Journal of Artificial Intelligence Practice (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2022.050305
ISSN 2371-8412 Vol. 5 Num. 3

Design of an AI Health Risk Assessment System for Dietary Hygiene of Key Groups Based on IoT Wearable Devices

Boyuan Wang[1],*, Hai Lin[1], Shenglin Xia[1]

[1]*Zhongshan Center for Disease Control and Prevention, Zhongshan, China*
**Corresponding author*

Keywords: Deep Learning, Artificial Intelligence, Big Data, Wearable IoT Devices, Diseases of food origin

Abstract: Population Spatio-temporal big data mining and analysis techniques have been applied to risk assessment of disease transmission, which can describe disease transmission pathways and high-risk areas in fine detail. Based on spatial statistical analysis and artificial intelligence technology, this study seeks to break through the previous risk warning model of a single data source from medical institutions in the era of small data and designs an AI health risk assessment system for the dietary hygiene of key populations. The system is designed to collect multi-source Spatio-temporal big data consisting of urban population positioning, a sanitary inspection of restaurant premises, foodborne disease cases in medical institutions, and environmental monitoring. Spatial location attributes are assigned to the monitoring data, and food and multi-source data are fused across borders. Through the Internet of Things (IoT) technology, the system is designed with an IoT system consisting of sensors for automatic monitoring and wearable devices for real-time warning. Based on the spatial and artificial intelligence models, the system designs personalized and real-time early warning information for critical populations to prevent dietary health risks and provide scientific basis and support for public health departments to prevent foodborne diseases.

1. Introduction

A healthy and hygienic diet is an essential element of good health. Diseases of food origin refer to infection or poisoning caused by the entry of food into the body's causative factors [1]. Bacteria and viruses cause most infectious diseases of food origin. Many factors influence the distribution, outbreak, or epidemic of these diseases. The apparent influencing factors include traditional, local, or behavioral factors such as kitchen environment, food procurement, food processing, and eating habits. There are few studies on the connection between external objective environments such as geospatial or space-time. The big data era has produced numerous big spatial data with time and space markers that can describe individual behaviors, such as mobile phone data, taxi data, and social media data. These data provide a new way for people to understand the socio-economic environment quantitatively. In recent

The study was supported by the Zhongshan Social Public Welfare Science and Technology Research Project (No. 2019B1106).

years, scholars in computer science, geography and science, and complexity science have conducted many investigations based on different data and tried to find the spatiotemporal dynamic mode of massive, big data and establish a reasonable interpretative way. Studies have shown urban population lifestyles influence that disease transmission. These spatially meaningful data with positioning characteristics provide valuable clues for epidemiological studies of health and the environment. It is a prerequisite for establishing a risk assessment system for diseases of food origin. The system is designed to extract the characteristics of environmental surveillance objects automatically. It intelligently classifies the monitoring objects through artificial intelligence neural network models and calculates the risk values of foodborne diseases in each street area. It then drives low, medium, and high-risk areas for foodborne diseases in the monitoring area through machine learning correlation analysis methods generates risk assessment maps and provides early warning to users in high-risk areas through wearable IoT devices [2].

2. System Architecture

2.1. Data Collection Layer

The framework of the system design is shown in Figure 1. The system first collects and monitors the city's population location big data, restaurant health inspection big data, medical institution foodborne disease case data, and environmental monitoring big data (chemical fertilizer and pesticide pollution, water pollution, etc.), which collectively make up multi-source Spatio-temporal big data. The system is designed with an acquisition layer that collects the above multi-source data through automated methods of IoT devices. Among them, the city's population location data is obtained through the location and heat interface of mobile operators and map service providers, the restaurant hygiene inspection data is gained through the "Bright Kitchen" project automatic monitoring equipment installed in restaurants [3], the foodborne disease case data is automatically received from the information system of monitoring medical institutions, and the environmental monitoring data and the pollution data is got through the IoT sensors installed at the monitoring points.

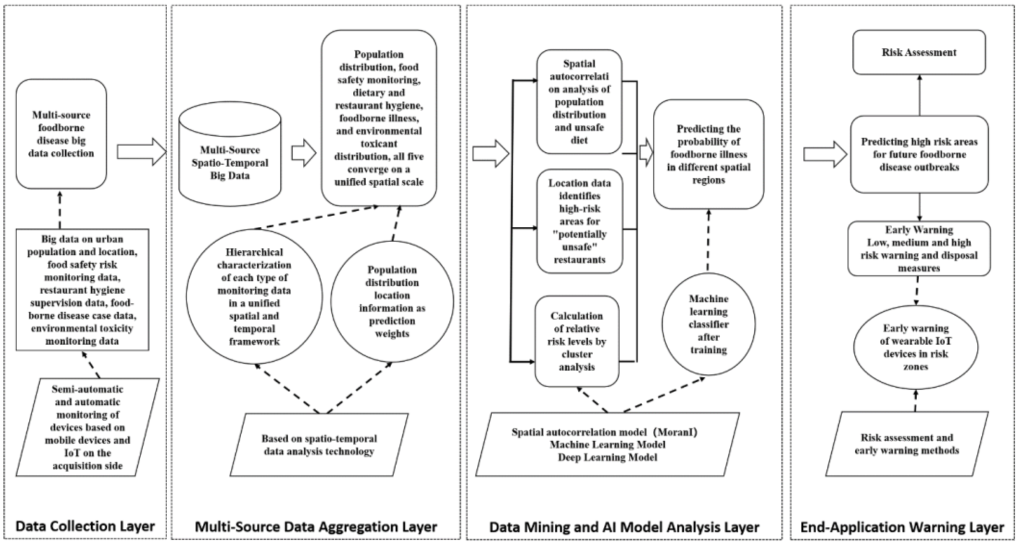
Multi-source foodborne disease big data collection
Big data on urban population and location, food safety risk monitoring data, restaurant hygiene supervision data, food-borne disease case data, environmental toxicity monitoring data
Semi-automatic and automatic monitoring of devices based on mobile devices and IoT on the acquisition side
Data Collection Layer
Multi-Source Spatio-Temporal Big Data
Population distribution, food safety monitoring, dietary and restaurant hygiene, foodborne illness, and environmental toxicant distribution, all five converge on a unified spatial scale
Hierarchical characterization of each type of monitoring data in a unified spatial and temporal framework
Population distribution location information as prediction weights
Based on spatio-temporal data analysis technology
Multi-Source Data Aggregation Layer
Spatial autocorrelation analysis of population distribution and unsafe diet
Location data identifies high-risk areas for "potentially unsafe" restaurants
Calculation of relative risk levels by cluster analysis
Predicting the probability of foodborne illness in different spatial regions
Machine learning classifier after training
Spatial autocorrelation model (MoranI)
Machine Learning Model
Deep Learning Model
Data Mining and AI Model Analysis Layer
Risk Assessment
Predicting high risk areas for future foodborne disease outbreaks
Early Warning
Low, medium and high risk warning and disposal measures
Early warning of wearable IoT devices in risk zones
Risk assessment and early warning methods
End-Application Warning Layer

Figure 1: Overall Architecture of the System

2.2. Multi-Source Data Aggregation Layer

On the unified Spatio-temporal framework, combined with 3S (GIS, RS, GPS) technology [4], the multi-source data aggregation layer takes the multi-source Spatio-temporal big data obtained from the data collection layer. It classifies each monitoring data as low, medium, and high qualitatively under the unified Spatio-temporal framework with the street of the monitoring city as the smallest unit.

2.3. Data Mining and AI Model Analysis Layer

Data mining and artificial intelligence model analysis layer, based on spatial autocorrelation model (Moran’s I), machine learning methods (Machine Learning), and deep learning methods (Deep Learning), to model and analyze multi-source big data on a unified spatial and temporal scale, including the application of spatial autocorrelation analysis methods to derive maps of population distribution and dietary disease characteristics. The unsupervised machine learning model trained from big data can automatically derive risk levels for food-borne diseases at different spatial and temporal levels.

2.4. End-Application Warning Layer

The end-application warning layer is aimed at warning the predicted Spatio-temporal areas of low, medium, and high risk of dietary hygiene. The system is designed to warn the user of the corresponding dietary risk level in the risk area through wearable IoT devices (bracelets, etc.) that scan GPS location information in real-time.

3. Key technologies for foodborne disease assessment models

Spatial statistical models and artificial intelligence models [5] are used for intelligent analysis of multi-source collected data, including key technologies such as spatial autocorrelation analysis algorithms, unsupervised machine learning clustering analysis algorithms, and deep learning convolutional neural network models.

3.1. Spatial Autocorrelation Analysis Method

Spatial autocorrelation analysis mainly refers to the degree of correlation between an attribute value on a geospatial region and the same attribute value in its neighboring spatial regions. The spatial autocorrelation coefficient is usually used as a primary metric to test whether a particular attribute value in a unit region has high-high adjacency, low-low adjacency, or high-low adjacency. Spatial autocorrelation analysis is mainly divided into global and local spatial autocorrelation analysis [6]. The commonly used spatial autocorrelation analysis methods are Moran's I, Geary's C, Getis, and Moran scatter plot.

The global spatial autocorrelation analysis focuses on whether the attribute variables are aggregated from the whole study area, and its formula is as follows [7].

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}\right) \sum_{j}^{n} = 1 \, (X_i - \bar{X})^2} \tag{1}$$

Where n denotes the number of regions in the space of studied attribute variables; Xi denotes the value of the attribute variable within the ith region (e.g., disease incidence) and Xj denotes the value of the attribute variable within the jth region, indicating the mean value of the attribute variable in the region under study; Wij denotes the spatial weight matrix, determined as follows.

$$W_{ij} = \begin{cases} 1\text{，When region i is adjacent to region j,} \\ 0\text{，Other situation.} \end{cases} \tag{2}$$

Under the z hypothesis, the expected value of Moran's I is as follows.

$$E(I) = \frac{-1}{n-1} \tag{3}$$

Under the assumption of normal distribution of spatial objects, the variance of Moran's I is as follows.

$$Var(I) = \frac{1}{S^2_{0(n-1)(n+1)}}(n^2 s_1 - n s_2 + 3 s_0^2) - E(I)^2 \tag{4}$$

Under the assumption of random distribution of spatial objects, the variance of Moran's I is as follows

$$Var(I) = \frac{n\left((n^2-3n+3)s_1 - n s_2 + 3 s_0^2\right) - k\left((n^2-n)s_1\right) - E(I)^2}{s_0^2(n-1)(n-2)(n-3)} \tag{5}$$

The Z-score statistics for Moran's I are determined as follows.

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} \tag{6}$$

If |Z|<1.96, P<0.05, the zero hypotheses are rejected, the overall spatial autocorrelation coefficient is not zero, and the attribute variables are considered spatial autocorrelation. When there is a negative spatial correlation, the value is less than 0, and the closer to -1, the stronger the negative correlation is, that is, the greater the spatial variability of the object of study; when there is a random distribution, the value is close to 0, that is, there is no autocorrelation.

Local spatial autocorrelation analyzes whether there is aggregation in spatial distribution among attribute variables from a specific local area within the overall geospatial scope. The results can be used to explain and detect "hot spots" or "cold spots" in the spatial aggregation of the attribute variables. Moran's I > 0 indicates the existence of a positive spatial correlation between local spatial units and neighboring spatial units, which is expressed as "high-high" or "low-low" aggregation. When Moran's I < 0, the spatial correlation between local spatial units and neighboring spatial units is negative, manifesting as "low-high" and "high-low" aggregation.

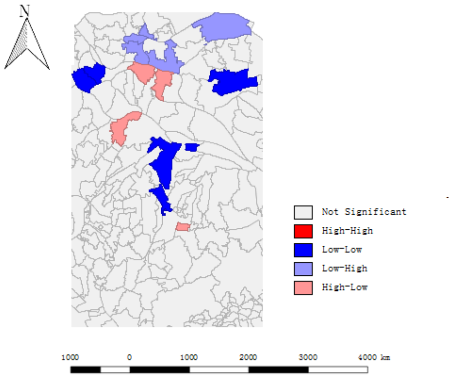

Figure 2: Schematic diagram of the results of spatial autocorrelation analysis

The results of the spatial autocorrelation simulation analysis constructed at the street scale with a city as the research target are shown in Figure 2. The simulation results showed that four locations presented a meaningful aggregated distribution of high and low values (High-Low), showing a high incidence of foodborne illness in this street and a low incidence of foodborne illness in the surrounding community, suggesting that this community and the surrounding community belong to the intersection of hot and cold spots of foodborne illness, which is of further research significance.

3.2. Unsupervised Machine Learning Clustering Method

The cluster analysis method [8] is an unsupervised machine learning method and has many applications in biological and medical classification problems. A large number of observations can be grouped into several classes. Here, a class is defined as a group of several observations, and the similarity of observations within a group is higher than the similarity between groups. The k-means algorithm commonly used in cluster analysis methods is a typical method of dividing clusters and belongs to unsupervised machine learning. The idea is to obtain the classification of each sample point by minimizing the sum of squares of errors within the group given the number of clusters k.

The advantage of the AP cluster analysis method [9] applied in this system design compared with K-means clustering is that the number of clusters does not need to be given in advance, and the optimal number of clusters can be automatically analyzed and obtained by the AP algorithm. The AP cluster analysis method applied in the design of this system is an unsupervised machine learning clustering algorithm. The basic idea of the AP algorithm is to treat all samples as network nodes and calculate the clustering center of each sample through the message transmission of each edge in the network. In the clustering process, two kinds of messages are passed among the nodes: responsibility and availability. By passing messages between points, figurative elements are finally selected to complete the clustering by continuously passing messages. The AP algorithm continuously updates each point's availability and attribution values through an iterative process until m high-quality Exemplars are produced. In contrast, the remaining data points are assigned to the corresponding clusters. AP clustering is a continuous iterative process. The iterative process mainly updates two matrices, Responsibility matrix R = r(i,k) and Availabilities matrix A = a(i,k), which are determined as follows.

$$r(i,k) = s(i,k) - \max_{k' \neq k}(a(i,k') + s(i,k'))$$

$$a(i,k) = \begin{cases} \min\left\{0, r(k,k) + \sum_{i' \notin \{i,k\}} \max\left(0, r(i',k)\right)\right\}, i \neq k \\ \sum_{i' \neq k} \max\left(0, r(i',k)\right), \ i = k \end{cases} \tag{7}$$

$$r(i,k) \leftarrow s(i,k) - \max_{k' s.t. k' \neq k} \{a(i,k') + s(i,k')\}$$

$$a(i,k) \leftarrow \min\{0, r(k,k) + \sum_{i' s.t. i' \notin \{i,k\}} \max\{0, r(i',k)\}\} \tag{8}$$

Where s(i,k) denotes similarity, indicating the likelihood of k when the representative element of i. The results of the AP clustering simulation analysis constructed at the district scale for one city as a study are shown in Figure 3.

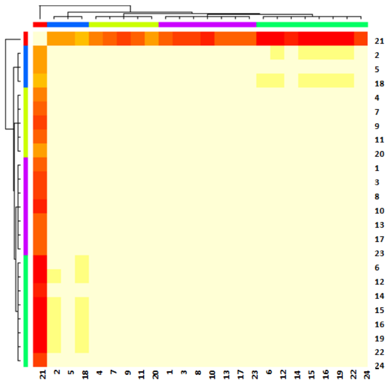

Figure 3: Schematic diagram of the results of AP clustering analysis

A heat map shows the degree of similarity of each district relative to other districts, with colors ranging from light yellow to dark red, darker colors showing greater differences between the district and other cities, and no colors revealing minimal differences. The results of Figure 3 demonstrate that the optimal number of clusters derived from AP analysis is 5, which are red (zone 21,Tanz), blue (zone 2, zone 5, zone 18), yellow (zone 4, zone 7, zone 9, zone 11, zone 20), purple (zone 1, zone 3, zone 8, zone 10, zone 13, zone 17, zone 23) and green: (zone 6, zone 12, zone 14, zone 15, zone 16, zone 19, zone 22, zone 24). The unsupervised clustering machine learning results show that zone 21 is separated into a separate class, demonstrating that it is incredibly different from the other zones and is therefore classified separately with further attention.

4. Early warning based on wearable IoT devices

Traditional food safety risk warning methods are limited, and the warning is mainly for professionals in food supervision departments, which require dedicated personnel to operate. It is difficult for the general public to use it in their lives. The terminal warning of the wearable IoT device designed in this study realizes personalized and real-time food-borne disease warning information for key groups of people and can effectively scan and suggest possible dietary hygiene problems in the scope of daily life with a real-time warning function based on the geographic location information of individual users.

The structure of the wearable IoT device is based on a smart bracelet, including the main ring with a data transmission module, 5G SIM card, GPS module, microprocessor, and alarm light stuck on the ring. The smart bracelet receives information on the dietary hygiene of people through the data transmission module and 5G SIM card, and the AI health risk assessment system sends information on dietary hygiene through the server, including the coordinates of the spatial location of areas with a high risk of food contamination. When the distance between the smart bracelet and the coordinates of the food contamination source sent by the server is less than the preset distance, the microprocessor sends a command to drive the alarm light to send out the specified alarm message to remind critical groups of people, such as students, the elderly, and wild plant pickers, to check the food safety warning information and avoid the problematic food in time.

Acknowledgment

This work was supported by the Zhongshan Social Public Welfare Science and Technology Research Project (Fund Number: 2019B1106).

References

[1] E. Abebe, G. Gugsa, and M. Ahmed, "Review on Major Food-Borne Zoonotic Bacterial Pathogens," Journal of Tropical Medicine, vol. 2020, pp. 4674235, 2020.
[2] M. U.Farooq, M. Waseem, S. Mazhar et al., "A Review on Internet of Things (IoT)," International Journal of Computer Applications, vol. 113, no. 1, pp. 1-7, 2015.
[3] M. Bhatia, and T. A. Ahanger, "Intelligent decision-making in Smart Food Industry: Quality perspective," Pervasive and Mobile Computing, vol. 72, pp. 101304, 2021.
[4] J. Zhang, and L. Wei, "Design and Application of 3S-based Resettlement Project Information Platform," in 2021 2nd International Conference on Artificial Intelligence and Information Systems, Chongqing, China, pp. Article 98, 2021.
[5] C. Zhang, and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," Journal of Industrial Information Integration, vol. 23, pp. 100224, 2021.
[6] L. Chen, L. Sun, R. Zhang et al., "Epidemiological analysis of wild mushroom poisoning in Zhejiang province, China, 2016-2018," Food Science & Nutrition, vol. 10, no. 1, pp. 60-66, 2022.
[7] L. Qin, R. Lin, R. Gao et al., "Correlation between Population Structure and Regional Innovation Ability Based on Big Data Analysis," Mathematical Problems in Engineering, vol. 2022, pp. 7000390, 2022.

[8] A. Saxena, M. Prasad, A. Gupta et al., "A review of clustering techniques and developments," Neurocomputing, vol. 267, pp. 664-681, 2017.

[9] B. J. Frey, and D. Dueck, "Clustering by passing messages between data points," Science, vol. 315, no. 5814, pp. 972-6, 2007.