

Article



# Prediction of Safety Risk Levels of Benzopyrene Residues in Edible Oils in China Based on the Variable-Weight Combined LSTM-XGBoost Prediction Model

Cheng Hao <sup>1,2</sup>, Qingchuan Zhang <sup>1,2,\*</sup>, Shimin Wang <sup>1,2</sup>, Tongqiang Jiang <sup>1,2</sup> and Wei Dong <sup>1,2</sup>

- <sup>1</sup> National Engineering Research Centre for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing 100048, China; hc2680400194@163.com (C.H.)
- <sup>2</sup> China Food Flavor and Nutrition Health Innovation Center, School of E-Business and Logistics, Beijing Technology and Business University, Beijing 100048, China
- Correspondence: zhangqingchuan@btbu.edu.cn

**Abstract:** To assess and predict the food safety risk of benzopyrene (BaP) in edible oils in China, this study collected national sampling data of edible oils from 20 Chinese provinces and their prefectures in 2019, and constructed a risk assessment model of BaP in edible oils with consumption data. Initially, the k-means algorithm was used for risk classification; then the data were pre-processed and trained to predict the data using the Long Short-Term Memory (LSTM) and the eXtreme Gradient Boosting (XGBoost) models, respectively, and finally, the two models were combined using the inverse error method. To test the effectiveness of the prediction model, this study experimentally validated the model according to five evaluation metrics: root mean square error (RMSE), mean absolute error (MAE), precision, recall, and F1 score. The variable-weight combined LSTM-XGBoost prediction model proposed in this paper achieved a precision of 94.62%, and the F1 score value reached 95.16%, which is significantly better than other neural network models; the results demonstrate that the prediction model has certain stability and feasibility. Overall, the combined model used in this study not only improves the accuracy but also enhances the practicality, real-time capabilities, and expandability of the model.

Keywords: risk assessment; LSTM; XGBoost; risk prediction; edible oil; BaP

# 1. Introduction

Edible oil plays an indispensable role in daily life, enhancing the taste of food when frying and providing us with essential fatty acids. China is a large consumer of edible oil, with consumption reaching about 35.11 million tons in 2019 [1]. As China's economy continues to develop steadily and rapidly, coupled with population growth, improved living standards, and accelerated urbanization, people's consumption demand for edible oil will continue to grow steadily. In addition, the consumption landscape changed significantly with the onset of the COVID-19 epidemic. Quarantine and extended holiday initiatives were carried out across the country in response to the sudden spread of the epidemic. This led to many families stockpiling necessary living materials, including edible oil, which saw a corresponding increase in household consumption. Edible oil, being an essential item in Chinese kitchens, was particularly affected. In this article, edible oil generally refers to edible vegetable oil, edible animal oil, and edible oil products.

BaP is a polycyclic aromatic compound known for its toxic effects on the reproductive, blood, heart, nervous, and immune systems, and its ability to induce various cancers [2]. It is widespread in the environment, present in the atmosphere, surface water, sediment, soil, food, and fatty tissues, and can enter the food chain through various pathways, including biotransformation, impacting the metabolic processes of organisms. Some relevant studies have reported the addition of medicinal oil to extra virgin olive oil as a way to potentially mitigate BaP contamination, thereby improving health benefits and extending shelf



Citation: Hao, C.; Zhang, Q.; Wang, S.; Jiang, T.; Dong, W. Prediction of Safety Risk Levels of Benzopyrene Residues in Edible Oils in China Based on the Variable-Weight Combined LSTM-XGBoost Prediction Model. *Foods* **2023**, *12*, 2241. https://doi.org/10.3390/ foods12112241

Academic Editor: Robert L. Buchanan

Received: 15 April 2023 Revised: 26 May 2023 Accepted: 30 May 2023 Published: 1 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). life [3–6]. Therefore, the accurate determination of BaP content in edible oils is essential to assess their quality and safety and to safeguard the health of those who consume them [7].

The presence of BaP can be attributed to various sources [8,9]. One common route of BaP intake in humans is through the consumption of vegetable oil [10,11] which can contain a large amount of BaP-contaminated residues. Such contamination can occur during mechanical harvesting, transportation, and processing of the oilseeds, as these activities can lead to the oilseeds directly contacting pollutants, thus triggering the migration of BaP into the edible oil. In order to improve the oil yield and increase the aroma of the finished oil, such as peanut oil and sesame oil, the seeds are often fried before pressing, and the pressing process involves heating. This heating phenomenon during the vegetable oil pressing process, particularly when using the hot-pressing method, can produce a series of chemical reactions due to the high temperature, which may directly lead to the production of BaP, a carcinogen in edible oil [12]. In addition to direct contamination, there is also indirect contamination, such as asphalt contamination. Asphalt contains polycyclic aromatic hydrocarbons (PAHs), and farmers may contaminate soybeans by drying them on asphalt roads, resulting in oil pressed from these contaminated soybeans containing BaP [13,14].

Jiang et al. [15] conducted a health risk assessment of 75 randomly collected edible oils from Shandong Province, China, to evaluate the presence and hazards of PAH contamination, using Incremental Lifetime Cancer Risk (ILCR) as an evaluation metric. Their results indicated a widespread PAH contamination among the samples. Jang et al. [16] estimated the chronic daily exposure to BaP for the total population group and the consumer-only group using food consumption data from the fifth Korean National Health and Nutrition Examination Survey in 2011. Ref. [17] investigated 303 edible oils from Korea and used Margins of Exposure (MOEs) to understand the contamination levels of PAHs in them. Li et al. [18] used ILCR for assessing the risk of BaP in doughnuts. Gelavizh Barzegar et al. [19] used Monte Carlo simulations to characterize the daily intake MOEs and ILCR of edible oils sold in southwestern Iran. Bomi Kang et al. [20] employed MOEs to assess the risk of PAHs in Korean edible oils and found that despite the detection of PAHs, their effects on human exposure were not significant. The above studies only assessed the safety risk of BaP residues in edible oils by a single evaluation index and did not combine it with relevant food consumption data.

In recent years, the widespread use of deep learning prediction models in various fields, such as stock ticket price prediction [21–23], short-term traffic flow prediction [24–26], and urban air pollutant concentration prediction [27–29], has been enabled by the rapid development of artificial intelligence. Deep learning prediction models are also applicable to the requirements of food safety risk prediction. Jiang et al. [30] utilized deep learning to grade and predict the safety risk level of carbofuran pesticide residues in vegetables in China. Jiang et al. [31] proposed a risk prediction model for veterinary drug residues in freshwater products in China based on transform. Wang et al. [32] predicted the risk hazard of heavy metals in processed grain products using a voting integrated deep learning approach.

In this study, we used the national sampling data of BaP in edible oil in China in 2019 and the weekly consumption data of edible oil in each prefecture-level city as the basis for the in-depth calculation of evaluation indicators to build the data set. Firstly, we used the k-means algorithm to classify the evaluation indicators of edible oil by risk level, and then predicted the safety risk assessment indicators of edible oil in each prefecture-level city using the variable-weight combined LSTM-XGBoost prediction model, and classified these indicators according to the pre-defined risk level. The model proposed in this paper provides scientific and technical assistance for government regulatory authorities to monitor the safety of edible oils more effectively.

#### 2. Materials and Methods

2.1. Materials

# 2.1.1. Data Source

The data of BaP residues in edible oils in this study were obtained from the sampling data of the State Administration of Market Supervision of China 2019, covering 20 provinces, and contains a total of 12,826 samples. The consumption data of edible oils were obtained from the National Bureau of Statistics China Statistical Yearbook 2020. According to the national standard of China "Food Safety National Standard Limits of Contaminants in Food" (GB 2762-2017) [33], the maximum limit value of BaP in edible oil is 10  $\mu$ g/kg.

#### 2.1.2. Data Pre-Processing

In this study, the substitution method recommended by the World Health Organization (WHO) [34] was used for samples below the detection limit of the method. When the proportion of non-detects was less than or equal to 60%, the results of all samples with detection results less than the LOD were calculated as 1/2 of the LOD. When the proportion of non-detects was greater than 60%, the results of all samples with assays less than the LOD were calculated as the LOD. Since the data of samples with undetected BaP residues in this study were much lower than 60%, the undetected BaP levels in this study were calculated as 1/2 of the LOD value. LOD is the minimum limit of detection for BaP in edible oils according to the database used in this study and was taken as  $0.2 \mu g/kg$  [35].

# 3. Food Safety Risk Grading Assessment and Prediction Model

Considering that this study focused solely on the contamination status of BaP in edible oils and based on the basic principles of food safety risk assessment and sampling data of food products, three assessment indexes were selected for the risk assessment of edible oils: ILCR, MOE, and the Nemerow Integrated Pollution Index (NIPI).

## 3.1. Evaluation Indicators

## 3.1.1. Carcinogenic Risk Factor Method

*ILCR* [18,36–38] is the increased likelihood of developing cancer over a lifetime due to exposure to potential carcinogens. It is commonly used to assess the carcinogenic risk of a pollutant to humans and is calculated as:

$$ILCR_{BaP} = \frac{C \times TEF \times Ir \times Ep \times SF \times CF}{BW \times TA}$$
(1)

where *ILCR* is the Incremental Lifetime Cancer Risk that evaluates the carcinogenic risk of contaminants to humans; *C* is the concentration of chemical contaminants in edible oil, and the median *BaP* content in edible oil, mg/kg, is used in this study; *TEF* is the toxicity equivalence factor of *BaP*, *TEF* = 1; *Ir* is the daily intake of edible oil, kg/d; *Ep* is the exposure frequency, 365 d/a; *Ed* is the duration of exposure over the average human lifespan, 70 a (25,550 d); *SF* is the *BaP* carcinogenicity slope factor, 7.3 kg·d/mg; *CF* is the conversion factor,  $10^{-6}$  mg/ng; *BW* is the body weight, 60 kg; and *TA* is the average exposure time to chemical pollutants,  $70 \times 365$  d.

With reference to the interval of potential carcinogenic risk between  $10^{-6}$  and  $10^{-4}$  proposed by the US Environmental Protection Agency (US EPA) for *ILCR* [39], the carcinogenic risk is divided into three categories: *ILCR* < 1 × 10<sup>-6</sup>, the carcinogenic risk is negligible;  $1 \times 10^{-6} \le ILCR \le 1 \times 10^{-4}$ , the carcinogenic risk is acceptable; and *ILCR* > 1 × 10<sup>-4</sup>, the carcinogenic risk is not negligible.

### 3.1.2. Margin of Exposure Method

The *MOE* method [17,40] is used to evaluate the risk of *BaP* intake in the population, using the toxicity endpoint of primary hepatocellular carcinoma [41]. The calculation formula is as follows:

$$MOE = \frac{BMDL_{10}}{Exp}$$
(2)

where  $BMDL_{10}$  is the toxicity reference point, referring to the lower limit of the 95% benchmark dose confidence interval for a 10% incidence of hepatocellular carcinoma in animal toxicology experiments; this value is 0.07 mg/(kg·BW) for *BaP*.

$$Exp = \frac{F_i \times C_i}{BW \times 1000} \tag{3}$$

where *Exp* refers to the daily intake of *BaP* per kilogram of body mass due to the consumption of edible oils;  $F_i$  refers to edible oil consumption, kg/d;  $C_i$  refers to the average content of *BaP* in edible oil, µg/kg; and *BW* refers to the body mass, taken as 60 kg. According to the "Report on the Nutrition and Chronic Disease Status of Chinese Residents (2015)" [42], the average body mass of Chinese residents is 60 kg.

According to the recommendations of the European Food Safety Authority (EFSA) [43]: MOE > 10,000 is of very low health risk and does not require attention in public health, while MOE < 10,000 is of some health risk and requires attention.

## 3.1.3. Nemerow Integrated Pollution Index

The *NIPI* [44] reflects the characteristics of food contamination. Based on the sampling data of each province, the integrated contamination index was applied to calculate the contamination level of each sample, and the expression is as follows:

$$NIPI = \sqrt{\frac{P_{max(i,j)}^2 + P_{ave(i,j)}^2}{2}}$$
(4)

where *NIPI* is the integrated pollution index of food j in province *i*;  $P_{max(i,j)}$  is the maximum value of pollution index of food *j* in province *i*; and  $P_{ave(i,j)}$  is the average value of pollution index  $P_{i,j}$  of food *j* in province *i*.

$$P_{i,j} = \frac{X_{i,j}}{S_j} \tag{5}$$

where  $P_{i,j}$  is the contamination index of food *j* in province *i*;  $X_{i,j}$  is the detection value of *BaP* content in food *j* in province *i* (mg/kg); and *S<sub>j</sub>* is the national limit standard of *BaP* in food *j* (mg/kg), taken as 0.01 mg/kg here.

#### 3.2. Food Safety Grading Based on k-Means

The k-means clustering algorithm is a commonly used method of cluster analysis that divides a data set into k clusters so that the data points within a cluster are as similar as possible, while those between clusters are as different as possible. In the food safety risk classification, the three evaluation indicators of edible oil (ILCR, MOE, and NIPI) were clustered and analyzed as a way to assess the safety risk level of edible oil in different prefecture-level cities in each province over time. By dividing the food samples into different clusters, food samples with similar characteristics can be placed in the same cluster, thus providing more refined and targeted control measures for food safety management. The specific process of the algorithm is shown in Figure 1.

- (1) Select k objects from the data as the initial clustering centers;
- (2) Calculate the distance from each clustering object to the cluster center to divide the clusters;
- (3) Calculate each clustering center again;

(4) Calculate the standard measure function until the maximum number of iterations is reached and then stop; otherwise, continue the operation.





### 3.3. Food Safety Risk Level Prediction Model

Considering that food sampling data are time-series and non-linear, we selected the LSTM model and XGBoost which have been frequently applied to such problems and have achieved better results. However, the LSTM model is a neural network model and the XGBoost model is a tree model. The principles of the two models differ greatly and the correlation of the prediction results is weak, so this study used these two theories to propose a food safety risk level prediction model that combines the LSTM model and the XGBoost model together to improve the overall prediction accuracy of the model, as shown in Figure 2. The results of this model are weighted using the inverse error method, a process that has been shown to significantly improve the accuracy of the combined model.



Figure 2. Flow chart of the variable-weight combined LSTM-XGBoost prediction model.

## 3.3.1. LSTM Model

LSTM [45,46], also known as Long Short-Term Memory, is a variation of the traditional RNN, which can effectively capture the semantic association between long sequences and mitigate the gradient disappearance or explosion phenomenon compared with classical RNN. The structure of LSTM is complex; it includes an input layer, a hidden layer, and an output layer, each with many cells. Every cell in the hidden layer has memory cells, and the input gate, forgetting gate, and output gate collectively determine the output value. The internal structure of one of the hidden layer cells is shown in Figure 3 below.



Figure 3. Hidden layer cell of LSTM model.

In the figure,  $x_t$  is the input value at moment t;  $C_t$  is the memory cell at moment t;  $h_t$  is the hidden cell at moment t;  $f_t$  is the forgetting gate at moment t;  $i_t$  is the input gate at moment t;  $c_t$  is the candidate memory cell at moment t;  $O_t$  is the output gate at moment t; and tanh and  $\sigma$  are both activation functions.

The forgetting gate, with inputs from the previous module output  $h_{t-1}$  and the current time input data  $x_t$ , determines how much of  $C_{t-1}$  is retained or forgotten from the previous cell module input. The input gate decides which new inputs can be retained to  $C_t$ , while the output gate, with control of  $O_t$ , determines what information is output and what needs to be transferred to the next module. The mathematical principle is to multiply the long-term memory input  $C_{t-1}$  at t-1 by a forgetting factor  $f_t$ . The forgetting factor is calculated from the short-term memory  $h_{t-1}$  as well as the event information  $x_t$ .

The formula for the calculation process of the forgetting gate is as follows:

$$f_t = \sigma \Big( W_f \cdot [h_{t-1}, x_t] + b_f \Big) \tag{6}$$

The input gate determines the amount of input  $x_t$  saved to the unit state  $C_t$  at the current moment, and it determines the corresponding new attribute information in this unit module for the attribute information discarded in the forgetting gate, and adds it to supplement the discarded attribute information. The mathematical principle is to accept

the long-term memory  $i_t$  from the forgetting gate and the short-term memory  $C_t$  from the learning gate and then directly merge the two. The computational procedure for the input gate is given by:

$$i_t = \sigma \left( W_f \cdot [h_{t-1}, x_t] + b_f \right) \tag{7}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
(8)

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{9}$$

The current output gate  $O_t$  determines the extent to which the state of the control unit  $C_t$  is input to the current output value  $h_t$ . The mathematical principle is that  $O_t$  is obtained

using a Sigmoid function, and  $O_t$  is multiplied by  $tanh(C_t)$  to obtain the final output  $h_t$ . The output is calculated by the formulas

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

$$h_t = o_t \times \tanh(C_t) \tag{11}$$

3.3.2. XGBoost Model

The XGBoost model [25,47,48], which improves upon the Gradient Boosting Decision Tree (GBDT) model, utilizes a second-order Taylor expansion, unlike the traditional GBDT model which only uses a first-order Taylor expansion. This tends to complicate the GBDT model and makes it prone to overfitting. The XGBoost model incorporates features such as regularization, learning rate, column sampling, and the approximation of optimal splitting points, all of which help in preventing overfitting. XGBoost, being an integrated model comprising multiple trees, derives its prediction for a sample from the aggregate of the predicted values of each tree for that sample. The equation of the XGBoost model is as follows:

$$\hat{y}_i = \sum_{k=1}^{\kappa} f_k(x_i) \tag{12}$$

where *k* is the total number of trees;  $f_k(x_i)$  is the prediction result of the *k*th tree for the *i*th  $x_i$ ; and  $\hat{y}_i$  is the prediction result of the XGBoost model for the *i*th sample.

The objective function is expressed as.

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k)$$
(13)

where  $l(y_i, \hat{y}_i)$  denotes the training error of indicator sample  $x_i$  in the original sample and  $\sum_{k=1}^{k} \Omega(f_k)$  denotes the regularization term of the *k*th tree to prevent overfitting of the model.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^T \omega_j^2$$
(14)

where *T* is the number of leaf nodes in each tree;  $\omega_j^2$  is the weight of the *j*th leaf node; and  $\gamma$  and  $\lambda$  are coefficients, which need to be adjusted for the parameters in practical applications.

#### 3.3.3. LSTM Model Construction

In this study, LSTM and Dropout were chosen for the hidden layer to build two layers to prevent overfitting.

As shown in Figure 4, the model, in terms of parameter settings, includes one input layer, two hidden layers, and one output layer. The default sigmoid activation function serves as the activation function, and the LSTM uses 7 as its Timesteps. The model selects Mean Absolute Error (MAE) as its loss function and adopts the Adam optimization algorithm for network training. The initial learning rate is set to 0.05, with a gradient threshold set to 1. The output layer employs a fully connected layer to reduce the dimensionality of the results. Upon obtaining the prediction data, the model performs inverse normalization, thereby obtaining the final prediction results.



Figure 4. LSTM model construction.

## 3.3.4. XGBoost Model Construction

The XGBoost model was built starting with the tuning of the tree parameters. The parameters are initialized based on the default values. The choice of parameters refers to the Mean Squared Error (MSE) as the loss function and Gamma as the objective function. When using the XGBoost model for temporal prediction, it is necessary to consider the following algorithmic parameters:

Learning\_rate: The learning rate boosts the model's robustness by reducing the weights at each gradient descent step, and the value typically ranges from 0.01 to 0.2. If the value is too low, it might cause underfitting in the model.

Gamma: A node only splits if the value of the loss function decreases post-split. Gamma determines the minimum decrease in the loss function required to split the node. The larger this parameter, the more conservative the algorithm will be, as a larger gamma value necessitates a more substantial decrease in the loss function before the node can split, reducing the likelihood of node splitting during tree generation.

Subsample: This parameter determines the proportion of random samples for each tree. By lowering this value, the algorithm will be more conservative and prevent overfitting. However, setting this value too low can lead to underfitting. It generally ranges between 0.5 and 1, with 0.5 representing an average sampling.

Colsample\_bytree: This parameter is utilized to control the percentage of columns sampled randomly for each tree (each column corresponds to a feature).

Max\_depth: This is the maximum depth of the tree, typically set between 3 and 10. A larger value allows the model to quickly identify the features of local samples, but it also increases the likelihood of overfitting and slows down the model's training speed.

#### 3.3.5. Model Tuning Method

In the model tuning stage, this study employed ten-fold cross-validation combined with a grid search approach. Firstly, the cross-validation is used to assess the model's performance, followed by a grid search to select the optimal parameters. The ten-fold cross-validation initially splits the edible oil data set into ten non-overlapping segments. Nine of these are used as training segments and one as a testing segment to enhance the model's performance by reducing the variance in data partitioning.

Grid search, a commonly used method for parameter tuning, applies an exhaustive search method. After a set of hyperparameters is provided, an exhaustive search is carried out among all the hyperparameter combinations, aiming to select the optimal set from all combinations.

Given the differences in data across provinces, each province's indicators were tuned separately during the tuning session. Following this, the experiment was conducted; the tuning results are depicted in Figure 5.

		colsample_bytree	learning_rate	max_depth	min_child_weight	n_estimators	reg_alpha	reg_lambda	subsample
Beijing	ILCR	0.6	0.05	5	3	300	0	0.5	0.8
	MOE	0.6	0.1	5	3	300	0	0.5	0.8
	NIPI	0.6	0.1	5	5	300	0	1	0.7
Fujian	ILCR	0.6	0.05	3	3	100	0.1	2	0.6
	MOE	0.6	0.05	3	5	200	0.2	1	0.7
	NIPI	0.6	0.05	3	5	200	0.2	1	0.7
Guangdong	ILCR	0.6	0.05	3	3	200	0	1	0.7
	MOE	0.6	0.05	3	5	200	0	1	0.7
	NIPI	0.6	0.05	5	3	200	0.2	0.5	0.8
Guangxi	ILCR	0.6	0.1	5	3	200	0	0.5	0.7
	MOE	0.6	0.05	3	3	200	0.2	1	0.7
	NIPI	0.6	0.05	5	5	200	0.2	0.5	0.8

Figure 5. XGBoost model tuning results.

3.3.6. The Inverse Error Method

The prediction results of each single model are obtained by the LSTM model and XGBoost Model, and for the analysis results, the following formula is applied for the inverse error method analysis to process the LSTM and XGBoost time series data.

$$f_t = \omega_1 f_{1t} + \omega_2 f_{2t}, t = 1, 2, \dots, n \tag{15}$$

$$\omega_1 = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} \tag{16}$$

$$\omega_2 = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2} \tag{17}$$

where  $\omega_i$  denotes the weight coefficient,  $f_{it}$  denotes the prediction data of LSTM and XGBoost, and  $\varepsilon_1$  and  $\varepsilon_2$  refer to the LSTM and XGBoost errors, respectively.

#### 3.3.7. The Variable-Weight Combined LSTM-XGBoost Prediction Model

Considering the substantial differences in principles between the LSTM model, which is a neural network model, and the XGBoost model, which is a tree model, and the relatively weak correlation between their prediction results, this study proposed integrating these two models using the inverse error method to improve the overall prediction accuracy. The primary process is as follows, and the corresponding flowchart is depicted in Figure 6.

- (1) The pre-processed data are input into the LSTM and XGBoost models for predictive analysis, resulting in the prediction outcomes of each individual model.
- (2) The prediction results of the obtained LSTM and XGBoost models are weighted and combined using the inverse error method to obtain the final prediction results of the combined LSTM-XGBoost model.
- (3) The evaluation metrics RMSE and MAE are utilized to compare each individual model and the combined model.



Figure 6. LSTM-XGBoost-based variable-weight combined prediction model.

# 4. Results and Discussion

# 4.1. Data Set Processing

The data set used in this study includes three evaluation indicators for edible oils (ILCR, MOE, and NIPI) for each prefecture-level city in 20 Chinese provinces in 2019. This data set contains a total of 12,826 records, and the total length of the time series for each city is 53 weeks. The data were divided in a 6:4 ratio for subsequent analysis or processing.

#### 4.2. Experimental Environment

The computer configurations used for the experiments in this paper are shown in Table 1.

	Operating System	Windows 10 64-bit		
Computer Information	CPU	AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz		
	GPU	Nvidia GeForce GTX1650		
_	Memory	16 GB		
		Numpy 1.18.5		
		Pandas 1.2.2		
Toolkit	Python 3.7.11	Keras 2.9.0		
		Torch 1.8.3		
		Matplotlib 3.5.3		

Table 1. Experimental platform and environmental parameters.

## 4.3. Model Evaluation Metrics

In order to scientifically measure the prediction effectiveness of this combined model, the five evaluation metrics were used to evaluate the model: *RMSE*, *MAE*, precision, recall, and *F*1 score [49,50].

*MAE* stands for Mean Absolute Error. It calculates the average absolute difference between the true and predicted values, preventing the errors from being cancelled out by positive or negative discrepancies. Generally, the lower the *MAE* value, the better the prediction ability of the model.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$
(18)

*RMSE* is the abbreviation for Root Mean Square Error. This metric computes the square root of the averaged squared differences between predicted and actual observations. It is particularly sensitive to outliers and serves as a robust measure of the predictive capability of the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$
(19)

where  $y_i$  denotes the actual value of a single assessment indicator for week *i*;  $\hat{y}_i$  denotes the predicted value of a single assessment indicator for week *i*; and *n* denotes the total number of data points to be measured.

Precision refers to the proportion of samples with a predicted value of 1 and a true value of 1 among all samples with a predicted value of 1.

$$P = \frac{TP}{TP + FP} \tag{20}$$

Recall, also known as the full rate, refers to the proportion of samples with a predicted value of 1 and a true value of 1 out of all samples with a true value of 1.

$$R = \frac{TP}{TP + FN} \tag{21}$$

where *TP* indicates the number of risk levels that the model correctly predicted and *FP* indicates the number of levels that the model predicted that are incorrectly predicted as being at that risk level. *FN* indicates the number of levels of that risk level that the model incorrectly predicted as other risk levels.

The *F*1 score, also known as the Balanced Score, is defined as the summed average of the precision and recall. To better evaluate the performance of the prediction model, this study used the *F*1 score as an evaluation criterion to measure the comprehensive performance of the model.

$$F1 = \frac{2 \times P \times R}{P + R} \tag{22}$$

# 4.4. Edible Oil Safety Risk Classification and Assessment 4.4.1. Risk Classification

In this study, the k-means algorithm was used to cluster and grade the three-evaluation metrics (*ILCR*, *MOE* and *NIPI*) in the assessment model. The elbow method was used to determine the value of k among them, and the core index of the elbow method is Sum of the Squared Errors (*SSE*).

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2$$
 (23)

where  $C_i$  is the *i*th cluster, *p* is the number of sample points in  $C_i$ ,  $m_i$  is the center of mass of  $C_i$  (the mean of all samples in  $C_i$ ), and *SSE* is the clustering error of all samples, which represents the performance of the clustering.

The core idea of the elbow method is that as the number of k clusters increases, the sample division becomes finer, and the degree of aggregation of each cluster gradually increases, causing the *SSE* to decrease progressively. When k is less than the true number of clusters, the decrease in *SSE* is significant as the increase in k substantially enhances the degree of aggregation of each cluster. However, when k reaches the true number of clusters, the returns on the degree of aggregation obtained by further increasing k rapidly diminish, leading to a sharp decrease in *SSE*, which then levels off as k continues to increase. Therefore, the graph of the relationship between *SSE* and k forms an 'elbow' shape, and the value of k corresponding to this elbow is considered to be the true number of clusters in the data.

By observing Figure 7, we can see that the elbow corresponds to a k value of 3 (maximum curvature), and the *SSE* decreases quite significantly from 1 to 2 and from 2 to 3, while the *SSE* decreases very little from 3 to 4 and even after 4; therefore, the optimal k value should be 3. The results for each cluster center are shown in Table 2, and the distance of the cluster center from the origin is calculated based on the specific normalized index. Then, the risk level is defined as low, medium, or high based on the distance.

Table 2. Clustering centers and ranking of the 3 clusters.

Category	ILCR	MOE	NIPI	Sample Size	Risk Level
1	0.026709	0.021994	0.047369	9177	Low
2	0.911746	0.001170	0.074586	3095	Medium
3	0.483147	0.000756	0.089037	554	High



Figure 7. Elbow method to determine k value.

## 4.4.2. Analysis of Grading Results

The distribution of indicators for the different security risk levels can be analyzed in Figures 8–10.



Figure 8. Probability density distribution of subgroup 1.



Figure 9. Probability density distribution of subgroup 2.



Figure 10. Probability density distribution of subgroup 3.

Based on the above analysis, the following conclusions can be drawn.

- (1) The ILCR values of subcluster 1 were distributed between 0 and 0.1, the MOE was less spaced and distributed between 0 and 0.001, and the NIPI values were concentrated between 0 and 0.1, indicating a low risk level;
- (2) The ILCR values of subgroup 2 were distributed between 0 and 0.4, the MOE was distributed between 0 and 0.002, and the NIPI values were less spaced and concentrated between 0 and 0.2, indicating a medium risk level;
- (3) Subgroup 3 had the highest ILCR values distributed between 0 and 0.9, the MOE values were distributed between 0 and 0.005, and the NIPI values had smaller intervals and were concentrated between 0 and 0.4, indicating the highest risk level.

The results of the above analysis show that, for the three indicators of the food safety risk assessment model, the k-means clustering algorithm can cluster the edible oil safety risks in each prefecture-level city at different time periods, and the samples could be divided into three groups, namely, subgroup 1 with a low risk level, subgroup 2 with an intermediate risk level, and subgroup 3 with the highest risk level. The *ILCR* values of subgroup 1 were small, the interval of MOE values was small, and the distribution of *NIPI* values was concentrated and small; the *ILCR* values of subgroup 2 were higher, the interval of MOE values was larger, and the *NIPI* values were at an intermediate level; the *ILCR* values of subgroup 3 were the largest, the interval of MOE values was also relatively the largest, and the *NIPI* values were at a high level. These results can provide targeted management measures and risk control programs for food safety regulatory authorities.

#### 4.4.3. Predicted Results of BaP Safety Risk Level in Edible Oil

In order to scientifically evaluate the prediction performance of the variable-weight combined LSTM-XGBoost prediction model in this paper, a comparative analysis was conducted. Considering that the predictive effect of this combined model will rely heavily on the predictive power of each single model, we conducted a comparative analysis of the LSTM model, the XGBoost model, and the variable-weight combined LSTM-XGBoost prediction model, and evaluated the prediction performance of the three models by comparing the errors between their prediction results and the actual values, and whether the variable-weight combined LSTM-XGBoost prediction model is superior to the individual models. We used a single-step prediction method with a step size of seven for the three-evaluation metrics of edible oils mentioned previously, and performed a preliminary analysis of the prediction results using RMSE and MAE.

Figures 11 and 12 show the RMSE and MAE values of the food safety risk assessment indicators predicted by the three models. The result plots show that the evaluation indicators predicted by the combined model proposed in this paper had the smallest RMSE and MAE values. The RMSE measures the mean error between the model's predicted and true values, while the MAE measures the mean absolute error between the model's predicted and true values. Smaller RMSE and MAE values imply that the model has a higher prediction accuracy and is better able to adapt to changes in the test data set. We observed that the variable-weight combined LSTM-XGBoost prediction model had the smallest RMSE and MAE values compared to the LSTM and XGBoost models alone, indicating that the combined model can significantly improve the prediction accuracy.



Figure 11. RMSE for ILCR, MOE, and NIPI indicators.





After the models predicted the weekly ILCR, MOE, and NIPI indicators for different prefecture-level cities in each province, the distance between this rating indicator and the three clustering centers was measured, and the risk level rating for that week in that city was determined, and the precision (P%), recall (R%), and F1 scores (F1%) of the risk rating predicted by the three models were tallied, as shown in Table 3.

Table 3. Experimental results of risk level prediction.

Model	P%	<b>R%</b>	F1%
LSTM	81.23%	78.66%	79.92%
XGBoost	80.08%	82.42%	82.19%
LSTM-XGBoost	94.62%	95.71%	95.16%

The experimental results show that the variable-weight combined LSTM-XGBoost prediction model proposed in this paper outperforms the other two models in terms of accuracy, and this model can provide a new approach to aid the government in regulating risky edible oils. In addition, the F1 value is significantly better than that of a single model. The F1 value shows that this model is able to balance the accuracy rate and recall rate, so the government can better capture potential food safety problems based on the model's prediction results, target and strengthen the regulation of specific products, specific supply chain links or specific regions, and optimize resource allocation, thus improving the overall food safety level.

#### 5. Conclusions

BaP is one of the most representative carcinogens among the more than 20 known carcinogenic PAHs. Fats and oils containing PAHs can intensify absorption in the intestinal tract, thus posing a great threat to human health. To reduce dietary intake of BaP, we should maintain a balanced and diverse diet that includes a variety of fruits and vegetables, avoid excessive intake of grilled meats, especially charcoal-grilled and smoked meats, and remove burnt parts of foods. Whenever possible, choose fats and oils rich in monounsaturated fatty acids (e.g., canola and olive oils) and polyunsaturated fatty acids (e.g., corn and soybean oils).

In order to thoroughly assess the safety risk of BaP in edible oils in China and to carry out precise regulation to effectively protect the health and safety of residents' food, we introduced an innovative prediction model, namely, the variable-weight combined

LSTM-XGBoost prediction model. This model combines two leading algorithms, LSTM and XGBoost, combining their respective strengths to improve the prediction accuracy; the LSTM algorithm is a modeling approach that is well suited for handling serial data, which can effectively capture long-term dependencies in time series; the XGBoost algorithm is effective at handling nonlinear relationships and high-dimensional data. Therefore, the variable-weight combined LSTM-XGBoost prediction model has the ability to better handle data with time-series and high-dimensional features, which helps to improve the accuracy of prediction. The advantages of this combined model are further enhanced by the adoption of the inverse error method, which adjusts the combined weights of the model by optimizing the inverse of the prediction error so that the prediction results of both algorithms can be optimally combined to further improve the prediction accuracy.

Experimentally, by comparing the variable-weight combined LSTM-XGBoost prediction model with the LSTM and XGBoost models alone, we found that the former performed better in terms of prediction accuracy, as evidenced by the lowest values of two crucial metrics, RMSE and MAE. Thus, the variable-weight combined LSTM-XGBoost prediction model is undoubtedly an efficient and effective way to combine algorithms to provide more accurate prediction results for data with time-series and high-dimensional characteristics. More importantly, this model also shows strong utility in food safety risk assessment, and the experimental results show that its F1 score was as high as 95.16%, which is a good balance of accuracy and recall. This means that the model is able to meet the high demand of food regulatory authorities to monitor the safety of edible oils in different prefecture-level cities in each province and strengthen early warnings and control of food safety, which also provides guidance in optimization of the allocation of resources, thereby more effectively preventing the occurrence of food safety risks.

**Author Contributions:** Conceptualization, Q.Z. and W.D.; methodology, Q.Z. and W.D.; software, C.H.; validation, W.D. and S.W.; formal analysis, C.H. and T.J.; investigation, Q.Z. and C.H.; resources, C.H.; data curation, C.H.; writing—original draft preparation, C.H.; writing—review and editing, C.H. and W.D.; visualization, C.H.; supervision, Q.Z.; project administration, T.J. and Q.Z.; funding acquisition, T.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Key Technology R&D Program of China (grant No. 2019YFC1606401).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data were available with the permission of the State Administration for Market Regulation Statistics.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

- 1. Meng, G.; Xu, Z.; Zhan, X.; Zhou, J. Development strategy and analysis of production and consumption demand of plant oilseeds and oils in China. *China Oils Fats* **2020**, *45*, 1–4, 27. [CrossRef]
- Bogdanović, T.; Pleadin, J.; Petričević, S.; Listeš, E.; Sokolić, D.; Marković, K.; Ozogul, F.; Šimat, V. The occurrence of polycyclic aromatic hydrocarbons in fish and meat products of Croatia and dietary exposure. *J. Food Compos. Anal.* 2019, 75, 49–60. [CrossRef]
   Orecchio S.: Amorello D.: Indelicato R.: Barreca S.: Orecchio S. A Short Review of Simple Analytical Methods for the Evaluation
- 3. Orecchio, S.; Amorello, D.; Indelicato, R.; Barreca, S.; Orecchio, S. A Short Review of Simple Analytical Methods for the Evaluation of PAHs and PAEs as Indoor Pollutants in House Dust Samples. *Atmosphere* **2022**, *13*, 1799. [CrossRef]
- Orecchio, S.; Bianchini, F.; Bonsignore, R.; Blandino, P.; Barreca, S.; Amorello, D. Profiles and Sources of PAHs in Sediments from an Open-Pit Mining Area in the Peruvian Andes. *Polycycl. Aromat. Compd.* 2016, 36, 429–451. [CrossRef]
- Barreca, S.; Bastone, S.; Caponetti, E.; Martino, D.F.C.; Orecchio, S. Determination of selected polyaromatic hydrocarbons by gas chromatography–mass spectrometry for the analysis of wood to establish the cause of sinking of an old vessel (Scauri wreck) by fire. *Microchem. J.* 2014, 117, 116–121. [CrossRef]
- 6. Barreca, S.; Mazzola, A.; Orecchio, S.; Tuzzolino, N. Polychlorinated Biphenyls in Sediments from Sicilian Coastal Area (Scoglitti) using Automated Soxhlet, GC-MS, and Principal Component Analysis. *Polycycl. Aromat. Compd.* **2014**, *34*, 237–262. [CrossRef]
- Shi, L.-K.; Zhang, D.-D.; Liu, Y.-L. Incidence and survey of polycyclic aromatic hydrocarbons in edible vegetable oils in China. *Food Control* 2016, 62, 165–170. [CrossRef]
- Ji, J.; Jiang, M.; Zhang, Y.; Hou, J.; Sun, S. Polycyclic Aromatic Hydrocarbons Contamination in Edible Oils: A Review. *Food Rev. Int.* 2022, 1–27. [CrossRef]

- 9. Yao, Z.; Li, J.; Wu, B.; Hao, X.; Yin, Y.; Jiang, X. Characteristics of PAHs from deep-frying and frying cooking fumes. *Environ. Sci. Pollut. Res.* **2015**, *22*, 16110–16120. [CrossRef]
- 10. Guerreiro, C.B.B.; Horálek, J.; de Leeuw, F.; Couvidat, F. Benzo(a)pyrene in Europe: Ambient air concentrations, population exposure and health effects. *Environ. Pollut.* **2016**, 214, 657–667. [CrossRef]
- Ge, Y.; Yan, H.; Shi, X.; Wu, Z.; Wang, Y.; Zhang, Z.; Luo, Q.; Liu, W.; Liang, L.; Peng, L.; et al. Study on dietary intake, risk assessment, and molecular toxicity mechanism of benzo[α]pyrene in college students in China Bashu area. *Food Sci. Nutr.* 2022, 10, 4155–4167. [CrossRef] [PubMed]
- 12. Bukowska, B.; Mokra, K.; Michałowicz, J. Benzo[*a*]pyrene—Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. *Int. J. Mol. Sci.* 2022, *23*, 6348. [CrossRef]
- 13. Hou, L.; Qiu, J. Review on Polycyclic Aromatic Hydrocarbons (PAHS) in Edible Oils. J. Henan Univ. Technol. 2017, 38, 115–122. [CrossRef]
- Huang, F.; Zhang, L.; Zhou, M.; Li, J.; Liu, Q.; Wang, B.; Deng, K.; Zhou, P.; Wu, Y. Polycyclic aromatic hydrocarbons in the Chinese diet: Contamination characteristics, indicator screening, and health risk assessment. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Asses.* 2023, 40, 625–640. [CrossRef] [PubMed]
- 15. Jiang, D.; Xin, C.; Li, W.; Chen, J.; Li, F.; Chu, Z.; Xiao, P.; Shao, L. Quantitative analysis and health risk assessment of polycyclic aromatic hydrocarbons in edible vegetable oils marketed in Shandong of China. *Food Chem. Toxicol.* **2015**, *83*, 61–67. [CrossRef]
- 16. Jang, M.-R.; Hong, M.-S.; Jung, S.-Y.; Choi, B.-C.; Lee, K.-A.; Kum, J.-Y.; Kim, I.-Y.; Kim, J.-H.; Chae, Y.-Z. Analysis and Risk Assessment of Benzo(a)pyrene in Edible Oils. *J. Food Hyg. Saf.* 2014, 29, 141–145. [CrossRef]
- 17. Lee, J.-G.; Suh, J.-H.; Yoon, H.-J. Occurrence and risk characterization of polycyclic aromatic hydrocarbons of edible oils by the Margin of Exposure (MOE) approach. *Appl. Biol. Chem.* **2019**, *62*, 51. [CrossRef]
- 18. Li, G.; Wu, S.; Wang, L.; Akoh, C.C. Concentration, dietary exposure and health risk estimation of polycyclic aromatic hydrocarbons (PAHs) in youtiao, a Chinese traditional fried food. *Food Control* **2016**, *59*, 328–336. [CrossRef]
- Barzegar, G.; Rezaei Kalantary, R.; Bashiry, M.; Jaafarzadeh, N.; Ghanbari, F.; Shakerinejad, G.; Khatebasreh, M.; Sabaghan, M. Measurement of polycyclic aromatic hydrocarbons in edible oils and potential health risk to consumers using Monte Carlo simulation, southwest Iran. *Environ. Sci. Pollut. Res.* 2023, *30*, 5126–5136. [CrossRef]
- 20. Kang, B.; Lee, B.-M.; Shin, H.-S. Determination of Polycyclic Aromatic Hydrocarbon (PAH) Content and Risk Assessment from Edible Oils in Korea. *J. Toxicol. Environ. Health Part A* **2014**, *77*, 1359–1371. [CrossRef]
- Yu, S.; Tian, L.; Liu, Y.; Guo, Y. LSTM-XGBoost Application of the Model to the Prediction of Stock Price. In *Artificial Intelligence and Security*; Lecture Notes in Computer Science; Sun, X., Zhang, X., Xia, Z., Bertino, E., Eds.; Springer International Publishing: Cham, Switzerland, 2021; Volume 12736, pp. 86–98. ISBN 978-3-030-78608-3.
- 22. Liwei, T.; Li, F.; Yu, S.; Yuankai, G. Forecast of LSTM-XGBoost in Stock Price Based on Bayesian Optimization. *Intell. Autom. Soft Comput.* 2021, 29, 855–868. [CrossRef]
- 23. Ding, G.; Qin, L. Study on the prediction of stock price based on the associated network model of LSTM. *Int. J. Mach. Learn. Cyber.* **2020**, *11*, 1307–1317. [CrossRef]
- Zhang, X.; Zhang, Q. Short-Term Traffic Flow Prediction Based on LSTM-XGBoost CombinationModel. *Comput. Model. Eng. Sci.* 2020, 125, 95–109. [CrossRef]
- 25. Ma, Y.; Zhang, Z.; Ihler, A. Multi-Lane Short-Term Traffic Forecasting with Convolutional LSTM Network. *IEEE Access* 2020, *8*, 34629–34643. [CrossRef]
- Wang, K.; Ma, C.; Qiao, Y.; Lu, X.; Hao, W.; Dong, S. A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction. *Phys. A Stat. Mech. Appl.* 2021, 583, 126293. [CrossRef]
- Qin, D.; Yu, J.; Zou, G.; Yong, R.; Zhao, Q.; Zhang, B. A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM <sub>2.5</sub> Concentration. *IEEE Access* 2019, 7, 20050–20059. [CrossRef]
- Xayasouk, T.; Lee, H.; Lee, G. Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models. Sustainability 2020, 12, 2570. [CrossRef]
- 29. Wu, X.; Liu, Z.; Yin, L.; Zheng, W.; Song, L.; Tian, J.; Yang, B.; Liu, S. A Haze Prediction Model in Chengdu Based on LSTM. *Atmosphere* **2021**, *12*, 1479. [CrossRef]
- Jiang, T.; Liu, T.; Dong, W.; Liu, Y.; Hao, C.; Zhang, Q. Prediction of Safety Risk Levels of Veterinary Drug Residues in Freshwater Products in China Based on Transformer. *Foods* 2022, 11, 1690. [CrossRef]
- 31. Jiang, T.; Liu, T.; Dong, W.; Liu, Y.; Zhang, Q. Security Risk Level Prediction of Carbofuran Pesticide Residues in Chinese Vegetables Based on Deep Learning. *Foods* **2022**, *11*, 1061. [CrossRef]
- Wang, Z.; Wu, Z.; Zou, M.; Wen, X.; Wang, Z.; Li, Y.; Zhang, Q. A Voting-Based Ensemble Deep Learning Method Focused on Multi-Step Prediction of Food Safety Risk Levels: Applications in Hazard Analysis of Heavy Metals in Grain Processing Products. *Foods* 2022, 11, 823. [CrossRef] [PubMed]
- GB 2762-2017; Food Safety National Standard—Contaminant Limits in Food. National Standard of the People's Republic of China: Beijing, China, 2017. Available online: https://www.chinesestandard.net/PDF.aspx/GB2762-2017 (accessed on 20 May 2023)(PDF in English).
- GEMS/Food-EURO Second Workshop on Reliable Evaluation of Low-Level Contamination of Food Report on a Workshop in the Frame of GEMS/Food-EURO Kulmbach. 1999. Available online: https://www.semanticscholar.org/paper/GEMS-Food-EURO-Second-Workshop-on-Reliable-of-of-on/7d5162794a407ce3361458649750a63b6bda3381 (accessed on 16 February 2023).

- 35. Zhang, Y.; Kuang, F.; Liu, C.; Ma, K.; Liu, T.; Zhao, M.; Lv, G.; Huang, H. Contamination and Health Risk Assessment of Multiple Mycotoxins in Edible and Medicinal Plants. *Toxins* 2023, *15*, 209. [CrossRef] [PubMed]
- Niu, B.; Zhang, H.; Zhou, G.; Zhang, S.; Yang, Y.; Deng, X.; Chen, Q. Safety risk assessment and early warning of chemical contamination in vegetable oil. *Food Control* 2021, 125, 107970. [CrossRef]
- Taghizadeh, S.F.; Rezaee, R.; Boskabady, M.; Mashayekhi Sardoo, H.; Karimi, G. Exploring the carcinogenic and non-carcinogenic risk of chemicals present in vegetable oils. *Int. J. Environ. Anal. Chem.* 2022, 102, 5756–5784. [CrossRef]
- Ma, J.-K.; Li, K.; Li, X.; Elbadry, S.; Raslan, A.A.; Li, Y.; Mulla, Z.S.; Tahoun, A.B.M.B.; El-Ghareeb, W.R.; Huang, X.-C. Levels of polycyclic aromatic hydrocarbons in edible and fried vegetable oil: A health risk assessment study. *Environ. Sci. Pollut. Res.* 2021, 28, 59784–59791. [CrossRef]
- Risk Assessment Guidance for Superfund (RAGS): Part A. Available online: https://www.epa.gov/risk/risk-assessment-guidance-superfund-rags-part (accessed on 25 May 2023).
- Lu, F.; Shen, B.; Li, S.; Liu, L.; Zhao, P.; Si, M. Exposure characteristics and risk assessment of VOCs from Chinese residential cooking. *J. Environ. Manag.* 2021, 289, 112535. [CrossRef] [PubMed]
- Benford, D.; DiNovi, M.; Setzer, R.W. Application of the margin-of-exposure (MoE) approach to substances in food that are genotoxic and carcinogenic e.g.: Benzo[a]pyrene and polycyclic aromatic hydrocarbons. *Food Chem. Toxicol.* 2010, 48, S42–S48. [CrossRef]
- 42. Liu, Y. Report on the state of nutrition and chronic diseases in China (2020). Food Nutr. China 2020, 26, 2.
- 43. Opinion of the Scientific Committee on a request from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic. *EFSA J.* **2005**, *282*, 1–31. [CrossRef]
- 44. Sawut, R.; Kasim, N.; Maihemuti, B.; Hu, L.; Abliz, A.; Abdujappar, A.; Kurban, M. Pollution characteristics and health risk assessment of heavy metals in the vegetable bases of northwest China. *Sci. Total Environ.* **2018**, *642*, 864–878. [CrossRef]
- Frame, J.M.; Kratzert, F.; Raney, A.; Rahman, M.; Salas, F.R.; Nearing, G.S. Post-Processing the National Water Model with Long Short-Term Memory Networks for Streamflow Predictions and Model Diagnostics. J. Am. Water Resour. Assoc. 2021, 57, 885–905. [CrossRef]
- 46. Li, G.; Zhao, X.; Fan, C.; Fang, X.; Li, F.; Wu, Y. Assessment of long short-term memory and its modifications for enhanced short-term building energy predictions. *J. Build. Eng.* **2021**, *43*, 103182. [CrossRef]
- 47. Feng, C.; Chen, Z. Application of Weighted Combination Model Based on XGBoost and LSTM in Sales Forecasting. Computer. *Syst. Appl.* **2019**, *28*, 226–232. [CrossRef]
- 48. Yan, Z.; Chen, H.; Dong, X.; Zhou, K.; Xu, Z. Research on prediction of multi-class theft crimes by an optimized decomposition and fusion method based on XGBoost. *Expert Syst. Appl.* **2022**, 207, 117943. [CrossRef]
- Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 2014, 7, 1247–1250. [CrossRef]
- 50. Nasiri, A.; Omid, M.; Taheri-Garavand, A.; Jafari, A. Deep learning-based precision agriculture through weed recognition in sugar beet fields. *Sustain. Comput. Inform. Syst.* 2022, 35, 100759. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.