. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 27(4), 412-423.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

EC-European Commission. (2002). Directive (2002/32/EC) of the European Parliament and of the Council of 7 May 2002 on undesirable substances in animal feed. Off J Eur Union C, 221, 232.

European Commission. (2007). Commission Regulation (EC) No 333/2007 of 28 March 2007 laying down the methods of sampling and analysis for the official control of the levels of lead, cadmium, mercury, inorganic tin, 3-MCPD and benzo (a) pyrene in foodstuffs. *Official Journal of the European Union*, *88*, 29.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22. https://www.jstor.org/stable/2984875

Elkan, C. (2001). The foundations of cost-sensitive learning. *International Joint Conference on Artificial Intelligence*, *17*(1), 973–978. https://www.researchgate.net/profile/Charles_Elkan/publication/2365611_The_Foundations_of_Cost-Sensitive_Learning/links/02bfe513e192c02262000000/The-Foundations-of-Cost-Sensitive-Learning.pdf

Focker, M., van der Fels-Klerx, H. J., & Oude Lansink, A. G. J. M. (2018). Systematic review of methods to determine the cost-effectiveness of monitoring plans for chemical and biological hazards in the life sciences. *Comprehensive Reviews in Food Science and Food Safety*, *17*(3), 633–645. https://doi.org/10.1111/1541-4337.12340

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*(2–3), 131–163. https://geiger.net.technion.ac.il/files/2016/08/friedman1997Bayesian.pdf

Ghosh, J. K., Delampady, M., & Samanta, T. (2006). Bayesian inference and decision theory. An introduction to Bayesian analysis. *Theory and* methods (pp. 29–63). Springer.

jayech, K., & mahjoub, mohamed ali. (2012). New approach using Bayesian Network to improve content based image classification systems (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1204.1631

Jiang, L., Li, C., & Wang, S. (2014). Cost-sensitive Bayesian network classifiers. *Pattern Recognition Letters*, *45*, 211–216. https://doi.org/10.1016/j.patrec.2014.04.017

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5). Prentice hall.

Juang, B.-H., Hou, W., & Lee, C.-H. (1997). Minimum classification error rate methods for speech recognition. *IEEE Transactions on Speech and Audio Processing*, *5*(3), 257–265. https://doi.org/10.1109/89.568732

Julia Flores, M., Nicholson, A. E., Brunskill, A., Korb, K. B., & Mascaro, S. (2011). Incorporating expert knowledge when learning Bayesian network structure: A medical case study. Artificial Intelligence in Medicine, 53(3), 181–204. https://doi.org/10.1016/j.artmed.2011.08.004

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). Springer.

Liu, C., Manstretta, V., Rossi, V., & der Fels-Klerx, V. (2018). Comparison of three modelling approaches for predicting deoxynivalenol contamination in winter wheat. *Toxins*, *10*(7), 267.

Liu, N., Liu, C., Dudaš, T., Loc, M., Bagi, F., & Der Fels-Klerx, V. (2021). Improved aflatoxin and fumonisin forecasting models for maize (PREMA and PREFUM), using combined mechanistic and Bayesian network modelling—Serbia as a case study. *Frontiers in Microbiology*, *12*, 630.

Madden, M. G. (2009). On the Classification Performance of TAN and General Bayesian Networks. Research and Development in Intelligent Systems XXV, 3–16. https://doi.org/10.1007/978-1-84882-171-2_1

Marvin, H. J., & Bouzembrak, Y. (2020). A system approach towards prediction of food safety hazards: Impact of climate and agrichemical use on the occurrence of food safety hazards. *Agricultural Systems*, *178*, 102760.

Marvin, H. J., van Asselt, E., Kleter, G., Meijer, N., Lorentzen, G., Johansen, L.-H., Hannisdal, R., Sele, V., & Bouzembrak, Y. (2020). Expert-driven methodology to assess and predict the effects of drivers of change on vulnerabilities in a food supply chain: Aquaculture of Atlantic salmon in Norway as a showcase. *Trends in Food Science & Technology*, *103*, 49–56.

Marvin, H. J. P., Bouzembrak, Y., Janssen, E. M., van der Fels- Klerx, H. J., van Asselt, E. D., & Kleter, G. A. (2016). A holistic approach to food safety risks: Food fraud as an example. Food Research International, 89, 463–470. https://doi.org/10.1016/j.foodres.2016.08.028

Nashnush, E., & Vadera, S. (2017). Learning cost-sensitive Bayesian networks via direct and indirect methods. *Integrated Computer-Aided Engineering*, *24*, 17–26. https://doi.org/10.3233/ICA-160514

Neapolitan, R. E. (2004). *Learning Bayesian networks* (Vol. 38) Pearson Prentice Hall. http://www.cs.technion.ac.il/~dang/books/Learning%20Bayesian%20Networks(Neapolitan,%20Richard).pdf

Pearl, J. (1987). *Embracing causality in formal reasoning*. AAAI. https://www.aaai.org/Papers/AAAI/1987/AAAI87-066.pdf

Sheng, V. S., & Ling, C. X. (2006). *Thresholding for making classifiers cost-sensitive*. AAAI. https://www.aaai.org/Papers/AAAI/2006/AAAI06-076.pdf

Sheng, V. S., & Ling, C. X. (2007). Roulette sampling for cost-sensitive learning. In J. N. Kok, J. Koronacki, R. L. D. Mantaras, S. Matwin, D. Mladenič, & A. Skowron (Eds.), *European conference on machine learning* (pp. 724–731). Springer. https://doi.org/10.1007/978-3-540-74958-5_73

Sheng, V. S., & Ling, C. X. (2006, July). Thresholding for making classifiers cost-sensitive. In Aaai (Vol. 6, pp. 476-481).

Ting, K. M. (2002). An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering*, *14*(3), 659–665.

van Asselt, E., Noordam, M., Pikkemaat, M., & Dorgelo, F. (2018). Risk-based monitoring of chemical substances in food: Prioritization by decision trees. *Food Control*, *93*, 112–120.

Wang, X., Bouzembrak, Y., Marvin, H. J., Clarke, D., & Butler, F. (2022). Bayesian Networks modeling of diarrhetic shellfish poisoning in *Mytilus edulis* harvested in Bantry Bay, Ireland. *Harmful Algae*, *112*, 102171.

Wang, X., Bouzembrak, Y., Oude Lansink, A. G. J. M., & van der Fels-Klerx, H. J. (2022). Designing a monitoring program for aflatoxin B1 in feed products using machine learning. Npj Science of Food, 6(1). https://doi.org/10.1038/s41538-022-00154-2

Wang, X., Bouzembrak, Y., Lansink, A. O., & Fels-Klerx, H. J. (2021). Application of machine learning to the monitoring and prediction of food safety: A review. Comprehensive Reviews in Food Science and Food Safety, 21(1), 416–434. Portico. https://doi.org/10.1111/1541-4337.12868

Weiss, G. M., & Provost, F. (2001). The effect of class distribution on classifier learning: An empirical study. https://doi.org/10.7282/t3-vpfw-sf95

Xu, M., Gan, D., Shen, J., & An, B. (2018a). Bayesian network retrieval discrimination criteria model based on unbalanced information. In H. Chen, Q. Fang, D. Zeng, & J. Wu (Eds.), *International* conference on smart health (pp. 211–223). Springer. https://doi.org/10.1007/978-3-030-03649-2_21

EFSA. (2004a). Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the Commission related to lead as undesirable substance in animal feed. *EFSA Journal*, *71*, 1–20.

EFSA. (2004b). Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the Commission related to cadmium as undesirable substance in animal feed. *EFSA Journal*, *72*, 1–24.

EFSA. (2005). Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the Commission related to arsenic as undesirable substance in animal feed. *EFSA Journal*, *180*, 1–35.

EFSA. (2008). Scientific opinion of the Panel on Contaminants in the Food Chain. Mercury as undesirable substance in animal feed. *EFSA Journal*, *654*, 1–76.

How to cite this article: Wang, X., Bouzembrak, Y., Lansink, A. G. J. M. O., & van der Fels-Klerx, H. J. (2023). Weighted Bayesian network for the classification of unbalanced food safety data: Case study of risk-based monitoring of heavy metals. *Risk Analysis*, 1–13. https://doi.org/10.1111/risa.14120

APPENDIX

Figure A1, Figure A2, Figure A3

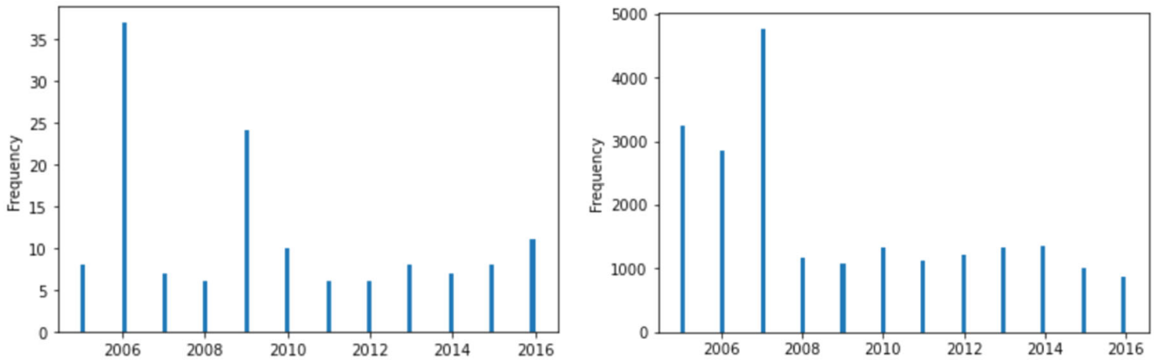

FIGURE A1 The yearly variation of heavy metals presented in feed "above legal limit" (left) and "below legal limit" (right) over the 12-year period. X axis represents the years, and Y axis represents the count of records.

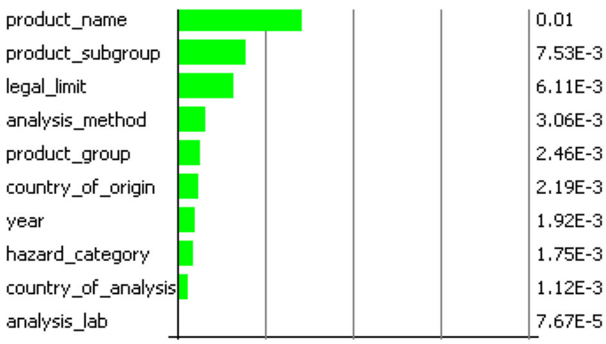

FIGURE A2 Value of information of parameters. The variables product name (0.01) was identified as having the greatest influence on the results (concentration of heavy metal [HM] in feed product above the legal limit) of the model.

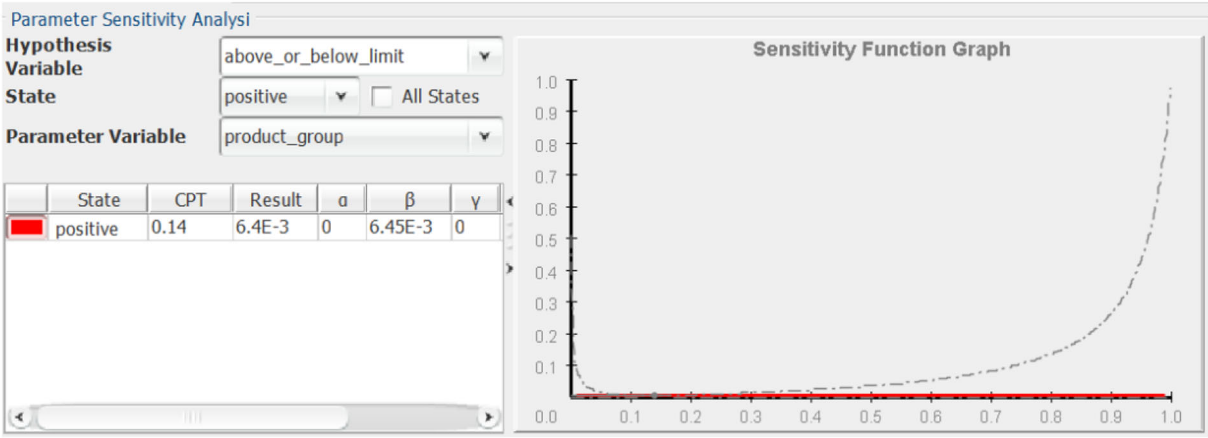

FIGURE A3 An example of the sensitivity analysis related to the parameter "grains" of the variable "product group" that affects the results (concentration of heavy metal [HM] in feed product being above the legal limit). If a reduction in the likelihood of the causal factor attached to the parameter grain is assumed, no reduction in the probability of the concentration of HM in the feed product of being above the legal limit was observed.