Monitoring and Evaluation of Human Nutrition and Health Based on Big Data Analysis

Sujuan Jia[1]

Abstract

The balanced consumption of nutrients is essential to human health. This research examines human nutrition and health monitoring and evaluation using extensive data analysis (HHH). Few attempts have been made thus far to harvest and evaluate HHH monitoring data. To ensure the efficacy of these data, extensive data analysis was conducted to analyze and monitor HHH. The authors created a Bayesian network (BN) to monitor HHH after describing the HHH optimization method. The suggested network serves four primary purposes: extracting features from dietary nutrients, classifying foods according to their best nutritional value, diagnosing undernutrition, and offering replacement suggestions. In conclusion, a strategy for massive data processing and selecting the primary controlling element was provided for nutrition monitoring. Experiments demonstrated the relative test error of the proposed network and confirmed its efficacy.

Keywords: human nutrition monitoring; human health evaluation; extensive data analysis

1. Introduction

The balanced consumption of essential nutrients is crucial, particularly for children, adolescents, and the elderly (Bommarito et al., 2019; Lu et al., 2020; Silva et al., 2019; Sivakumaran, Huffman, & Sivakumaran, 2018; Welis, 2017). Lack of essential nutrients will result in severe diseases and organ deterioration, leading to significant health issues. It is crucial to automatically monitor the nutrient content of the foods delivered to people to preserve their physical health (Comber et al., 2012; Hoefkens et al., 2010; Jones, 2010; Levenhagen et al., 2001; Nicklas et al., 2012; Ohara et al., 2008; Tordoff, 2002; Velasco-Ryenold et al., 2008). The quality and amount of personal monitoring data have been enhanced for quite some time. The vast digital data reflects individual and social characteristics (Gupta, Chakraborty, & Gupta, 2019; Li, Lin, & Xu, 2019; Rahaman et al., 2019; Yan et al., 2019). Regarding human nutrition and health (HHH) monitoring, big data and data mining have increased the sensing range and execution effectiveness of healthy diet solutions.

Ng and Jin (2017) built a personalized recipe suggestion system for infants that pushes dishes to infants based on the original data acquired from user archives and MyPlate. Not only do the words offer newborns the required nutrients from various food categories, but they also help them develop healthy eating habits. Tucker (2016) suggested that, as people age, nutritional status is a key factor in the decline of cognitive function. He also provided evidence for the essential roles of vitamin B. He demonstrated that a healthy diet model (i.e., maximizing the intake of fish, fruits, vegetables, nuts, and seeds while limiting the intake of sugar) could significantly slow down and mitigate the decline of cognitive function. Villalobos et al. (2011) developed an intelligent method to track obese individuals' caloric consumption. The system analyzes the surveillance images to identify food classes and components and then calculates calorie counts based on the detection results. Hoefkens et al. (2009) viewed dietary intake evaluation as a two-step process, including collecting and evaluating data on food components. They combined these data with food consumption data to create a database about the nutrients and pollutants in organic and conventional vegetables and potatoes. Shimbo et al. (1996) evaluated the daily dietary intake of 28 trace elements in food consumption data, compared the estimates with the reported values of 15 elements, and then drew the appropriate conclusions. Rodríguez-Palmero et al. (1998) evaluated the components in 43 food samples from a nursing home and estimated the calorie, fat, polyunsaturated fatty acid (PUFA), cholesterol, potassium, and phosphorus intakes with precision.

Currently, local research on healthy diets and food nutrition analysis is comparatively developed. Few, however, have mined or studied the HHH monitoring data. To ensure the efficacy of these data, this research monitors and evaluates HHH using extensive data analysis.

Our research offers a framework for exchanging and sharing HHH monitoring data. This work concludes the construction of an application platform for HHH monitoring services using extensive data analysis. Using the Internet as a medium, the suggested platform can offer a personalized healthy diet to exercise enthusiasts and allow administrators to query for and statistically evaluate the health test results from various locations.

[1] School of Film and Television, Hebei University of Science and Technology, Shijiazhuang, China. Email: imeyou@163.com

Section 2 describes the HHH optimization approach in full. In Section 3, a Bayesian network (BN) is constructed to monitor HHH, which can extract characteristics from food nutrients, classify foods by their highest nutritional value, diagnose undernutrition, and provide replacement recommendations. Section 4 presents a system for massive data processing and a method for selecting the primary controlling factor for human nutrition monitoring. Experiments validated the efficiency of the proposed approach.

2. Methodology

2.1 Optimization of HHH

Nutritional sufficiency results from adhering to the fundamental criteria of a healthy diet. According to sources such as the *European Journal of Epidemiology*, an unregulated diet is the primary cause of most human disorders that diminish life expectancy. Monitoring the calorie input and output of daily foods, a crucial aspect of health care, addresses the nutrition imbalance. This research focuses on monitoring and evaluating HHH using extensive data analysis.

Based on the nutrients of various foods in Table 1, which contains more than 20 nutrients, as well as the nutrients and calories of over 580 typical foods, the general situation of food nutrients was reviewed. Here, the common foods are separated into 12 categories: staple food FC1, meat and eggs FC2; soybeans and their products FC3; vegetables, algae, and fungi FC4; fruits FC5, milk FC6, lipid FC7, nuts FC8, condiments FC9, beverages FC10, snacks and cold drinks FC11, and others FC12. Table 1 also includes the amount of proteins, lipids, carbs, and cholesterol in 100g of each food and the glycemic index (GI) and glycemic load (GL) for each food. The suitable amount of each food should be defined based on food class and characteristics so that large data sets of food nutrients may be applied to dietary health evaluation and recommendations for the general population.

Table 1

Nutrients of some foods

Food	Calorie	Proteins	Carbohydrates	Fats	Cholesterol	Iron	Calcium	…
Egg (100g)	139	13.1	2.4	8.6	0.648	1.6	0.056	…
Pork (100g)	143	20.3	1.5	6.2	0.081	0.003	0.006	…
Bread (100g)	313	8.3	58.6	5.1	/	0.002	0.049	…
Spinach (100g)	28	2.6	4.5	0.3	/	0.0029	0.066	…
Soybean sprouts (100g)	47	4.5	4.5	1.6	/	0.0009	0.021	…
…	…	…	…	…	…	…	…	…

Table 2

Critical features of food nutrients

Nutrient	Description
PRO_u	Protein content k
LIP_u	Lipid content k
LIP_u	Carbohydrate content k
CHO_u	Cholesterol content k
SOD_u	Sodium content k
SAT_u	Saturated fat content k
…	…

Table 2 characterizes the core food nutrients with protein content k PRO_u, lipid content LIP_u, carbohydrate content LIP_u, cholesterol content CHO_u, sodium content SOD_u, and saturated fat content SAT_u. Then, the key features of type u food can be described as:

$$\Phi_u = (PRO_u, LIP_u, CARB_u, CHO_u, SOD_u, SAT_u, \cdots) \quad (1)$$

The body state, the number of adipose tissues, and the number of muscle tissues of the object were estimated against the standard reference model of daily nutrient intake. The minimum deviation of indices (nutrients and calories) from the standard reference model was defined as the objective function of human nutrient and health evaluation. Let b_j be the mass of food j; r_{lj} be the content of element l in food j; R_l^0 be the daily content of element l required by the human body. Then, we have:

$$\sum_{l=1}^{m}\left(R_l^0 - \sum_{j=1}^{n} r_{lj}\, b_j\right)^2 \to min \quad (2)$$

The total mass U of the daily food intake by the human body can be derived from the mass of each food consumed on each day:

$$U = \sum_{j=1}^{n} b_j \quad (3)$$

where, b_j^{min} and b_j^{max} are the upper and lower limits on the daily allowable intake volume of food j, respectively. Then, we have:

$$b_j^{min} \le b_j \le b_j^{max} \quad (4)$$

Besides food mass and element content, it is important to fully consider whether the digestibility, edible value, and health value of foods meet the dietary and age needs of the human body. This paper estimates the deviations of the object from specific groups of people, i.e., children, adolescents, adult males, adult females, and the elderly, in terms of the daily nutrient intake measured by the standard reference model (Table 3).

Table 3

The standard reference model for daily nutrient intake of the human body (adolescents)

Food class	Food Name
FC1	Rice (80g), noodles (100G), steamed buns (50g)
FC2	Eggs (50g), fish (60g), pork (60g)
FC4	Spinach (100g), lettuce (50g), broccoli (100g), mushrooms (50g)
FC5	Apples (100g), kiwi fruits (100g)
FC6	Yoghurt (250g)
FC8	Walnuts (20g), cashews (20g), almonds (10g)

Let a_{ij}, a^0_{ij} and Δa^0_{ij} be the actual value, reference value, and tolerance of index j of standard reference model i, respectively; $c_{ij}=(a_{ij}-a_{ij}^0)/\Delta a^0_{ij}$ be the relative deviation of index j of standard reference model i; g_i and r_{ij} be the set of significant coefficients and index j of standard reference model i, respectively; c_l be the coefficient of the relative deviation for the content of element l from the overall standard reference model. Then, we have:

$$H = \prod_{l=1}^{n_l}(1-c_l^2)\cdot\left[\sum_{i=1}^{n} g_i\left(1-\sqrt{\frac{1}{m}\sum_{j=1}^{m_j} r_{ij}c_{ij}^2}\right)\right] \quad (5)$$

2.2 BN-based HHH monitoring

Constantly evolving with the times, information technology has a significant impact on the economic and social growth of the world. Globally, a new generation of information technology, including big data and machine learning, has arisen in the Internet framework. Advanced information-based health assessments are gradually replacing the old nutrition and health tests at all grade levels and classrooms. Numerous nations exert great effort to actualize data sharing and exchange, informatize health tests, and promote scientific health management. To implement scientific HHH monitoring, enhance relevant databases, and encourage scientific exercise among the populace, it is essential to leverage big data and improve the following areas of national nutrition and health monitoring: data interchange and sharing and nutrition and health tests. This work offers a BN-based algorithm for HHH monitoring that covers four essential functions: extracting characteristics from food nutrients, classifying foods by their best nutritional value, diagnosing undernutrition, and offering replacement suggestions. Figure 1 demonstrates the flow of the suggested method.

The structure of BN can be formed using either the search for a scoring function or the determination of condition dependency. The latter technique was used for the HHH monitoring application scenario. Let P(CD) represent the likelihood that food class C and food name D occur together. The conditional probability P(C|D) of food classes C and D is the chance that the object requires food class D, given the likelihood that the object requires food class C. The connection between P(CD) and P(C|D) can be expressed as follows:

$$P(C|D) = \frac{P(C\cap D)}{P(D)} \quad if\ P(D)\neq 0 \quad (6)$$

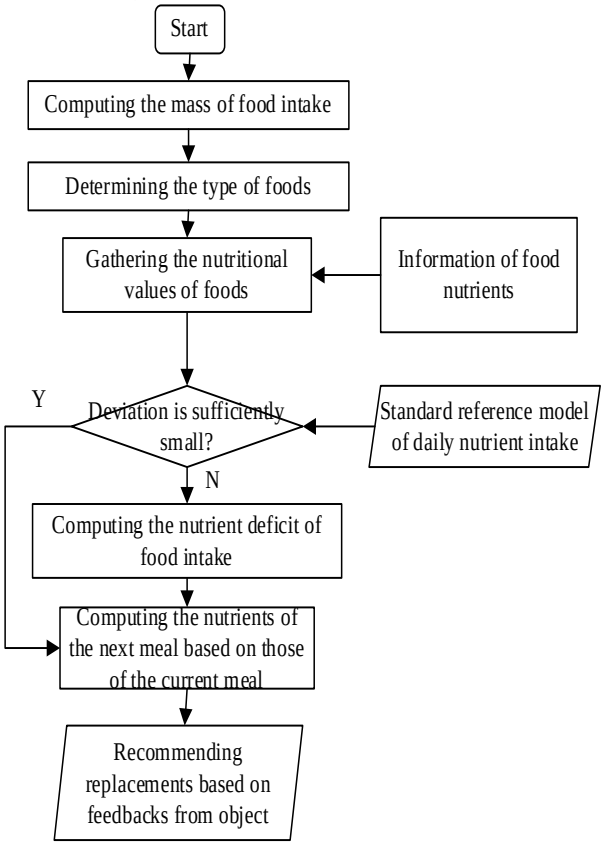

Figure 1. The flow of our HHH monitoring algorithm

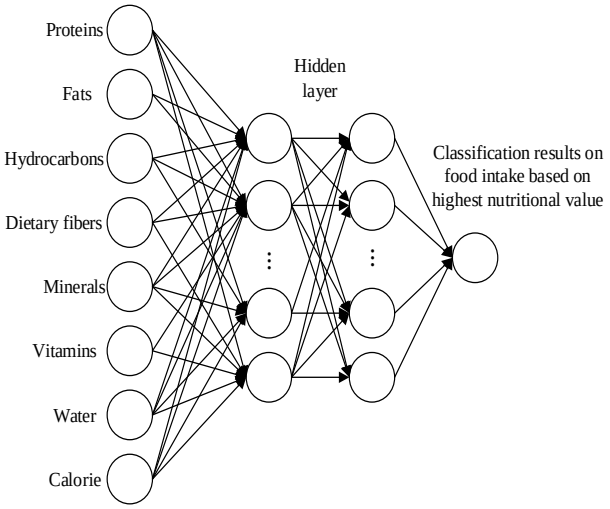

Figure 2. Topology of our neural network

Formula (6) characterizes the joint distribution of the discrete variables and conditional probabilities of a group of food nutrients. Let $P(\omega)$ be the prior probability; $P(\omega|A)$ be the posterior probability, which characterizes the assumed association of each nutrient variable A in the model. Then, $P(\omega|A)$ can be expressed as:

$$P(\omega|A) = \frac{P(A|\omega)}{P(A)}P(\omega) \quad (7)$$

The likelihood ratio is the factor linking the two probabilities $P(\omega)$ and $P(\omega|A)$. The key parameters were determined based on the prior distribution for the BN-based 4-layer sensor neural network. Figure 2 shows the topology of the proposed neural network. Each layer owns interconnected nodes with an activation function in the layered network.

Let A_i be the output of the input layer, Q_{Li} be the connection weight of hidden layer nodes, and E_i is the input of the output layer. The input layer is connected with the output layer via the hidden layer. The relationship between A_i and E_i can be expressed as:

$$E_i = \sum_{i=1}^{m} Q_{Li}A_i \quad (8)$$

The output B_j of the neural network is connected with E_i via activation function AF:

$$B_j = AF(E_i) \quad (9)$$

When AF is a linear activation function, the neural network can be viewed as a multilayer sensor. The weights can be updated based on target output and actual output. Let $Q(m)$ and $Q(m\text{-}1)$ be the weight vectors before and after the update, respectively; $a(m)$ be the input corresponding to the output $b(m)$; $\varepsilon(m)$ be the target vector; δ be the preset step length of network learning. Then, $Q(m)$ can be updated by:

$$Q(m) = Q(m-1) + \delta\big(\varepsilon(m) - b(m)\big) \cdot a(m) \quad (10)$$

During network training, the connection weights of network nodes are updated continuously to minimize the deviation in formula (5) and make it converge quickly. This paper adopts a gradient descent algorithm to train the neural network for nutritional balance. The actual gradient is approximated in the neural network by the point gradient. Each time, the gradient descent algorithm only processes the data of one row of nodes and makes predictions through network training. Let O_n be the hidden layer output; M is the number of input-output pairs. Then, the network with the input a_n and the output b_n is iteratively updated to minimize the mean squared error $[MSE(m)]$:

$$[MSE(A)] = \frac{1}{2M}\left(\sum_{i=1}^{M}(O_n - b_n)^2\right) \quad (11)$$

Let η be the learning rate. Based on formula (11), the increment of the weight $Q^o{}_{nm}$ of node m on layer o in sensor n can be calculated by the gradient descent method:

$$\Delta q_{nm}^o = -\eta \frac{\partial MSE(A)}{\partial q_{nm}^o} \quad (12)$$

The deviation of layer o in sensor n can be calculated by:

$$\Delta r_n^o = -\eta \frac{\partial MSE(A)}{\partial r_n^o} \quad (13)$$

The proposed HHH monitoring model combines the merits of the probability model and functional approximation. The model structure classifies foods based on the highest nutritional value. It judges the nutritional balance of the object's food intake under the structural advantage of a multilayer sensor neural network. The relationship between food nutritional features can be described as a directed edge. That is, the possibility of meeting the goal of food intake depends on the intake of a class of foods. For example, the probability of yoghurt belonging to milk foods can be determined by the conditional probability below:

$$P(d-g|Yog) = \frac{P(Yog|d-g)}{P(Yog)}P(d-g) \quad (14)$$

The conditional probability in formula (14) is a posterior probability, which depends on the probability of intaking milk foods.

2.3 Big data processing and primary control factor selection

(1) Abnormal data processing

The collection and transmission of big data on human nutrition monitoring could be disturbed by some non-human factors, causing data abnormality. From the data mining angle, the monitoring model's final upper bound, to a certain degree, hinges on the effect of data preprocessing and feature selection. It is necessary to take some measures to ensure the completeness and accuracy of the dataset.

Following traditional statistics, this paper defines the abnormal values of the big data on human nutrition monitoring as the data beyond 1.5 times the interquartile range from the upper and lower quartiles. Let R_{IQ}, Q_L, and Q_R be the interquartile range, lower quartile, and upper quartile, respectively. Then, we have:

$$R_{IQ} = Q_U - Q_L \quad (15)$$

The abnormal values of the big data on human nutrition monitoring satisfy the following screening conditions:

$$VAL_{OUT} < Q_L - 1.5R_{IQ} \quad (16)$$

$$VAL_{OUT} > Q_U + 1.5R_{IQ} \quad (17)$$

The identified abnormal values in the big data can be replaced or removed. During the human nutrition monitoring, almost all data were measured from the diets of all objects in the same period. There were overlaps and couplings between human nutrition indices owing to the difference in food intake. Therefore, handling abnormal values should not focus on only the independent features of a single nutrient. Hence, this paper resorts to the density-based outlier detection algorithm to judge the abnormal values of human nutrition monitoring with obvious overlaps and couplings.

Let W be an n-dimensional set of m human nutrition samples ($\forall A_i=(a_{i1}\ a_{i2}, \ldots, a_{in})\in\mathbb{R}$, $i = 1, 2, \ldots, m$). For any two samples A_i and A_j in W, the Euclidean distance between the two can be defined as:

$$DIS(A_i, A_j) = \sqrt{\sum_{l=1}^{m}(A_{ij}, A_{jl})^2} \quad (18)$$

For convenience, the K-th nearest sample to a given sample T was defined as the K-th distance of that sample; the K closest sample to sample T was defined as the K-th distance neighborhood of that sample; the maximum of the distance between samples T and O and the k-th distance of T was defined as the reachability distance of T.

In the outlier detection algorithm, the K-th distance of a given sample T can be defined. Assuming that $\xi_K(T)=\xi(T, O)$, then the K-th distance of T needs to satisfy at least two constraints:

(1) There exist at least K samples $O'\in W\backslash\{T\}$ making $\xi(T, O') < \xi(T, O)$;

(2) There exist at most K-1 samples $O'\in W\backslash\{T\}$ making $\xi(T, O') < \xi(T, O)$, i.e., sample O is the K-th nearest sample of T. Figure 3 shows the 6th distance neighborhood of sample T.

The K-th distance neighborhood $NA_K(T)$ of sample T can be defined as:

$$NA_K(T) = \{O' \in W\{T\} | \xi(T, O') \le \xi_K(T)\} \quad (19)$$

The K-th reachability distance $\xi_K(T, O)$ from the given sample O to T can be calculated by:

$$\xi_K(T, O) = max(\xi_K(T, O), \xi(T, O)) \quad (20)$$

Figure 4 shows the 9th reachability distance of sample T.

Formula (20) shows that $\xi_K(T, O)$ is at least the K-th distance of sample T. The local reachability density (LRD) can be calculated by:

$$\sigma_l(T) = \frac{1}{\sum_{O\in NA_K(T)}\xi_K(T,O)/K} = \frac{K}{\sum_{O\in NA_K(T)}\xi_K(T,O)} \quad (21)$$

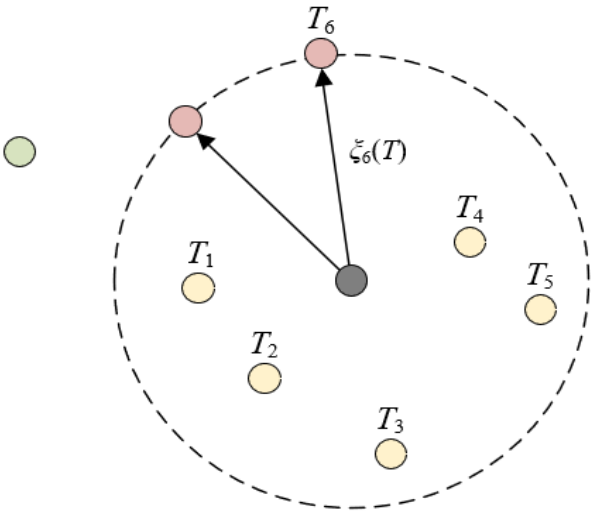

Figure 3. The 6th distance neighborhood of sample T

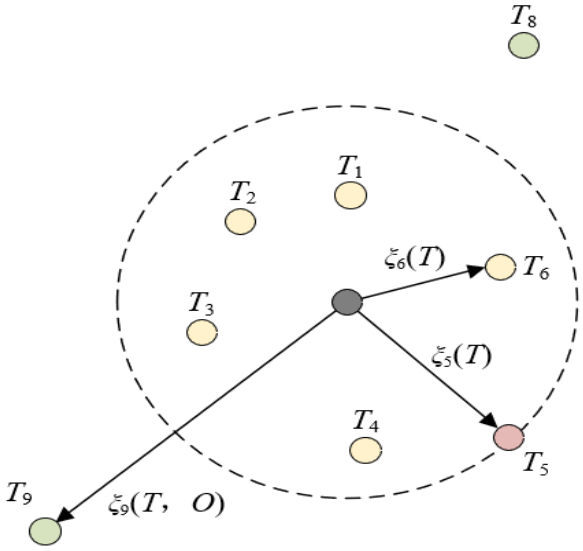

Figure 4. The 9th reachability distance of sample T

Formula (21) shows that the mean reachability density (MRD) of sample T is the reciprocal of the MRD of all samples in the K-th distance neighborhood of T. If the MRD of all samples in the K-th distance neighborhood of T is large, then the reachability density will be small. It is impossible to judge if the data in T are abnormal based on reachability density alone. The densities of the other samples in the K-th distance neighborhood of T should be considered to facilitate the judgement.

The local outlier factor (LOF) of sample T can be calculated by:

$$OF_K(T) = \frac{\sum_{O\in NA_K(T)}\frac{\sigma_K(O)}{\sigma_K(T)}}{K} \quad (22)$$

Formula (22) shows that the LOF of sample T is the mean of the sum of the density ratios of the other samples in the K-th distance neighborhood of T.

(2) Factor analysis

The most prominent difference from other data types for human nutrition data is the apparent overlaps and couplings between human nutrition indices, owing to the difference in food intake. If the monitoring model is directly trained on the collected data, the features of some indices might be extracted repeatedly. For example, digestibility and protein content are closely correlated in human nutrition data. Suppose the two indices are imported to the monitoring model for training. In that case, the model will become much more complex, the training will become inefficient, and the prediction and monitoring will be less stable and accurate. To solve the problem, this paper carries out factor analysis to reduce the dimension of the big data of human nutrition. The information of the original sample set was expressed with a few common factors, and the nutrient intake was predicted and evaluated against the available features of food nutrients. Figure 5 shows the flow of factor analysis-based dimension reduction.

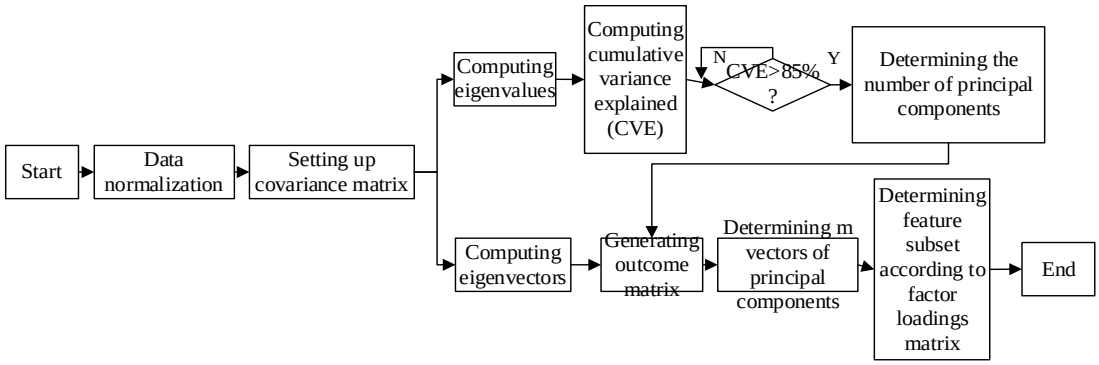

Figure 5. The flow of factor analysis-based dimension reduction

It is assumed that each sample has w indices for the given human nutrition sample set W. Then, W can be written as an $m \times w$- dimensional matrix.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1w} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mw} \end{bmatrix} = [a_1, a_2, a_3, \ldots, a_w] \quad (23)$$

Column i of matrix A can be described as $a_i=(a_{1i}, a_{2i}, a_{3i}, \ldots, a_{mi})^T$. After normalization, the matrix can be rewritten as:

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1w} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mw} \end{bmatrix} \quad (24)$$

Let c_{ij} be any element in the normalized matrix; a'_{ij} and E_j be the mean and standard deviation of index j, respectively. The dimensional and value differences can be eliminated through normalization:

$$c_{ij} = \frac{a_{ij} - a'_{ij}}{E_j} \quad i = 1,2,\ldots,m; j = 1,2,\ldots,w \quad (25)$$

where, a'_{ij} and E_j can be calculated by:

$$a'_{ij} = \frac{1}{m}\sum_{i=1}^{m} a_{ij}, E_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(a_{ij} - a'_{ij}\right)^2} \quad (26)$$

The factor loadings matrix G can be solved by the principal components of the covariance matrix based on matrix C. Let τ be the principal components obtained after dimension reduction of the w indices of the original sample set. Then, matrix C can be expressed as $C=GA+\phi$, where special factor $\phi=D\times S$ ($D= diag(d_1, d_2, \ldots, d_w)$, and $V=(V_1, V_2, \ldots, V_w)^T$). Let element g_{ij} of factor loadings matrix G denote the loading of index i on factor j. Then, matrix C contains the following elements:

$$c_i = g_{i1}\tau_1 + g_{i2}\tau_2 + \ldots + g_{in}\tau_n + d_i V_i \quad (27)$$

The CVE f_i^2 of common factor n to index i can be calculated by:

$$f_i^2 = g_{i1}^2 + g_{i2}^2 + \ldots + g_{in}^2 \quad (28)$$

Let γ_i be the variance of the special factor. The variance of factors other than factor i can be expressed as:

$$VAR_{c_i} = f_i^2 + \gamma_i^2 \quad (29)$$

The correlation coefficient matrix Z of the normalized matrix (24) can be solved as follows:

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1w} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mw} \end{bmatrix} = E \quad (30)$$

Each element in matrix Z can be calculated by:

$$z_{ij} = \frac{\sum_{l=1}^{m}(a_{li} - a'_i)(a_{lj} - a'_i)}{\sqrt{\sum_{l=1}^{m}(a_{li} - a'_i)^2 \sum_{l=1}^{m}(a_{lj} - a'_i)^2}} \quad (31)$$

The next is solving the covariance matrix's eigenvalues and eigenvectors (30). Suppose $Z-\mu I=0$, it is possible to derive the eigenvectors $v_1, v_2, \ldots, v_w$ of eigenvalues $\mu_1, \mu_2, \ldots, \mu_w$. Then, factor loadings matrix G can be expressed as:

$$G = \left(\sqrt{\mu_1}v_1, \sqrt{\mu_2}v_2, \ldots, \sqrt{\mu_n}v_n\right) \quad (32)$$

Finally, γ_i can be estimated according to the elements on the diagonal $Z—GG^T$:

$$\gamma_i^2 = 1 - \sum_{j=1}^{n} g_{ij}^2 \quad (33)$$

3. Results

The original human nutrient data include 26 indices: calories, proteins, fats, carbohydrates, dietary fibers, vitamins A, C, E, carotene, B1, B2, niacin, cholesterol, magnesium, calcium, iron, zinc, copper, manganese, potassium, phosphorus, sodium, selenium, rate of consumption, edible value, and health value. The dimensions of these indexes' data were lowered by factor analysis. Table 4 summarises the eigenvalues and CVEs of the covariance matrix for the original sample set.

Table 4

Eigenvalues and CVEs

Principal component	Eigenvalue	Variance explained	CVE
1	3.526	31.679	30.125
2	2.572	21.524	51.762
3	1.765	14.346	65.234
4	1.274	10.729	75.826
5	1.031	8.694	84.894
6	0.475	3.642	87.092
7	0.362	3.005	92.854
8	0.286	2.46	95.748
9	0.218	1.634	95.246
10	0.162	1.72	97.634
11	0.197	1.426	98.912
12	0.0523	0.475	100.2

The first five main components explained 84.89 percent of the total variation, as seen in Table 4. Figure 6's scree plot of principal components indicates that the first five principal components account for most of the information in the original sample set, as their eigenvalues are all greater than 1.5. Therefore, this research selects the first five major components to decrease the original sample set's dimensions.

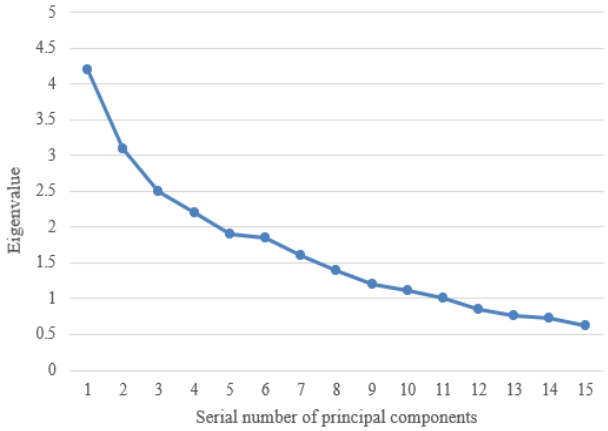

Figure 6. Scree plot of principal components

Table 5's factor loadings matrix displays the links between each index in the initial sample set and the five selected main components. The first principal component has a characteristic root of 3.526 and is substantially associated with calories, proteins, lipids, and vitamin A. The second principal component has a characteristic root of 2.572 and is closely associated with vitamin B and calcium. The third principal component has a characteristic root of 1.765 and is closely associated with carbs. The fourth principal component has a characteristic root of 1.274 and is closely associated with carotene and cholesterol. The fifth main component has a characteristic root of 1.031 and is closely associated with proteins and dietary fibre.

Table 5

Factor loadings matrix

Principal component					
Index	**1**	**2**	**3**	**4**	**5**
Calorie	0.678	0.457	0.312	0.043	-0.035
Proteins	0.524	-0.162	0.249	-0.125	0.706
Fats	0.679	0.138	-0.164	0.219	0.421
Carbohydrates	0.023	0.263	0.762	0.202	-0.385
Dietary fibers	-0.579	0.451	0.185	-0.238	0.456
Vitamin A	0.873	0.289	-0.271	-0.56	0.253
Vitamin C	-0.529	0.354	0.439	-0.193	0.419
Vitamin E	-0.374	0.367	-0.251	0.256	0.042
Carotene	0.026	-0.235	0.435	0.762	0.351
Vitamin B	-0.343	0.741	-0.192	-0.63	0.047
Cholesterol	-0.271	0.270	-0.276	0.785	-0.065
Calcium	-0.293	0.526	-0.361	0.271	-0.271
Iron	-0.213	0.501	-0.331	0.215	-0.215

Using factor analysis to reduce the dimensionality of the original sample set, eleven human nutrition parameters were chosen to analyze the primary indices of HHH: calories, proteins, lipids, carbs, dietary fibers, vitamins A, B, and C, and cholesterol.

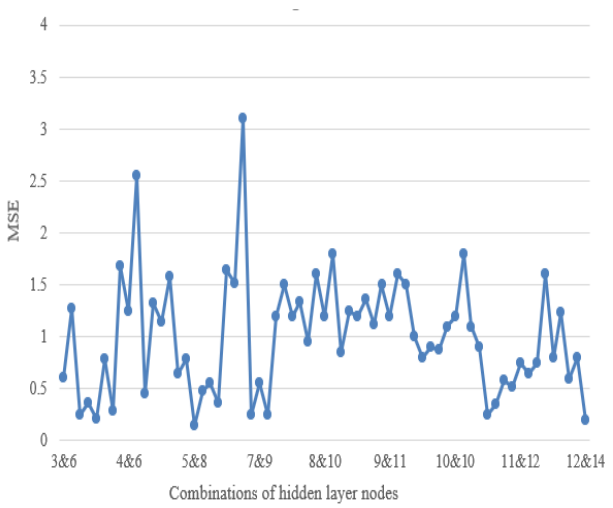

Figure 7. MSEs at different settings of hidden layer nodes

Figure 7 depicts the model prediction impacts evaluated by the MSEs for various hidden layer node settings. Combinations of the number of nodes on two hidden levels comprise the x-axis. The model's MSE was reduced, indicating that the prediction was most accurate when the two hidden layers included 12 and 14 nodes, respectively. Consequently, our model's architecture was configured as follows: an input layer of 10 nodes, an invisible layer of 12 nodes, a hidden layer of 14 nodes, and an output layer of 1 node.

Through Bayesian regularization, the weights and thresholds of the created neural network were calculated in this study. Figure 8 demonstrates the MSE fluctuation of the BN-based HHH monitoring model as a function of the number of iterations. After the 34th iteration, the suggested BN was largely stable.

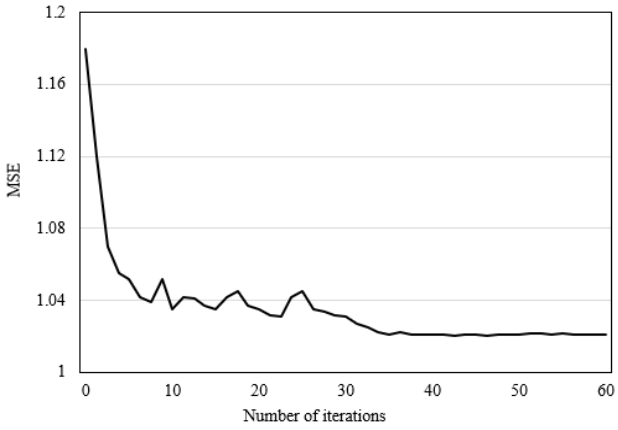

Figure 8. MSE variation with the number of iterations

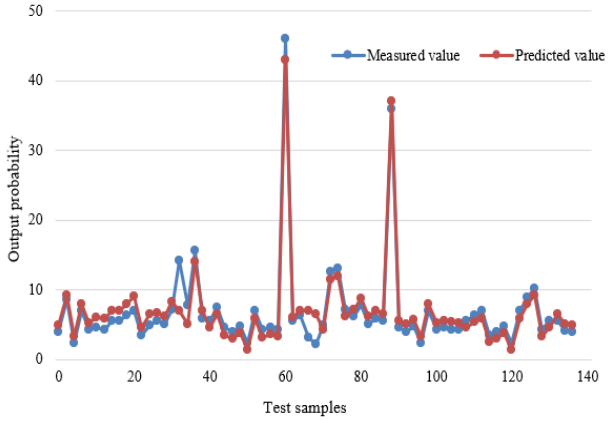

Figure 9. Test effect of our model

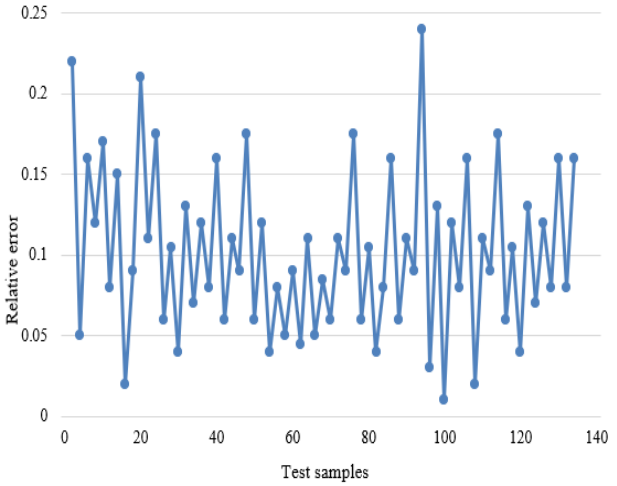

Figure 10. Relative errors of our model on the test set

To further validate the class prediction effect of our model, 100 sets of human nutritional samples were imported into the trained BN-based model as test samples. Figure 9 depicts the outcomes of our model's application to the test set. Figure 10 illustrates the relative mistakes of the classification based on the highest nutritional value relative to the ground truth. The suggested model achieved a favorable prediction effect on the test set: the mean relative error was less than 15%, and the prediction accuracy was satisfactory.

The above numbers indicate that our model's errors were within 10%. On the same dataset, the maximum relative error of our model was 9.7%, which is barely half that of the genetic algorithm-optimized deep learning network. Thus, the suggested model is more resistant to sounds and disturbances and makes more accurate predictions than the reference model.

4. Conclusions

This research assesses and monitors HHH using big data analysis. After describing the HHH optimization method, the authors developed a BN-based model capable of extracting characteristics from dietary nutrients, classifying foods by their highest nutritional value, recognizing undernutrition, and recommending replacements. The authors then explained how to process massive amounts of data on human nutrition and pick the primary control variables. The dimensions of the data on 26 indices were reduced through factor analysis. The eigenvalues and CVEs of the original sample set's covariance matrix were calculated and summarized. To examine the primary index of HHH, eleven human nutrition parameters were selected: calories, proteins, lipids, carbs, dietary fibers, vitamin A, carotene, vitamin B, cholesterol, and calcium. After obtaining the MSEs of our model under various configurations of hidden layer nodes, the topology of the proposed neural network was completed using these values. The fluctuation of the MSE concerning the number of iterations indicates that our model achieves a good prediction effect and a high prediction accuracy on the test set.

Our research results not only expand the theoretical underpinnings of big data but also give a standard scheme for HHH monitoring data, thereby setting the groundwork for data sharing and interchange. In addition, the authors built a thorough evaluation model for HHH, a recommendation model for exercises, and a platform for HHH monitoring services. These tools support the efficient use of data and enhance the government's service capabilities.

Future research will delve deeply into the following issues to advance the related work and improve HHH monitoring: (1) The BN was chosen to evaluate HHH data for large data scenarios, and the accuracy of various models was tested. It is suggested that future studies study the genetic algorithm's search process, properly design the algorithm's hyperparameters, and enhance the prediction accuracy. (2) Our tests were conducted in rapid succession (1 day). After the program has been operational for an adequate period, there will be sufficient data for research. The study duration should then be chosen by adjusting the intervals based on the actual circumstances.

References

Bommarito, P. A., Xu, X., González-Horta, C., Sánchez-Ramirez, B., Ballinas-Casarrubias, L., Luna, R. S., Pérez, S. R., Ávila, J. E. H., García-Vargas, G. G., & Del Razo, L. M. (2019). One-carbon metabolism nutrient intake and the association between body mass index and urinary arsenic metabolites in adults in the Chihuahua cohort. *Environment international, 123*, 292-300. https://doi.org/10.1016/j.envint.2018.12.004

Comber, R., Weeden, J., Hoare, J., Lindsay, S., Teal, G., Macdonald, A., Methven, L., Moynihan, P., & Olivier, P. (2012). Supporting visual assessment of food and nutrient intake in a clinical care setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 919-922). Association for Computing Machinery, New York, NYUnited States. https://doi.org/10.1145/2207676.2208534

Gupta, A. K., Chakraborty, C., & Gupta, B. (2019). Monitoring of Epileptical Patients Using Cloud-Enabled Health-IoT System. *Traitement du Signal, 36*(5), 425-431. https://doi.org/10.18280/ts.360507

Hoefkens, C., Sioen, I., Baert, K., De Meulenaer, B., De Henauw, S., Vandekinderen, I., Devlieghere, F., Opsomer, A., Verbeke, W., & Van Camp, J. (2010). Consuming organic versus conventional vegetables: The effect on nutrient and contaminant intakes. *Food and Chemical Toxicology, 48*(11), 3058-3066. https://doi.org/10.1016/j.fct.2010.07.044

Hoefkens, C., Sioen, I., De Henauw, S., Vandekinderen, I., Baert, K., De Meulenaer, B., Devlieghere, F., & Van Camp, J. (2009). Development of vegetable composition databases based on available data for probabilistic nutrient and contaminant intake assessments. *Food chemistry, 113*(3), 799-803. https://doi.org/10.1016/j.foodchem.2008.06.049

Jones, J. (2010). Added sugars, nutrient intakes, and grain-based foods. *Cereal foods world, 55*(5), 226-230. https://www.proquest.com/openview/d19379fd004ff979859a00dcc3f862f8

Levenhagen, D. K., Gresham, J. D., Carlson, M. G., Maron, D. J., Borel, M. J., & Flakoll, P. J. (2001). Postexercise nutrient intake timing in humans is critical to recovery of leg glucose and protein homeostasis. *American Journal of Physiology-Endocrinology And Metabolism, 280*(6), E982-E993. https://doi.org/10.1152/ajpendo.2001.280.6.E982

Li, X., Lin, C., & Xu, X. (2019). A Target Tracking Model for Enterprise Production Monitoring System Based on Spatial Information and Appearance Model. *Traitement du Signal, 36*(4), 369-375. https://doi.org/10.18280/ts.360410

Lu, Y., Stathopoulou, T., Vasiloglou, M. F., Christodoulidis, S., Stanga, Z., & Mougiakakou, S. (2020). An artificial intelligence-based system to assess nutrient intake for hospitalised patients. *IEEE transactions on multimedia, 23*, 1136-1147. https://doi.org/10.1109/TMM.2020.2993948

Ng, Y.-K., & Jin, M. (2017). Personalized recipe recommendations for toddlers based on nutrient intake and food preferences. In *Proceedings of the 9th international conference on management of digital ecosystems* (pp. 243-250). Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3167020.3167057

Nicklas, T. A., O'Neil, C. E., Zanovec, M., Keast, D. R., & Fulgoni III, V. L. (2012). Contribution of beef consumption to nutrient intake, diet quality, and food patterns in the diets of the US population. *Meat science, 90*(1), 152-158. https://doi.org/10.1016/j.meatsci.2011.06.021

Ohara, N., Naito, Y., Nagata, T., Tachibana, S., Okimoto, M., & Okuyama, H. (2008). Dietary intake of rapeseed oil as the sole fat nutrient in wistar rats-Lack of increase in plasma lipids and renal lesions. *The Journal of Toxicological Sciences, 33*(5), 641-645. https://doi.org/10.2131/jts.33.641

Rahaman, A., Islam, M. M., Islam, M. R., Sadi, M. S., & Nooruddin, S. (2019). Developing IoT Based Smart Health Monitoring Systems: A Review. *Revue d'Intelligence Artificielle, 33*(6), 435-440. https://doi.org/10.18280/ria.330605

Rodríguez-Palmero, M. a., Castellote-Bargalló, A. I., López-Sabater, C., de la Torre-Boronat, C., & Rivero-Urgell, M. (1998). Assessment of dietary nutrient intakes: analysed vs calculated values. *Food chemistry, 61*(1-2), 215-221. https://doi.org/10.1016/S0308-8146(97)00106-4

Shimbo, S., Hayase, A., Murakami, M., Hatai, I., Higashikawa, K., Moon, C. S., Zhang, Z. W., Watanabe, T., Iguchi, H., & Ikeda, M. (1996). Use of a food composition database to estimate daily dietary intake of nutrient or trace elements in Japan, with reference to its limitation. *Food Additives & Contaminants, 13*(7), 775-786. https://doi.org/10.1080/02652039609374465

Silva, R. E., Simões-e-Silva, A. C., Miranda, A. S., Justino, P. B., Brigagão, M. R., Moraes, G. O., Gonçalves, R. V., & Novaes, R. D. (2019). Potential role of nutrient intake and malnutrition as predictors of uremic oxidative toxicity in patients with end-stage renal disease. *Oxidative Medicine and Cellular Longevity, 2019*, 7463412. https://doi.org/10.1155/2019/7463412

Sivakumaran, S., Huffman, L., & Sivakumaran, S. (2018). The New Zealand food composition database: a useful tool for assessing New Zealanders' nutrient intake. *Food chemistry, 238*, 101-110. https://doi.org/10.1016/j.foodchem.2016.12.066

Tordoff, M. G. (2002). Obesity by choice: the powerful influence of nutrient availability on nutrient intake. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 282*(5), R1536-R1539. https://doi.org/10.1152/ajpregu.00739.2001

Tucker, K. L. (2016). Nutrient intake, nutritional status, and cognitive function with aging. *Annals of the New York Academy of Sciences, 1367*(1), 38-49. https://doi.org/10.1111/nyas.13062

Velasco-Ryenold, C., Navarro-Alarcón, M., De La Serrana, H. L.-G., Perez-Valero, V., & Lopez-Martinez, M. C. (2008). Total and dialyzable levels of manganese from duplicate meals and influence of other nutrients: Estimation of daily dietary intake. *Food chemistry, 109*(1), 113-121. https://doi.org/10.1016/j.foodchem.2007.12.025

Villalobos, G., Almaghrabi, R., Hariri, B., & Shirmohammadi, S. (2011). A personal assistive system for nutrient intake monitoring. In *Proceedings of the 2011 international ACM workshop on Ubiquitous meta user interfaces* (pp. 17-22). Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/2072652.2072657

Welis, W. (2017). The quality of nutrient intake of table tennis athlete. *IOP Conference Series: Materials Science and Engineering, 180*(1), 012184. https://doi.org/10.1088/1757-899X/180/1/012184

Yan, Y., Wu, Z., Wu, X., Zhou, X., & Weng, C. (2019). A Linux-Based Integrated Structural Health Monitoring System for Bridges in Remote Regions. *Instrumentation, Mesures, Métrologies, 18*(6), 527-534. https://doi.org/10.18280/i2m.180603